

Лекция 1.

Тема 1. ПРЕДМЕТ И ЗАДАЧИ КУРСА

- 1.1. Понятие об исследовании операций.
- 1.2. Системный подход – это научная основа принятия решений (понятие системы, строение системы, система и среда, классификация систем).
- 1.3. Модели исследования операций (понятие модели, классификация моделей, принципы построения экономико-математических моделей, типовые модели исследования операций).
- 1.4. Этапы исследования операций.
- 1.5. Понятие о критерии эффективности.

1.1. Понятие об исследовании операций

Исследование операций представляет собой комплекс научных методов для решения задач эффективного управления организационными системами.

Как самостоятельное научное направление исследование операций сформировалось в начале 40-х годов. Среди первых исследований в этом направлении можно назвать работу Л.В. Канторовича «Математические методы организации и планирования производства», которая вышла в 1939 году, и работу Дж. Данцига, посвященную решению линейных экстремальных задач, которая вышла в 1947 году.

В 1939–40 гг. методы исследования операций применялись для решения в основном военных задач, в частности для анализа и исследования боевых операций. Отсюда и возникло название дисциплины. За разработку метода линейного программирования в 1965 году Л.В. Канторовичу, В.С. Немчинову, В.В. Новожилову была присуждена Ленинская премия. В 1975 году Л.В. Канторович и Т. Купманс получили Нобелевскую премию за вклад в теорию оптимизации распределения ресурсов. С увеличением масштабов производства, с развитием форм и методов организации управления экономическими системами совершенствовались и методы исследования операций, расширялся круг решаемых этой наукой задач.

Большой вклад в формирование и развитие этой науки внесли зарубежные ученые Р. Акоф, Р. Беллман, Дж. Данциг, Г. Кун, Дж. Нейман, Т. Саати, Р. Черчмен, А. Кофман, Р. Фор и др. Среди русских

ученых можно назвать таких, как Л.В. Канторович, В.В. Новожилов, Д.Б. Юдин, Н.П. Федоренко и др.

В современных условиях для решения практических задач исследования операций успешно применяются нейросетевые модели, позволяющие преобразовать с помощью весов обученной нейросети входной вектор в выходной. В сложных практических задачах обученная нейросеть выступает как эксперт, обладающий большим опытом и способный правильно ответить на трудные вопросы. Примерами исследования нейросетей служат медицинская диагностика, задачи экспертной оценки, поиска зависимостей в исследуемой базе данных, прогнозирования показателей и т.д.

Исследование операций – это наука, занимающаяся разработкой и практическим применением методов наиболее эффективного (оптимального) управления экономическими организационными системами. Из определения следует, что *предметом* исследования операций являются экономические системы организационного управления (организации, объекты, процессы), которые состоят из большого числа взаимодействующих между собой подразделений (подсистем), причем интересы подсистем (или подразделений) не всегда согласуются между собой и могут быть противоположны.

Целью исследования операций является количественное обоснование принимаемых решений по организации управления системами, объектами или процессами.

Решение, которое оказывается наиболее выгодным для всего объекта (системы), называется *оптимальным*, а решение, наиболее выгодное одному или нескольким подразделениям (подсистем) этого объекта, называется *субоптимальным*.

1.2. Системный подход – это научная основа принятия решений (понятие системы, строение системы, система и среда, классификация систем)

Системный подход предполагает рассмотрение любого объекта или процесса, протекающего в АПК, в качестве целостной сложной динамической системы.

Системный подход представляет собой совокупность методологических принципов и теоретических положений, позволяющих рассматривать каждый элемент системы в его связи и взаимодействии с другими. Его сущность состоит в определенной направленности и последовательности анализа систем путем изучения

их структуры, состава систем, характера взаимосвязей между элементами, между подсистемами, их устойчивости и чувствительности.

Понятия системы связано с наличием некоторого множества совокупности элементов. Но не всякое множество элементов образует систему. Система – это не механический набор, а совокупность взаимодействующих между собой элементов.

При этом важно уметь отделить систему от среды, с которой взаимодействует система, т.е. любая система – это конечное множество функциональных элементов и отношений между ними, выделенное из среды в соответствии с определенной целью.

Таким образом, *система* есть отражение в сознании исследователя, наблюдателя свойств системы, ее элементов при решении задач исследования операций. Так, изучение товара или прогнозирование рынка не дает должного эффекта, а только комплексный подход позволяет эффективно прорваться на рынок с товарами и услугами, особенно с новыми товарами и оригинальными услугами. Следовательно, к решению любой проблемы необходимо подходить с позиций системного анализа.

Любую систему можно подразделить на подсистемы. Подсистема представляет собой выделенное по определенному правилу из системы целенаправленное подмножество взаимосвязанных элементов, которые взаимодействуют между собой, реализуют определенную функцию, необходимую для достижения цели, поставленной перед системой в целом.

Расчленяя систему на подсистемы, следует иметь в виду, что выделение подсистем зависит от цели и может меняться по мере ее уточнения и развития представлений исследователя об анализируемом объекте или процессе.

Каждая подсистема состоит из упорядоченной совокупности элементов, имеющих свою постоянную структуру. Подсистему можно расчленять на элементы различными способами в зависимости от формулировки задачи, цели ее уточнения в процессе системного анализа. При необходимости можно изменять принцип расчленения, выделяя другие элементы подсистемы, и получать с помощью этого нового расчленения более адекватное представление об анализируемом объекте или процессе. Таким образом, *элемент* – это предел расчленения подсистем с точки зрения аспекта рассмотрения системы, решения конкретной задачи, поставленной цели.

Все элементы подсистемы подразделяются на первичные и вторичные. Вторичные элементы можно не рассматривать при исследовании системы. Не включение хотя бы одного первичного элемента в подсистему и систему нарушает установившиеся связи и превращает ее в качественно новый объект.

Каждый элемент выполняет свое собственное назначение в подсистеме. Наряду с этим все элементы находятся между собой в теснейшем взаимодействии и неразрывном единстве. Но отдельные элементы еще не создают целостности структуры подсистемы. Органическое целое образуется тогда лишь, когда между отдельными элементами устанавливаются устойчивые внутренние связи, благодаря которым подсистема и система приобретают целостный характер и новые качества.

Связь одновременно характеризует и строение (статику), и функционирование (динамику) системы. *Связь* – это ограничение степени свободы элементов, т.е. элементы, вступая в связь друг с другом, утрачивают часть своих свойств, которыми они обладали в свободном состоянии.

Связи в конкретных системах могут быть одновременно охарактеризованы несколькими признаками.

Характеристика связей:

1. По направлению связи делят на направленные и ненаправленные;
2. По силе связи делят на сильные и слабые;
3. По характеру связи подразделяют на связи подчинения (часть-целое), связи порождения (т.е. причинно-следственные связи), равноправные связи, связи управления;
4. По месту приложения связи подразделяют на внутренние и внешние;
5. По направленности процессов в системе в целом или в отдельных ее подсистемах связи делят на прямые и обратные.

Очень важную роль в моделировании систем играет понятие обратной связи.

Обратная связь может быть положительной, т.е. сохраняющей тенденции происходящих в системе изменений того или иного выходного параметра, и отрицательной, т.е. направленной против сохранения требуемого значения этого параметра.

Обратная связь является основой саморегулирования, развития систем, приспособления их к меняющимся условиям существования.

Теоретически, для того чтобы система не распалась на части, суммарная сила внутренних связей (т.е. связей между элементами системы) должна быть больше, чем суммарная сила внешних связей (т.е. связей между элементами системы и элементами среды).

Среда – это совокупность всех объектов, изменение свойств которых влияет на систему, а также тех объектов, чьи свойства меняются в результате поведения системы.

При исследовании объекта требуется выяснить, что собой представляет объект, что в нем обеспечивает выполнение поставленной цели. При этом цель рассматривают как заранее мыслимый результат сознательной деятельности человека. В этих случаях систему отображают путем расчленения на подсистемы, элементы со взаимосвязями.

При этом любую экономическую систему рассматривают как открытую, постоянно взаимодействующую со средой, учитывая этот факт при исследовании. Но при этом невозможно учесть все объекты, не включенные в систему и отнесенные к среде. Например, среда – это то окружение, в котором действует предприятие (т.е. система). Среда состоит главным образом из участников рыночных отношений. От их поведения, целевых установок и интересов в большей или меньшей степени зависят результаты хозяйственной и коммерческой деятельности предприятия. Все факторы среды подразделяют на управляемые и неуправляемые.

При этом при исследовании систем отражают лишь наиболее существенные элементы, подсистемы и связи, которые мало меняются при текущем функционировании системы и обеспечивают существование системы и ее основных свойств, т.е. организованность системы, устойчивую упорядоченность ее элементов и связей. Определенные взаимосвязи, взаиморасположение составных частей системы, ее устройство отражает структура системы (в переводе с латинского означает строение, расположение, порядок). Одна и та же система может быть представлена по-разному в зависимости от стадии познания объекта, от его цели.

С одной стороны, *элементом* считается структурная единица, способная к относительно самостоятельному осуществлению определенной операции, не подлежащая дальнейшему расчленению на части на данном уровне анализа подсистемы.

С другой стороны, *элемент* подсистемы или системы представляет собой динамическую ячейку, изменяющую с течением времени свое

состояние под воздействием внешних и внутренних факторов, воспринимающую входные и выдающую выходные сигналы в процессе взаимодействия с другими элементами подсистемы.

Элементы характеризуются следующими особенностями.

1. В каждый данный момент времени состояние элемента может быть количественно описано с помощью некоторой величины, называемой его мгновенной характеристикой. Например, курс акций компании на фондовой бирже составил 100 долларов на 01.07.2011 года.

2. Мгновенная характеристика элемента изменяется с течением времени по определенным законам функционирования. Например, колебание курса акций компании на фондовой бирже за 3 месяца составило 25 %.

3. Характер изменения мгновенной характеристики элемента может описываться вероятностным процессом. Например, неблагоприятные погодные условия в Латинской Америке могут повлиять на мировые цены на кофе.

Известно много различных вариантов *классификации систем*.

Количественно все подсистемы некоторой системы подразделяются на моноподсистемы (которые характеризуются одним свойством, одним элементом, одной связью) и полиподсистемы (характеризующиеся многими свойствами, связями).

По составу системы подразделяются на статические и динамические. Для статической системы характерно то, что она находится в состоянии покоя, ее состояние с течением времени остается постоянным. Динамическая система изменяет свое состояние во времени. Динамические системы подразделяются на функционирующие и развивающиеся. *Функционирующие* – это те, у которых процесс перехода из состояния в состояние не сопровождается сменой качества, цели. *Развивающиеся* – это те, у которых изменение состояния приводит к смене качества.

Структурно, т.е. по характеру взаимоотношений между элементами системы, а также между системой и средой системы могут подразделяться на:

- 1) открытые и закрытые;
- 2) детерминированные и вероятностные;
- 3) простые и сложные.

Системы делятся на открытые и закрытые по характеру их взаимоотношений со средой. Большинство систем открытые, так как

они постоянно обмениваются энергией, информацией со средой, например, открытая рыночная система. Система называется закрытой, если в нее не поступает и из нее не выделяется энергия или информация.

Система называется детерминированной, если ее поведение полностью объяснимо и предсказуемо на основе информации об элементах системы и отношениях между ними. Для вероятностной системы (т.е. случайной, стохастической) знание элементов и отношений между ними в данный момент времени позволяет только предсказать вероятность нахождения системы в том или ином состоянии в последующие моменты времени.

Характерных черт «сложности» много и до сих пор еще нет общепринятого определения понятия «сложная система».

Несмотря на различия выделенных систем, все они обладают общими свойствами или *характеристиками*. Системы характеризуются следующими свойствами:

1. Свойство целостности проявляется через взаимодействие элементов системы в соответствии с целью ее функционирования. Свойство целостности характеризуется появлением качественно новых характеристик у системы, не присущих ее элементам.

2. Свойство связанности системы проявляется в форме упорядоченности отношений между элементами определенной внутренней структуры.

3. Свойство разнообразия системы зависит от числа элементов системы, возможных состояний каждого элемента и вероятности этих состояний. Каждый элемент обладает разными свойствами и проявляет их по-разному. Целенаправленное функционирование системы возможно только благодаря ограничению разнообразия элементов в силу их взаимодействия между собой.

4. Свойство сложности системы зависит от ее величины (числа элементов, образующих систему, степени разветвленности внутренней структуры, характера функционирования (т.е. одноцелевое или многоцелевое)).

5. Свойство организованности системы проявляется в изменении соотношения между нарастающей сложностью и совершенствованием структуры. Благодаря совершенствованию структуры и организованности повышается управляемость системы.

Принципиальной *особенностью экономических систем* является участие в них человека, как пользователя ресурса труда, так и носителя

и преобразователя информации. Но в то же время человек стоит над экономической системой, определяя цель ее функционирования, что приводит к проявлению у экономических систем особых свойств.

1. Изменчивость отдельных параметров системы и стохастичности ее поведения, так как масштабы агропромышленного производства как управляемой системы несравненно больше, чем любой управляемой технической системы. Причем с развитием производительных сил параметры системы изменяются, что вызывает необходимость исследования новых закономерностей развития производства и использования их в управлении.

2. Уникальность и непредсказуемость поведения системы в конкретных экономических условиях. Благодаря наличию активного элемента, т.е. человека, у системы проявляется «свобода воли». Но в то же время проявляется наличие у нее предельных возможностей, которые определяются ограниченными ресурсами (т.е. элементами, их свойствами).

3. Способность изменять свою структуру, сохраняя целостность, и формировать варианты поведения. Способность адаптироваться к изменяющимся условиям.

4. Способность и стремление к целеобразованию. В отличие от технических систем, которым цели задаются извне, в экономических системах цели формируются внутри системы.

Таким образом, все выявленные особенности экономических систем необходимо учитывать при моделировании, которое является основным способом исследования систем.

1.3. Модели исследования операций (понятие модели, классификация моделей, принципы построения экономико-математических моделей, типовые модели исследования операций)

Для изучения процессов управления производством, нахождения наилучшего решения хозяйственной ситуации в конкретных экономических условиях создаются модели.

Модель позволяет имитировать поведение системы в различных условиях, включая и такие, которые в действительности редко встречаются или сопряжены с большими затратами ресурсов или риском. В общем смысле слова *модель* – это некоторый аналог той системы, которой мы должны управлять, получая знания из исследования данного аналога.

По своей природе модели могут быть физическими и

математическими. Физические модели похожи на оригинал по физической природе, но отличаются от оригинала размерами, скоростями и т.д. Математические – это те модели, которые не похожи на оригинал, но с помощью математических уравнений или неравенств описывают протекающие в оригинале экономические процессы. Математическая модель – это способ описания операции, позволяющий исследовать ее математическими методами. При этом под операцией понимают управляемое мероприятие, направленное исследователем на выбор стратегии (параметров, характеризующих изучаемую систему), позволяющей достичь поставленную цель.

В операционных исследованиях применяются только математические модели.

Математическая модель – это концентрированное выражение наиболее существенных взаимосвязей и закономерностей поведения исследуемой системы, записанное в математической форме. Исследование систем на их моделях и перенесение полученных знаний на оригинал при управлении его поведением называется *моделированием*.

Классификация моделей. По временным характеристикам (т.е. по периоду планирования) все модели делятся на:

- 1) долгосрочные (срок планирования – 5–10 лет);
- 2) среднесрочные (период планирования – 3–5 лет);
- 3) краткосрочные (1–3 года);
- 4) модели оперативного планирования (до 1 года).

По характеру взаимосвязей компонентов модели делятся на:

- 1) детерминированные – это модели, в которых результат полностью и однозначно определяется набором независимых переменных.
- 2) стохастические – это модели, которые описывают случайные процессы, подчиненные законам теории вероятности.

В зависимости от учета фактора времени модели делятся на:

- 1) статические;
- 2) динамические.

Модель носит динамический характер, если в процессе решения задачи ряд технико-экономических коэффициентов изменяет свое значение. Если же при решении задачи технико-экономические коэффициенты остаются неизменными, то имеем статическую модель.

В зависимости от уровня управления системами в АПК модели делятся на:

1) межотраслевые, т.е. модели, описывающие взаимоотношения между отдельными отраслями;

2) отраслевые, т.е. модели, описывающие взаимоотношения внутри отрасли;

3) региональные;

4) внутривозрастные.

Исходя из применяемого математического аппарата модели делятся на:

1) статистические, которые описывают зависимость результата производства от влияния на него одного или нескольких факторов (например, производственные функции).

2) балансовые, т.е. модели, представляющие собой систему балансов производства и распределения продукции, которая записывается в виде шахматных матриц;

3) оптимизационные модели базируются на методах математического программирования. Данные модели представляют собой систему уравнений и неравенств, подчиненную целевой функции, т.е. цели решения задачи.

4) игровые модели – это модели в виде игры, описывающие конфликтную ситуацию, анализ которой осуществляется по определенным правилам, в результате чего определяется наилучшая стратегия игрока, т.е. такие его действия, которые при многократном повторении игры обеспечивают данному игроку максимально возможный средний выигрыш или минимально возможный средний проигрыш, причем каждому участнику игры ясно, что результат игры зависит не только от него, но и от действий партнера, т.е. он принимает решения в условиях неопределенности;

5) имитационные модели получают широкое распространение, так как их решение выполняется на персональном компьютере в диалоговом режиме, с их помощью проводят имитацию. Модель представляет собой программу для персонального компьютера, а эксперимент над ней состоит в наблюдении за результатами расчетов по этой программе при различных значениях вводимых переменных;

6) модели сетевого планирования и управления служат для управления производственной деятельностью коллективов людей, выполняющих комплекс взаимосвязанных работ.

7) модели массового обслуживания применяются для улучшения качества обслуживания, т.е. сокращения времени ожидания обслуживания, сокращения очередей, уменьшения стоимости обслуживания.

8) модели управления запасами служат для определения оптимальных значений уровня запасов (точки заказа) и размера заказа.

По степени детализации модели делятся на:

1) структурные;

2) развернутые.

Развернутая модель – это сама задача, которая описывает функционирование конкретной системы, она составлена на основе конкретного цифрового материала.

Структурная же модель описывает систему в виде символов и математических выражений, каждое выражение структурной модели объединяет группу однородных ограничений развернутой модели.

Принципы построения моделей. Так как реальные процессы или объекты АПК представляют собой сложные динамические системы, то для описания их функционирования в современных условиях чаще всего используют интегрированную систему моделей, представляющую собой совокупность логически, информационно и алгоритмически связанных моделей, которые отражают экономические, организованные и технологические процессы воспроизводства в моделируемой системе.

При построении системы моделей необходимо соблюдать следующие принципы.

1. Принцип развития требует постоянного совершенствования системы моделей, включения в ее состав новых моделей, использование которых становится необходимым по мере совершенствования планирования и управления.

2. Принцип единства означает представление системы моделей в единой структуре блоков, которые взаимосвязаны между собой логически, информационно и алгоритмически.

3. Принцип относительной автономии позволяет выделить из общей системы моделей относительно самостоятельные модели, результаты решения по которым можно внедрять в производство, не ожидая расчетов по всей системе моделей.

4. Принцип соответствия и адаптации означает соответствие системы моделей реальной действительности.

5. Принцип ориентации на выходные плановые показатели означает, что система моделей и ее решение должны обеспечивать выход на утвержденные плановые показатели.

6. Принцип разнообразия состоит в том, что для адекватного отражения действительности в состав системы моделей должны быть

включены разнообразные модели (статистические, оптимизационные, сетевые и т.д.).

7. Принцип взаимного дополнения требует, чтобы модели, различающиеся по своему функциональному назначению, дополняли друг друга и были увязаны в единую систему логически, алгоритмически и информационно.

Наибольшее распространение получили следующие *типы моделей исследования операций*:

1) задачи теории управления запасами составляют самый распространенный тип задач исследования операций. В зависимости от условий задачи управления запасами делятся на 3 группы:

– моменты поставок или оформление заказов на пополнение запасов фиксированы. Необходимо определить объемы производимой или закупаемой партии запасов;

– объемы производимой или закупаемой партии заказов фиксированы. Следует определить моменты оформления заказов;

– моменты оформления заказов и объемы производимых или закупаемых партий не фиксированы. Необходимо их определить;

2) задачи распределения также решаются методами исследования операций. Они делятся на 3 основные группы:

– к первому типу относятся задачи такого распределения ресурсов по работам, при котором достигается наибольший эффект (максимум прибыли или минимум издержек) (транспортная задача). Эти задачи усложняются, если для выполнения некоторых работ требуется более одного вида ресурсов или если один и тот же ресурс может пойти для выполнения нескольких работ;

– второй тип задач связан с распределением ограниченных ресурсов, которых не хватает для выполнения всех наличных работ. При этом могут быть использованы следующие подходы: а) заявки на ограниченные ресурсы урезаются пропорционально величине заявленной потребности; б) ограниченные ресурсы распределяются путем последовательного удовлетворения различных направлений в порядке убывания их значимости, определенной экспертами; в) продукция распределяется с учетом потерь от дефицитности. При этом строится функция дефицитности, выражающая потери, которые несет система при недопоставке продукции;

– в третьем типе задач имеется возможность в некоторой степени регулировать состав ресурсов;

3) задачи теории массового обслуживания рассматривают вопросы

образования и функционирования очередей.

4) задачи теории расписаний изучают порядок следования операций.

5) в задачах сетевого планирования рассматриваются моменты начала и окончания работ всего комплекса или программы. Требуется так выбрать сроки начала работ и сроки их выполнения, чтобы общие затраты по реализации работ всей программы были бы самые наименьшие. С помощью задач этого типа также можно определить потребность в ресурсах при заданных сроках выполнения работ.

6) при решении ряда задач приходится анализировать ситуации, в которых две или более конфликтующие стороны преследуют противоположные цели. Причем результат одной из сторон зависит от того, какие действия выберет противник. Такие задачи решаются с помощью теории игр. Цель игры – выработать рекомендации по рациональным действиям каждого из противников в ходе конфликтной ситуации;

7) задачи смешанного типа представляют собой системы взаимосвязанных моделей различного типа.

Но данная классификация задач исследования операций не является окончательной. Детализация некоторых типов задач приводит к появлению новых задач. Некоторые же типы задач объединяются и решаются совместно.

1.4. Этапы исследования операций

Для исследования системы необходимо выполнить следующие этапы.

1. *Постановка (формулировка) задачи.* От решения вопросов первого этапа во многом зависит качество получаемых результатов. Постановка задачи включает решение следующих вопросов:

а) выбор и формулировка цели задачи, решение которой наиболее важно в данный момент времени;

б) выбор периода планирования (т.е. краткосрочный, среднесрочный, долгосрочный или текущего планирования);

в) определение объемов основных ресурсов моделируемой системы и тех параметров, которые оказывают влияние на функционирование системы;

г) выявление возможных альтернатив решения применительно к исследуемой конкретной ситуации.

2. *Построение математической модели функционирования*

системы. На этом этапе изучаются уже имеющиеся математические модели, которые позволяют описать данную проблему. Для построения модели надо определить множество известных и неизвестных параметров, которые необходимы для записи зависимостей исследуемой операции. Цель решения задачи выражается с помощью критерия эффективности. При построении модели следует учитывать только основные факторы и отбрасывать второстепенные. При этом математическое описание системы должно отражать все ее основные закономерности и множество возможных стратегий функционирования.

3. *Определение алгоритма решения модели*. Алгоритмом называется система правил, указывающих, как и в какой последовательности эти правила применять к исходным данным модели, чтобы получать ее решение. Здесь может быть три варианта: а) модель известна и известен ее алгоритм решения; б) модель новая, но ее можно решить, сведя к какой-либо известной модели; в) модель новая и алгоритм решения ее неизвестен и ее, например, можно решить методами имитационного моделирования. На этом этапе, кроме нахождения оптимального решения, целесообразно провести анализ модели на чувствительность, который показывает, как изменяется решение задачи при изменении значений входной информации системы.

4. *Проверка адекватности модели и экономическая интерпретация решения задачи*. Проверка адекватности модели состоит в сопоставлении оптимального решения при заданных входных параметрах с характеристиками системы, которые она имела в прошлом.

5. *Внедрение результатов решения в производство*. На этом этапе может возникнуть необходимость корректировки модели и улучшения полученного решения.

1.5. Понятие о критерии эффективности

Исследователь операций должен иметь возможность оценивать различные варианты функционирования системы, соответствующие разным стратегиям. Для оценки этих стратегий используется критерий эффективности. Под *критерием эффективности* понимается экономическая категория, выражающая предельную меру экономического эффекта принимаемого хозяйственного решения. Критерий эффективности должен обладать следующими свойствами:

- быть простым, т.е. не содержать большого количества факторов;
- быть представительным, т.е. отражать основную цель поставленной задачи;
- быть критичным, т.е. сильно реагировать на изменения параметров исследуемых стратегий функционирования системы;
- быть единственным, т.е. каждой модели соответствует единственный критерий эффективности.

Критерий эффективности подразделяется на глобальный и локальный.

Глобальный критерий эффективности является народнохозяйственным, он вытекает из действия основного экономического закона любой системы, т.е. интенсивного использования ресурсов с целью максимального производства продукции, снижения издержек производства, создания условий для нормального функционирования общества.

Локальный – это частный критерий эффективности, он связан с детализацией глобального. Этот критерий используют для решения задач более низкого уровня. Требуется, чтобы локальный критерий эффективности учитывал основные положения глобального критерия эффективности и не противоречил ему.

Критерий эффективности может быть выражен как количественно, так и качественно. Математически критерий эффективности, соответствующий качественной цели, записывается следующим образом:

$$\Phi = \begin{cases} 1, & \text{если цель достигается;} \\ 0, & \text{если цель не достигается.} \end{cases}$$

В литературе такие критерии называются порядковыми или ранговыми критериями эффективности. Они определяют, какая стратегия лучше или хуже других, но не поясняют насколько.

В основном критерий эффективности носит количественный характер, состоящий в стремлении к увеличению (максимизации) или уменьшению (минимизации) показателя, характеризующего уровень достижения поставленной цели и зависящего от рассматриваемых стратегий и входных параметров модели.

Количественным выражением критерия эффективности в экономико-математических моделях является целевая функция. Особенность ее в том, что она однозначна. Это означает, что если при одних и тех же условиях задачи изменить целевую функцию, то получим новое ее решение. В силу специфики своего развития

агропромышленное производство многокритериально, т.е. общество заинтересовано в получении максимальной прибыли, в росте производительности труда, снижении издержек и т.д.

Возникает необходимость поиска многих решений, отвечающих разным критериям эффективности. Многоцелевой характер критерия эффективности чаще всего выражается в модели следующим образом:

1) применяют *прием ведущего критерия*, т.е. один из наиболее предпочтительных критериев используется в качестве целевой функции задачи, а требования всех оптимальных учитываются при составлении ограничений задачи;

2) *прием последовательных уступок*. Сущность данного приема состоит в замене многокритериальной задачи оптимизации последовательностью однокритериальных задач. Вначале исследуемые критерии ранжируются в порядке убывания их значимости. Задача решается с первым по значимости критерием f_1 и определяется его экстремальное значение f_1^* . Затем назначается величина допустимого отклонения критерия от его оптимального значения, т.е. уступка Δf_1 , и решается задача еще раз, но уже со вторым по значимости критерием f_2 , при условии, что отклонение первого критерия от его оптимального значения не превзойдет величины уступки. Далее назначается уступка для второго критерия и задача решается с третьим критерием и т.д. Таким образом, решение каждой исследуемой задачи основано на решении предыдущей, так как оно содержит дополнительные ограничения, характеризующие величину уступки по критериям;

3) используют *прием скаляризации векторного критерия* (приведения его к скаляру), которая может быть осуществлена следующими способами:

– аддитивная свертка критериев –

$$\Phi = \sum_{i=1}^n v_i f_i ;$$

– мультипликативная свертка –

$$\Phi = \prod_{i=1}^n f_i^{v_i} ;$$

– логарифмически-аддитивная свертка –

$$\ln \Phi = \sum_{i=1}^n v_i \ln f_i,$$

где f_i – локальный критерий вида i , $i = \overline{1, n}$;

v_i – вес критерия вида i , $i = \overline{1, n}$.

При этом $\sum_{i=1}^n v_i = 1$ и $v_i \geq 0$, $i = \overline{1, n}$.

Вектор весов критериев $v_i = (v_1, v_2, \dots, v_n)$ обычно определяется на основе экспертных оценок.

Сущность экспертных методов заключается в выработке коллективного мнения группы специалистов (экспертов). Формируется экспертная группа из специалистов в конкретной области. Путем анонимного анкетирования может быть сделан отбор специалистов, которые, по мнению большинства, не могут выступать экспертами в данной области. Опрос экспертов является существенным элементом получения качественной информации. В зависимости от целей и методов обработки результатов опроса применяются различные способы организации работы экспертов, касающиеся их взаимных контактов, анонимности опроса, дозирования информации и т.п. Коллективно выбираются критерии эффективности, характеризующие цель исследуемой системы. В практике наиболее часто применяются следующие методы установления весов критериев.

1. *Метод непосредственной оценки* состоит в том, что эксперт каждому критерию присваивает определенную оценку (балл), например от 1 до 10. Балльные оценки нормируются, для этого определяется сумма оценок, выставленных каждым экспертом, по всем критериям, а затем каждая из оценок делится на полученную сумму. Далее нормированные оценки всех экспертов по каждому критерию суммируются, а полученная сумма делится на число экспертов.

Таким образом, вес i -го критерия определяется по формуле

$$v_i = \frac{\sum_{j=1}^m B_{ij}}{m},$$

где v_i – вес критерия вида i ;

i – номер критерия, $i = \overline{1, n}$;

j – номер эксперта, $j = \overline{1, m}$;

B_{ij} – балл, присвоенный i -му критерию j -м экспертом;

B_j – сумма баллов, присвоенных всем критериям j -м экспертом.

2. *Метод последовательных сравнений* позволяет не только оценить вес каждого критерия, но и выявить зависимости между их количественными оценками. Эксперт должен оценить исследуемые критерии по их относительной важности. При этом наиболее важному критерию дается оценка $v_i = 1$, а остальным критериям присваиваются оценки v_i в пределах от 0 до 1 в зависимости от относительной важности критериев.

Затем эксперт сравнивает значимость более важного критерия (критерия, получившего оценку 1) с комбинацией остальных критериев. Если этот критерий более значим, то его оценка увеличивается так, чтобы она была больше, чем суммарная оценка остальных критериев:

$$v_1 > \sum_{i=2}^n v_i .$$

Далее аналогичную процедуру проделывают со вторым и последующими критериями, получившими более низкие оценки. Последовательное сравнение продолжается до $(n-1)$ -го критерия.

Информация, полученная от экспертов, может считаться достаточно надежной только при условии хорошей согласованности оценок экспертов.

Степень согласованности оценок двух экспертов или двух групп экспертов характеризуется *коэффициентом ранговой корреляции Ч. Спирмена* (p):

$$p = 1 - \left(\frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n \cdot (n^2 - 1)} \right),$$

где $\bar{x}_i - \bar{y}_i$ – разность между рангами (оценками);

n – число критериев.

Коэффициент корреляции рангов Спирмена (p) равен $+1$, если все ранги совпадают, и равен -1 , если ранговые ряды имеют обратное направление. Чем ближе p к единице, тем более согласованы решения. Обычно согласованность считается удовлетворительной при $p=0,85-0,9$ и хорошей при $p \geq 0,95$.

Для оценки согласованности мнений группы из m экспертов применяется коэффициент конкордации Кендалла W (общий коэффициент ранговой корреляции для группы, состоящей из m экспертов).

В случае отсутствия равных рангов (оценок) в оценках любого из экспертов коэффициент конкордации определяется по формуле

$$W = \frac{12 \sum_{j=1}^n \left(\sum_{i=1}^m x_{ij} - \frac{1}{2} m(n+1) \right)^2}{m^2(n^3 - n)},$$

где m – число экспертов;

n – число критериев;

\bar{x}_{ij} – оценка i -м экспертом j -го критерия.

Коэффициент конкордации принимает значения в интервале от 0 до 1. При отсутствии согласованности мнений экспертов $W = 0$, при полной согласованности $W=1$. Практически, согласованность считается удовлетворительной, если $W \geq 0,5$, и хорошей, если $W \geq 0,7$.

Лекция 2.

Тема 2. ПРИНЯТИЕ РЕШЕНИЙ И ЭЛЕМЕНТЫ ТЕОРИИ ИГР

- 2.1. Предмет и задачи теории игр. Классификация игр
- 2.2. Статистические игры
- 2.3. Решение матричных игр в чистых стратегиях
- 2.4. Решение матричных игр геометрическим способом
- 2.5. Решение матричных игр в смешанных стратегиях
- 2.6. Позиционные игры
- 2.7. Биматричные игры
- 2.8. Кооперативные игры

2.1. Предмет и задачи теории игр. Классификация игр

На практике часто появляется необходимость согласования интересов двух или более разумных противников, каждый из которых

стремится оптимизировать свои решения за счет других. Теория, занимающаяся принятием решения в условиях конфликта, получила название теории игр. Возникновение теории игр как науки относится к 1944 году, когда вышла монография Джона фон Неймана и Оскара Моргенштерна «Теория игр и экономическое поведение». В ней была обоснована возможность анализа различных экономических вопросов с помощью игровых моделей. В 1950 году Джон Нэш ввел понятие ситуации равновесия как метода решения бескоалиционных игр. Равновесие по Нэшу гласит, что ситуация, образованная в результате выбора всеми игроками некоторых своих стратегий, является равновесной, если ни одному из игроков невыгодно изменять свою стратегию при условии, что остальные игроки придерживаются равновесных стратегий. В настоящее время главное внимание в теории игр уделяется ее экономическим приложениям.

Теория игр – это математическая теория конфликтных ситуаций, разрабатывающая рекомендации по наиболее рациональному образу действий каждого из участников в ходе конфликтной ситуации, т.е. таких действий, которые обеспечивали бы игроку наилучший результат при многократном повторении игры. При этом *игра* рассматривается как упрощенная математическая модель конфликтной ситуации, отличающаяся от реального конфликта тем, что ведется по определенным правилам, которые устанавливают:

- выбор действий игроков на каждом этапе игры;
- информацию, которой обладает каждый игрок при осуществлении своих выборов;
- выигрыши или проигрыши каждого игрока после завершения игры.

Суть игры состоит в том, что каждый из участников принимает такие решения в развивающейся конфликтной ситуации, которые, как он полагает, могут обеспечить ему наилучший ход игры, т.е. величина выигрыша зависит от стратегии, применяемой игроком.

Стратегия – это совокупность правил, однозначно определяющих последовательность действий игрока в каждой конкретной ситуации, складывающейся в процессе игры.

При этом оптимальной считается та стратегия, которая обеспечивает игроку при многократном повторении игры максимально возможный выигрыш или минимально возможный проигрыш.

Принятие решений может происходить в условиях определенности, неопределенности и риска.

Конфликтные ситуации, встречающиеся на практике, порождают различные виды игр.

В зависимости от количества игроков игры подразделяются на парные (с двумя игроками) и множественные (имеющие не менее трех игроков).

По количеству стратегий игры делятся на конечные (где каждый из игроков имеет конечное число возможных стратегий) и бесконечные (в которых хотя бы один из игроков имеет бесконечное число возможных стратегий).

По взаимоотношениям между конфликтующими сторонами игры подразделяются на: кооперативные, коалиционные и бескоалиционные. Бескоалиционными считают такие игры, в которых игроки не имеют права вступать в соглашения. Если же участники конфликта вступают в соглашения и создают коалицию, то данную ситуацию описывают коалиционной игрой. Кооперативная игра описывает конфликт, в котором заранее определены группы участников, т.е. коалиции.

По характеру выигрышей игры делятся на игры с нулевой и ненулевой суммой. В игре с нулевой суммой сумма выигрышей равна сумме проигрышей всех игроков, т.е. общая сумма перераспределяется между игроками, но величина ее не изменяется. Такие игры относятся к классу антагонистических игр. Игра, в которой вносится взнос за право участия в ней, является игрой с ненулевой суммой.

В зависимости от вида функции выигрышей игры подразделяются на: матричные, биматричные, непрерывные, выпуклые, сепарабельные и т. д.

По количеству ходов игры можно подразделить на одношаговые (которые заканчиваются после одного хода каждого игрока) и многошаговые (в которых игрок делает более одного хода).

По информированности сторон различают игры с полной (т. е. каждый игрок на каждом ходу знает ранее примененные другими игроками стратегии) и неполной информацией (т.е. игроку не все стратегии предыдущих ходов других игроков известны).

По степени неполноты информации игры подразделяются на статистические (в условиях частичной неопределенности) и стратегические (в условиях полной неопределенности).

2.2. Статистические игры

В отдельную группу игр выделяют статистические игры. Их особенностью является то, что сознательный игрок A (его еще

называют статистиком) заинтересован в наиболее выгодном для него исходе игры и играет он против игрока B , который совершенно безразличен к результату игры (его еще называют природой и обозначают P). Стратегии игрока P обозначают совокупность внешних условий, в которых игрок A выбирает свою стратегию. Поэтому при решении статистической игры находят только наилучшие рекомендации для игрока A .

При поиске оптимальных стратегий для игрока A в условиях неопределенности и риска обращаются к различным критериям. Так как критерии формируются на основе здравого смысла, интуиции и практической целесообразности, то они помогают оценить принимаемое решение с различных позиций.

Используется несколько критериев определения оптимальной стратегии игрока.

1. По критерию Лапласа оптимальной является та стратегия, которая обеспечивает максимальную среднюю прибыль или минимальный средний риск.

Для выбора оптимальной стратегии по второму случаю предварительно рассчитаем матрицу риска. Риск – это разность между результатом, который можно получить, если знать состояние «природы», и результатом, который будет получен при j -й стратегии игрока.

2. По критерию Байеса оптимальная стратегия игрока характеризуется максимальным математическим ожиданием выигрыша или минимальным математическим ожиданием риска и определяется с учетом вероятностей всех возможных исходов.

3. По критерию Вальда оптимальной будет стратегия, которая в наихудших условиях обеспечивает наибольшую прибыль. Этот критерий опирается на принцип «наибольшей осторожности».

4. При использовании критерия Сэвиджа строят матрицу рисков. Если в платежной матрице игры элементы характеризуют потери игрока, то для определения коэффициентов матрицы риска от элементов каждого столбца отнимают наименьший элемент этого столбца.

Оптимальной является стратегия с наименьшим максимальным риском.

5. Для проверки вышеизложенных выводов используют критерий Гурвица, по которому, если в качестве результата игры выступают

прибыль, полезность, доход и т.д., выбирают оптимальную стратегию по формуле

$$\max_j [\lambda \min_i a_{ij} + (1 - \lambda) \max_i a_{ij}],$$

где λ – коэффициент, характеризующий степень доверия, $0 \leq \lambda \leq 1$.

Значение λ определяется в зависимости от склонности игрока к пессимизму или к оптимизму.

Если элементы платежной матрицы игры представляют затраты или потери игрока, то по критерию Гурвица оптимальную стратегию игрока выбирают следующим образом:

$$\min_j [\lambda \min_i a_{ij} + (1 - \lambda) \max_i a_{ij}].$$

С помощью критерия Гурвица устанавливают баланс между случаями крайнего пессимизма и крайнего оптимизма, используя определенную величину коэффициента λ .

При $\lambda=0$ имеем критерий крайнего оптимизма. Если $\lambda=1$, то имеем критерий крайнего пессимизма:

Анализ игр по нескольким критериям позволяет более достоверно принять ту или иную стратегию игрока с наилучшей функцией выигрыша.

2.3. Решение матричных игр в чистых стратегиях

Рассмотрим игры, в которых у каждого из двух игроков A и B конечное число возможных действий, т.е. чистых стратегий. Пусть игрок A располагает m -чистыми стратегиями – A_1, A_2, \dots, A_m , а игрок B – n -чистыми стратегиями – B_1, B_2, \dots, B_n . Рассмотрим антагонистическую игру с нулевой суммой, в которой выигрыш одного игрока равен проигрышу другого, тогда число a_{ij} примем за выигрыш игрока A за счет игрока B или проигрыш игрока B .

Если известны значения a_{ij} для каждой пары (A_i, B_j) чистых стратегий, то можно составить матрицу игры, которая получила название платежной матрицы.

Если выигрыши выражаются отрицательными числами, то это означает, что фактически выигрывает игрок B , а игрок A проигрывает.

Игра протекает следующим образом: игрок A выбирает одну из строк платежной матрицы (т.е. свою чистую стратегию, характеризуемую числом α_i), а игрок B , не зная результата его выбора, выбирает один из столбцов (свою чистую стратегию, характеризуемую числом β_j). На пересечении строки и столбца будет

стоять элемент матрицы, определяющий выигрыш игрока A или проигрыш игрока B :

A_i	B_j				α_i
	B_1	B_2	...	B_n	
A_1	a_{11}	a_{12}	...	a_{1n}	α_1
A_2	a_{21}	a_{22}	...	a_{2n}	α_2
...
A_m	a_{m1}	a_{m2}	...	a_{mn}	α_m
β_i	β_1	β_2	...	β_n	

При поиске оптимальных стратегий игроки опираются на основной принцип теории игр – принцип осторожности, в соответствии с которым каждый игрок считает, что играет с очень умным противником, который может воспользоваться любой его ошибкой в своих интересах. Поэтому игрок A находит для каждой стратегии минимальный выигрыш $\alpha_i = \min_j a_{ij} (i = \overline{1, m})$, а затем из всех α_i выбирает наибольшее значение $\alpha = \max_i \alpha_i$, которое определяет чистую стратегию игрока A , ее называют максиминной стратегией, так как она определяется по формуле

$$\alpha = \max_i \min_j a_{ij}.$$

При этом число α называют нижней чистой ценой игры (максимином). Оно показывает, какой минимальный выигрыш может получить игрок A , правильно применяя свои чистые стратегии при любых действиях игрока B .

Игрок B , стремясь минимизировать проигрыш, тоже при выборе стратегии использует принцип осторожности. Он сначала выбирает по столбцу максимально возможный проигрыш $\beta_j = \max_i a_{ij} (j = \overline{1, n})$, а затем среди β_j минимальное значение $\beta = \min_j \beta_j$, которое определит его чистую стратегию, т.е. минимаксную стратегию, так как она выбирается по формуле

$$\beta = \min_j \max_i a_{ij}.$$

Число β называется верхней чистой ценой игры, или минимаксом. Оно показывает, какой максимальный проигрыш может быть у игрока B при правильном выборе стратегии независимо от действий игрока A .

Если в матричной игре нижняя и верхняя чистые цены игры совпадают, т.е. $\alpha = \beta$, то говорят, что игра имеет *седловую точку* и чистую цену игры, равную значению или седловому элементу платежной матрицы:

$$v = \max_i \min_j a_{ij} = \min_j \max_i a_{ij}.$$

2.4. Решение матричных игр геометрическим способом

Матричная игра описывает конфликт, отвечающий следующим условиям:

1. Конфликт определяется антагонистическим взаимодействием двух сторон (игроков), каждая из которых располагает лишь конечным числом возможных действий (стратегий);

2. Свои действия (стратегии) игроки предпринимают независимо друг от друга, т.е. каждый из игроков не имеет информации о действии, совершаемом другой стороной. Результат этих действий оценивается числом, равным элементу платежной матрицы, который определяет полезность (выгодность) сложившейся ситуации для одного из игроков (игрока A);

3. Каждый из игроков оценивает как для себя, так и для противника выгодность применения любой возможной ситуации, которая сложится в результате применения той или иной стратегии;

4. Действия конфликтующих сторон отличаются друг от друга лишь по степени полезности сложившейся ситуации, которая характеризуется величиной элементов платежной матрицы.

Если игра не имеет седловой точки, то она может быть решена в смешанных стратегиях. *Смешанная стратегия* игрока представляет собой полный набор его чистых стратегий при многократном повторении игры в одних и тех же условиях с заданными вероятностями.

Алгоритм геометрического решения игры.

1. Матрицу игры, если это необходимо и есть такая возможность, приводим к размерности $2 \times n$ или $m \times 2$. Это достигается путем сравнения между собой почленно элементов столбцов или строк платежной матрицы. При этом вычеркивается, если имеются, дублирующие и заведомо невыгодные стратегии.

2. Проверяем платежную матрицу на наличие седловой точки. Если она есть – решение игры найдено, если ее нет – то находим оптимальные смешанные стратегии игроков.

3. Строим прямые, соответствующие стратегиям первого (второго) игрока.

4. Определяем нижнюю (верхнюю) границу выигрыша.

5. Находим две стратегии первого (второго) игрока, которым соответствуют две прямые, пересекающиеся в точке с максимальной (минимальной) ординатой.

6. Определяем цену игры и оптимальные смешанные стратегии игроков.

2.5. Решение матричных игр в смешанных стратегиях

Если игра не имеет седловой точки, то она должна быть решена в смешанных стратегиях. Согласно основной теореме игр – теореме Джона фон Неймана – каждая матричная игра двух лиц с нулевой суммой имеет решение в смешанных стратегиях. Это значит, что игроки будут применять не одну чистую стратегию, а несколько, и будут смешивать их случайным образом. Пусть $p_i (p_1, p_2, \dots, p_m)$ вероятности, с которыми игрок A использует в ходе игры свои чистые стратегии A_1, A_2, \dots, A_m .

При этом $p_i \geq 0 (i = \overline{1, m})$, $\sum_{i=1}^m p_i = 1$ и

$q_j (q_1, q_2, \dots, q_n)$ – вероятности, с которыми игрок B использует в ходе игры свои чистые стратегии B_1, B_2, \dots, B_n . При этом $q_j \geq 0 (j = \overline{1, n})$ и

$$\sum_{j=1}^n q_j = 1.$$

При использовании смешанных стратегий игра приобретает случайный характер, случайной становится и величина выигрыша игрока A (проигрыша игрока B), которая является функцией от смешанных стратегий (обозначим их через \bar{p}, \bar{q}) \bar{p} и \bar{q} и рассчитывается по формуле

$$f(\bar{p}, \bar{q}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} p_i q_j.$$

По аналогии с решением матричных игр в чистых стратегиях нижней ценой игры будет число α , которое определяется по формуле

$$\alpha = \max_{\bar{p}} \min_{\bar{q}} f(\bar{p}, \bar{q}).$$

Верхней ценой игры будет число β , которое вычисляется по формуле

$$\beta = \min_{\bar{q}} \max_{\bar{p}} f(\bar{p}, \bar{q}).$$

Цена игры при этом будет равна

$$\nu = f(\bar{p}^*, \bar{q}^*) = \alpha = \beta, \text{ или}$$

$$\nu = \max_{\bar{p}} \min_{\bar{q}} f(\bar{p}, \bar{q}) = \min_{\bar{q}} \max_{\bar{p}} f(\bar{p}, \bar{q}).$$

При этом \bar{p}^*, \bar{q}^* называются оптимальными смешанными стратегиями игроков A и B . Цена игры ν характеризует средний выигрыш первого игрока или средний проигрыш второго игрока при использовании обоими игроками смешанных стратегий.

Практически матричные игры в смешанных стратегиях сводят к решению двух взаимно симметричных двойственных задач линейного программирования.

Так как оптимальная смешанная стратегия игрока A определяется по формуле $\alpha = \max_{\bar{p}} \min_{\bar{q}} f(\bar{p}, \bar{q})$, то пусть $\nu = \min_{\bar{q}} f(\bar{p}, \bar{q})$. Так как при оптимальной стратегии средний выигрыш не меньше ν (цены игры) при любой стратегии противника, то математическое ожидание цены игры по каждой стратегии для игрока A будет равно:

$$1) \sum_{i=1}^m a_{ij} p_i \geq \nu, j = \overline{1, n}$$

при условии

$$2) p_i \geq 0, i = \overline{1, m}.$$

При этом требуется максимизировать выигрыш игрока A :

$$F_{\max} = \nu.$$

Разделив левые и правые части ограничений на ν , получим:

$$1) \sum_{i=1}^m a_{ij} \frac{p_i}{\nu} \geq 1, j = \overline{1, n};$$

$$2) \frac{p_i}{\nu} \geq 0, i = \overline{1, m}.$$

Обозначим $\frac{p_i}{\nu} = x_i$.

Получим:

$$1) \sum_{i=1}^m a_{ij} x_i \geq 1, j = \overline{1, n};$$

$$2) x_i \geq 0, i = \overline{1, m}.$$

Известно, что $\sum_{i=1}^m p_i = 1$. Разделив данное выражение на ν , получим

$$\sum_{i=1}^m \frac{p_i}{\nu} = \frac{1}{\nu}. \text{ Обозначив } \frac{p_i}{\nu} = x_i, \text{ получим } \sum_{i=1}^m x_i = \frac{1}{\nu}. \text{ Так как } F_{\max} = \nu,$$

то $F_{\max} = \nu$ равнозначно $F_{\min} = \frac{1}{\nu}$, т.е. $F_{\min} = \sum_{i=1}^m x_i$.

Таким образом, задачу линейного программирования для игрока A запишем следующим образом:

$$1) \sum_{i=1}^m a_{ij} x_i \geq 1, j = \overline{1, n};$$

$$2) x_i \geq 0, i = \overline{1, m}.$$

$$F_{\min} = \sum_{i=1}^m x_i.$$

Аналогично и для игрока B .

Таким образом, задачу линейного программирования для игрока B запишем следующим образом:

$$1) \sum_{j=1}^n a_{ij} y_j \leq 1, i = \overline{1, m};$$

$$2) y_j \geq 0, j = \overline{1, n}.$$

$$F_{\max} = \sum_{j=1}^n y_j.$$

2.6. Позиционные игры

В реальной действительности игроки наблюдают за динамикой конфликта, имеют информацию о фактически складывающейся обстановке. Конфликты, в которых детализируется поведение участников конфликта во времени, можно моделировать позиционными играми.

Позиция – это элемент дерева игры. В каждой позиции делает ход один из игроков (1, 2 или 0), при этом цифрой «0» обозначен фиктивный игрок (природа), который делает ход согласно заданному распределению вероятностей.

Дерево игры изображается графически с помощью плоской фигуры, состоящей из конечного числа вершин (рис. 2.6.1). Вершины, или позиции, соединяются ребрами. Ребра, соединяющие некоторую позицию с непосредственно следующей за ней, называются *альтернативами* этой позиции.

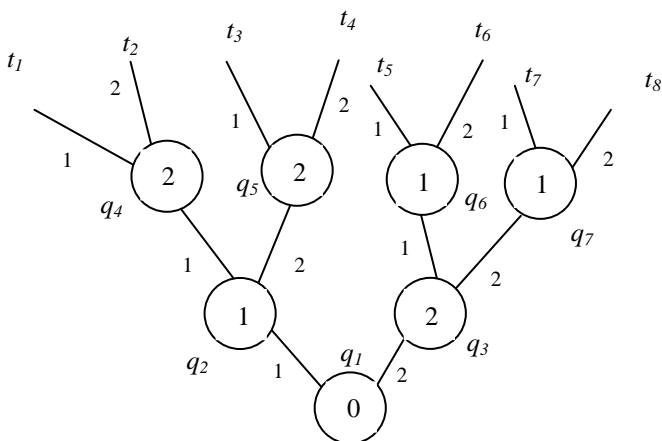


Рис. 2.6.1 Дерево игры

Альтернативы каждой позиции нумеруются натуральными числами по часовой стрелке (1 или 2).

Позиции, не имеющие последующих, называются *окончательными*, а остальные позиции считаются *неокончательными*. Построение диаграммы игры начинается с начальной позиции, в которой указывается, какой из игроков делает ход. Из вершины q_1 проводят ребра, соответствующие альтернативам игрока, который делает первый ход. Они соединяют вершину q_1 с вершинами (позициями) игроков, которые делают следующий ход, и так далее до тех пор, пока не будут обозначены окончательные вершины.

Каждая окончательная вершина характеризуется элементом t_i , который указывает, сколько выиграет игрок 1, если игра закончится в данной вершине.

Если в игре используется ход, осуществляемый не игроком, а случайным механизмом (природой), то позиции (или узлу), соответствующему данному ходу, присваивается номер «0».

Ветвью дерева называется ломаная линия, состоящая из отрезков дерева (т.е. ребер), идущая с нижней позиции последовательно через соответствующие позиции до вершины дерева. Каждая ветвь отображает партию игры. Для изображения необходимых сведений о сделанных выборах при определенных ходах игроков на дереве игры отмечают кружком, овалом и т.д. информационные множества позиций или узлов конкретного игрока. В каждое информационное множество входят те позиции (узлы), для которых соответствующий игрок не может точно указать, в какой точке дерева он находится, делая этот ход.

Таким образом, позиционная игра состоит из:

- дерева игры;
- функций, определяющих в каждой вершине сумму, которая должна быть уплачена игроку 1, если партия заканчивается в точке i ;
- набора чисел (количество которых равно числу игроков игры), указывающих в каждой позиции, какой из игроков делает очередной ход;
- сопоставления каждой позиции альтернатив, т.е. выборов конкретных игроков (обозначенных ребрами), и помеченных натуральными числами (1 и 2);
- разбивки позиций на информационные множества, которые должны удовлетворять следующим условиям:
 - а) все позиции, принадлежащие данному информационному множеству, относятся к одному игроку;
 - б) все позиции, принадлежащие одному информационному множеству, имеют одинаковое число альтернатив, которые нумеруют натуральными числами;
 - в) для игрока 0 (природы), если он участвует в игре, информационное множество всегда состоит из одной позиции (начальной);
 - г) для конкретной партии игры (т.е. ломаной линии, идущей от основания дерева к одной из его вершин) имеется одна позиция из конкретного информационного множества (для этой партии).

Для того чтобы найти решение позиционной игры, необходимо ее свести к матричной игре. Процесс сведения позиционной игры к матричной называется *нормализацией позиционной игры*.

2.7. Биматричные игры

С помощью биматричных игр разрабатывают рекомендации для игроков, интересы которых необязательно являются противоположными. Допустим, игрок A может в процессе игры выбрать любую из своих стратегий $A_i = (A_1, A_2, \dots, A_m)$, где $i = \overline{1, m}$, а игрок B – любую из стратегий $B_j = (B_1, B_2, \dots, B_n)$, где $j = \overline{1, n}$. Выигрыши игроков A и B характеризуются соответствующими платежными матрицами:

A_i	B_j				
	B_1	...	B_k	...	B_n
A_1	a_{11}	...	a_{1k}	...	a_{1n}
...
A_r	a_{r1}	...	a_{rk}	...	a_{rn}
...
A_m	a_{m1}	...	a_{mk}	...	a_{mn}

A_i	B_j				
	B_1	...	B_k	...	B_n
A_1	b_{11}	...	b_{1k}	...	b_{1n}
...
A_r	b_{r1}	...	b_{rk}	...	b_{rn}
...
A_m	b_{m1}	...	b_{mk}	...	b_{mn}

Первая платежная матрица описывает выигрыш игрока A , а вторая – выигрыш игрока B . Когда интересы игроков не противоположны, а только различны, то при выборе игроком A r -й стратегии, а игроком B – k -й стратегии выигрыш игрока A будет характеризоваться элементом a_{rk} , а выигрыш игрока B – элементом b_{rk} . Так как интересы игроков не совпадают, то необходимо найти такое решение, которое в одинаковой мере удовлетворяло бы обоим игрокам. Другими словами, требуется определить такую равновесную ситуацию (точку равновесия), отклонение от которой уменьшает выигрыш игрока.

Средние выигрыши игроков A и B определяются таким образом:

$$f_A(p, q) = a_{11}pq + a_{12}p(1-q) + a_{21}(1-p)q + a_{22}(1-p)(1-q),$$

$$f_B(p, q) = b_{11}pq + b_{12}p(1-q) + b_{21}(1-p)q + b_{22}(1-p)(1-q),$$

где $0 \leq p \leq 1$ и $0 \leq q \leq 1$.

Согласно теореме Дж. Нэша, всякая биматричная игра имеет хотя бы одну равновесную ситуацию (точку равновесия) в смешанных стратегиях.

Считается, что пара чисел (p^*, q^*) , где $0 \leq p^* \leq 1$, $0 \leq q^* \leq 1$, характеризует равновесную ситуацию, если для любых p и q , где $0 \leq p \leq 1$ и $0 \leq q \leq 1$, одновременно выполняются следующие неравенства:

$$f_A(p, q^*) \leq f_A(p^*, q^*),$$

$$f_B(p^*, q) \leq f_B(p^*, q^*).$$

Иными словами, отклонение от равновесной ситуации невыгодно самому игроку.

Практически, для того чтобы убедиться, что пара чисел (p^*, g^*) определяет равновесную ситуацию, достаточно проверить справедливость неравенства $f_A(p, q^*) \leq f_A(p^*, q^*)$ только для двух чистых стратегий игрока A – для $p=0$ и $p=1$, а неравенство $f_B(p^*, q) \leq f_B(p^*, q^*)$ проверяют для двух чистых стратегий игрока B – для $q=0$ и $q=1$.

Рассчитаем среднее значение выигрыша игрока A :

$$\begin{aligned} f_A(p, q) &= a_{11}pq + a_{12}p(1-q) + a_{21}(1-p)q + \\ &+ a_{22}(1-p)(1-q) = a_{11}pq + a_{12}p - a_{12}qp + a_{21}q - \\ &- a_{21}pq + a_{22} - a_{22}q - a_{22}p + a_{22}pq = (a_{11} - a_{12} - \\ &- a_{21} + a_{22})pq + (a_{12} - a_{22})p + (a_{21} - a_{22})q + a_{22}. \end{aligned}$$

Аналогичным образом запишем среднее значение выигрыша игрока B :

$$\begin{aligned} f_B(p, q) &= b_{11}pq + b_{12}p(1-q) + b_{21}(1-p)q + \\ &+ b_{22}(1-p)(1-q) = (b_{11} - b_{12} - b_{21} + b_{22})pq + \\ &+ (b_{12} - b_{22})p + (b_{21} - b_{22})q + b_{22}. \end{aligned}$$

Приведем расчеты для игрока A :

при $p=1$ получим:

$$f_A(1, q) = (a_{11} - a_{12} - a_{21} + a_{22})q + a_{12} + (a_{21} - a_{22})q;$$

при $p=0$ получим:

$$f_A(0, q) = (a_{21} - a_{22})q + a_{22}.$$

Рассмотрим разности:

$$\begin{aligned} f_A(p, q) - f_A(1, q) &= (a_{11} - a_{12} - a_{21} + a_{22})pq + \\ &+ (a_{12} - a_{22})p + (a_{21} - a_{22})q + a_{22} - (a_{11} - a_{12} - \\ &- a_{21} + a_{22})q - a_{12} - (a_{21} - a_{22})q = (a_{11} - a_{12} - \\ &- a_{21} + a_{22})pq + (a_{12} - a_{22})p - (a_{11} - a_{12} - a_{21} + \\ &+ a_{22})q + a_{22} - a_{12}; \end{aligned}$$

$$\begin{aligned} f_A(p, q) - f_A(0, q) &= (a_{11} - a_{12} - a_{21} + a_{22})pq + \\ &+ (a_{12} - a_{22})p + (a_{21} - a_{22})q + a_{22} - (a_{21} - a_{22})q - \\ &- a_{22} = (a_{11} - a_{12} - a_{21} + a_{22})pq + (a_{12} - a_{22})p. \end{aligned}$$

Обозначим через $C = a_{11} - a_{12} - a_{21} + a_{22}$, а через $\alpha = a_{22} - a_{12}$.

Подставив значения C и α в вышеизложенные выражения, получим:

$$\begin{aligned} f_A(p, q) - f_A(1, q) &= Cpq - \alpha p - Cq + \alpha = Cq(p-1) - \\ &- \alpha(p-1) = (p-1)(Cq - \alpha); \end{aligned}$$

$$f_A(p, q) - f_A(0, q) = \tilde{N}pq - \alpha p = p(Cq - \alpha).$$

Если пара чисел (p^*, q^*) определяет равновесную ситуацию, то $(p-1)(Cq - \alpha) \geq 0$ и $p(Cq - \alpha) \geq 0$.

Для игрока B аналогично вычисляем средний выигрыш при $q=1$ и $q=0$:

$$f_B(p, 1) = (b_{11} - b_{12} - a_{21} + a_{22})p + b_{12} + (b_{21} - b_{22})p;$$

$$f_B(p, 0) = (b_{21} - b_{22})p + b_{22}.$$

Определим следующие разности, обозначив через

$$D = b_{11} - b_{12} - b_{21} + b_{22} \text{ и } \beta = b_{22} - b_{21};$$

$$f_B(p, q) - f_B(p, 1) = (q-1)(Dp - \beta);$$

$$f_B(p, q) - f_B(p, 0) = q(Dp - \beta).$$

Если пара чисел (p^*, q^*) определяет точку равновесия, то $(q-1)(Dq - \beta) \geq 0$ и $q(Dp - \beta) \geq 0$.

Таким образом, для того чтобы в биматричной игре 2×2 пара чисел (p, q) определяла равновесную ситуацию, необходимо и достаточно выполнение следующих неравенств:

$$(p-1)(Cq - \alpha) \geq 0,$$

$$p(Cq - \alpha) \geq 0,$$

$$(q-1)(Dp - \beta) \geq 0,$$

$$q(Dp - \beta) \geq 0,$$

$$0 \leq p \leq 1,$$

$$0 \leq q \leq 1,$$

где $C = a_{11} - a_{12} - a_{21} + a_{22}$,

$$\alpha = a_{22} - a_{12},$$

$$D = b_{11} - b_{12} - b_{21} + b_{22}$$

$$\beta = b_{22} - b_{21}.$$

2.8. Кооперативные игры

Кооперативные игры – это игры с нулевой суммой, в которых игроки могут принимать решения по согласованию друг с другом, вправе вступать в коалицию. Кооперативные игры отличаются от коалиционных игр тем, что могут и не содержать коалиций. По форме платежа кооперативные игры подразделяются на игры с побочными платежами и без побочных платежей. В первом случае допускается заключение взаимосвязывающих соглашений о стратегиях, а платежи

могут перераспределяться между игроками. Во втором случае игроки согласуют свои стратегии, а платежи делят в соответствии с оптимумом по Парето, т.е. не существует такого решения игры (такой набор платежей), которое было бы лучше, чем данное. На рис. 2.8.1 графически представлено решение кооперативной игры.

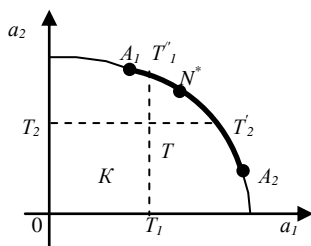


Рис. 2.8.1. Результаты решения кооперативной игры

Исходы кооперативной игры представлены выпуклым, замкнутым и ограниченным сверху множеством K . Кривая A_1A_2 содержит парето-оптимальные решения, при которых увеличение выигрыша одного из игроков возможно только за счет уменьшения выигрыша другого игрока. Точки T_1 и T_2 соответствуют выигрышам игроков, которые они могут получить без кооперации с партнером. Точки кривой T'_1 и T'_2 обозначают переговорное множество N , т.е. игроки, ведя переговоры, могут улучшить положение одного из них без ущерба для партнера. На переговорном множестве выделяется точка N^* , соответствующая равновесию по Нэшу (точка Нэша), в которой достигается максимум произведения:

$$\max(a_1 - T_1)(a_2 - T_2).$$

Смножители произведения характеризуют превышение выигрышей игроков над платежами, которые они могли бы получить без кооперации. Точка Нэша является решением кооперативной игры.

Лекция 3.

Тема 3. ЛИНЕЙНЫЕ МОДЕЛИ

- 3.1. Общая характеристика линейных моделей.
- 3.2. Примеры моделей планирования производства и макроэкономики.
- 3.3. Двойственные оценки и их экономическая интерпретация.

- 3.4. Устойчивость оптимального плана.
- 3.5. Иерархические системы и методы декомпозиции.
- 3.6. Целочисленные линейные модели.

3.1. Общая характеристика линейных моделей

Под *экономико-математической линейной моделью* понимают программу вычислений, обеспечивающую нахождение наилучшего, т.е. оптимального решения задачи, условия которой заданы в виде линейных уравнений или неравенств, сведены в единую систему, подчиненную цели решения задачи (т.е. целевой функции, записанной в виде линейного уравнения).

В зависимости от характера моделируемых объектов и процессов структура моделей может быть различной. Но имеются и общие элементы модели, включающие следующие группы:

1) неизвестные величины, значения которых определяются в результате решения задачи. Обычно их обозначают x_j , где $i=1 \dots m$ или y_j , где $j=1 \dots n$. Решить задачу, значит найти величины неизвестных переменных;

2) технико-экономические коэффициенты, т.е. известные величины при переменных, они служат для отображения закономерных взаимосвязей ресурсов с результатами решения задачи. Технико-экономические коэффициенты обычно характеризуются двумя индексами и обозначаются малыми латинскими буквами;

3) известные величины, стоящие в правой части ограничений (т.е. уравнений или неравенств). Они отображают возможные объемы ресурсов и ограничивающие условия, влияющие на результаты решения задачи. Эти элементы обозначаются большими латинскими буквами;

4) коэффициенты целевой функции, или коэффициенты F -строки, которая определяет цель решения задачи. Они обозначаются малыми латинскими буквами.

Элементы второй, третьей и четвертой групп составляют исходную информацию экономико-математической задачи. Чтобы получить достоверное решение, необходимо правильно количественно описать моделируемый объект, т.е. следует правильно обосновать исходную информацию задачи.

Используя приведенные четыре группы элементов, запишем общую задачу линейного программирования.

3.2. Примеры моделей планирования производства и макроэкономики

Линейные модели нашли применение в планировании производства и макроэкономики.

В экономической литературе выделяются следующие базовые модели:

1. *Модели оптимизации производственной программы* (программы развития перерабатывающей, сельскохозяйственной организаций). Общая постановка этой задачи состоит в обосновании оптимального плана производства нескольких видов продукции, обеспечивающего наиболее рациональное использование имеющихся ресурсов и максимизирующего конечные результаты деятельности.

2. *Модели оптимального составления смеси* (пищевого рациона; рецепта комбикорма скота, рецептуры производства продовольственных товаров и т.д.).

Требуется оптимизировать выбор наилучшего способа смешивания исходных ингредиентов для получения смеси с заданными свойствами при минимизации стоимости ингредиентов.

3. *Модели оптимального раскрытия материала*. Необходимо оптимизировать способы раскрытия материала, получив запланированное количество заготовок, с целью минимизации расхода материала или минимизации его отходов.

4. *Модель максимальной загрузки промышленного оборудования*.

Необходимо максимально загрузить промышленное оборудование с целью минимизации неиспользуемых остатков его полезного фонда рабочего времени.

5. *Транспортные (распределительные) модели*. Для удовлетворения спроса во всех пунктах потребления требуется оптимизировать план перевозок с целью минимизации суммарных транспортных затрат.

6. *Модели планирования финансов*.

а) Требуется обосновать портфель срочных вкладов для выплаты по займу с целью минимизации размера целевого фонда.

б) при фиксированном размере целевого фонда необходимо обосновать портфель срочных вкладов с целью максимизации дохода от их использования.

7. *Модели рынка ценных бумаг*.

а) Требуется минимизировать риск портфеля ценных бумаг при заданной ожидаемой доходности.

б) Требуется максимизировать ожидаемую доходность портфеля

ценных бумаг при заданном уровне риска.

8. Модель оптимального портфеля инвестиционных проектов.

Требуется определить максимум доходности портфеля инвестиционных проектов.

3.3. Двойственные оценки и их экономическая интерпретация

Любой задаче линейного программирования можно определенным образом сопоставить некоторую другую задачу такого же класса, называемую двойственной (обратной) по отношению к исходной (прямой) задаче. Принято считать, что прямой будет та задача, в результате решения которой получают размеры отраслей и другие параметры. Решение *двойственной задачи* дает систему двойственных оценок (объективно обусловленных оценок, теневых цен ресурсов). Прямая и двойственная задачи линейного программирования могут образовывать пару как симметричных, так и несимметричных задач. Несимметричная пара задач может быть в том случае, если одно или несколько ограничений прямой задачи представлено уравнением, а в двойственной задаче система ограничений сформирована неравенствами одного вида. При этом двойственная оценка, относящаяся к данному уравнению, может принимать как положительное, так и отрицательное значение.

В теории линейного программирования доказывается несколько теорем о взаимосвязи решений прямой и двойственной задач. Знание этих взаимосвязей важно не только с теоретической, но и с прикладной точки зрения. Важнейшие из них такие:

а) взаимодвойственность: прямая задача и двойственная к ней являются взаимодвойственными;

б) теорема двойственности: если взаимодвойственные задачи имеют хотя бы одно допустимое решение, то они имеют одинаковые значения целевых функций в оптимуме;

в) полнота симплекс-таблицы: последняя симплексная таблица, соответствующая прямой задаче, содержит всю информацию о решении двойственной и наоборот.

Свойство взаимодвойственности заключается в следующем. Если двойственную задачу рассматривать в качестве прямой и применить к ней правила построения двойственной, то мы получим исходную прямую задачу.

Свойство полноты симплекс-таблицы продемонстрируем на основе изучения последней симплексной таблицы рассматриваемой

двойственной задачи:

1. дополнительные небазисные переменные двойственной задачи равны соответствующим основным переменным прямой задачи и их значение, взятое со знаком плюс, находится в строке целевой функции:

$$y_i^g = x_j^n = (+)c_j;$$

2. двойственные переменные, стоящие в небазисных переменных, равны соответствующим дополнительным переменным прямой задачи и их значение, взятое со знаком плюс, находится в строке целевой функции:

$$u_i^g = y_i^n = (+)c_j,$$

Свойства объективно обусловленных оценок.

Использование двойственных оценок в анализе связано со следующим их значением:

1. *Двойственные оценки индивидуальны.* В современных рыночных условиях важно оценить роль отдельных ресурсов в формировании результата конкретного товаропроизводителя, так как одинаковые ресурсы в каждой организации АПК играют различную роль и их двойственные оценки будут различны. Это обусловлено такими факторами, как климатические условия, формы собственности, организации и оплаты труда, применяемые производственные технологии и т.д. Поэтому двойственные оценки рассчитывают индивидуально для каждой аграрной организации при решении задачи линейного программирования.

2. *Двойственные оценки устойчивы.* Они имеют единицы измерения целевой функции. Двойственные оценки получают только лимитированные, т.е. недостающие ресурсы. Избыточные ресурсы, то есть те, которые в данных производственных условиях используются неполностью в рассматриваемой организации, имеют нулевые двойственные оценки. Но этот факт не означает отсутствия хозяйственной ценности таких ресурсов, а лишь указывает на их нерациональное использование. При изменении производственных условий избыточный ресурс может стать недостающим и иметь двойственную оценку.

3. *Двойственные оценки позволяют соизмерить затраты и результаты производства,* определить конечный эффект от принятия того или иного управленческого решения, так как двойственная оценка показывает насколько изменится величина целевой функции, если ресурс изменится на единицу сверх имеющегося объема. Нулевые двойственные оценки по ресурсам свидетельствуют о том, что

изменение их объема на единицу не повлияет на конечные результаты целевой функции. Ненулевые двойственные оценки подсказывают, объем какого ресурса необходимо увеличить для повышения конечных результатов организации. При этом ресурсы, получившие ненулевую оценку, должны быть ранжированы по своей эффективности, которую определяют с учетом затрат, необходимых для привлечения единицы дополнительного ресурса.

4. *Двойственные оценки позволяют определить нормы взаимозаменяемости между ресурсами.* В данном случае речь идет только об относительной заменяемости с учетом влияния ресурсов на конечные результаты.

5. *Двойственные оценки позволяют расположить отрасли сельскохозяйственной организации и производимую продукцию по степени ее эффективности* и тем самым достоверно обосновать направления ее дальнейшего развития. Аналогично, если анализируется роль расчетного или планового задания, то двойственная оценка может рассматриваться как цена увеличения задания в плане потерь при достижении требуемого эффекта.

Свойства двойственных оценок базируются на содержании первой и второй теорем двойственности. Сформулируем *первую теорему двойственности*:

если одна из задач, или прямая или двойственная, имеет оптимальное решение, то и другая задача имеет оптимальное решение, причем экстремальные значения целевых функций равны, то есть, если целевая функция прямой задачи записана в виде

$$F_{\max} = \sum_{j \in J_0} c_j x_j,$$

а двойственной задачи – в виде

$$F_{\min} = \sum_{i \in I_0} A_i u_i,$$

то согласно первой теореме двойственности

$$\max \sum_{j \in J_0} c_j x_j = \min \sum_{i \in I_0} A_i u_i.$$

Суть *второй теоремы двойственности* в следующем:

1. если двойственные оценки положительны, то производственные ресурсы, для которых они рассчитаны, используются полностью, т. е. если

$$u_i > 0,$$

то

$$\sum_{j \in J_0} a_{ij} x_j = A_i, i \in I_0;$$

2. двойственные оценки равны нулю, если производственные ресурсы, к которым они относятся, недоиспользуются, т. е.

$$u_i = 0,$$

при условии

$$\sum_{j \in J_0} a_{ij} x_j < A_i, i \in I_0.$$

В свою очередь ресурсы, получившие ненулевые двойственные оценки, должны быть ранжированы по своей эффективности. При этом эффективность определяют с учетом затрат, необходимых для привлечения единицы ресурса.

В практическом отношении для задач линейного программирования важно выявление альтернативных оптимальных решений.

Оно основывается на том теоретическом положении, что если

$$\sum_{i \in I_0} a_{ij} u_i > c_j,$$

то отрасль не выгодно вводить в базис, а если

$$\sum_{i \in I_0} a_{ij} u_i < c_j,$$

то отрасль для организации будет прибыльна.

Таким образом, двойственные оценки становятся важным инструментом анализа оптимальных решений и в этом качестве могут служить для обоснования действий, направленных на повышение эффективности производства.

3.4. Устойчивость оптимального плана

Анализ устойчивости (чувствительности) оптимального плана основывается на изменении параметров задачи: a_{ij} , A_i и p_j , где $i = 1, 2 \dots m$; $j = 1, 2 \dots n$.

Влияние изменения параметров задачи на оптимальный план можно посмотреть, если найти интервалы устойчивости оценок, в пределах которых они измеряют влияние ограничений на целевую функцию задачи.

Допустимый интервал устойчивости оценок имеет вид:

$$\left[A_i - \Delta A_i^{\min}; A_i + \Delta A_i^{\max} \right], i = 1, 2 \dots m,$$

где ΔA_i^{\min} – нижний предел уменьшения ресурса вида i , который

находится по формуле:

$$\Delta A_i^{\min} = \min_{\substack{j \\ d_{ij} > 0}} \left\{ \frac{x_j^*}{d_{ij}} \right\},$$

ΔA_i^{\max} – верхний предел увеличения ресурса вида i , который находится по формуле:

$$\Delta A_i^{\max} = \left| \max_{\substack{j \\ d_{ij} < 0}} \left\{ \frac{x_j^*}{d_{ij}} \right\} \right|,$$

где d_{ij} – элементы матрицы $A^{-1} = \|d_{ij}\|$, обратной к матрице $A = \|a_{ij}\|$ в оптимальном решении прямой задачи;

x_j^* – оптимальное решение.

В послеоптимизационном анализе результатов решения задач линейного программирования чаще всего используют информационные технологии LPX.88 и Excel.

3.5. Иерархические системы и методы декомпозиции

При моделировании параметров развития сложных систем исследователь сталкивается с трудностью поиска оптимального решения задач линейного программирования большой размерностью. С целью упрощения процедуры расчета применяют методы декомпозиции, которые основаны на рациональном расчленении сложной экономико-математической задачи и решением отдельных подзадач с последующим согласованием частных решений для получения общего оптимального решения. При этом каждую сложную систему, к которой относится и экономическая система (для микроэкономики – это предприятие, для макроэкономики – отрасли, регионы), состоящую из большого числа взаимосвязанных объектов, представляют в виде совокупности подсистемы изучаемой системы. Таким образом, любая сложная система имеет иерархическую структуру, в которой существует множество составляющих ее подсистем и элементов разного уровня, обладающих определенными взаимосвязями и взаимоотношениями. Но, превратившись в относительно самостоятельную подсистему или элемент системы, такое звено обретает и собственные интересы, которые могут не совпадать с интересами руководства системой. Для устранения таких

расхождений в экономике применяются меры экономического стимулирования, которые направлены на объединение личных, коллективных и общегосударственных интересов в экономике. Обосновать мероприятия по экономическому стимулированию интересов системы и подсистем можно с помощью методов декомпозиции.

Методы декомпозиции или разложения эффективны для задач линейного программирования большой размерности, у которых матрица задачи может быть приведена к блочной структуре.

Блочную экономико-математическую модель можно записать следующим образом: требуется найти максимальное значение целевой функции

$$F_{\max} = \sum_{j \in J_0} c_j x_j, \quad (3.1)$$

При условиях:

по использованию ресурсов

$$\sum_{j \in J_0} a_{ij} x_j = A_i, i \in I_0;$$

по использованию глобальных ресурсов

$$\sum_{j \in J_0} b_{ij} x_j = B_i, i \in I_1;$$

по неотрицательности неизвестных величин

$$x_j \geq 0, j \in J_0.$$

Индексация:

i – номер ресурса;

I_0 – множество видов ресурсов;

I_1 – множество видов глобальных ресурсов;

j – номер технологического способа производства;

J_0 – множество видов технологических способов производства;

Неизвестные величины:

x_j – размер технологического способа производства вида j ;

Известные величины:

a_{ij} – расход ресурса вида i на единицу технологического способа производства вида j ;

A_i – объем ресурса вида i ;

b_{ij} – расход глобального ресурса вида i на единицу технологического способа производства вида j ;

B_i – объем глобального ресурса вида i ;

c_j – эффект (прибыль, выручка от реализации продукции и т.д.) с единицы технологического способа производства вида j .

Существует два основных подхода к декомпозиции выше изложенной системы:

1. *Прямая декомпозиция* путем перераспределения между секторами глобальных ресурсов. При прямой декомпозиции строится итеративный процесс определения лимитов глобальных ресурсов вида i , т.е. y_i , удовлетворяющих условию:

$$\sum_{i \in I_1} y_j \leq B_i, i \in I_1$$

и таких, что оптимальные решения подзадач (блоков матрицы):

$$\sum_{j \in J_0} a_{ij} x_j = A_i, i \in I_0 \quad (3.2)$$

$$\sum_{j \in J_0} b_{ij} x_j = y_i, i \in I_1$$

$$F_{\max} = \sum_{j \in J_0} c_j x_j$$

являются одновременно решением задачи (3.1).

2. *Двойственная декомпозиция* или декомпозиция через двойственные оценки. Схема построена по принципу сравнения централизованно определенных цен с децентрализованным определением наилучших возможностей.

Секторам экономики задаются зависящие от цен (двойственных оценок) целевые функции, которые заставляют сектора выбирать решения, способствующие достижению общей цели прямой задачи (3.1). Т.е. строится итеративный процесс определения двойственных оценок u_i таких, что оптимальные решения подзадач (задач блоков):

$$\sum_{j \in J_0} a_{ij} x_j = A_i, i \in I_0 \quad (3.3)$$

$$F_{\max} = (c_j + u_i B_i) \cdot x_j$$

являются в тоже время решением исходной прямой задачи (3.1).

В общих подходах алгоритм решения общей задачи (3.1) состоит из двух этапов: на первом этапе находится допустимое решение для системы в целом (как совокупность допустимых решений подзадач), на втором этапе полученное решение шаг за шагом улучшается, приближаясь к оптимальному.

Первоначальные значения лимитов глобальных ресурсов y_i или цен u_i выбираются априорно, и на каждом шаге корректируются и скорректированные подзадачи (3.2) или (3.3) решаются заново. Критерием корректировок является минимизация отклонений двойственных оценок глобальных ограничений в локальных подзадачах.

Практически, через 4–5 шагов полученное приближенное решение становится экономически приемлемым. При этом оценки одноименных ограничений в локальных задачах приблизительно равны и совокупность решений локальных подзадач образует решение исходной задачи (3.1).

Следует отметить, что рассмотрение методов декомпозиции представляет интерес в теоретическом плане, в практических целях они теряют свою значимость в связи с развитием вычислительной техники и пакетов прикладных программ, позволяющих в короткие сроки решать экономико-математические задачи большой размерности.

3.6. Целочисленные линейные модели

Экстремальные задачи, в которых на переменные накладываются условия целочисленности, а область допустимых решений конечна являются предметом изучения целочисленного или дискретного программирования. Они возникают в случае, когда искомые переменные определяют неделимые объекты.

В общем виде задача целочисленного программирования имеет следующий вид.

Требуется найти экстремальное (максимальное или минимальное) значение целевой функции:

$$F_{\max(\min)} = \sum_{j=1}^n c_j x_j .$$

При условиях:

$$1) \sum_{j=1}^n a_{ij} x_j \{ \leq, =, \geq \} A_i, i = \overline{1, m} \quad (3.4)$$

$$2) x_j \geq 0 \text{ и } x_j - \text{целые для всех } j = \overline{1, n}.$$

Можно выделить следующие основные классы задач дискретного программирования:

1. *транспортная задача и ее варианты*: требуется обосновать план перевозки грузов от поставщиков до потребителей с целью минимизации транспортных затрат. При этом оптимальный объем груза x_{ij} перевозимого от поставщиков вида i к потребителю вида j есть число целое ($x_{ij} \geq 0, x_{ij}$ – целые числа).

К этому классу задач можно отнести и задачу о назначениях. Требуется так распределить работников по работам, чтобы общая их выработка была наибольшей.

Введем переменные задачи:

$$x_{ij} = \begin{cases} 1, & \text{если работник вида } i \text{ выполняет работу вида } j; \\ 0, & \text{в противном случае.} \end{cases}$$

Переменные, принимающие только два значения – 0 или 1, называются булевыми переменными.

Математическая запись задачи:

$$F_{\max} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij}.$$

При ограничениях

по закреплению работ за работниками:

$$\sum_{j=1}^n x_{ij} = 1, i = \overline{1, n}$$

по выполнению работ работниками:

$$\sum_{i=1}^n x_{ij} = 1, j = \overline{1, n},$$

где c_{ij} – доход при выполнении работником вида i работы вида j ;

2. *задачи с неделимостью*. Классическим примером задач данного класса является задача о ранце. Суть которой состоит в следующем: школьник, собираясь в школу, складывает ранец, вес которого не должен превышать A_i килограмм. В ранец можно положить n предметов, каждый из которых весит a_{ij} килограмм и характеризуются

полезностью p_j . Требуется выбрать так предметы, чтобы их общий вес не превышать максимально допустимый (A_i) и суммарная полезность содержимого ранца была максимальной.

Обозначим через x_j неизвестные параметры задачи, при этом

$$x_j = \begin{cases} 1, & \text{если } j\text{-ый предмет помещают в ранец } (j = \overline{1, n}) \\ 0 & \text{в противном случае.} \end{cases}$$

Математическая формулировка задачи имеет вид:

$$F_{\max} = \sum_{j=1}^n p_j x_j .$$

При условиях:

1) по предельному весу ранца:

$$\sum_{j=1}^n a_{ij} x_j \leq A_i, i = \overline{1, m}$$

2) по неотрицательности и целочисленности переменных:

$$x_j \geq 0; x_j \in \{0; 1\}, j = \overline{1, n} ;$$

3) *комбинаторные задачи*. Классическим представителем комбинаторных задач является задача о коммивояжере. Торговый агент должен выехать из определенного города и вернуться в него, побывав в каждом городе один раз и при этом проехав минимальное расстояние.

Введем переменные задачи:

$$x_{ij} = \begin{cases} 1, & \text{если в маршрут входит переезд из города } i \text{ в город } j; \\ 0 & \text{в противном случае } (i, j = \overline{1, n}; i \neq j) \end{cases}$$

Математическая запись задачи:

$$F_{\min} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} .$$

При условиях:

1) по въезду в город:

$$\sum_{j=1}^n x_{ij} = 1, i = \overline{1, n}$$

2) по выезду из города:

$$\sum_{i=1}^n x_{ij} = 1, j = \overline{1, n} .$$

К приведенным ограничениям добавляют условия на недопустимость подциклов (повторного посещения городов, за исключением исходного):

$$u_i - u_j + nx_{ij} \leq n - 1, i = \overline{2, n}; j = \overline{1, n} (i \neq j),$$

где x_{ij} – решение коммивояжера о переезде из города i в город j ;

c_{ij} – расстояние между городами i и j ;

4) задачи с разрывными целевыми функциями. Экономические системы характеризуются наличием постоянных затрат, величина которых не зависит от объема производства продукции. Учет такого факта в модели приводит к появлению целевых функций, не обладающих свойством непрерывности. В качестве примера рассмотрим транспортную задачу с фиксированными доплатами. Она отличается от транспортной задачи, рассмотренной в вопросе 2 тем, что в ней затраты по перевозке груза от поставщика i к потребителю вида j определяются следующим образом:

$$\bar{c}_{ij} x_{ij} = \begin{cases} c_{ij} x_{ij} + d_{ij}, & \text{если } x_{ij} > 0, \\ 0, & \text{если } x_{ij} = 0, \end{cases}$$

где c_{ij} – издержки на перевозку груза от поставщика вида i к потребителю вида j ;

d_{ij} – фиксированная доплата за аренду транспорта при перевозке груза от поставщика вида i к потребителю вида j .

Целевая функция суммарных транспортных затрат имеет вид:

$$F_{\min} = \sum_{i=1}^m \sum_{j=1}^n \bar{c}_{ij} x_{ij}$$

Следует отметить, что она изменяется скачкообразно, что затрудняет нахождение ее минимума.

Для решения такой задачи вводят вспомогательные переменные y_{ij} , которые принимают два значения – 0 и 1:

$$y_{ij} = 0 \vee y_{ij} = 1,$$

тогда $x_{ij} \leq (\min\{a_i, b_j\}) y_{ij}$

и целевая функция имеет вид:

$$F_{\min} = \sum_{i=1}^m \sum_{j=1}^n (c_{ij} x_{ij} + d_{ij} y_{ij}).$$

Данная задача является задачей частично-целочисленного программирования.

Для решения задач линейного целочисленного программирования может использоваться симплексный метод, с помощью которого получают оптимальное решение задачи. Если значения переменных задачи получены нецелочисленными, их округляют до ближайших целых чисел. Этот метод применяется только тогда, когда отдельная единица совокупности составляет относительно малую часть объема совокупности. В противном случае округление может привести к неоптимальному решению задачи.

Поэтому для решения линейных задач целочисленного программирования используются следующие методы:

1. методы отсечения;
2. комбинаторные методы;
3. приближенные методы.

Сущность методов отсечения состоит в том, что сначала задача решается симплексным методом без условия целочисленности. Если полученные значения переменных не являются целочисленными, то к ограничениям задачи добавляется новое ограничение, обладающее следующими свойствами:

1. оно должно быть линейным;
2. отсекает найденное оптимальное нецелочисленное решение задачи;
3. не затрагивать ни одного целочисленного решения.

Алгоритм Р. Гомери позволяет за конечное число шагов (итераций) прийти к оптимальному целочисленному решению, если оно существует. Главное – сформировать дополнительное ограничение, называемое правильным отсечением.

Допустим, симплексным методом без условия целочисленности решим задачу (3.4).

Система ее ограничений, после получения оптимального решения будет иметь вид:

$$x_i = A_i - \sum_{j=1}^n a_{ij} x_j, i = \overline{1, m} \quad (3.5)$$

где x_i – базисная переменная, стоящая в i -ой строке, последней симплексной таблицы;

x_j – небазисная переменная, стоящая в j -м столбце, последней симплексной таблицы;

a_{ij} – коэффициент пропорциональности, стоящий в i -ой строке j -го столбца последней симплексной таблицы;

A_i – свободные члены, стоящие в i -ой строке последней симплексной таблицы.

Предположим, что некоторые A_i и a_{ij} – нецелые числа. Обозначим наибольшую часть чисел A_i и a_{ij} их не превосходящую, через $[A_i]$ и $[a_{ij}]$, а дробную положительную часть – через $\{A_i\}$ и $\{a_{ij}\}$. При этом:

$$A_i = [A_i] + \{A_i\} \quad (3.6)$$

$$a_{ij} = [a_{ij}] + \{a_{ij}\}$$

$$\text{и } 0 < \{A_i\} < 1; \quad \{A_i\} \geq 0;$$

$$0 < \{a_{ij}\} < 1; \quad \{a_{ij}\} \geq 0.$$

Подставим (3.6) в уравнение (3.5), получим:

$$x_i = ([A_i] - \sum_{j=1}^n [a_{ij}] \cdot x_j) + (\{A_i\} - \sum_{j=1}^n \{a_{ij}\} \cdot x_j). \quad (3.7)$$

Так как выражение $[A_i] - \sum_{j=1}^n [a_{ij}] \cdot x_j$ – есть целое число, то для того чтобы x_i было целым числом, необходимо, чтобы величина $\{A_i\} - \sum_{j=1}^n \{a_{ij}\} \cdot x_j$ тоже была целым числом. Предположим, что

$$\{A_i\} - \sum_{i=1}^n \{a_{ij}\} \cdot x_j < 0, \quad (3.8)$$

зная, что $0 \leq \{A_i\} \leq 1$; $0 \leq \{a_{ij}\}$ и $\{a_{ij}\} \geq 0$, $x_j \geq 0$, должно выполняться неравенство:

$$\{A_i\} - \sum_{i=1}^n \{a_{ij}\} \cdot x_j < 1 \quad (3.9)$$

Но неравенства (3.8) и (3.9) противоречат требованиям целочисленности дробной части неравенства (3.7). Следовательно, для целочисленных решений должно выполняться условие, противоположное неравенству (3.8):

$$\{A_i\} - \sum_{j=1}^n \{a_{ij}\} \cdot x_j \leq 0 \text{ или } \sum_{j=1}^n \{a_{ij}\} \cdot x_j \geq \{A_i\} \quad (3.10)$$

Соотношение (3.10) определяет правильное отсечение Гомори.

Таким образом, для решения задач целочисленного линейного программирования *методом Гомори* используется следующий алгоритм:

1. Решают искомую задачу (3.4) симплексным методом без учета условия целочисленности. Если все переменные задачи – целочисленные, то получено искомое решение. Если задача без условия целочисленности не имеет решения, то и целочисленная задача решения не имеет.

2. Если среди оптимальных значений переменных есть нецелые, то выбирают компоненту с наибольшей целой частью и по соответствующему уравнению системы формируют правильное отсечение.

3. В неравенство, формирующее правильное отсечение, вводят дополнительную переменную и, превращая его в равенство, включают его в систему ограничений задачи.

4. Полученную расширенную задачу решают симплексным методом до тех пор, начиная с пункта 2, пока значения базисных переменных не будут целочисленными.

Одним из комбинированных методов является метод ветвей и границ. Суть метода состоит в упорядоченном переборе вариантов и рассмотрении только перспективных, отбрасывая бесперспективные варианты.

Алгоритм метода ветвей и границ:

1. Решаем искомую задачу (3.4) симплексным методом без учета условия целочисленности.

2. Если в полученном симплексном решении некоторые переменные имеют дробные значения, то выбираем любую из них и по ней строим два ограничения.

3. В одном ограничении величина переменной меньше или равна наибольшему целому числу, не превышающему значения дробной переменной в оптимальном решении, в другом ограничении она больше или равна наименьшему целому значению, но не меньше значение дробной переменной (например: $x_1 = 3,5$, первое ограничение будет $x_1 \leq 3$, а второе $x_1 \geq 4$, что исключает промежуток с дробным значением x_1).

4. В каждую из искомым задач добавляем по выше изложенному ограничению, в результате получаем две задачи (подзадачи) линейного программирования и решаем их.

5. Если снова получены оптимальные решения с дробным значением переменной, то, сравнив значения целевых функций задач, выбираем задачу с большим значением целевой функции и с пункта 2 продолжаем до тех пор пока не получим целочисленные значения переменных.

В результате получим ветви (рис. 3.6.1).

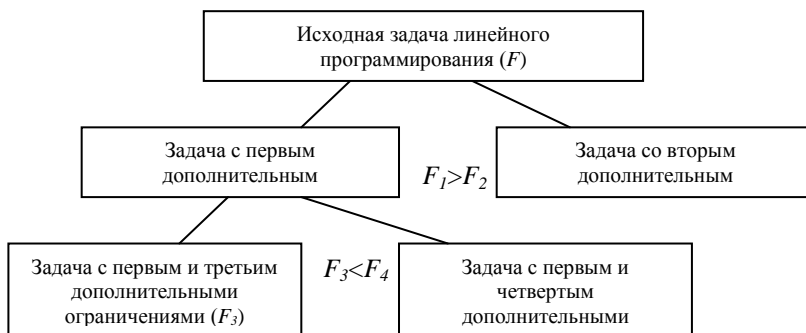


Рис. 3.6.1 Алгоритм решения целочисленной задачи линейного программирования методом ветвей и границ.

Решение линейной задачи целочисленного программирования можно найти средствами Excel.

Лекция 4.

Тема 4. СЕТЕВЫЕ МОДЕЛИ

- 4.1. Экстремальные задачи на графах.
- 4.2. Задача о минимальных покрывающих деревьях.
- 4.3. Задача о кратчайших цепях.
- 4.4. Задача о максимальном потоке в сетях и ее обобщения.
- 4.5. Элементы сетевого и календарного планирования.
- 4.6. Сетевые графики и их параметры.
- 4.7. Задачи распределения ресурсов на сетях.
- 4.8. Задачи оптимизации сетей во времени.
- 4.9. Задачи оптимизации сетей по стоимости.
- 4.10. Варианты задачи о назначениях.
- 4.11. Задача коммивояжера и ее приложения.

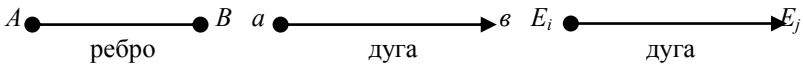
4.1. Экстремальные задачи на графах

Для схематического представления зависимостей между различными работами проекта используется теория графов.

Основы теории графов были заложены Л. Эйлером в 1736 г. Теория графов как математическая дисциплина сформировалась к середине 30-х годов прошлого века благодаря работам венгерского математика Д. Кенига.

Графом называется совокупность двух конечных множеств: множества точек, которые называются вершинами, и множества соединенных пар вершин, которые называются ребрами.

Пары точек, для которых установлено соответствие, соединяются непрерывной линией, которая называется *ребром*, если направление линии не указано, или *дугой* – если ее направление указано стрелкой:



Оба конца ребра или дуги принадлежат множеству вершин графа.

Ребра или дуги обозначаются парами вершин (A, B) ; $(1, 3)$; $(a, в)$; $(E_i, E_j) = \vec{e}$. Если множество вершин обозначить через X или E , множество дуг (ребер) графа через U или \vec{e} , то граф обозначается символом $G = (X, U)$ или $G = (E, \vec{e})$.

Если пары вершин графа соединены направленными линиями (т.е. дугами), то такой граф называется ориентированным или *орграфом*. Любые две вершины графа, соединенные ребром или дугой, называются *смежными*. В этом случае говорят, что вершины инцидентны ребру (дуге), а ребро (дуга) инцидентно вершинам. Если начало и конец ребра совпадают, то получается *петля*.

Граф называется *полный*, если каждые две различные вершины его соединены только одним ребром.

Граф, состоящий только из изолированных вершин называется *нуль-граф*. Граф, у которого любые две вершины соединены в одном направлении называется *полным ориентированным графом*.

Число ребер графа, которым принадлежит (инцидентна) вершина называется *степенью вершины*. Степень вершины X_i (или E_i) может

быть четной или нечетной, она обозначается $\rho(X_i)$ или $\rho(E_i)$. В орграфе *степень вершины* X_i (или E_i) определяется количеством дуг, выходящих из вершины X_i (или E_i), и количеством дуг, входящих в X_i (или в E_i).

Путь в орграфе называется последовательность сцепленных дуг, позволяющих переместиться из одной вершины в другую. Путь, у которого начальная вершина совпадает с конечной, называется контуром или *циклом* в неориентированном графе.

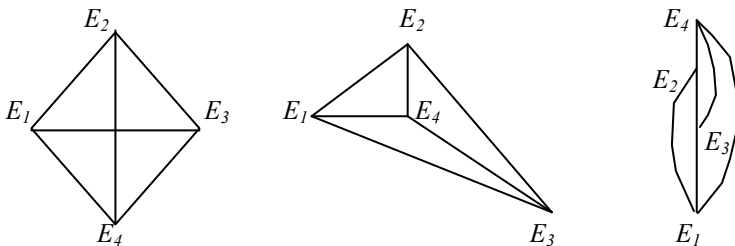
Аналогом пути в неориентированном графе служит *цепь*. *Граф* называется *связным*, если для каждой пары вершин существует соединяющая их цепь или путь.

Длина пути определяется числом его ребер или дуг. *Петля* – это путь имеющий длину равную единице.

Если граф не имеет циклов, то он называется *деревом*.

Для каждой пары вершин дерева существует только один соединяющий их путь. *Вершина* дерева степень, которой равна единице, называется *висячей*, а *ветвями дерева* называются ребра графа, входящие в дерево.

Изображают графы на плоскости произвольно и один и тот же граф может выглядеть на плоскости по-разному. Эти графы изоморфны, так как между множествами их вершин наблюдается взаимно однозначное соответствие, т.е. вершины одного графа соединены ребрами также как и соответствующие им вершины другого графа:



Аналитически граф можно задать двумя способами: списковым и матричным.

Списковый способ используют в виде: 1) перечня списка вершин и множество ребер или дуг; 2) перечня списка ребер или дуг и информации о их следовании. Граф можно задать в виде: 1) матрицы смежности вершин; 2) матрицы смежности дуг (ребер) 3) матрицы

инцидентий.

Для практического использования теории графов важную роль играет задача: разбить множество всех вершин связного графа без контуров на слои (ранги) так, что: 1) элементы первого слоя не должны иметь предшествующих вершин, а элементы последнего слоя не должны иметь последующих вершин; 2) все вершины рассматриваемого слоя не должны иметь предыдущих вершин в последующем слое; 3) порядок вершин внутри одного и того же слоя безразличен, т.е. вершины не соединяются между собой дугами.

Существует несколько способов упорядочения вершин (и дуг) графа.

Суть графического способа упорядочения вершин графа состоит в последовательном нахождении вершин, степень входящих дуг которых равна нулю: $\rho^+(E) = 0$. Вершины, для которых $\rho^+(E) = 0$, отбрасываются и вычеркивают выходящие из них дуги. Находят вершины, в которые входят вычеркнутые дуги, они составляют второй ранг. Вычеркивают выходящие из них дуги и т. д. Процесс продолжают до тех пор, пока не доберутся до конечной вершины. Таким образом, мы двигаемся слева направо. Можно двигаться и справа налево. Тогда находят вершины, для которых степень выходящих дуг равна нулю: $\rho^-(E) = 0$, с последующим вычеркиванием входящих в них дуг и т. д. Результат будет тот же.

Для упорядочения дуг графика можно воспользоваться матрицей смежности дуг.

4.2. Задача о минимальных покрывающих деревьях

Покрывающее дерево – это дерево, содержащее все вершины (узлы) сети.

Алгоритм построения минимального покрывающего дерева предполагает соединение всех вершин сети с помощью путей наименьшей длины.

Пусть сеть имеет множество вершин равное N . Через E_k – обозначим множество вершин сети, соединенных алгоритмом после выполнения k -ой итерации алгоритма, а через \bar{E}_k – множество вершин сети, не соединенных с узлами множества E_k после выполнения k -ой итерации алгоритма.

Алгоритм построения покрывающего дерева.

1. Из множества всех вершин сети выбираем любую вершину i и считаем $E_1 = \{i\}$, тогда $\bar{E}_1 = N - \{i\}$.

2. В множестве вершин E_1 выбираем вершину j , которая соединена самой короткой дугой (ребром) с какой-либо вершиной из множества \bar{E}_1 . Вершину j присоединяем к множеству E_1 и удаляем из множества \bar{E}_1 , получим новые множества вершин: $E_2 = E_1 + \{j\}$; $\bar{E}_2 = \bar{E}_1 - \{j\}$.

3. Алгоритм продолжаем выполнять с пункта 2 до тех пор, пока множество \bar{E}_k будет пусто.

4.3. Задача о кратчайших цепях

Задача состоит в определении кратчайшего маршрута, который необходимо проложить от исходного пункта до пункта назначения, используя существующую сеть дорог.

Кратчайший путь можно найти, используя:

- 1) алгоритм Дейкстры,
- 2) алгоритм Флойда.

Алгоритм Дейкстры применим для поиска кратчайшего пути между заданной исходной вершиной и любой другой вершиной сети. В процессе определения кратчайшего пути по алгоритму Дейкстры помечают вершины сети. Метка для вершины j определяется следующим образом:

$$[u_j, i] = [u_i + t_{ij}, i], \text{ при } t_{ij} \geq 0,$$

где i – номер начальной вершины дуги (ребра) (i, j) ;

j – номер конечной вершины дуги (ребра) (i, j) ;

u_j – расстояние к вершине j ;

u_i – кратчайшее расстояние к вершине i ;

t_{ij} – длина дуги (ребра) (i, j) .

Метки вершин подразделяются на временные и постоянные. Временная метка может заменяться на другую временную метку, если к данной вершине найден более короткий путь. Если такого пути не найдено, то временная метка заменяется на постоянную.

Алгоритм Дейкстры состоит из следующих этапов.

1. Исходной вершине присваиваем постоянную метку $[0, -]$.

2. Считаем, что $i = 1$. Определяем временные метки для всех вершин j , которых можно достичь непосредственно из вершины i и которые не имеют постоянных меток: $[u_i + t_{ij}; i]$.

3. Если вершина j имеет метку $[u_j, r]$, полученную от вершины r , то эту метку заменяют на метку $[u_i + t_{ij}; i]$, если $u_i + t_{ij} < u_j$.

4. Процесс пометок вершин продолжают до тех пор, пока все вершины не будут иметь постоянных меток.

5. Если какая-то вершина s имеет временные метки $[u_r, s]$, то среди них выбираем ту, которая имеет наименьшее значение расстояния u_r и процедуру повторяем с п. 1.

6. Определяем кратчайшую цепь, проходя этот путь в обратном направлении, используя постоянные метки.

Алгоритм Флойда более общий, так как он позволяет одновременно найти минимальные пути между любыми двумя вершинами сети.

Алгоритм Флойда представляет сетевой график в виде двух матриц: матрицы расстояний U_0 и матрицы последовательности вершин E_0 , диагональные элементы в вычислениях не участвуют и отсутствуют. Над элементами матрицы расстояний выполняется процедура замены $u_{ik} + u_{kj}$ на u_{ij} , если $u_{ij} > u_{ik} + u_{kj}$ (рис. 4.3.1). Такая замена получила название *треугольного оператора*.

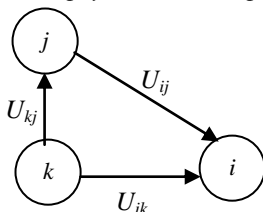


Рис. 4.3.1 Треугольный оператор Флойда.

Алгоритм Флойда.

1. Принимаем $k = 1$. В матрице расстояний U_0 задаем ведущую строку r и ведущий столбец k . При этом рассматриваем возможность применения треугольного оператора ко всем элементам u_{ij} матрицы U_0 .

Принимаем за ведущие первую строку и первый столбец, рассчитываем какие элементы матрицы U_0 можно улучшить, применив треугольный оператор.

2. Рассчитываем новые элементы матрицы расстояний U_1 , возьмем эти элементы и соответствующие элементы матрицы последовательности вершин E_0 путем замены в матрице расстояний элемента u_{ij} на сумму элементов $u_{ik} + u_{kj}$.

3. Определяем новые элементы матрицы последовательности вершин E_1 , меняя в матрице последовательности вершин E_0 , элемент e_{ij} на k .

4. Принимаем $k = k + 1$ и повторяем расчеты с пункта 1 до тех пор, пока ни один элемент матрицы расстояний улучшить нельзя.

5. Используя конечные матрицы расстояний и последовательности вершин можно найти кратчайшее расстояние как между двумя

соседними вершинами, так и между исходной и завершающей вершиной сети. При этом сегмент маршрута (i, j) матрицы последовательности вершин состоит из ребра (i, j) только тогда, когда элемент матрицы e_{ij} равен j , в противном случае, вершины i и j связаны между собой через одну или более промежуточные вершины.

Найти кратчайший путь между исходной и завершающей вершинами сети можно, используя задачи линейного программирования.

При составлении прямой задачи предполагают, что в исходную вершину сети входит одна единица внешнего потока, которая выходит из завершающей вершины сети.

В качестве неизвестных величин выступает x_{ij} – величина потока, проходящего по дуге (i, j) длина которой равна c_{ij} .

Переменные x_{ij} – двоичные, они могут принимать значения 0 или 1: $x_{ij} = 0 \cup 1, (i = \overline{1, n})$.

Прямая задача линейного программирования состоит в том, чтобы минимизировать длину сети:

$$F_{\min} = \sum_{(i,j) \in \bar{e}} c_{ij} x_{ij}.$$

В качестве ограничений задачи записывают баланс потока, проходящего через каждую вершину сети: общий входной поток минус общий выходной поток равен нулю.

Кратчайший путь между исходной и завершающей вершинами сети можно найти, используя двойственную задачу.

В качестве неизвестных величин двойственной задачи выступают u_i – расстояние от исходной вершины до вершины i .

Требуется максимизировать величину:

$$F_{\max} = u_n - u_1,$$

при условиях: $u_j - u_i \leq c_{ij}$.

Это ограничение показывает, что расстояние от вершины i до вершины j не может превышать длину дуги (i, j) . Это расстояние может быть меньше длины этого маршрута, если вершину j можно достичь из вершины i через другие промежуточные вершины.

Кратчайший путь можно определить по значениям соответствующих двойственных переменных данной задачи, значение которых равны 1, так как ресурсы, согласно, второй теореме двойственности, использованы полностью.

4.4. Задача о максимальном потоке в сетях и ее обобщения

Задача состоит в обосновании максимального потока из источника или источников до стока или стоков. При этом под источником понимают исходную вершину сети, под стоком – завершающую вершину сети. По путям сети направляется однородное вещество из источников в стоки. Каждая дуга сети характеризуется числом b_{ij} , называемым пропускной способностью дуги, под которой понимают максимальное количество вещества, пропускаемое за единицу времени. Ставится задача определить для данной сети максимальную величину потока из источника в сток, под которым понимают совокупность потоков x_{ij} по всем дугам сети, равный количеству вещества, перемещаемого по ней в единицу времени.

Важную роль в решении этой задачи играет понятие разреза. Разрез определяет множество дуг (ребер), при удалении которых из сети полностью прекращается поток от источника к стоку. При этом пропускная способность разреза равна сумме пропускных способностей дуг (ребер) разреза. Среди всех разрезов сети разрез с минимальной пропускной способностью определяет максимальный поток в сети.

Но перебор всех разрезов сети – непростая задача, поэтому для поиска максимального потока целесообразно использовать алгоритм Форда. Для его применения предварительно формируется матрица пропускных способностей дуг (ребер) сети. В таблицу в клетку (i, j) записывают пропускную способность дуги b_{ij} , если она больше нуля, а если пропускная способность симметричной ей дуги равна нулю, то в клетку (j, i) ставим нуль. Если $b_{ij} = b_{ji} = 0$, то клетки (i, j) и (j, i) не заполняются.

Алгоритм Форда:

1. Используя матрицу пропускных способностей, находим путь из исходной вершины E_1 в завершающую E_n с пропускной способностью больше нуля. Для этого столбец E_1 помещаем значком *. И в строке E_1 ищем положительные элементы матрицы ($b_{ij} > 0$) и столбцы, в которых они находятся помечаем номером просматриваемой строки. Таким образом, выбрали дуги, которые являются первыми дугами пути из E_1 в E_n .

Процедуру пометок продолжаем до тех пор, пока:

а) не будет помечен столбец E_n , т.е. сток;

б) нельзя пометить новые столбцы, что означает отсутствие пути из E_1 в E_n , проходящего по дугам с положительной пропускной способностью.

2. Находим путь из E_1 в E_n , используя пометки столбцов. При этом соответствующий элемент b_{ij} искомой дуги помечаем знаком « \leftarrow », а симметричный ему элемент b_{ji} – знаком « \rightarrow ». Данный процесс продолжаем до тех пор, пока не придем к истоку (вершине E_1) и не отметим знаком « \leftarrow » элемент этой строки и знаком « \rightarrow » – симметричный ему элемент.

3. Определяем пропускную способность пути, она равна наименьшей из пропускных способностей дуг, входящих в этот путь:

$$Q_i = \min_{(i,j)} \{b_{ij}^-\}.$$

4. Определяем остаточные пропускные способности дуг пути и симметричных к ним дуг. Для этого из элементов таблицы b_{ij}^- , получивших знак « \leftarrow » вычитаем выбранный минимальный элемент Q_i , а к элементам b_{ij}^+ , получивших знак « \rightarrow » прибавляем элемент Q_i . Все изменения заносим в новую матрицу пропускных способностей дуг V_1 и вычисления повторяем сначала до тех пор пока не получим таблицу, в которой нет ни одного пути из E_1 в E_n с пропускной способностью больше нуля.

5. Пометив столбцы табл., расставив знаки, находим путь из E_1 в E_n . Среди элементов матрицы, получивших знак « \leftarrow » выбираем наименьший. Изменяем пропускную способность дуг на данное число, получим новую табл. и т.д.

6. Выполнив выше изложенный алгоритм, находим искомый путь. Определяем величину потока по этому пути. Находим новые элементы матрицы пропускных способностей дуг и т.д.

7. Помечаем столбец E_1 знаком *. Просматриваем строку E_1 , убеждаемся, что никакие столбцы пометить нельзя, так как не существует ни одного пути с положительной пропускной способностью из вершины E_1 в вершину E_n .

9. Из элементов первоначальной таблицы вычитаем соответствующие элементы последней таблицы, получим новую таблицу, положительные элементы которой характеризуют величины дуговых потоков. Для определения максимального потока сети необходимо просуммировать элементы строки E_1 (источника) или элементы столбца E_n (стока), выделив разрез с минимальной

пропускной способностью. При этом отмечаем, что удалив дуги разреза, блокируем все пути из источника в сток.

Если имеется сеть с несколькими источниками и стоками, то для ее решения с помощью алгоритма Форда необходимо свести данную задачу с одним источником и одним стоком путем введения фиктивного источника (вершины E_0) и фиктивного стока (вершины E_{n+1}), а также фиктивных дуг (E_0, E_j) и (E_i, E_{n+1}) .

Задачу определения максимального потока в сети можно свести к задаче линейного программирования. Обозначим через x_{ij} – поток по дуге (E_i, E_j) , равный количеству вещества, перемещаемого по ней в единицу времени.

Требуется найти значения x_{ij} , максимизирующие одну из целевых функций:

1) максимальный поток, равный количеству вещества, вытекающего из источника –

$$F_{\max} = \sum_{j=1}^n x_{0j},$$

2) или максимальный поток, равный количеству вещества, притекающего в сток –

$$F_{\max} = \sum_{i=1}^{n-1} x_{in}.$$

При условиях:

1. По предельной пропускной способности дуг

$$0 \leq x_{ij} \leq \overline{b_{ij}}, \quad i, j = \overline{0, n}, \quad i \neq j.$$

2. По балансу вещества, притекающего в любую промежуточную вершину и вытекающего из нее

$$\sum_{i=0}^{n-1} x_{ik} - \sum_{j=1}^n x_{kj} = 0, \quad k = \overline{1, n-1}.$$

3. Неотрицательность переменных

$$x_{ij} \geq 0.$$

Положительные элементы характеризуют величины дуговых потоков.

Задача о потоке минимальной стоимости.

Задача нахождения потока минимальной стоимости в сети с ограниченной пропускной способностью обобщает задачу

определения максимального потока, так как каждой дуге соответствует определенная стоимость прохождения единицы потока по этой дуге (c_{ij}).

Требуется найти поток по дугам, заданной величины B , минимизирующий общую стоимость прохождения потока по сети. При этом должны удовлетворяться ограничения на пропускные способности дуг и на баланс вещества, притекающего в промежуточную вершину и вытекающего из нее.

Используя условные обозначения предыдущей задачи, запишем структурную модель определения потока минимальной стоимости в сети:

$$F_{\min} = \sum_{(i,j) \in \vec{e}} c_{ij} x_{ij}.$$

При условиях:

1. По предельной пропускной способности дуг:

$$0 \leq x_{ij} \leq b_{ij}, \quad i, j = \overline{0, n}, \quad i \neq j.$$

2. По балансу вещества, притекающего в любую промежуточную вершину и вытекающего из нее:

$$\sum_{i=0}^{n-1} x_{ik} - \sum_{j=1}^n x_{kj} = 0, \quad k = \overline{1, n-1}.$$

3. По количеству вещества, вытекающего из источника и притекающего в сток:

$$\text{а) } \sum_{j=1}^n x_{0j} = B; \quad \text{б) } \sum_{i=0}^{n-1} x_{in} = B.$$

4. Неотрицательность переменных:

$$x_{ij} \geq 0.$$

Лекция 5

4.5. Элементы сетевого и календарного планирования

Математический аппарат теории сетевого планирования и управления базируется на теории графов. Граф $G = (X, U)$ считается заданным, если заданы все его вершины и дуги. отождествим вершины орграфа с событиями, а дуги – с работами. События и работы, составляющие сетевой график являются основными

понятиями в сетевом планировании и управлении (СПУ).

Сетевой график – это динамическая модель, в которой моделируется совокупность взаимосвязанных работ и событий, отображающих процесс достижения определенной цели. *Работа* характеризует материальное действие, требующее использования ресурсов или времени, или логическое действие, требующее лишь взаимосвязи событий. Работа обозначается парой заключенных в скобки чисел (i, j) , где i – номер события, из которого работа выходит, а j – номер события, в которое она входит. Работа не может начаться раньше, чем свершится событие, из которого она выходит. Каждая работа имеет определенную продолжительность $t(i, j)$. К работам также относятся также такие процессы, которые не требуют ни ресурсов, ни времени для их выполнения. Это *фиктивные работы*, которые показывают, что одна работа не может совершиться пока не закончена другая, связанная с ней логически. На сетевых графиках фиктивные работы изображают пунктирными стрелками. Работа на графике соединяет два события.

Событиями называют результаты выполнения одной или нескольких работ. Событие свершается в тот момент, когда оканчиваются все работы, входящие в него. Оно становится предпосылкой для начала следующих за ним работ. События обозначаются одним числом, на графике изображаются кружком (реже точкой, ромбом и т.д.), внутри которого проставляют его порядковый номер ($i = 1, 2, \dots, N$). Событие с номером 1 называется *исходным*. Событие под номером n называется *завершающим*.

Любая непрерывная логическая (технологическая) последовательность работ от исходного события до завершающего называется *путем*.

При построении сетевых графиков надо соблюдать следующие правила:

1) в сетевом графике не должно быть тупиков, т.е. событий, из которых не выходит ни одна работа (за исключением завершающего события);

2) в сетевом графике не должно быть и событий (кроме исходного), которым не предшествует хотя бы одна работа;

3) не должно быть двух событий, связанных двумя или большим количеством работ;

4) в сети не должно быть контуров, т.е. цепей, соединяющих некоторые события с ними же самими;

5) не должно быть петель, т.е. начало выполнения работы является и условием ее окончания.

Длина пути определяется суммой продолжительности лежащих на нем работ. От исходного события до завершающего может быть много путей. В результате анализа сетевого графика определяют такой путь, суммарная продолжительность работ на котором будет максимальной. Он называется *критическим* и характеризует время, необходимое для выполнения всех работ, включенных в сетевой график. Работы, лежащие на критическом пути, не имеют резервов времени. Пути, близкие по времени к критическому, называются *подкритическими*, остальные пути являются *некритическими* или *ненапряженными*. Работы, лежащие на некритическом пути, имеют резерв времени. Наличие резервов времени у некритических работ дает возможность маневрировать внутренними ресурсами и этим ускорять выполнения критических и подкритических работ. На этом основана оптимизация сетевых графиков.

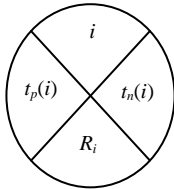
4.6. Сетевые графики и их параметры

Имеется 10 основных временных параметров сетевого графика:

Элемент сети	Наименование параметра	Условное обозначение
Событие	Ранний срок свершения	$t_p(i)$
	Поздний срок свершения	$t_n(i)$
	Резерв времени события	$R(i)$
Работа	Продолжительность работы	$t_{(i,j)}$
	Ранний срок начала работы	$t_{pi}(i, j)$
	Ранний срок окончания работы	$t_{po}(i, j)$
	Поздний срок начала работы	$t_{ni}(i, j)$
	Поздний срок окончания работы	$t_{no}(i, j)$
	Полный резерв времени	$R_n(i, j)$
	Свободный резерв времени	$R_c(i, j)$
	Независимый резерв времени	$R_H(i, j)$

Когда рассчитываем ранние сроки свершения событий $t_p(i)$ движемся с начала в конец графика, когда рассчитываем поздние сроки свершения событий $t_n(i)$ движется с конца в начало сетевого графика.

Расчет временных параметров событий выполняется на сетевом графике, для этого каждое событие укрупняется и разбивается на 4 сектора:



i – номер события;
 $t_p(i)$ – ранний срок свершения события i ;
 $t_n(i)$ – поздний срок свершения события i ;
 R_i – резерв времени события i ;

Приведем формулы для расчета временных параметров событий.

Если какому-то событию j предшествует свершение нескольких событий i , то ранний срок свершения события j определяется как максимальная сумма ранних сроков свершения событий i и продолжительность работ, входящих в событие j :

$$t_p(j) = \max_{(i,j) \in u_j^+} \{t_p(i) + t_{(i,j)}\}.$$

Если нескольким событиям j предшествуют свершение одного события i , то поздний срок свершения события i определяется как минимальная разность поздних сроков свершения событий j и продолжительности работ, выходящих из события i :

$$t_n(j) = \min_{(i,j) \in u_i^-} \{t_n(i) - t_{(i,j)}\}.$$

Резерв времени события определяется как разность между поздним и ранним срокам его свершения: $R_i = t_n(i) - t_p(i)$.

У критических работ резерв времени начального и конечного событий равен «0».

Работа лежит на критическом пути, если выполняются следующие требования:

1) ранний и поздний сроки свершения события, из которых выходит работа-стрелка, совпадают, т. е.:

$$t_p(i) = t_n(i);$$

2) ранний и поздний сроки свершения события, в которое входит работа-стрелка, совпадают, т. е.:

$$t_p(j) = t_n(j);$$

3) разница между ранними или поздними сроками свершения предшествующего и последующего событий работы равна продолжительности работы:

$$t_p(j) - t_p(i) = t_i(i) - t_n(i) = t_{(i,j)}.$$

Зная сроки свершения событий, можно определить временные параметры работ.

Ранний срок начала работы совпадает с ранним сроком свершения начального события:

$$t_{pi}(i, j) = t_p(i).$$

Ранний срок окончания работы равен раннему сроку начала работы плюс продолжительность работы:

$$t_{po}(i, j) = t_{pi}(i, j) + t_{(i,j)}.$$

При расчете раннего срока начала работы $t_{pi}(i, j)$ и раннего срока окончания работы $t_{po}(i, j)$ мы двигались по таблице сверху вниз. При расчете позднего срока начала работы $t_{pi}(i, j)$ и позднего срока окончания работы $t_{po}(i, j)$ необходимо двигаться снизу вверх.

Поздний срок окончания работы совпадает с поздним сроком свершения конечного события, т. е.:

$$t_{pi}(i, j) = t_i(i, j).$$

Поздний срок начала работы равен позднему сроку окончания работы минус продолжительность работы:

$$t_{po}(i, j) = t_{pi}(i, j) - t_{(i,j)}.$$

Полный резерв времени определяется как разность между поздним сроком окончания и ранним сроком начала работы и продолжительностью работы:

$$R_i(i, j) = t_{pi}(i, j) - t_{po}(i, j) - t_{(i,j)}$$

$$R_n(i, j) = t_i(j) - t_p(i) - t_{(i,j)}.$$

Полный резерв времени работы – это максимально возможный запас времени, на который можно отсрочить начало работы или увеличить продолжительность ее выполнения при условии, что конечное для данной работы событие наступит не позднее своего позднего срока. Работы, лежащие на критическом пути полного резервного времени не имеют.

Полный резерв времени может быть использован частично или полностью для выполнения данной работы или для любой другой работы, лежащей на данном пути.

Если растянуть сроки выполнения работ, лежащих на критическом пути на одни сутки, то критический путь возрастет на одни сутки и,

следовательно, на такое же время увеличится срок выполнения всего комплекса работ.

Свободный резерв времени – это запас времени, которым можно располагать при выполнении данной работы при условии, что начальное и конечное ее событие наступят в свои ранние сроки:

$$R_c(i, j) = t_p(j) - t_p(i) - t_{(i, j)}.$$

Свободный резерв присущ только данной работе, и его использование никак не повлияет на выполнение последующих работ.

Независимый резерв времени – это запас времени, которым можно располагать при выполнении данной работы при условии, что начальное ее событие наступит в свой поздний срок, а конечное – в свой ранний срок:

$$R_f(i, j) = t_p(j) - t_f(i) - t_{(i, j)}.$$

Для небольших проектов удобным дополнением к сетевому графику является *линейный график или график Ганта*. На нем каждая работа (i, j) изображается с учетом оси времени $0t$ горизонтальным отрезком, длина которого в соответствующем масштабе равна продолжительности работы $t_{(i, j)}$. Начало каждой работы совпадает с ранним сроком свершения ее начального события. Работы изображаются в той же последовательности, что и на сети.

4.7. Задачи распределения ресурсов на сетях

Рассмотрим задачу распределения ресурсов на сетевом графике. Допустим на предприятии необходимо выполнить комплекс проектных работ, последовательность которых изображена на сетевом графике. На его дугах проставлена продолжительность выполнения работ и в скобках (необходимое для выполнения работы число исполнителей, т.е. интенсивность потребления ресурса $v(i, j)$).

Используем *эвристический метод распределения ресурсов*.

Алгоритм решения задачи следующий.

1. Рассчитываем продолжительность критического пути.
2. Рассчитаем на графике ранний и поздний срок свершения событий и резерв времени событий.
3. Рассчитаем в таблице временные характеристики работ.
4. Строим сетевой график в календарной шкале времени по ранним и поздним срокам начала и окончания работ (график Ганта).
5. Просуммируем количество человек, которые выполняют одновременно разные виды работ, и построим эпюру интенсивности

потребления ресурсов. Запишем наверху каждой проекции – работы ее интенсивность, т.е. количество человек, которые должны выполнять работу.

Проецируем на ось времени ($0t$) начало и конец каждой работы и обозначим проекцию, совпадающую с началом координат через τ_0 , следующую за ней τ_1 и т.д.

6. Из анализа графика Ганта и эпюры интенсивности потребления видно, что существуют интервалы, для которых интенсивность ресурсов превышает их наличное количество, следовательно, график необходимо оптимизировать, установив порядок выполнения работ. В первую очередь выполняются работы, начатые в предыдущем промежутке. Оставшиеся работы нумеруют в порядке возрастания их полных резервов, а если резервы одинаковы, то в порядке убывания интенсивностей. Таким образом, выбираем работу, начало выполнения, которой сдвигаем к началу следующего временного промежутка и т.д. Если есть возможность выбора, то растягиваем сроки выполнения работы, задействовав на ее выполнении меньше работников.

Данный алгоритм выполняем до тех пор, пока не определим сроки начала и окончания работ, численность работников, занятых на всех работах при условии, что проект будет выполнен в заданные сроки.

Лекция 5

4.8. Задачи оптимизации сетей во времени

Необходимо обосновать минимальную величину дополнительных вложений x_{ij} в отдельные работы проекта с тем, чтобы общий срок его выполнения не превышал заданной величины времени t_0 .

Пусть задан сетевой график выполнения проекта. Продолжительность каждой работы равна t_{ij} . Вложение дополнительных средств x_{ij} в работу (i,j) сокращает время ее выполнения до t'_{ij} , причем $t'_{ij} = t_{ij} - k_{ij} \cdot x_{ij}$, где k_{ij} – технологический коэффициент использования дополнительных средств. Но время выполнения каждой работы можно сократить до минимально возможного времени ее выполнения. Требуется определить количество дополнительных средств x_{ij} , которые надо вложить в работы (i,j) , а также время начала t_{ij}^H и время окончания t_{ij}^O выполнения этих работ,

чтобы проект должен быть выполнен в срок. При этом суммарные дополнительные затраты должны быть минимальными.

Запишем *структурную модель задачи оптимизации проекта по времени*.

I. Индексация:

i – номер предыдущего события (начального события работы);

j – номер последующего события (конечного события работы);

r – номер промежуточного события;

n – номер завершающего события работы;

1 – номер исходного события работы;

E – множество вершин орграфа;

\vec{e} – множество дуг орграфа.

II. Неизвестные величины:

x_{ij} – величина дополнительных вложений в работу ij , позволяющая сократить время ее выполнения;

t_{ij}^H – время начала работы ij ;

t_{ij}^O – время окончания работы ij .

III. Известные величины:

t_0 – срок выполнения проекта;

d_{ij} – минимально возможное время выполнения работы ij ;

k_{ij} – технологический коэффициент использования дополнительных средств, позволяющих сократить время выполнения работы ij ;

t_{ij} – время выполнения работы ij .

Целевая функция модели – суммарные затраты дополнительных вложений должны быть минимальными:

$$F_{\min} = \sum_{(i,j) \in \vec{e}} x_{ij}.$$

При условиях:

1) по времени завершения проекта, т.е. время завершения проекта не должно превышать заданного времени –

$$t_{i,n}^0 \leq t_0, (i, n) \in \vec{e},$$

2) по продолжительности работ, т.е. продолжительность работы должна быть не менее минимально возможной ее продолжительности –

$$t_{ij}^0 - t_{ij}^H \geq d_{ij}, (i, j) \in \vec{e},$$

3) по сокращению времени продолжительности работ, т.е. в зависимости от величины вложенных средств, продолжительность работ может быть сокращена –

$$t_{ij}^0 - t_{ij}^H = k_{ij} x_{ij}, (i, j) \in \vec{e},$$

4) по последовательности выполнения работ, т.е. время начала выполнения каждой работы должно быть не меньше времени окончания непосредственно предшествующих ей работ –

$$a) t_{1,j}^H = 0, (1, j) \in \vec{A};$$

$$б) t_{jr}^H \geq t_{ij}^0, (i, j, r) \in E,$$

5) неотрицательность переменных –

$$t_{ij}^H, t_{ij}^0, x_{ij} \geq 0, (i, j) \in \vec{e}.$$

Требуется найти такие значения $t_{ij}^H, t_{ij}^0, x_{ij}$, чтобы:

1. суммарное количество дополнительных средств было минимальным;
2. время выполнения всего комплекса работ не превышала t_0 ;
3. продолжительность выполнения каждой работы была не меньше заданной величины d_{ij} .

Составляем развернутую задачу и решаем ее на компьютере симплексным методом.

Рассмотрим второй вариант этого типа задач.

Задача состоит в сокращении срока выполнения комплекса работ на столько, насколько это возможно за счет вложения определенной суммы дополнительных средств (т. е. не более B). Время выполнения каждой работы должно быть не меньше минимально возможного времени (d_{ij}). Требуется определить время начала t_{ij}^H и окончания t_{ij}^O каждой работы и величину дополнительных средств x_{ij} , которые необходимо выделить для сокращения продолжительности выполнения работы (ij).

Используем ранее изложенные три группы условных обозначений: индексацию, неизвестные и известные величины. В известные величины добавим величину B – количество дополнительных средств, выделяемых для сокращения продолжительности работ.

Структурная экономико-математическая модель имеет следующий вид:

$$F_{\min} = t_i^0, \quad (i,n) \in \vec{e}$$

т. е. срок выполнения проекта должен быть минимальным.

При условиях:

1. по использованию дополнительных вложений в работы, т.е. суммарное количество дополнительных вложений в работы проекта не должно превышать выделяемых для сокращения продолжительности работ дополнительных средств:

$$\sum_{(i,j) \in \vec{e}} x_{ij} \leq B,$$

2. по продолжительности работ:

$$t_{ij}^0 - t_{ij}^H \geq d_{ij}, (i, j) \in \vec{e},$$

3. по сокращению продолжительности работ:

$$t_{ij}^0 - t_{ij}^H = t_{ij} - k_{ij} x_{ij}, (i, j) \in \vec{e}.$$

4. по последовательности выполнения работ:

$$\text{а) } t_{1j}^H = 0, (1, j) \in E;$$

$$\text{б) } t_{jr}^H \geq t_{ij}^0, (i, j, r) \in E,$$

5. неотрицательность переменных:

$$t_{ij}^H, t_{ij}^0, x_{ij} \geq 0, (i, j) \in \vec{e}.$$

Требуется найти такие значения $t_{ij}^H, t_{ij}^0, x_{ij}$, чтобы:

1. срок выполнения комплекса работ был минимальным;
2. суммарное количество дополнительных средств не превышало значения B ;
3. продолжительность выполнения каждой работы не превышала заданной величины d_{ij} .

Лекция 6

4.9. Задачи оптимизации сетей по стоимости

Рассмотрим *минимизацию стоимости проекта при фиксированной его продолжительности*. Пусть задан сетевой график проекта $G = (E, \vec{e})$. Для каждой работы в скобках на графике дана минимальная

продолжительность работ d_{ij} (т.е. срочный режим их выполнения), которому соответствуют наибольшие затраты средств C_{ij} . Также известна нормальная (или наибольшая) продолжительность работ D_{ij} , которой соответствуют наименьшие затраты средств c_{ij} .

Известно, что затраты на выполнение отдельных работ находятся в обратной зависимости от продолжительности их выполнения.

Рассчитывают коэффициент дополнительных затрат $h_{ij} = \frac{C_{ij} - c_{ij}}{D_{ij} - d_{ij}}$,

который показывает насколько увеличится стоимость работы (i,j) при уменьшении ее продолжительности на единицу времени.

Алгоритм решения задачи:

1) рассчитаем коэффициент дополнительных затрат для каждой работы;

2) определим продолжительность критического пути, минимальную стоимость проекта.

3) определим продолжительность критического пути, максимальную стоимость проекта.

Требуется найти начало и окончание работ при необходимости выполнения проекта за фиксированное время t_0 при наименьшей стоимости затрат.

Целевую функцию экономико-математической задачи можно записать в следующем виде:

$$F_{\min} = \sum_{(i,j) \in E} [C_{ij} - h_{ij}(t_{ij}^0 - t_{ij}^H - d_{ij})].$$

При условиях:

1. по предельной продолжительности работ:

$$d_{ij} \leq t_{ij}^0 - t_{ij}^H \leq D_{ij}, (i, j) \in \vec{E}$$

2. по окончанию проекта:

$$t_{i,n}^0 \leq t_0, (i, n) \in \vec{E}$$

3. по последовательности выполнения работ:

$$а) t_{1,j}^H = 0, (i, j) \in E$$

$$б) t_{jr}^H \geq t_{ij}^0, (i, j, r) \in E.$$

4. неотрицательность переменных:

$$t_{jr}^H, t_{ij}^0 \geq 0, (i, j) \in \vec{E}.$$

4) составляем развернутую экономико-математическую задачу. Решаем ее на компьютере. Проанализируем результаты решения задачи.

5) Рассчитаем оптимальные затраты средств в работы (i, j) :

$$C_{ij\text{опт}} = C_{ij} - h_{ij}(t_{ij}^0 - t_{ij}^H - d_{ij}).$$

6) Определим количество средств, необходимых для сокращения срока выполнения проекта.

4.10. Варианты задачи о назначениях

Ставится задача так распределить исполнителей по работам, чтобы суммарная стоимость выполненных работ была максимальной или суммарные затраты на выполнение работ были минимальными.

Сформулированную таким образом задачу о назначениях можно представить как транспортную задачу, в которой исполнители соответствуют пунктам отправления груза A_i , а работы – пунктам назначения B_j . Если число исполнителей не равно числу работ, то открытая задача о назначениях должна быть приведена к закрытому виду путем ввода фиктивных исполнителей или фиктивных работ. Особенность задачи о назначениях состоит в том, что каждый исполнитель может работать только на одной работе, следовательно, $A_i=1; B_j=1$, и неизвестные величины задачи могут принимать только два значения:

$$x_{ij} = \begin{cases} 1 - \text{если } i - \text{й исполнитель выполняет работу вида } j \\ 0 - \text{в противном случае, } (i, j = \overline{1, n}; i \neq j) \end{cases}$$

Назначения исполнителей на работы x_{ij} должны удовлетворять ограничениям:

1. по распределению работ для выполнения

$$\sum_{i=1}^n x_{ij} = 1, j = \overline{1, n}$$

2. по распределению исполнителей по работам

$$\sum_{j=1}^n x_{ij} = 1, i = \overline{1, n}$$

3. неотрицательность, целочисленность переменных

$$x_{ij} \geq 0; x_{ij} \in \{0, 1\}.$$

При этом, в зависимости от постановки, требуется максимизировать или минимизировать целевую функцию:

$$F_{\max} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} \text{ или } F_{\min} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij},$$

где c_{ij} – соответственно производительность исполнителя вида i при выполнении работы вида j или затраты на выполнение работы вида j исполнителем вида i .

Для решения задачи используем «Поиск решения» в электронных таблицах Excel. При этом x_{ij} – двоичные, т. е. принимают значение 1 или 0.

Для решения задач о назначениях можно использовать *венгерский метод*, названный в честь венгерского математика Кёнига.

Алгоритм венгерского метода.

1. Среди элементов каждой строки матрицы c_{ij} (табл. 4.29) находим наименьшие элементы ($\min_j c_{ij}$).

2. Из элементов каждой строки матрицы c_{ij} вычитаем выбранные минимальные элементы, результаты заносим в новую матрицу (табл. 4.30).

3. Выбираем среди элементов каждого столбца новой матрицы наименьшие элементы ($\min_i c_{ij}$).

4. Из элементов каждого столбца матрицы почленно вычитаем выбранные минимальные элементы, в результате получим таблицу, в которой каждая строка и каждый столбец содержит хотя бы по одному нулевому значению.

5. Используя нулевые значения элементов, получаем допустимое решение задачи о назначениях.

6. Анализируя допустимое решение, находим оптимальное решение задачи о назначениях.

4.11. Задача коммивояжера и ее приложения

Задачей схожей с задачей о назначениях является *задача о коммивояжере*, постановка которой состоит в том, что имеется n городов, которые должен посетить коммивояжер только один раз, выезжал из первого города и возвращаясь в исходный пункт.

Цель решения задачи – найти кратчайший замкнутый маршрут, проходящий через каждый пункт только один раз, называемый полным циклом.

Введем неизвестные величины:

$$x_{ij} = \begin{cases} 1 - \text{если в маршрут входит перезд из города } i \text{ в город } j; \\ 0 - \text{в противном случае.} \end{cases}$$

Требуется минимизировать маршрут коммивояжера:

$$F_{\min} = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij},$$

где c_{ij} – расстояние между городом i и городом j .

При условиях:

$$1. \text{ По въезду в города: } \sum_{i=1}^n x_{ij} = 1, j = \overline{1, n}.$$

$$2. \text{ По выезду из городов: } \sum_{j=1}^n x_{ij} = 1, i = \overline{1, n}$$

3. Неотрицательность, целочисленность переменных

$$x_{ij} \geq 0, x_{ij} \in \{0, 1\}.$$

Основными методами решения задачи коммивояжера являются методы решения целочисленных задач линейного программирования: методы ветвей и границ и отсекающих плоскостей.

Идея применения метода отсекающих плоскостей для решения задачи коммивояжера состоит в том, чтобы в первоначальную задачу ввести дополнительные ограничения, позволяющие исключить возможность разрыва пути коммивояжера и появления нескольких не связанных между собой подмаршрутов. Общее количество таких ограничений равно $(n-1) \cdot (n-2)$. Они имеют вид:

по формированию маршрута коммивояжера –

$$u_i - u_j + n x_{ij} \leq n - 1, i, j = \overline{2, n}, i \neq j.$$

Основным недостатком применения метода отсекающих плоскостей к решению данной задачи является увеличение размерности задачи при введении дополнительных ограничений.

Лекция 7

Тема 5. ЗАДАЧИ ОПТИМАЛЬНОГО УПОРЯДОЧЕНИЯ

5.1. Понятие теории расписаний. Классификация задач теории расписаний.

- 5.2. Системы с одним обслуживающим устройством.
5.3. Последовательное обслуживание (Общая задача Джонсона.
Задача Джонсона для двух и трех машин).

5.1. Понятие теории расписаний.

Классификация задач теории расписаний

Теория расписаний возникла на базе оперативно-календарного планирования производства в начале 20 в. Ее основоположником считается Гант, впервые предложивший оптимизировать планирование с помощью специальных графиков (Гант-Карт).

Под *операцией* в теории расписаний понимают какое-либо действие, направленное на достижение цели. Операции могут производиться над деталями, узлами, машинами, работами, которые принято называть *требованиями*. *Операция* – это детализированное мероприятие. На производстве с точки зрения технологии часто бывает безразлично, в каком порядке выполняются те или иные операции, но в интересах конкретного исполнителя этот порядок играет важную роль. Это вызвано приоритетностью заказов, затратами, связанными с различным порядком их выполнения на имеющемся оборудовании. Раздел исследования операций, который изучает эффективность выполнения операций в зависимости от порядка их следования, называется *теорией расписания*.

Операции выполняются на машинах или оборудовании, которые принято называть обслуживающим устройством. Под машиной или оборудованием понимают любое обслуживающее устройство, способное выполнять операцию. Множество машин, которые могут выполнять некоторое множество операций, называются *системой обслуживания*. Операции назначаются на машины согласно некоторой дисциплине обслуживания. Совокупность машин, операций и дисциплин назначения операций на машины называются *процессом обслуживания*. Для него составляется расписание, т.е. порядок обслуживания требований обслуживающим устройством. Дисциплина обслуживания, т.е. система выполнения работ на машинах (требований на обслуживающих устройствах) может быть конвейерной, случайной или произвольной.

В основном модели, рассматриваемые в теории расписаний, могут быть отнесены к классу детерминированных задач, т.е. наилучшее решение принимается в условиях определенности, когда четко известны операции и все значения неконтролируемых факторов.

В общем случае для задачи упорядочения должны быть известны:

- 1) подлежащие выполнению операции;
- 2) количество и типы обслуживающих устройств;
- 3) трудоемкость или время и порядок выполнения операций;
- 4) критерий эффективности расписания.

Задачи упорядочения различаются числом выполняемых операций, характером поступления требований в систему обслуживания (одновременно или в некоторые фиксированные моменты времени), количеством и последовательностью участия обслуживающих устройств в выполнении конкретных операций (системы с одним, двумя, тремя и более обслуживающими устройствами).

В настоящее время более глубоко изучены модели простых процессов обслуживания. Процесс называется *простым*, если:

- 1) каждое обслуживающее устройство может быть назначено в любой момент времени;
- 2) работы представляют строго упорядоченную последовательность требований. Для заданного (конкретного) требования существует не более одного требования, непосредственно следующего за ним, и не более одного, непосредственно предшествующего ему;
- 3) каждое требование обслуживается только на одном обслуживающем устройстве;
- 4) имеется только по одному обслуживающему устройству каждого вида;
- 5) отсутствует прерывание операций обслуживания;
- 6) одновременно не может реализоваться более одной операции одной и той же работы;
- 7) в каждый момент времени обслуживающее устройство может выполнять не более одной операции.

5.2. Системы с одним обслуживающим устройством

На практике выделяют три вида задач оптимального упорядочения систем с одним обслуживающим устройством. Дадим описание данной системы обслуживания. В систему, состоящую из одного обслуживающего устройства, для выполнения поступает конечное множество требований $N=\{1,2,\dots,n\}$. Предполагается, что все эти требования поступают одновременно в нулевой момент времени. Под требованием подразумевается любой объект, например, детали, обрабатываемые на одном станке; сельскохозяйственные операции,

выполняемые одним механизмом, работы, выполняемые бригадой строителей и т. д.

Задача состоит в том, чтобы указать расписание обслуживания требований, доставляющее оптимум тому или другому критерию эффективности.

Под *расписанием* понимают такое предписание, по которому в каждый момент времени можно установить, простаивает обслуживающее устройство или нет. Если оно не простаивает, то можно указать, какое из требований оно обслуживает. Т.е. *расписание* – это последовательность выбора требований на обслуживание, которое обозначается $\pi(n) = (i_1, i_2 \dots i_n)$, где i_2 – элемент из множества N , занимающий в последовательности $\pi(n)$ второе место.

Предполагают, что обслуживающее устройство всегда готово для обслуживания требований. И если имеются ожидающие обслуживания требования (т. е. очередь), то обслуживающее устройство не простаивает. Обслуживание каждого последующего требования начинается сразу после окончания обслуживания предыдущего. Если нужна настройка обслуживающего устройства, то ее продолжительность присоединяется к длительности обслуживания требования.

Прерывание процесса обслуживания требования не допускается, т. е. требование, начав обслуживаться, будет занимать обслуживающее устройство до тех пор, пока не будет полностью обслужено. Задачи теории расписаний оцениваются определенным критерием эффективности.

I. По определению системы обслуживания все требования поступают в систему одновременно в нулевой момент времени и одно из требований обслуживается, а остальные стоят в очереди, следовательно, *первая задача теории расписаний состоит в том, чтобы минимизировать суммарный штраф, связанный с ожиданием всех требований в очереди.*

Очевидно, что суммарная длительность обслуживания всех требований равна сумме обслуживания каждого требования. Но время обслуживания каждого требования заранее известно и, следовательно, суммарная длительность обслуживания всех требований одинакова для всех $n!$ возможных расписаний. Поэтому в качестве критерия эффективности непригодна данная величина. В качестве критерия нельзя принимать максимальное или минимальное количество

требований в системе, так как эти критерии не зависят от порядка обслуживания требований.

Введем условные обозначения.

t_i – продолжительность обслуживания требования вида i ;

\underline{t}_i – время начала обслуживания требований вида i ;

\bar{t}_i – время окончания обслуживания требований вида i , $i \in N$.

Пусть d_i – штраф за ожидание требования вида i в очереди в течение единицы времени. Тогда суммарный штраф, связанный с ожиданием всех требований в очереди будет зависеть от расписания и для заданного расписания. $\Phi_1(\pi_n)$ будет вычисляться по формуле:

$$\Phi_1(\pi_n) = \sum_{i=1}^n d_i \cdot \underline{t}_i.$$

Первая задача состоит в том, чтобы построить такое расписание, при котором $\Phi_1(\pi_n)$ критерий эффективности будет минимальным.

Рассмотрим *алгоритм решения первой задачи*.

Для решения первой задачи, т.е. построения такого расписания, которое минимизирует критерий эффективности $\Phi_1(\pi_n) = \sum_{i=1}^n d_i \cdot \underline{t}_i$

необходимо для всех требований вычислить отношение $\frac{t_i}{d_i}$, $i \in N$ и

упорядочить требования по неубыванию этих отношений, т. е.:

$$\frac{t_i}{d_i} \leq \frac{t_{i+1}}{d_{i+1}}, \quad i \in N.$$

Алгоритм решение задачи.

Для информации задачи определим:

а) время начала обслуживания требований по формуле:

$$\underline{t}_{i+1} = \underline{t}_i + t_i,$$

т. е. время начала обслуживания последующего требования равно сумме времени начала обслуживания предыдущего требования и непосредственного времени его обслуживания;

б) определим время начала обслуживания требований при неоптимальном порядке обслуживания.

в) найдем штраф, связанный с ожиданием каждого требования в очереди при неоптимальном порядке их обслуживания:

$$\alpha_i \cdot \underline{t}_i;$$

г) рассчитаем суммарный штраф:

$$\sum_{i=1}^n \alpha_i \cdot \check{t}_i;$$

д) вычислим для каждого требования отношение:

$$\frac{t_i}{\alpha_i};$$

е) упорядочим требования по неубыванию этих отношений:

$$\frac{t_i}{\alpha_i} \leq \frac{t_{i+1}}{\alpha_{i+1}}, \quad i \in N;$$

и определим оптимальный порядок обслуживания требований;

ж) найдем время начала обслуживания требований для оптимального порядка их обслуживания:

$$(\check{t}_{i+1} = \check{t}_i + t_i);$$

з) рассчитаем штраф, связанный с ожиданием каждого требования в очереди при оптимальном порядке их обслуживания:

$$\alpha_i \check{t}_i;$$

и) определим суммарный штраф:

$$\sum_{i=1}^n \alpha_i \cdot \check{t}_i.$$

к) определим эффект от оптимизации расписания обслуживания требований для первой задачи.

II. Рассмотрим вторую задачу теории расписаний системы с одним обслуживающим устройством.

Допустим, после завершения обслуживания требование остается в системе и ожидает до тех пор, пока не будет обслужено последнее из множества N , т. е. до момента времени:

$$T = \sum_{i=1}^n t_i.$$

Если γ_i – количество средств, связываемых требованием вида i в единицу времени после завершения его обслуживания, то суммарное количество средств связанных всеми требованиями после завершения их обслуживания будет равно:

$$\Phi_2(\pi_n) = \sum_{i=1}^n \gamma_i (T - \bar{t}_i),$$

где \bar{t}_i – время окончания обслуживания i -го требования. Оно равно сумме времени начала его обслуживания и времени его непосредственного обслуживания:

$$\bar{t}_i = \underline{t}_i + t_i.$$

Вторая задача состоит в том, чтобы построить расписание минимизирующее суммарную величину средств, связываемых требованиями в связи с их пребыванием в системе после завершения обслуживания.

Для решения второй задачи необходимо упорядочить требования по невозрастанию отношений $\frac{t_i}{\gamma_i}$, $i \in N$, т. е.:

$$\frac{t_i}{\gamma_i} \geq \frac{t_{i+1}}{\gamma_{i+1}}, \quad i \in N.$$

Решение задачи:

а) определим время окончания обслуживания каждого требования при неоптимальном порядке их обслуживания:

$$\bar{t}_i = \underline{t}_i + t_i.$$

б) рассчитываем время окончания обслуживания всех требований в системе по формуле:

$$T = \sum_{i=1}^n t_i.$$

в) определим время ожидания каждого требования в системе после завершения его обслуживания при неоптимальном порядке обслуживания:

$$T - \bar{t}_i.$$

г) найдем количество средств, связываемых каждым требованием после завершения обслуживания при неоптимальном порядке их обслуживания:

$$\gamma_i(T - \bar{t}_i);$$

д) рассчитаем суммарное количество средств, связываемых требованиями после их обслуживания при неоптимальном порядке их обслуживания:

$$\sum_{i=1}^n \gamma_i(T - \bar{t}_i).$$

е) вычислим для каждого требования отношение:

$$\frac{t_i}{\gamma_i}$$

ж) упорядочим требования по невозрастанию этих отношений:

$$\frac{t_i}{\gamma_i} \geq \frac{t_{i+1}}{\gamma_{i+1}}, \quad i \in N$$

и установим оптимальный порядок обслуживания требований

з) найдем время ожидания каждого требования в системе после завершения его обслуживания при оптимальном порядке обслуживания требований:

$$T - \bar{t}_i;$$

и) рассчитаем количество средств, связываемых каждым требованием после завершения обслуживания при условии оптимального порядка их обслуживания:

$$\gamma_i(T - \bar{t}_i);$$

к) определим суммарное количество средств, связываемых требованиями после их обслуживания при оптимальном порядке обслуживания требований:

$$\sum_{i=1}^n \gamma_i(T - \bar{t}_i).$$

л) определим эффект от оптимизации расписания обслуживания требований для второй задачи.

III. Рассмотрим третью задачу оптимального упорядочения.

Допустим D_i – директивный срок обслуживания требования вида i , $i \in N$, т.е. это момент времени, к которому желательно завершить процесс обслуживания этого требования. Но не всегда удастся построить расписание, при котором обслуживание каждого требования вида i будет завершено не позднее директивного срока D_i . Тогда величина z_i будет обозначать задержку (т.е. превышение директивного срока окончания обслуживания) в обслуживании требования вида i по сравнению с его директивным сроком:

$$z_i = \max\{0, \bar{t}_i - D_i\}, \quad i \in N.$$

Пусть δ_i – штраф за задержку в обслуживании требования вида i на единицу времени. Тогда, критерий эффективности, связанный с задержкой требований, запишем так:

$$\Phi_3(\pi_n) = \max_{i \in N} \delta_i \cdot z_i.$$

Он позволяет вычислять величину максимального штрафа связанного с задержкой обслуживания требований.

Третья задача состоит в том, чтобы построить такое расписание π_n , которое будет доставлять минимум критерию эффективности $\Phi_3(\pi_n)$.

Третья задача построения расписания решается с помощью следующего алгоритма:

1. вычисляем время окончания обслуживания всех требований по формуле:

$$T = \sum_{i=1}^n t_i ;$$

2. среди всех неупорядоченных требований находим такое требование с номером ℓ , для которого будет выполняться условия:

$$\delta_\ell \cdot z_\ell = \min \delta_\ell z'_\ell ,$$

где z'_ℓ – задержка в обслуживании требования по сравнению с его директивным сроком при допущении, что последним обслуживается рассматриваемое требование.

При этом z'_ℓ определяется по формуле:

$$z'_\ell = \max\{0; T - D_i\},$$

а минимальное значение произведения $\delta_\ell z'_\ell$ вычисляется только по множеству неупорядоченных требований;

3. найденное требование с номером ℓ ставим выполняться последним среди рассматриваемого множества. Исключаем требование с номером ℓ из рассмотрения. Если множество оставшихся требований не пусто, то принимаем T равным $T - t_\ell$ и переходим к шагу 2, в противном случае – найден оптимальный порядок обслуживания требований.

Решение задачи. Рассчитаем параметры системы при неоптимальном порядке обслуживания требований:

а) определим время окончания обслуживания каждого требования ($\bar{t}_i = \underline{t}_i + t_i$) при неоптимальном порядке их обслуживания. Расчеты заносим в табл.;

б) вычислим задержку в обслуживании каждого требования по сравнению с его директивным сроком, используя формулу:

$$z_i = \bar{t}_i - D_i.$$

При этом, если $\bar{t}_i < D_i$, то z_i принимаем равным нулю;

в) находим штраф за задержку в обслуживании каждого требования при неоптимальном порядке обслуживания:

$$\delta_i z_i;$$

г) определим значение критерия эффективности задачи при неоптимальном расписании:

$$\Phi_3(\pi_n) = \max_{i \in N} \delta_i \cdot z_i$$

д) оптимизируем порядок обслуживания требований, для этого сначала вычислим время обслуживания всех требований в системе (или время обслуживания последнего требования):

$$T = \sum_{i=1}^n t_i.$$

е) определим задержку в обслуживании каждого требования по сравнению с его директивным сроком при условии, что последним обслуживается рассматриваемое требование с номером i . Данную задержку в обслуживании определяем по формуле:

$$z'_i = \max\{0; T - D_i\}$$

ж) рассчитаем штраф за задержку в обслуживании каждого требования при условии, что оно обслуживается последним:

$$\delta_i z'_i;$$

з) находим требование, для которого произведение $\delta_i z'_i$ самое минимальное. Исключаем его из рассмотрения и т.д.

и) определяем порядок обслуживания требований, которому соответствует минимальное значение Φ_3 ;

к) используя методику пунктов а, б, в, г, определим значение критерия эффективности задачи при оптимальном расписании:

$$\Phi_3(\pi_n) = \max_{i \in N} \delta_i \cdot z_i$$

л) рассчитаем эффект от оптимизации расписания обслуживания требований.

Для наглядности расписание обслуживания требований в системе с одним обслуживающим устройством можно представить графически.

Для этих целей наиболее часто пользуются графиком Ганта, который представляет собой программу занятости обслуживающего

устройства. Каждая операция на графике изображается масштабированным отрезком прямой, соответствующим длительности обслуживания требования.

Кроме графика Ганта для наглядного изображения расписания используются планировочный и ленточные графики. Планировочный график позволяет схематически изобразить процесс обслуживания требований. При этом каждому требованию отводится одна строка, в которой проводится линия, соответствующая времени обслуживания требования в выбранном масштабе времени.

Ленточный график – это изображение последовательности обслуживания требований, время обслуживания которых на графике изображается последовательно с учетом длительности их обслуживания.

5.3. Последовательное обслуживание (Общая задача Джонсона. Задача Джонсона для двух и трех машин)

Общая задача Джонсона. Типичной задачей теории расписаний является проблема составления расписания работы технологической линии, состоящей из m -станков ($i = 1, \bar{m}$), на которых нужно обработать партии из n -деталей ($j = 1, \bar{n}$). Критерием эффективности расписания станет минимальное время обработки всех n -деталей, каждая из которых должна последовательно пройти обработку на каждом станке. Исходными данными задачи служит продолжительность обработки на i -ом станке j -ой детали t_{ij} .

Т.е. необходимо определить порядок обработки n -деталей, минимизирующий общее время их изготовления.

При условиях:

- 1) обработка каждой детали на i -м станке должна начинаться не ранее, чем окончится на станке $i-1$;
- 2) на каждом станке одновременно может обрабатываться не более одной детали;
- 3) начавшаяся операция не прерывается до полного ее завершения.

Такая задача получила название задачи Джонсона. В 1954г. американский ученый С. М. Джонсон сформулировал и решил ее для двух станков.

Задача Джонсона для двух машин.

Рассмотрим алгоритм решения задачи.

1. Запишем матрицу размерностью $n \times 2$ коэффициенты которой равны времени обслуживания требований t_{ij} первым и вторым устройствами.

2. В матрице $\|t_{ij}\|$ находим минимальный элемент. Если он находится в первом столбце, соответствующим первому обслуживаемому устройству, то данное требование обслуживается первым, если во втором столбце, то – последним.

3. Исключаем из рассмотрения выбранное требование и работу продолжаем согласно п. 2, пока не упорядочим порядок обслуживания требований.

4. Если в одном столбце имеются несколько минимальных величин, то выбираем сначала требование с меньшим номером.

5. Если и в первом и во втором столбце есть несколько одинаковых минимальных величин, то выбираем сначала требования с первого столбца.

6. Определяем суммарный простой двух обслуживающих устройств при неоптимальном и оптимальном расписании.

7. Изобразим с помощью графика Ганта оптимальное расписание.

Задача Джонсона для трех машин или трех обслуживающих устройств.

При $m \geq 3$ алгоритм Джонсона не пригоден, а простых алгоритмов пока не найдено. При $m = 3$ можно воспользоваться алгоритмом Джонсона для двух обслуживающих устройств, так как оптимальный план задачи Джонсона произвольной размерности достигим на множество планов, в которых последовательность запуска деталей на первом станке совпадает с последовательностью обработки деталей на втором станке, а последовательность на последнем станке – с последовательностью на предпоследнем станке.

Отсюда следует, что для трех станков последовательность обработки деталей на всех станках одинакова. Джонсон доказал, что алгоритм двух станков можно применить к трем станкам в случае, если:

$$\min a_j \geq \max b_j \text{ или } \min c_j \geq \max b_j,$$

где a_j – время обработки j -ой детали на первом станке;

b_j – время обработки j -ой детали на втором станке;

c_j – время обработки j -ой детали на третьем станке.

Алгоритм решения задачи следующий: оптимальный порядок обслуживания требований находят с помощью сумм:

$$a_j + b_j \text{ и } b_j + c_j.$$

К ним применяют алгоритм Джонсона для двух станков.

Лекция 8

Тема 6. МОДЕЛИ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

6.1. Общая характеристика системы массового обслуживания (понятие системы массового обслуживания, ее элементы, классификация систем).

6.2. Потоки событий и предельные вероятности состояний системы (уравнения Колмогорова).

6.3. Процесс гибели и размножения.

6.4. Показатели эффективности работы системы массового обслуживания.

6.5. Одноканальная система массового обслуживания с отказами.

6.6. Одноканальная система массового обслуживания с ожиданием и ограничением очереди.

6.7. Одноканальная система массового обслуживания с ожиданием без ограничения очереди.

6.8. Многоканальная система массового обслуживания с отказами.

6.9. Многоканальная система массового обслуживания с ожиданием и неограниченной очередью.

6.10. Многоканальная система массового обслуживания с ожиданием и ограниченной очередью.

6.11. Замкнутая система массового обслуживания.

6.1. Общая характеристика системы массового обслуживания (понятие системы массового обслуживания, ее элементы, классификация систем)

Теория массового обслуживания начала развиваться в начале 20 века. В 1907 г. Йоханнсен сформулировал основные предпосылки новой теории. В 1909 г. А. К. Эрланг применил теорию вероятностей к исследованию зависимости обслуживания телефонных вызовов от числа поступающих на телефонную станцию вызовов. В 1963 г. А. Я. Хинчин, советский математик, систематизировал основные положения теории массового обслуживания в монографии «Работы по математической теории массового обслуживания».

Теория массового обслуживания представляет собой научное направление, изучающее системы, в которых возникают массовые запросы на выполнение определенных услуг и происходит удовлетворение этих запросов.

Предметом теории массового обслуживания является построение математических моделей, связывающих условия работы системы с показателями эффективности функционирования с целью определения наилучших вариантов управления данными системами.

Эти системы получили название *систем массового обслуживания*, а процессы, возникающие при этом, *называются процессами обслуживания*.

Элементами системы массового обслуживания являются входящий поток заявок, очередь, поток необслуженных (покинувших очередь) заявок, каналы обслуживания, выходящий поток обслуженных заявок.

Обслуживающие устройства системы массового обслуживания (пункты, станции, приборы, устройства, кассовые аппараты, продавцы, телефонные линии связи и т.д.) называются *каналами обслуживания*.

Заявка (требование) – это запрос на выполнение каких-либо услуг или удовлетворение определенной потребности.

Под входящим потоком заявок (требований) понимают последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени.

Если поток событий имеет свойства стационарности, ординарности и отсутствия последствия, то он является *простейшим или пуассоновским потоком*.

Если вероятностные характеристики потока событий не зависят от времени, т.е. он имеет постоянную интенсивность, то такой поток называется *стационарным*. Если события происходят поодиночке, а не два и более сразу, то такой поток событий является *ординарным*. В *потоке без последствий* события появляются в последовательные моменты времени независимо друг от друга, т. е. число событий, попадающих на любой один из произвольно выбранных промежутков времени, не зависит от числа событий, попавших на другой, произвольно выбранный промежуток времени.

Под потоком необслуженных заявок понимают заявки, поступившие в систему в тот момент, когда все каналы заняты и заявки получают отказ в обслуживании, покидают систему массового обслуживания и в дальнейшем процессе обслуживания не участвуют.

Заявки, стоящие в очереди могут обслуживаться любым освободившимся каналом обслуживания. В системах с последовательным расположением каналов каждый канал выступает как отдельная одноканальная система массового обслуживания или фаза обслуживания, т.е. выходной поток обслуженных заявок одним каналом обслуживания является входным потоком для последующего канала.

В зависимости от дисциплины очереди системы массового обслуживания подразделяются на системы *с приоритетами и без приоритетов*. При этом под *дисциплиной очереди* понимают порядок, который принят при поступлении заявок из очереди в канал обслуживания. Правило отбора требований на обслуживание может производиться в виде: случайного отбора; по критерию приоритетности; первый пришел – первый обслужен; последний пришел – первый обслужен.

По количеству этапов обслуживания системы массового обслуживания подразделяются на *однофазные* и *многофазные* системы. *Однофазными* называются те системы, которые выполняют одну и ту же операцию обслуживания. *Многофазные* системы массового обслуживания имеют неоднородные каналы, выполняющие разные операции обслуживания, осуществляемые с помощью последовательно расположенных каналов обслуживания.

6.2. Потоки событий и предельные вероятности состояний системы (уравнения Колмогорова)

Напомним, что потоком событий называется последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени. Поток характеризуется *интенсивностью* λ , под которой понимают частоту появления событий или среднее число событий, поступающих в систему массового обслуживания в единицу времени. Поток событий называется *регулярным*, если события следуют одно за другим через определенные равные промежутки времени. Простейший (пуассоновский) поток характеризуется стационарностью, ординарностью и отсутствием последствия.

При этом регулярный поток не является простейшим, так как моменты появления событий в таком потоке жестко зафиксированы и он обладает последствием.

Для простейшего потока число m событий, попадающих на произвольный участок времени τ , распределено по закону Пуассона:

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

для которого математическое ожидание случайной величины равно ее дисперсии:

$$\alpha = \sigma^2 = \lambda\tau.$$

А вероятность того, что за время τ не произойдет ни одного события ($m = 0$), равна:

$$p_0(\tau) = e^{-\lambda\tau}.$$

Найдем вероятность того, что на участке времени длиной t не появится ни одного из последующих событий. Она равна:

$$p(T \geq t) = e^{-\lambda t},$$

а вероятность противоположного события, т.е. функция распределения случайной величины T равна:

$$F(t) = p(T < t) = 1 - e^{-\lambda t}.$$

Плотность вероятности случайной величины равна производной ее функции распределения:

$$\varphi(t) = F'(t) = \lambda e^{-\lambda t}.$$

Данное распределение, т.е. интервал времени между двумя соседними произвольными событиями, называется *показательным или экспоненциальным*, для которого математическое ожидание равно среднему квадратическому отклонению случайной величины и обратно по величине интенсивности потока λ :

$$\alpha = \sigma = \frac{1}{\lambda}.$$

Важное свойство показательного распределения состоит в том, что для интервала времени T между двумя последовательными соседними событиями потока, любые данные о том, сколько времени протекал этот интервал, не влияют на закон распределения оставшейся части.

Согласно распределения случайной величины T вероятность попадания на элементарный (малый) отрезок времени Δt хотя бы одного события потока равна:

$$p_{\Delta t} = p(T < \Delta t) = 1 - e^{-\lambda\Delta t} \approx \lambda\Delta t.$$

Рассмотрим процесс работы простейшей системы массового обслуживания.

Допустим система S состоит из двух каналов обслуживания, каждый из которых в случайный момент времени может выйти из строя, после чего мгновенно начинается ремонт канала, продолжающийся неизвестное случайное время.

Возможны следующие состояния системы:

S_0 – оба канала исправны;

S_1 – первый канал ремонтируется, второй исправен;

S_2 – второй канал ремонтируется, первый исправен;

S_3 – оба канала ремонтируются.

Стрелка, направленная, например, из состояния S_0 в S_1 означает переход системы в момент отказа первого канала, из S_1 в S_0 – переход в момент окончания ремонта этого канала. Так как каналы независимы друг от друга, то вероятностью одновременного выхода из строя двух каналов (переход из S_0 в S_3) или одновременного окончания ремонта двух каналов (переход из S_3 в S_0) пренебрежем.

Предположим, что все переходы системы из состояния S_i в S_j происходят под воздействием простейших потоков событий с интенсивностями:

$$\lambda_{ij} \quad (i = \overline{0,3}; j = \overline{0,3}) \text{ с вероятностью } p_i.$$

Вероятность i -го состояния называется вероятностью $p_i(t)$ того, что в момент времени t система будет находиться в состоянии S_i . При этом сумма вероятностей всех состояний системы для любого момента t равна единице:

$$\sum_{i=0}^3 p_i(t) = 1.$$

Зададим малый промежуток Δt и найдем вероятность $p_0(t + \Delta t)$ того, что система в момент $t + \Delta t$ будет находиться в состоянии S_0 .

Рассмотрим данную систему:

1. Система в момент t с вероятностью $p_0(t)$ находилась в состоянии S_0 и за время Δt не вышла из него. Вывести систему из состояния S_0 можно суммарным простейшим потоком с интенсивностью $(\lambda_{01} + \lambda_{02})$ с вероятностью, приближенно равной:

$$(\lambda_{01} + \lambda_{02})\Delta t.$$

А вероятность того, что система останется в состоянии S_0 равна:

$$[1 - (\lambda_{01} + \lambda_{02})\Delta t].$$

Применяя теорему умножения вероятностей, получим:

$$p_0(t)[1 - (\lambda_{01} + \lambda_{02})\Delta t].$$

2. Система в момент t с вероятностями $p_1(t)$ или $p_2(t)$ находилась в состоянии S_1 и S_2 и за время Δt перешла в состояние S_0 . Система может перейти в состояние S_0 в результате действия потока интенсивностью λ_{10} или λ_{20} с вероятностью $\approx \lambda_{10}\Delta t$ или $\lambda_{20}\Delta t$. Вероятность того, что система будет находиться в состоянии S_0 равна:

$$p_1(t)\lambda_{10}\Delta t \text{ или } p_2(t)\lambda_{20}\Delta t.$$

Применяя теорему сложения вероятностей, получим:

$$p_0(t + \Delta t) = p_1(t)\lambda_{10}\Delta t + p_2(t)\lambda_{20}\Delta t + p_0(t)[1 - (\lambda_{01} + \lambda_{02})\Delta t].$$

Отсюда найдем:

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = p_1(t)\lambda_{10} + p_2(t)\lambda_{20} - (\lambda_{01} + \lambda_{02})p_0(t).$$

При $\Delta t \rightarrow 0$, переходя к пределу, получим в левой части уравнения производную $p'_0(t)$. Обозначим ее как p'_0 :

$$p'_0 = \lambda_{10}p_1 + \lambda_{20}p_2 - (\lambda_{01} + \lambda_{02})p_0.$$

Получили дифференциальное уравнение первого порядка, содержащее и неизвестную функцию и ее производную первого порядка.

Рассуждая аналогичным образом для других состояний системы S , получим *систему дифференциальных уравнений Колмогорова* для вероятностей состояний:

$$\begin{cases} p'_0 = \lambda_{10}p_1 + \lambda_{20}p_2 - (\lambda_{01} + \lambda_{02})p_0, \\ p'_1 = \lambda_{01}p_0 + \lambda_{31}p_3 - (\lambda_{10} + \lambda_{13})p_1, \\ p'_2 = \lambda_{02}p_0 + \lambda_{32}p_3 - (\lambda_{20} + \lambda_{23})p_2, \\ p'_3 = \lambda_{13}p_1 + \lambda_{23}p_2 - (\lambda_{31} + \lambda_{32})p_3. \end{cases}$$

Уравнения Колмогорова составлены по следующему правилу: в левой части каждого уравнения стоит производная вероятности i -го состояния, в правой части – сумма произведений вероятностей всех состояний, из которых идут стрелки в данное состояние, на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного состояния вида i .

Для решения данной системы в нее добавляют следующее уравнение:

$$\sum_{i=0}^3 p_i(t) = 1.$$

Задают начальные условия, например $t=0$, тогда система находится в состоянии S_0 , т. е.:

$$p_0(0) = 1; p_1(0) = 0; p_2(0) = 0; p_3(0) = 0.$$

Уравнения Колмогорова позволяют найти все вероятности состояний системы как функции времени. В теории случайных процессов доказано, что если число состояний системы конечно и из каждого из них можно за конечное число шагов перейти в любое другое состояние, то предельные вероятности существуют и они постоянны, поэтому в уравнениях Колмогорова их производные можно заменить на нули:

$$\begin{cases} (\lambda_{01} + \lambda_{02})p_0 = \lambda_{10}p_1 + \lambda_{20}p_2, \\ (\lambda_{10} + \lambda_{13})p_1 = \lambda_{01}p_0 + \lambda_{31}p_3, \\ (\lambda_{20} + \lambda_{23})p_2 = \lambda_{02}p_0 + \lambda_{32}p_3, \\ (\lambda_{31} + \lambda_{32})p_3 = \lambda_{13}p_1 + \lambda_{23}p_2. \end{cases}$$

Данная система уравнений составлена по следующему правилу: слева в уравнениях стоит предельная вероятность данного состояния p_i , умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния, а справа – сумма произведений интенсивностей всех потоков, входящих в i -е состояние, на вероятности тех состояний, из которых эти потоки исходят.

6.3. Процесс гибели и размножения

Используя процесс гибели и размножения можно изучить множество состояний системы массового обслуживания. Термин «процесс гибели и размножения» связан с решением ряда биологических задач, в которых моделируется изменение численности биологических популяций. Рассмотрим упорядоченное множество состояний системы S_0, S_1, \dots, S_n .

Переходы от одного состояния системы в другое могут осуществляться из любого состояния только в состояния с соседними номерами. Допустим, что все потоки событий, переводящие систему из одного состояния в другое, являются простейшими с интенсивность λ .

Составим и решим уравнения для предельных вероятностей состояний. Например, для состояния S_0 уравнение имеет вид:

$$\lambda_{01}P_0 = \lambda_{10}P_1,$$

а для состояния S_1 –

$$(\lambda_{12} + \lambda_{10})P_1 = \lambda_{01}P_0 + \lambda_{21}P_2,$$

подставив в него уравнение $\lambda_{01}P_0 = \lambda_{10}P_1$ и преобразовав, получим:

$$\lambda_{12}P_1 = \lambda_{21}P_2.$$

Аналогично рассуждая, получим следующую систему уравнений:

$$\begin{cases} \lambda_{01}P_0 = \lambda_{10}P_1, \\ \lambda_{12}P_1 = \lambda_{21}P_2, \\ \dots \\ \lambda_{k-1,k} P_{k-1} = \lambda_{k,k+1} P_k, \\ \dots \\ \lambda_{n-1,n} P_{n-1} = \lambda_{n,n+1} P_n. \end{cases}$$

Добавим к системе нормировочное уравнение:

$$P_0 + P_1 + P_2 + \dots + P_k + \dots + P_n = 1.$$

Решая систему уравнений, получим:

$$P_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1,n} \dots \lambda_{12}\lambda_{01}}{\lambda_{n,n-1} \dots \lambda_{21}\lambda_{10}} \right)^{-1},$$

$$P_1 = \frac{\lambda_{01}}{\lambda_{10}} P_0,$$

$$P_2 = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} P_0,$$

$$\dots$$

$$P_n = \frac{\lambda_{n-1,n} \dots \lambda_{12}\lambda_{01}}{\lambda_{n,n-1} \dots \lambda_{21}\lambda_{10}} P_0.$$

Видно, что в формуле расчета p_1, p_2, \dots, p_n коэффициенты при p_0 есть слагаемые, стоящие после единицы в формуле расчета p_0 . Числители этих коэффициентов равны произведению всех интенсивностей, стоящих у стрелок, ведущих слева направо до данного состояния, а знаменатели – произведение всех

интенсивностей, стоящих у стрелок, ведущих справа налево до данного состояния.

6.4. Показатели эффективности работы системы массового обслуживания

Целью теории массового обслуживания является выработка рекомендаций по рациональному построению систем массового обслуживания, рациональной организации их работы и регулированию потока заявок для обеспечения высокой эффективности функционирования систем.

В качестве характеристик эффективности работы систем массового обслуживания чаще всего используют 3 группы показателей:

1. Показатели эффективности использования системы:

– абсолютная пропускная способность системы, т. е. среднее число заявок, которое может обслужить система массового обслуживания в единицу времени;

– относительная пропускная способность системы, т. е. отношение среднего числа заявок, обслуживаемых в системе массового обслуживания в единицу времени, к среднему числу поступивших заявок за это же время;

– средняя продолжительность периода занятости системы массового обслуживания;

– коэффициент использования системы массового обслуживания, т. е. средняя доля времени, в течение которого система занята обслуживанием заявок.

2. Показатели качества обслуживания заявок:

– среднее время ожидания заявки в очереди;

– среднее время пребывания заявки в системе массового обслуживания;

– вероятность отказа заявке в обслуживании без ожидания;

– вероятность того, что поступившая заявка немедленно будет принята к обслуживанию;

– закон распределения времени ожидания заявки в очереди;

– закон распределения времени пребывания заявки в системе массового обслуживания;

– среднее число заявок, находящихся в очереди;

– среднее число заявок, находящихся в системе массового обслуживания.

3. Показатели эффективности функционирования пары «система массового обслуживания – потребитель», где под «потребителем» понимают всю совокупность заявок или их источник:

- средний доход, приносимый системе массового обслуживания в единицу времени;
- затраты на обслуживание системы массового обслуживания;
- доход, приносимый системе массового обслуживания.

6.5. Одноканальная система массового обслуживания с отказами

Простейшей одноканальной моделью с вероятностным входным потоком и процедурой обслуживания является модель, характеризуемая показательным распределением длительностей интервалов между поступлением заявок и их обслуживанием. Допустим, в систему массового обслуживания поступает простейший поток требований с заданной интенсивностью λ . Время обслуживания требований для одного канала экспоненциальное. Если требование поступает в систему в момент, когда канал занят, получает отказ и покидает систему обслуживания. Если же канал свободен, то требование принимается к обслуживанию и обслуживается с интенсивностью μ .

Система имеет два состояния:

S_0 – канал свободен,

S_1 – канал занят, т. е. идет обслуживание заявки.

Пусть p_0 – вероятность состояния S_0 , а p_1 – вероятность состояния S_1 .

Составим систему дифференциальных уравнений Колмогорова для вероятностей состояний:

$$\begin{cases} p_0' = -\lambda p_0 + \mu p_1 \\ p_1' = -\mu p_1 + \lambda p_0 \end{cases}$$

Введем в систему нормировочное уравнение:

$$p_0 + p_1 = 1$$

и найдем значение вероятностей состояний:

$$p_0 = \frac{\mu}{\lambda + \mu};$$

$$p_1 = 1 - p_0.$$

Для одноканальной системы с отказами вероятность p_0 есть не что иное, как относительная пропускная способность системы q , так как

для конкретного момента времени t среднее отношение числа обслуженных заявок к числу поступивших равно вероятности p_0 того, что в момент t канал свободен и заявка будет обслужена:

$$q = p_0.$$

При $t \rightarrow \infty$ достигается стационарный (установившийся) режим и:

$$q = p_0 = \frac{\mu}{\mu + \lambda}.$$

Зная относительную пропускную способность, можно найти абсолютную, т. е. среднее число заявок, которое может обслужить система массового обслуживания в единицу времени:

$$A = \lambda \cdot q = \frac{\lambda \cdot \mu}{\lambda + \mu}.$$

Вероятность отказа в обслуживании заявки будет равна вероятности p_1 :

$$p_{отк} = p_1 = 1 - p_0 = 1 - \frac{\mu}{\mu + \lambda} = \frac{\lambda}{\lambda + \mu}.$$

6.6. Одноканальная система массового обслуживания с ожиданием и ограничением очереди

Система массового обслуживания имеет один канал. Входящий поток заявок на обслуживание является простейшим потоком с интенсивностью λ . Интенсивность потока обслуживания равна μ . Длительность обслуживания – случайная величина, подчиненная показательному закону распределения. Поток обслуживания является простейшим пуассоновским потоком событий. Заявка, поступившая в момент, когда канал занят, становится в очередь и ожидает обслуживания. Очередь плюс обслуженные заявки не превышают n заявок, а требования, не попавшие в очередь, покидают систему.

Система имеет следующие состояния:

S_0 – канал свободен;

S_1 – канал занят и очереди нет;

S_2 – канал занят и одна заявка стоит в очереди;

...

S_k – канал занят и k заявок стоят в очереди;

...

S_n – канал занят и $(n-1)$ заявок стоят в очереди.

Стационарный процесс в такой системе будет описываться

следующей системой алгебраических уравнений:

$$\begin{cases} -\rho \cdot p_0 + p_1 = 0, & i = 0 \\ \dots \\ -(1-\rho)p_n + p_{n+1} + \rho \cdot p_{n-1} = 0, & 0 < i < n \\ \dots \\ -p_n + \rho \cdot p_{n-1} = 0, & i = n, \end{cases}$$

где ρ – относительная нагрузка на систему:

$$\rho = \frac{\lambda}{\mu}.$$

Решение системы уравнений имеет вид:

$$p_n = \begin{cases} \left(\frac{1-\rho}{1-\rho^{m+2}} \right) \cdot \rho^n, & \rho \neq 1, i = 0, 1, \dots, n \\ \frac{1}{(m+2)}, & \rho = 1 \end{cases}$$

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

$$\text{Тогда } p_n = \begin{cases} p_0 \cdot \rho^n, & \rho \neq 1, i = 0, 1, \dots, n \\ \frac{1}{(m+2)}, & \rho = 1. \end{cases}$$

Определим характеристики одноканальной системы массового обслуживания с ожиданием и ограниченной длиной очереди m :

1) p_0, p_n – предельные вероятности системы (вероятность свободного состояния (p_0)):

$$p_0 = \begin{cases} \frac{1-\rho}{1-\rho^{m+2}}, & \text{если } \rho \neq 1 \\ \frac{1}{m+2}, & \text{если } \rho = 1 \end{cases}$$

$$p_n = \rho^n \cdot p_0; n = 1, \dots, m+1,$$

где n – состояние системы;

m – максимальное число мест в очереди;

ρ – относительную нагрузку (трафик) на систему: $\rho = \frac{\lambda}{\mu}$

2) $p_{отк}$ – вероятность отказа в обслуживании требования:

$$p_{отк} = \rho^{m+1} \cdot p_0;$$

3) q – относительная пропускная способность системы обслуживания:

$$q = 1 - p_{отк};$$

4) A – абсолютная пропускная способность системы:

$$A = \lambda \cdot q;$$

5) \bar{N} – среднее число требований, находящихся в системе (т. е. на обслуживании и в очереди):

$$\bar{N}_{сис} = \begin{cases} \frac{\rho[1 - (m+2)\rho^{m+1} + (m+1)\rho^{m+2}]}{(1-\rho)(1-\rho^{m+2})}, & \text{если } \rho \neq 1; \\ \frac{m+1}{2}, & \text{если } \rho = 1 \end{cases}$$

6) $\bar{T}_{сис}$ – среднее время пребывания автомобиля в системе:

$$\bar{T}_{сис} = \frac{\bar{N}_{сис}}{\lambda(1-p_{m+1})};$$

7) $\bar{T}_{оч}$ – средняя продолжительность пребывания заявки в очереди на обслуживание:

$$\bar{T}_{оч} = \bar{T}_{сис} - \frac{1}{\mu};$$

8) $\bar{N}_{оч}$ – среднее число заявок в очереди (длину очереди):

$$\bar{N}_{оч} = \lambda \cdot (1 - p_{m+1}) \cdot \bar{T}_{оч}.$$

6.7. Одноканальная система массового обслуживания с ожиданием без ограничения очереди

Условия функционирования данной системы совпадают с условиями работы одноканальной системы массового обслуживания с ожиданием и ограничением очереди, кроме ограничения ожидания (т.е. $n \rightarrow \infty$). Стационарный режим функционирования этой системы существует при $t \rightarrow \infty$ для любого $n=0, 1, 2, \dots$ и когда $\lambda < \mu$. Система

уравнений в этом случае имеет вид:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0, n = 0 \\ \lambda p_{n-1} + \mu p_{n+1} - (\lambda + \mu) p_n = 0, n > 0. \end{cases}$$

Решение системы уравнений имеет вид:

$$p_n = (1 - \rho) \rho^n, n = 0, 1, 2, \dots,$$

где $\rho = \frac{\lambda}{\mu} < 1$.

Одноканальная система массового обслуживания с ожиданием, без ограничения на длину очереди, имеет следующие характеристики:

1) ρ – трафик системы:

$$\rho = \frac{\lambda}{\mu} < 1;$$

2) p_0, p_n – предельные вероятности системы:

$$p_0 = 1 - \rho;$$

$$p_n = (1 - \rho) \cdot \rho^n; n = 1, 2, \dots;$$

3) \bar{N}_{cuc} – среднее число требований, находящихся в системе (то есть на обслуживании и в очереди):

$$\bar{N}_{cuc} = \frac{\rho}{1 - \rho};$$

4) \bar{T}_{cuc} – среднее время пребывания требования в системе:

$$\text{а) } \bar{T}_{cuc} = \frac{\bar{N}_{cuc}}{\lambda};$$

$$\text{б) } \bar{T}_{cuc} = \frac{1}{\mu \cdot (1 - \rho)};$$

$$\text{в) } \bar{T}_{cuc} = \frac{\rho}{\lambda \cdot (1 - \rho)};$$

5) $\bar{N}_{оч}$ – число требований в очереди на обслуживание:

$$\text{а) } \bar{N}_{оч} = \bar{N}_{cuc} - \frac{\lambda}{\mu};$$

$$\text{б) } \bar{N}_{оч} = \frac{\rho^2}{(1 - \rho)};$$

б) $\bar{T}_{оч}$ – среднюю продолжительность пребывания требования в очереди (среднее время ожидания заявки в очереди):

$$а) \bar{T}_{оч} = \frac{\rho}{\mu(1-\rho)};$$

$$б) \bar{T}_{оч} = \frac{\bar{N}_{оч}}{\lambda};$$

$$в) \bar{T}_{оч} = \frac{\rho^2}{\lambda(1-\rho)};$$

7) A – абсолютную пропускную способность:

$$A = \lambda \cdot q,$$

где q – относительная пропускная способность системы, $q=1$, т. к. каждая заявка, пришедшая в систему, будет обслужена.

6.8. Многоканальная система массового обслуживания с отказами

На практике большинство систем массового обслуживания являются многоканальными. Процесс обслуживания имеет интенсивность входного потока λ , при этом параллельно может обслуживаться не более n заявок на n каналах обслуживания. Средняя продолжительность обслуживания одной заявки равна $\frac{1}{\mu}$. Входной и выходной потоки являются пуассоновскими. Режим функционирования того или иного обслуживающего канала не влияет на режим функционирования других обслуживающих каналов системы, а 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 заявок.

Состояния этой системы имеют следующую интерпретацию:

S_0 – все каналы свободны;

S_1 – занят один канал, остальные свободны;

...

S_k – заняты k каналов, остальные свободны;

...

S_n – заняты n каналов, заявка получает отказ в обслуживании.

Уравнения Колмогорова для вероятностей состояний системы будут иметь вид:

$$\begin{cases} p'_0 = \lambda p_0 + \mu p_1; \\ \dots \\ p'_k = \lambda p_{k-1} - (\lambda + k\mu) p_k + \mu(k+1) p_{k+1}, 1 \leq k \leq n-1; \\ \dots \\ p'_n = \lambda p_{n-1} - \mu \cdot n \cdot p_n. \end{cases}$$

Начальные условия решения системы:

$$p_0(0) = 1; p_1(0) = 0; p_2(0) = 0 \dots; p_k(0) = 0 \dots; p_n(0) = 0.$$

Стационарное решение системы имеет вид:

$$\begin{cases} p_k = \frac{\rho^k}{k!} = \frac{\rho^k}{k!} \cdot p_0, k = 1, 2, \dots, n \\ p_0 = \frac{1}{\sum_{k=0}^n \frac{\rho^k}{k!}}, k = 0, 1, 2, \dots, n, \end{cases}$$

где $\rho = \frac{\lambda}{\mu}$.

Формулы для выполнения вероятностей p_k называются формулами Эрланга.

Определим вероятностные характеристики функционирования многоканальной системы массового обслуживания с отказами в стационарном режиме:

1) p_0, p_k – предельные вероятности состояния системы:

$$p_0 = \frac{1}{\sum_{k=0}^n \frac{\rho^k}{k!}}; k = 0, 1, \dots, n;$$

$$p_k = \frac{\rho^k}{k!} \cdot p_0; k = 1, \dots, n;$$

2) $p_{отк}$ – вероятность отказа в обслуживании заявки:

$$p_{отк} = p_n;$$

3) q – относительная пропускная способность системы:

$$q = 1 - p_{отк};$$

4) A – абсолютная пропускная способность системы:

$$A = \lambda q = \lambda(1 - p_{отк});$$

5) \bar{k} – среднее число занятых каналов:

$$\bar{k} = \rho(1 - p_{отк}) = \rho q = \sum_{k=1}^n k \cdot p_k.$$

6.9. Многоканальная система массового обслуживания с ожиданием и неограниченной очередью

Поток заявок, поступающих в систему, имеет интенсивность λ , а поток обслуживаний – интенсивность μ . Система может находиться в одном из состояний $S_0, S_1, S_2, \dots, S_k, \dots, S_n, \dots$, нумеруемых по числу заявок, находящихся в системе массового обслуживания:

S_0 – в системе заявок нет (все каналы обслуживания свободны);

S_1 – занят один канал, остальные свободны;

S_2 – заняты два канала, остальные свободны;

...

S_k – занято k каналов, остальные свободны;

...

S_n – заняты все n каналов (очереди нет);

S_{n+1} – заняты все n каналов, в очереди одна заявка;

...

S_{n+r} – заняты все n каналов, r заявок стоят в очереди;

...

Интенсивность потока обслуживания данной системы не остается постоянной, а по мере увеличения числа заявок в системе от 0 до n увеличивается от величины μ до $n\mu$, так как увеличивается число каналов обслуживания (n каналов). При числе заявок в системе больше, чем n , интенсивность потока обслуживания сохраняется равной $n\mu$.

Используя формулы расчета $p_0, p_1, p_2, \dots, p_n$ для процесса гибели и размножения, можно получить следующие формулы для предельных вероятностей состояний n -канальной системы с неограниченной очередью:

$$p_0 = \left(1 + \frac{\rho^1}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n!(n-\rho)} \right)^{-1},$$

$$p_1 = \frac{\rho}{1!} p_0, \dots, p_k = \frac{\rho^k}{k!} p_0, \dots, p_n = \frac{\rho^n}{n!} p_0,$$

$$p_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} p_0, \dots, p_{n+r} = \frac{\rho^{n+r}}{n^r \cdot n!} p_0, \dots$$

Вероятность того, что заявка окажется в очереди:

$$P_{оч} = \frac{\rho^{n+1}}{n!(n-\rho)} p_0.$$

Найдем другие характеристики данной системы:

1) при $r \rightarrow \infty, p_{отк} = 0; q = 1; A = \lambda; \bar{k} = \rho;$

2) $\bar{N}_{оч}$ – среднее число заявок в очереди на обслуживание:

$$\bar{N}_{оч} = \frac{p_0 \cdot \rho^{n+1}}{n \cdot n! \left(1 - \frac{\rho}{n} \right)^2};$$

3) $\bar{N}_{сис}$ – среднее число находящихся в системе заявок:

$$\bar{N}_{сис} = \bar{N}_{оч} + \rho$$

4) $\bar{T}_{оч}$ – среднее время пребывания требования в очереди на обслуживание:

$$\bar{T}_{оч} = \frac{\bar{N}_{оч}}{\lambda};$$

5) $\bar{T}_{сис}$ – среднее время пребывания требования в системе:

$$\bar{T}_{сис} = \bar{T}_{оч} + \bar{t}_{об}.$$

6.10. Многоканальная система массового обслуживания с ожиданием и ограниченной очередью

Системы массового обслуживания с ограниченной очередью отличаются от рассмотренных в вопросе 9 тем, что число заявок в очереди ограничено и не может быть более m . Если новая заявка поступает в момент времени, когда все места в очереди заняты, она покидает систему необслуженной. Для вычисления предельных вероятностей состояний и показателей эффективности таких систем

суммируют не бесконечную, а конечную прогрессию.

Предельные вероятности состояния многоканальной системы массового обслуживания с ожиданием и ограниченной очередью:

$$p_0 = \left[1 + \frac{\rho}{1!} + \dots + \frac{\rho^n}{n!} + \dots + \frac{\rho^{n+1} \left(1 - \left(\frac{\rho}{n} \right)^m \right)}{n \cdot n! \left(1 - \frac{\rho}{n} \right)} \right]^{-1},$$

$$p_1 = \frac{\rho^1}{1!} p_0, \dots, p_k = \frac{\rho^k}{k!} p_0, \dots, p_n = \frac{\rho^n}{n!} p_0,$$

$$p_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} p_0, \dots, p_{n+r} = \frac{\rho^{n+r}}{n^r \cdot n!} p_0, (r = 1, \dots, m).$$

Рассчитаем другие показатели функционирования системы массового обслуживания:

1) $p_{отк}$ – вероятность отказа в обслуживании заявки:

$$p_{отк} = p_{n+m} = \frac{\rho^{n+m}}{n^m \cdot n!} \cdot p_0;$$

2) q – относительная пропускная способность системы:

$$q = 1 - p_{отк} = 1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0;$$

3) A – абсолютная пропускная способность системы:

$$A = \lambda q = \lambda \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0 \right)$$

4) \bar{k} – среднее число занятых каналов обслуживания:

$$\bar{k} = \rho \cdot \left(1 - \frac{\rho^{n+m}}{n^m \cdot n!} p_0 \right) = \rho q;$$

5) $\bar{N}_{оч}$ – среднее число заявок в очереди на обслуживание:

$$\bar{N}_{оч} = \frac{\rho_0 \rho^{n+1} \cdot \left[1 - \left(\frac{\rho}{n} \right)^m \cdot \left(m + 1 - \frac{m\rho}{n} \right) \right]}{n \cdot n! \cdot \left(1 - \frac{\rho}{n} \right)^2};$$

6) $\bar{N}_{сис}$ – среднее число находящихся в системе заявок:

$$\bar{N}_{сис} = \bar{N}_{оч} + \bar{k};$$

7) $\bar{T}_{оч}$ – среднее время пребывания требования в очереди на обслуживание:

$$\bar{T}_{оч} = \frac{\bar{N}_{оч}}{A};$$

8) $\bar{T}_{сис}$ – среднее время пребывания требования в системе:

$$\bar{T}_{сис} = \bar{T}_{оч} + \bar{t}_{об}.$$

6.11. Замкнутая система массового обслуживания

В замкнутых системах массового обслуживания или системах Энгсета источник требований находится внутри системы и интенсивность потока заявок зависит от состояния самой системы. Чаще всего потоком требований в такой системе является поток неисправностей от некоторой группы работающих устройств. Пусть имеется m работающих устройств, которые могут выходить из строя за счет неисправностей. Имеется n каналов обслуживания этих требований. Обычно $n < m$.

Система может находиться в одном из состояний:

S_0 – все устройства работают и каналы обслуживания свободны;

S_1 – одно устройство вышло из строя и обслуживается одним каналом;

...

S_n – n устройств не работают и все n каналов обслуживания заняты;

...

S_m – все устройства не работают, из них n – обслуживаются, а $m-n$ ожидают обслуживания в очереди.

Вероятности состояний и характеристики замкнутой системы

массового обслуживания определяются следующим образом:

1) p_0, p_k – предельные вероятности состояния системы для вариантов ее организации:

$$p_0 = \left(1 + \sum_{k=1}^n \frac{N!}{k!(N-k)!} \rho^k + \sum_{n+1}^N \frac{N!}{n^{k-n} \cdot n!(N-k)!} \rho^k \right)^{-1} =$$

$$= \frac{1}{N!} \left(\sum_{k=0}^n \frac{\rho^k}{k!(N-k)!} + \frac{n^n}{n!} \sum_{k=n+1}^N \frac{\rho^k}{n^k(N-k)!} \right)^{-1}.$$

$$p_k = \begin{cases} \frac{N! \rho^k p_0}{k!(N-k)!}, & 1 \leq k \leq n; \\ \frac{N! p^k p_0}{n^{k-n} n!(N-k)!}, & n+1 \leq k \leq N, \end{cases}$$

где N – число источников заявок.

2) $\bar{N}_{оч}$ – среднее число требований в очереди на обслуживание (для вариантов организации системы):

$$\bar{N}_{оч} = \sum_{n+1}^N (k-n) p_k;$$

3) $\bar{N}_{сис}$ – среднее число требований в системе (на обслуживании и в очереди) для вариантов организации системы:

$$\bar{N}_{сис} = \sum_{k=1}^N k p_k;$$

4) \bar{N}_{np} – среднее число каналов, простаивающих из-за отсутствия работы (для вариантов организации системы):

$$\bar{N}_{np} = \sum_{k=0}^{n-1} (n-k) p_k;$$

5) A – абсолютная пропускная способность системы:

$$A = (N - \bar{N}_{сис}) \lambda;$$

6) q – относительная пропускная способность системы:

$$q = 1;$$

7) a_1 – коэффициент простоя требования в очереди (для вариантов организации системы):

$$a_1 = \frac{\bar{N}_{оч}}{N};$$

8) a_2 – коэффициент использования устройства (для вариантов организации системы):

$$a_2 = 1 - \frac{\bar{N}_{сис}}{N};$$

9) a_3 – коэффициент простоя обслуживающих каналов (для вариантов организации системы):

$$a_3 = \frac{\bar{N}_{np}}{N};$$

10) $\bar{T}_{ож}$ – среднее время ожидания обслуживания требования (для вариантов организации системы):

$$\bar{T}_{ож} = \frac{1}{\lambda} \left(\frac{1 - a_2}{a_2} \right) - \frac{1}{\mu}.$$

Лекция 9

Тема 7. МОДЕЛИ ТЕОРИИ УПРАВЛЕНИЯ ЗАПАСАМИ

- 7.1. Постановка задачи управления запасами.
- 7.2. Статическая детерминированная однопродуктовая модель.
- 7.3. Обоснование точки заказа в модели Уилсона.
- 7.4. Учет дискретности спроса в модели Уилсона.
- 7.5. Модель оптимальной партии поставки с конечной интенсивностью поступления партии.
- 7.6. Модель оптимальной партии поставки с дефицитом при учете неудовлетворенных требований.
- 7.7. Обобщенная модель оптимальной партии поставки с учетом неудовлетворенных требований.
- 7.8. Многопродуктовая модель размера партий поставок при отсутствии взаимодействия между запасами различных видов.
- 7.9. Многопродуктовая модель размера партий поставок в случае нескольких ограничений.

7.10. Многопродуктовая модель размера партий поставок с периодическими проверками при полном совмещении заказов.

7.1. Постановка задачи управления запасами

Основными причинами создания производственных запасов служат необходимость обеспечения бесперебойного снабжения производственного процесса, периодичность производства различных видов продукции поставщиками, осуществление транспортировки большинства видов продукции от поставщика к потребителю партиями, а также несовпадение ритма производства с ритмом потребления.

Под запасами понимают все то, на что имеется спрос и что выключено временно из потребления. Запасы подразделяются на запасы средств производства и запасы предметов потребления.

Предприятия стремятся уменьшить свои запасы, так как оплачивают их хранение, отвлекают денежные средства на их оплату, несут потери от морального износа и естественной убыли запасов. Но слишком низкий уровень запасов связан с риском возникновения их дефицита, остановкой производства, изменением конъюнктуры, случайными колебаниями спроса, что приводит к увеличению упущенной прибыли.

Предметом теории управления запасами является отыскание такой организации поставок или производства, при которых суммарные затраты на функционирование системы были бы минимальными. *Под организацией поставок* понимается определение объемов поставок и периодичность заказов.

Возникновение теории управления запасами можно связать с работами Ф. Эджуорта и Ф. Харриса, опубликованными в конце XIX – начала XX века, в которых исследовалась простая модель для определения экономического размера партии поставки для складской системы с постоянным равномерным расходом и периодическим поступлением хранимого продукта. Как научная дисциплина теория управления запасами начала формироваться в середине 1950-х годов. В настоящее время в литературе имеется более 300 моделей управления запасами.

Существует четыре основных вида затрат, которые могут оказать влияние на решение задачи по управлению запасами, т. е. суммарные затраты системы управления запасами состоят из:

- 1) затрат на приобретение запасов;

- 2) затрат на организацию заказа;
- 3) затрат на хранение запас;
- 4) потерь от дефицита запаса.

Затраты на приобретение запасов характеризуются стоимостью единицы продукции, которая может быть постоянной или переменной. При этом цена единицы продукции может зависеть от величины партии заказа, что выражается в виде оптовых скидок. При этом затраты, которые не зависят от величины и периодичности заказов при решении задач не учитываются.

К *затратам на организацию заказа* относят постоянные расходы, связанные с размещением заказов: расходы на разъезды и командировки, почтово-телеграфные расходы, транспортные затраты. При размещении более мелких заказов и, следовательно, более частых эти затраты возрастают по сравнению с размещением более крупных, но менее частных заказов.

Затраты на хранение запаса включают затраты на амортизацию и эксплуатацию складов, на грузопереработку и содержание запаса на складах. Они возрастают с увеличением уровня запаса.

Потери от дефицита – это расходы, обусловленные отсутствием необходимых ресурсов или продукции, и возможные потери из-за утраты доверия покупателей. Эти потери рассматриваются как уменьшение прибыли за счет простоя мощности и обслуживающего персонала предприятия, переналадки производственного процесса, замены дефицитных материалов другими более дорогими, штрафа за нарушение сроков поставки продукции.

Классификация моделей управления запасами. Большое разнообразие моделей управления запасами объясняется многообразием реальных ситуаций, характером спроса. Спрос может быть детерминированным (достаточно известным) и вероятностным (случайным). Детерминированный спрос может быть:

- 1) статическим, т. е. неизменным во времени, интенсивность потребления постоянна во времени;
- 2) динамическим, когда спрос изменяется во времени.

Вероятностный спрос может быть:

- 1) стационарным, т. е. с неизменной во времени плотностью вероятности;
- 2) нестационарным, т. е. с изменяющейся во времени плотностью вероятности.

На разнообразие форм моделей управления запасов влияет:

1) число видов продукции, т. е. модели могут быть однопродуктовыми (однономенклатурными) и многопродуктовыми (многономенклатурными);

2) число пунктов накопления запасов (один или несколько);

3) период времени (интервал), в течение которого осуществляется регулирование уровня запасов (этот период может быть конечным и бесконечным).

Модели управления запасами различаются по:

1) пополнению запасов (этот процесс может происходить мгновенно или равномерно во времени);

2) запаздыванию поставок, т. е. заказ может быть поставлен немедленно или для его выполнения требуется некоторое время.

Интервалом времени между моментом размещения заказов и его поставкой называется *срок выполнения заказов*. Эта величина может быть детерминированной и случайной.

7.2. Статическая детерминированная однопродуктовая модель

Самая простая модель характеризуется постоянным во времени спросом, мгновенным пополнением запасов и отсутствием дефицита.

Введем условные обозначения:

v – интенсивность спроса;

K – затраты на организацию поставки;

q – величина заказанной партии;

s – издержки содержания единицы продукции в единицу времени;

$I(t)$ – уровень запасов в зависимости от времени;

τ – интервал между двумя поставками.

Простая модель оптимальной партии поставки строится при следующих предположениях:

1) спрос в единицу времени – постоянный;

2) заказанная партия доставляется мгновенно;

3) дефицит недопустим;

4) затраты на организацию поставки постоянны и не зависят от величины партии;

5) издержки содержания единицы продукции в течение единицы времени составляют s .

В простой модели уровень запасов уменьшается равномерно от q до 0, после чего подается заказ на доставку новой партии величиной q . Заказ выполняется мгновенно, и уровень запаса восстанавливается до величины q .

Интервал времени между поставками τ называется *циклом*. Издержки в течение цикла $L_{ци}$ состоят из стоимости заказа K и затрат на содержание запаса, которые пропорциональны средней величине запаса $I = \frac{q}{2}$ и длина цикла $\tau = \frac{q}{v}$, следовательно, издержки в течение цикла вычисляются по формуле: $L_{ци} = K + s \frac{q}{2} \cdot \frac{q}{v}$.

Разделим это выражение на длину цикла ($\tau = \frac{q}{v}$), получим издержки в единицу времени: $\frac{L_{ци}}{\tau} = L$ или $\frac{L_{ци}}{\tau} = K \frac{v}{q} + s \frac{q}{2}$.

Чтобы найти точку минимума, продифференцируем выражение по q , получим:

$$\frac{\partial L}{\partial q} = \frac{(Kv)'q - Kv(q)'}{q^2} + \frac{1}{2}(sq)' = \frac{0 \cdot q - Kv \cdot 1}{q^2} + \frac{1}{2} \cdot s \cdot 1 = -\frac{Kv}{q^2} + \frac{s}{2},$$

так как $-\frac{Kv}{q^2} + \frac{s}{2} = 0$, то найдем отсюда оптимальный размер

партии заказа (q^*): $\frac{Kv}{q^2} = \frac{s}{2}$ или $q^2 s = 2Kv$ или $q^* = \sqrt{\frac{2Kv}{s}}$.

Эта формула называется формулой Уилсона или формулой для определения экономной величины заказа или размера партии.

Рассчитаем остальные оптимальные параметры системы:

$$\tau^* = \frac{q^*}{v} = \sqrt{\frac{2Kv}{sv^2}} = \sqrt{\frac{2K}{sv}},$$

где τ^* – оптимальная продолжительность цикла потребления или оптимальный интервал между поставками.

Суммарные затраты по формированию оптимального размера партии заказа и содержанию запасов в единицу времени рассчитываются по формуле: $L^* = \sqrt{2Ksv} = sq^*$,

так как $\frac{d^2L}{dq^2} \geq 0$, то для всех $q > 0$ выражение $L^* = sq^*$ является минимумом функции затрат.

7.3. Обоснование точки заказа в модели Уилсона

В статической детерминированной однопродуктовой модели предполагали, что заказы выполняются мгновенно. На практике необходимо учитывать время их доставки. Пусть Θ – время от момента размещения заказа до его появления у потребителя. Заказ должен подаваться заранее, чтобы избежать дефицита наличного ресурса, которого должно быть достаточно для удовлетворения потребности за время реализации заказа. Рассмотрим три случая:

1) $\Theta = \tau^*$, т. е. в момент поступления очередной партии необходимо давать заказ на пополнение запаса. При этом точка заказа равна нулю: $r=0$.

Под *точкой заказа* понимают величину наличного запаса, при котором подается заказ на пополнение запаса. Если $\Theta = \tau^*$, то средний уровень фиктивного запаса равен: $\tau^*v + \frac{q^*}{2} = \frac{3}{2}q^*$.

Под *фиктивным уровнем текущего запаса* понимают сумму наличного запаса и заказанного ресурса или товара в любой момент времени. При $\Theta = \tau^*$ динамика изменения текущего уровня фиктивного запаса изображена на рис. 7.3.1 пунктирной линией.

2) Если $\Theta < \tau^*$, то точка размещения заказа: $r = \Theta v$, а средний уровень фиктивного запаса составляет: $\Theta v + \frac{q^*}{2}$.

Динамика изменения текущего уровня фиктивного запаса при $\Theta < \tau^*$ отражена пунктиром на рис. 7.3.2

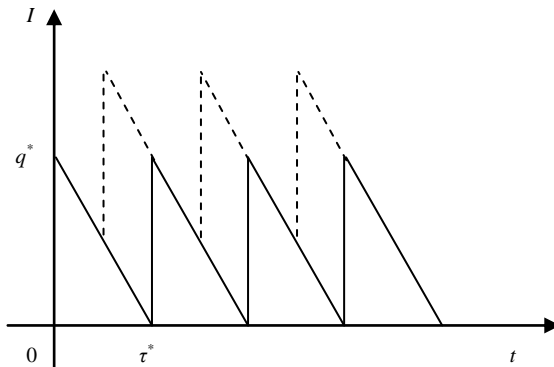


Рис. 7.3.1 Динамика изменения фиктивного уровня текущего запаса, если время выполнения заказа равно длительности одного цикла ($\Theta = \tau^*$).

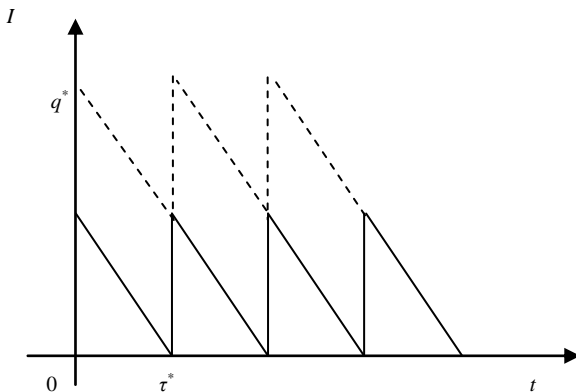


Рис. 7.3.2 Динамика изменения фиктивного уровня текущего запаса, если время выполнения заказа меньше длительности одного цикла ($\Theta < \tau^*$).

3) Если $\Theta > \tau^*$, то динамика изменения фиктивного уровня запаса изображена пунктиром на рис. 7.3.3.

Допустим $\Theta = 2,5\tau^*$, т.е. $\Theta > \tau^*$, то в интервале $[t_k, k\tau^*]$ средняя величина заказанного ресурса или товара равна:

$$2,5q^* = 2,5\tau^* v = \Theta v,$$

а это соответствует величине потребления за время реализации заказа.

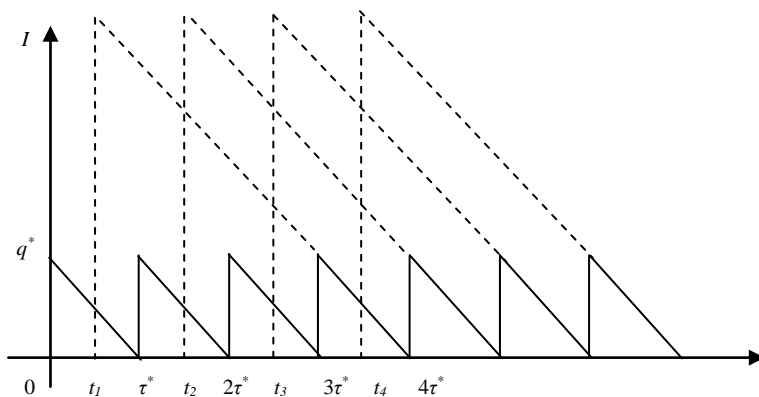


Рис. 7.3.3 Динамика изменения фиктивного уровня текущего запаса, если время выполнения заказа больше длительности одного цикла ($\Theta > \tau^*$).

Обобщим рассмотренные ситуации. Через $\left[\frac{\Theta}{\tau^*} \right]$ обозначим наибольшее целое число, не превышающее $\frac{\Theta}{\tau^*}$. Тогда в интервалах $[t_k, k\tau^*]$ перед поступлением очередной партии количество товара или ресурса в невыполненных заказах будет равно $\left[\frac{\Theta}{\tau^*} \right] \cdot q^*$, а в интервалах $(k\tau^*, t_{k+1})$ после поступления до момента размещения заказа – $\left(\left[\frac{\Theta}{\tau^*} \right] + 1 \right) \cdot q^*$. Если $\frac{\Theta}{\tau^*}$ – является целым числом, то имеется $\frac{\Theta}{\tau^*}$ невыполненных заказов. Отсюда следует, что точка размещения заказа определяется по формуле: $r = \Theta v - \left[\frac{\Theta}{\tau^*} \right] q^*$.

Если $\Theta = \tau^*$, то имеем $r = \Theta v - q^* = 0$, если $\Theta < \tau^*$, то $\left[\frac{\Theta}{\tau^*} \right] = 0$ и, следовательно, $r = \Theta v$.

Для бесперебойной работы системы необходимо иметь минимальный начальный запас I_0 , который можно рассчитать по формуле: $I_0 = \Theta v$.

Если I – наличный начальный запас, то для бездефицитной работы необходимо, чтобы $I \geq \Theta v$. Тогда отношение $\frac{I}{v}$ определит время потребления запаса I . Чтобы партия прибыла к моменту полного использования начального уровня, ее надо разместить в момент времени t_0 : $t_0 = \frac{I}{v} - \Theta$.

Остальные заказы выполнить в моменты времени t_k :

$$t_k = \frac{I}{v} - \Theta + k\tau^*, k = 0, 1, 2, \dots$$

7.4. Учет дискретности спроса в модели Уилсона

В простейшей модели Уилсона предполагали, что спрос является непрерывной величиной. Если же спрос – непрерывная величина, а на размер партии заказа налагается ограничение положительности и целочисленности, тогда зависимость суммарных издержек работы

системы от размера партии изображена на рис. 7.4,1 для которой выполняются следующие условия:

$$\begin{cases} \Delta L(q) = L(q) - L(q-1) \leq 0 \\ \Delta L(q+1) = L(q+1) - L(q) \geq 0. \end{cases}$$

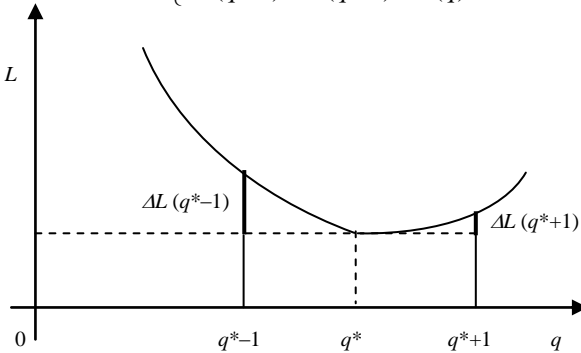


Рис. 7.4.1 Зависимость суммарных издержек работы системы от размера партии поставок.

Так как $L(q) = \frac{Kv}{q} + \frac{sq}{2}$, то найдем q :

$$-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2Kv}{s}} \leq q \leq \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2Kv}{s}},$$

$$q^* = \left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2Kv}{s}} \right],$$

где $[a]$ – целая часть числа a .

Обозначим $q_1^* = \left[-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2Kv}{s}} \right],$

$$q_2^* = \left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2Kv}{s}} \right],$$

тогда имеем два оптимальных решения:

$$q_1^* \leq q^* \leq q_2^*$$

Оптимальный интервал поставок $\tau^* = \frac{q^*}{v}$ может быть нецелочисленным, поэтому представим суммарные издержки работы системы как функцию от величины интервала возобновления поставки τ :

$$L(\tau) = \frac{K}{\tau} + \frac{2sv\tau}{2}.$$

Для данной функции выполняются условия:

$$\begin{cases} \Delta L(\tau) = L(\tau) - L(\tau - 1) \leq 0 \\ \Delta L(\tau + 1) = L(\tau + 1) - L(\tau) \geq 0 \end{cases}$$

Откуда найдем значение τ :

$$-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2K}{sv}} \leq \tau \leq \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2K}{sv}},$$

$$\tau^* = \left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2K}{sv}} \right],$$

где $[a]$ – целая часть числа a .

Обозначим

$$\tau_1^* = \left[-\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2K}{sv}} \right],$$

$$\tau_2^* = \left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{2K}{sv}} \right],$$

Имеем два оптимальных значения:

$$\tau_1^* \leq \tau^* \leq \tau_2^*.$$

7.5. Модель оптимальной партии поставки с конечной интенсивностью поступления партии

В модели Уилсона предполагалось, что вся партия поступает заказчику одновременно. На практике часто партия поступает по частям с интенсивностью λ . Например, один цех завода производит детали, которые поступают на склад с определенной интенсивностью и используются другим цехом для производства товара с интенсивностью потребления v . В этом случае, система может работать бездефицитно, если $\lambda \geq v$.

За время τ_1 запас одновременно поступает и расходуется, а в течение времени τ_2 запас только расходуется.

Длина цикла определяется по формуле:

$$\tau = \tau_1 + \tau_2.$$

Скорость пополнения запасов равна $(\lambda - v)$. Величина партии поставки равна q , но максимальный уровень запаса $I_{\max} < q$, так как продукция используется по мере изготовления.

Если производственный цикл длится t единиц времени, то общий объем продукции, произведенной за цикл равен:

$$q = \lambda t.$$

Найдем отсюда:

$$t = \frac{q}{\lambda}.$$

Тогда максимальный уровень запаса определяется по формуле:

$$I_{\max} = (\lambda - v)t = (\lambda - v) \frac{q}{\lambda} = \left(1 - \frac{v}{\lambda}\right)q.$$

Отсюда следует, что средний уровень запаса равен:

$$\bar{I} = \left(1 - \frac{v}{\lambda}\right) \frac{q}{2}.$$

Тогда издержки системы в единицу времени составят:

$$L = \frac{Kv}{q} + \frac{sq}{2} \left(1 - \frac{v}{\lambda}\right).$$

Для определения оптимальных параметров работы системы найдем величину оптимальной партии:

$$q^* = \sqrt{\frac{2Kv}{s}} \cdot \frac{1}{\sqrt{1 - \frac{v}{\lambda}}};$$

оптимальный период возобновления производства:

$$\tau^* = \sqrt{\frac{2K}{sv}} \cdot \frac{1}{\sqrt{1 - \frac{v}{\lambda}}};$$

время производства:

$$\tau_1^* = \frac{q^*}{\lambda};$$

время чистого потребления:

$$\tau_2^* = \sqrt{\frac{2K}{sv}} \cdot \sqrt{1 - \frac{v}{\lambda}};$$

минимальные издержки в единицу времени:

$$L^* = \sqrt{2Ksv} \cdot \sqrt{1 - \frac{v}{s}}$$

При определении оптимальной точки заказа рассматриваются два случая:

$$\text{а) } r = \Theta v - \left[\frac{\Theta}{\tau^*} \right] q^*, \text{ если } \Theta - \left[\frac{\Theta}{\tau^*} \right] \tau^* < \tau_2^*;$$

$$\text{б) } r = \Theta(v - \lambda) + \left(\left[\frac{\Theta}{\tau^*} \right] + 1 \right) \cdot \left(\frac{\lambda}{v} - 1 \right) q^*, \text{ если } \Theta - \left[\frac{\Theta}{\tau^*} \right] \tau^* > \tau_2^*;$$

моменты размещения заказа при $t_0 = 0$: $t_k = t_0 + k\tau^*$, $k = 1, 2, \dots$

7.6. Модель оптимальной партии поставки с дефицитом при учете неудовлетворенных требований

В модели Уилсона предполагается, что дефицит не допустим. Однако на практике требования, поступающие в моменты дефицита, ставятся на учет и при поступлении очередной партии в первую очередь удовлетворяется задолженный спрос, а затем пополняется запас. Изменение запаса показано на рис. 7.6.1

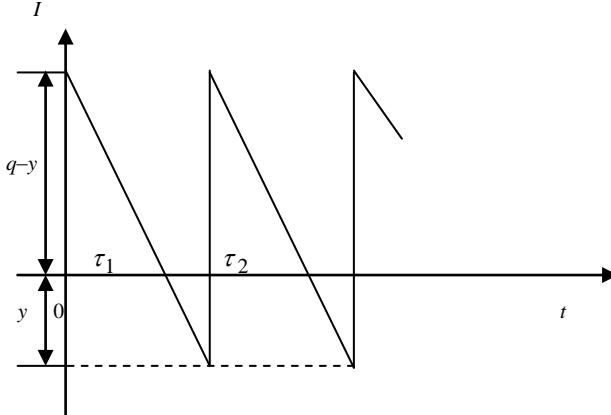


Рис. 7.6.1 Динамика изменения уровня запаса в модели при дефиците с учетом неудовлетворенных требований.

На рис. 7.6.1 видно, что:

y — максимальная величина задолженного спроса;

$q-y$ — максимальная величина наличного запаса;

τ_1 и τ_2 — соответственно время существования наличного запаса и

время дефицита.

Издержки d , вызванные дефицитом единицы запаса в единицу времени пропорциональны средней величине дефицита и времени его существования. Тогда издержки цикла, включающие затраты на размещение заказа, содержание запаса и потери от дефицита будут равны: $L_q = K + \frac{s(q-y)^2}{2v} + \frac{dy^2}{2v}$.

Разделим издержки цикла на его величину $\tau_1 + \tau_2$, получим средние издержки работы системы в единицу времени:

$$L = \frac{Kv}{q} + \frac{s(q-y)^2}{2q} + \frac{dy^2}{2q}.$$

Отсюда определяем оптимальные параметры системы:

оптимальный размер партии поставки: $q^* = \sqrt{\frac{2Kv}{s}} \cdot \sqrt{1 + \frac{s}{d}}$;

максимальная величина задолженного спроса (максимальный уровень дефицита): $y^* = \frac{s}{d} \sqrt{\frac{2Kv}{s}} \cdot \frac{1}{\sqrt{1 + \frac{s}{d}}}$;

максимальная величина наличных (текущих) запасов:

$$Y^* = q^* - y^* = \sqrt{\frac{2Kv}{s}} \cdot \frac{1}{\sqrt{1 + \frac{s}{d}}};$$

время существования наличного запаса:

$$\tau_1^* = \frac{Y^*}{v} = \sqrt{\frac{2K}{sv}} \cdot \frac{1}{\sqrt{1 + \frac{s}{d}}};$$

время существования дефицита:

$$\tau_2^* = \frac{y^*}{v} = \frac{s}{d} \sqrt{\frac{2K}{sv}} \cdot \frac{1}{\sqrt{1 + \frac{s}{d}}};$$

оптимальный период возобновления заказа (продолжительность цикла):

$$\tau^* = \tau_1^* + \tau_2^* = \frac{q^*}{v} = \sqrt{\frac{2K}{sv}} \cdot \sqrt{1 + \frac{s}{d}};$$

$$\text{точка заказа: } r = \Theta v - \left[\frac{\Theta}{\tau^*} \right] q^* - y^*;$$

Точка заказа может принимать отрицательное значение, что показывает, что заказ необходимо разместить в момент, когда величина требований, поставленных на учет, равна $|r|$.

минимальные затраты работы системы в единицу времени:

$$L^* = \sqrt{2Ksv} \cdot \frac{1}{\sqrt{1 + \frac{s}{d}}}.$$

7.7. Обобщенная модель оптимальной партии поставки с учетом неудовлетворенных требований

В данной модели предполагается, что товар или ресурс поступает на склад непосредственно с производственной линии с постоянной интенсивностью спроса v и поступления λ в единицу времени. По достижении некоторого уровня запаса производство товара или ресурса прекращается. Возобновление его производства и поставки на склад осуществляется в момент, когда неудовлетворенный спрос достигает некоторого значения (рис. 7.7.1).

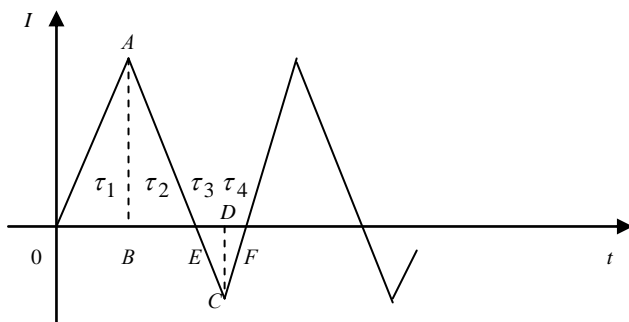


Рис. 7.7.1. Динамика изменения уровня запаса в обобщенной модели при дефиците с учетом неудовлетворенных требований.

Из рис. 7.7.1 видно, что запас пополняется и расходуется одновременно в течение интервала τ_1 каждого цикла. Причем в

течение τ_1 идет одновременное и пополнение запасов и их расходование, поэтому абсолютная интенсивность поступления определяется как разность между λ и ν , т.е. $(\lambda - \nu)$. Максимальная величина запаса AB равна: $(\lambda - \nu)\tau_1$.

Запас, накопленный в интервале τ_1 , т. е. $(\lambda - \nu)\tau_1$ полностью расходуется в течение интервала τ_2 , поэтому AB равно $\nu\tau_2$. В интервале τ_3 спрос не удовлетворяется, идет рост дефицита со скоростью, равной интенсивности потребления, поэтому CD равно $\nu\tau_3$. Неудовлетворенный спрос покрывается в течение интервала τ_4 , интенсивность поступлений равна λ , а интенсивность потребления ν , следовательно, разность $(\lambda - \nu)$ характеризует скорость ликвидации дефицита и CD равна: $(\lambda - \nu)\tau_4$.

Продолжительность цикла равна: $\tau = \tau_1 + \tau_2 + \tau_3 + \tau_4$ или $\tau = \frac{q}{\nu}$.

Если d – удельные издержки дефицита, а s – удельные издержки содержания, то имеем пропорцию $AB:CD=d:s$ и следующую систему уравнений:

$$\begin{cases} (\lambda - \nu)\tau_1 = \nu\tau_2 \\ \nu\tau_3 = (\lambda - \nu)\tau_4 \\ \frac{\tau_2}{\tau_3} = \frac{d}{s} \\ \tau_1 + \tau_2 + \tau_3 + \tau_4 = \frac{q}{\nu}. \end{cases}$$

Решим систему четырех уравнений и найдем значения τ_1 , τ_2 , τ_3 и τ_4 :

$$\tau_1^* - \text{время возрастания запаса: } \tau_1^* = \frac{q}{\lambda} \cdot \frac{1}{1 + \frac{s}{d}};$$

τ_2^* – время, в течение которого уровень запаса понижается до нуля:

$$\tau_2^* = \frac{q}{\nu} \cdot \frac{1 - \frac{\nu}{\lambda}}{1 + \frac{s}{d}};$$

τ_3^* – время роста дефицита (время накопления невыполненных

заказов):
$$\tau_3^* = \frac{q^*}{v} \cdot \frac{1 - \frac{v}{\lambda}}{1 + \frac{s}{d}} \cdot \frac{s}{d};$$

τ_4^* – время, в течение которого дефицит ликвидируется:

$$\tau_4^* = \frac{q^*}{\lambda} \cdot \frac{1}{1 + \frac{s}{d}} \cdot \frac{s}{d};$$

Найдем τ^* – продолжительность цикла:

$$\tau^* = \tau_1^* + \tau_2^* + \tau_3^* + \tau_4^* = \frac{q^*}{v} = \sqrt{\frac{2K}{sv}} \cdot \sqrt{\frac{1 + \frac{s}{d}}{1 - \frac{v}{\lambda}}}.$$

Определим τ_{np}^* – время, затраченное на производство партии:

$$\tau_{np}^* = \tau_1^* + \tau_4^* = \frac{q^*}{\lambda}.$$

При формировании издержек будем учитывать: издержки формирования запаса K и содержания запасов, которые пропорциональны величине текущего запаса и времени содержания L_1 , издержки дефицита, которые пропорциональны величине текущего дефицита и времени существования дефицита. Из рис. 6.19 видно, что площадь треугольника OAE определяет среднюю величину запаса в течение цикла, а треугольника ECF – среднюю величину дефицита, поэтому издержки содержания запасов рассчитываются произведением удельных издержек содержания s на площадь

треугольника OAE :
$$L_1 = \frac{s \cdot AB(\tau_1 + \tau_2)}{2} = \frac{sd^2(\lambda - v)q^2}{2}.$$

Издержки от дефицита равны величине d , умноженной на площадь треугольника ECF :

$$L_2 = \frac{d \cdot CD(\tau_3 + \tau_4)}{2} = \frac{s^2 d(\lambda - v)q^2}{2}.$$

Суммарные издержки работы системы в течение цикла равны:

$$L_y = K + L_1 + L_2 = K + \frac{s \cdot (1 - \frac{v}{\lambda})}{2 \cdot (1 + \frac{s}{d})} \cdot \frac{q^2}{v}.$$

Разделив общие издержки в течение цикла на τ , получим удельные издержки функционирования системы:

$$L = \frac{Kv}{q} + \frac{sq \cdot (1 - \frac{v}{\lambda})}{2 \cdot (1 + \frac{s}{d})}.$$

Из уравнения найдем $\frac{\partial L}{\partial q} = 0$ и определим значение оптимального размера партии поставки q^* :

$$q^* = \sqrt{\frac{2Kv}{s}} \cdot \sqrt{\frac{1 + \frac{s}{d}}{1 - \frac{v}{\lambda}}}.$$

Найдем другие параметры системы:

y^* – максимальный уровень дефицита:

$$y^* = I_{\min} = v\tau_3^* = \frac{s}{d} \sqrt{\frac{2Kv}{s}} \cdot \sqrt{\frac{1 - \frac{v}{\lambda}}{1 + \frac{s}{d}}};$$

Y^* – максимальный уровень наличных запасов:

$$Y^* = I_{\max} = v\tau_2^* = \sqrt{\frac{2Kv}{s}} \cdot \sqrt{\frac{1 - \frac{v}{\lambda}}{1 + \frac{s}{d}}};$$

L^* – минимальные затраты работы системы в единицу времени:

$$L^* = \sqrt{2Ksv} \cdot \sqrt{\frac{1 - \frac{v}{\lambda}}{1 + \frac{s}{d}}}.$$

7.8. Многопродуктовая модель размера партий поставок при отсутствии взаимодействия между запасами различных видов

Складские системы содержат множество видов товаров или ресурсов. Если взаимодействие между запасами различных видов отсутствует, то общие издержки в единицу времени, связанные с размещением заказов и содержанием запасов n видов, составляют:

$$L = \sum_{i=1}^n \left(\frac{K_i v_i}{q_i} + \frac{s_i q_i}{2} \right).$$

Найдем частные производные L по q_i , приравняем их к нулю и найдем q_i : $q_i = \sqrt{\frac{2K_i v_i}{s_i}}, i = \overline{1, n}$.

Минимальные издержки в единицу времени составляет:

$$L^* = \sum_{i=1}^n \sqrt{2K_i \cdot s_i \cdot v_i}.$$

В данную модель необходимо ввести ограничение на складские помещения, площадь которых составляет f , а на единицу товара или ресурса i -го вида необходимо f_i единиц складской площади. Ограничение на использование площадей склада имеет вид:

$$h \sum_{i=1}^n f_i \cdot q_i \leq f,$$

где h – нормировочный множитель, характеризующий уровень запаса или их поступление на склад в разное время.

Если $h=1$, то поставки поступили в одно время и имеется их максимальный уровень, а если $h = \frac{1}{2}$, то поставки прибыли в разное время и на складе имеется средний уровень запаса. Таким образом, считается, что $\frac{1}{2} \leq h \leq 1$.

Применим метод неопределенных множителей Лагранжа к решению задачи вида: 1) $L = \sum_{i=1}^n \left(\frac{K_i v_i}{q_i} + \frac{s_i q_i}{2} \right)$;

$$2) q_i = \sqrt{\frac{2K_i v_i}{s_i}}, i = \overline{1, n};$$

$$3) L^* = \sum_{i=1}^n \sqrt{2K_i \cdot s_i \cdot v_i};$$

$$4) h \sum_{i=1}^n f_i \cdot q_i \leq f.$$

Составим функцию Лагранжа:

$$L' = \sum_{i=1}^n \left(\frac{K_i v_i}{q_i} + \frac{s_i q_i}{2} \right) + \lambda \left(h \sum_{i=1}^n f_i \cdot q_i - f \right),$$

где λ – неопределенный множитель Лагранжа.

Из уравнений: $\frac{\partial L'}{\partial q_i} = 0$ и $\frac{\partial L'}{\partial \lambda} = 0$ найдем:

$$q_i = \sqrt{\frac{2K_i v_i}{s_i + 2\lambda h f_i}};$$

$$h \sum_{i=1}^n f_i \cdot q_i = f.$$

Так как $s_i \leq s_i + 2\lambda h f_i$, то размер оптимальной партии уменьшается, растет число партий поставок в плановом периоде и, следовательно, возрастают затраты на содержание запаса, т. е.:

$$\sum_{i=1}^n \sqrt{2K_i \cdot (s_i + 2\lambda h \cdot f_i) v_i} > \sum_{i=1}^n \sqrt{2K_i s_i v_i}.$$

Таким образом, если $q_i^* = \sqrt{\frac{2K_i v_i}{s_i + 2\lambda^* h f_i}}$, то минимальные затраты

работы системы в единицу времени с учетом ограничений на складские площади $L(q^*)$ равны:

$$L(q^*) = \sum_{i=1}^n \sqrt{2K_i \cdot (s_i + 2\lambda^* h \cdot f_i) v_i} = \sum_{i=1}^n \sqrt{s_i + 2\lambda^* h f_i} q_i^*.$$

При этом λ^* – неопределенный множитель Лагранжа, который показывает, на сколько можно сократить минимальные издержки функционирования системы в единицу времени, увеличив ограниченные складские площади на единицу. Его значение

определяется из уравнения:
$$h \sum_{i=1}^n f_i \sqrt{\frac{2K_i v_i}{s_i + 2\lambda^* h f_i}} = f .$$

7.9. Многопродуктовая модель размера партий поставок в случае нескольких ограничений

На практике чаще всего встречаются случаи, когда функционирование системы происходит при ограничениях на складские помещения и на оборотные средства, вложенные в запасы. Для обоснования оптимального решения такой модели определяют размеры партии поставок товаров или ресурсов вида i без учета ограничений, т. е. q_i^0 . Их значения подставляют в ограничения, и если они не нарушаются, то оптимальное решение задачи найдено. Если же нет, то поочередно в модель вводят ограничения и, используя метод неопределенных множителей Лагранжа, определяют оптимальные параметры системы: q_i^* , $L^*(q)$.

Используя информацию:

1) найдем s – издержки содержания единицы продукции в единице времени: $s = \alpha \cdot p$;

2) определим q_i^0 – оптимальные размеры партий поставок при отсутствии ограничений: $q_i^0 = \sqrt{\frac{2K_i v_i}{s_i}}, i = \overline{1, n}$;

3) проверим, существует ли ограничение на складские площади (при среднем уровне запаса $-h = \frac{1}{2}$): $h \sum_{i=1}^n f_i q_i^0 \leq f$

4) проверим существует ли ограничение по оборотным средствам (при среднем уровне запаса $-h = \frac{1}{2}$): $h \sum_{i=1}^n \alpha_i q_i^0 \leq A$.

5) так как оба ограничения существенны, то определим λ^* , введя сначала ограничение на складские площади, используя формулу:

$$h \sum_{i=1}^n f_i \sqrt{\frac{2K_i v_i}{s_i + 2\lambda^* h f_i}} = f, \text{ при } h = \frac{1}{2};$$

Из выше приведенного уравнения найдем значение λ^* .

6) найдем q_i^* – оптимальные размеры партий поставок с учетом ограничений на складские площади: $q_i^* = \sqrt{\frac{2K_i v_i}{s_i + 2\lambda^* h f_i}}$;

7) проверим, удовлетворяют ли q_i^* ограничению на оборотные средства: $h \sum_{i=1}^n \alpha_i q_i^* < A$, при $h = \frac{1}{2}$.

8) так как выше изложенное условие выполняется, то задача решена, в противном случае необходимо определить λ^* , используя формулу:

$$h \sum_{i=1}^n \alpha_i \sqrt{\frac{2K_i v_i}{s_i + 2\lambda^* h f_i}} = A, \text{ при } h = \frac{1}{2}.$$

Затем найти q_i^* – оптимальные размеры партий поставок с учетом ограничений на оборотные средства, используя формулу:

$$q_i^* = \sqrt{\frac{2K_i v_i}{s_i + 2\lambda^* h f_i}}.$$

И проверить, удовлетворяют ли найденные значения q_i^* требованию ограничения;

9) определим $L(q^0)$ – минимальные затраты работы системы в единицу времени без ограничений: $L(q^0) = \sum_{i=1}^n \sqrt{2K_i s_i v_i}$

10) определим $L(q^*)$ – минимальные затраты работы системы в единицу времени с учетом ограничений на складские площади и оборотные средства: $L(q^*) = \sum_{i=1}^n \sqrt{2K_i (s_i + 2\lambda^* h f_i) v_i}$

11) рассчитаем потери, связанные с ограничениями на складские помещения и оборотные средства: $L(q^*) - L(q^0)$

7.10. Многопродуктовая модель размера партий поставок с периодическими проверками при полном совмещении заказов

Рассмотрим многопродуктовую модель со стационарным детерминированным спросом, мгновенным получением заказа и не

допущением дефицита. Обозначим через τ период возобновления заказов, который является одинаковым для всех товаров или ресурсов вида i . Издержки размещения заказов K считаем пропорциональными числу заказов:

$$K = q(1 + \gamma \cdot n),$$

где q – фиксированные издержки, не зависящие от числа одновременно заказываемых товаров или ресурсов и от величины партий поставок;

γ – доля издержек, связанная с размещением заказа по каждому товару или ресурсу;

n – количество товаров или ресурсов.

Используя данные обозначения, запишем издержки размещения заказов и содержания запасов в единицу времени:

$$L = \frac{q(1 + \gamma \cdot n)}{\tau} + \frac{\tau}{2} \sum_{i=1}^n s_i v_i.$$

Решив уравнение $\frac{\partial L}{\partial \tau} = 0$, получим оптимальный период совместного размещения всех n товаров или ресурсов вида i :

$$\tau^* = \sqrt{\frac{2q(1 + \gamma \cdot n)}{\sum_{i=1}^n s_i v_i}}.$$

Величины оптимальных партий поставок определяются по формуле: $q_i^* = v_i \tau^*$, ($i = \overline{1, n}$);

Минимальные затраты работы системы в единицу времени составят L^* :

$$L^* = \sqrt{2q(1 + \gamma \cdot n) \sum_{i=1}^n s_i v_i} = \tau^* \sum_{i=1}^n s_i v_i.$$