ИССЛЕДОВАНИЕ ОПЕРАЦИЙ И ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

М. А. ТЫНКЕВИЧ
А. Г. ПИМОНОВ
С. А. ВЕРЕВКИН

г. Кемерово, 2015

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Кузбасский государственный технический университет
имени Т. Ф. Горбачева»

М. А. ТЫНКЕВИЧ А. Г. ПИМОНОВ С. А. ВЕРЕВКИН

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ И ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Учебное пособие

Рекомендовано Сибирским региональным учебно-методическим центром высшего профессионального образования для межвузовского использования в качестве учебного пособия для студентов, обучающихся по направлению подготовки 09.03.03 «Прикладная информатика»

Кемерово 2015

УДК 519.8:519.876.5

РЕЦЕНЗЕНТЫ

А. М. Гудов, доктор технических наук, доцент, декан математического факультета, заведующий кафедрой ЮНЕСКО по новым информационным технологиям федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Кемеровский государственный университет»

Кафедра вычислительной математики и компьютерного моделирования федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Национальный исследовательский Томский государственный университет» (заведующий кафедрой *А. В. Старченко*, доктор физико-математических наук, профессор)

Тынкевич, М. А.

Исследование операций и имитационное моделирование : учеб. пособие / М. А. Тынкевич, А. Г. Пимонов, С. А. Веревкин ; КузГТУ им. Т. Ф. Горбачева. – Кемерово, 2015. – 248 с.

ISBN 978-5-906805-12-6

Учебное пособие разработано на базе курсов лекций по математическому и имитационному моделированию, исследованию операций и методам оптимизации для студентов направления подготовки 09.03.03 «Прикладная информатика».

Авторы, ориентируясь на компетенции бакалавров и интерес читателей, не обладающих профессиональной математической подготовкой, постарались увязать математическую строгость изложения и алгоритмическое описание методов с многочисленными примерами экономической постановки решаемых задач. В учебном пособии приведен обзор основных методов исследования операций (линейное, нелинейное и динамическое программирование, теория игр, теория массового обслуживания), целей и средств имитационного моделирования экономических систем.

Пособие может быть полезно студентам, аспирантам и всем исследователям различных специальностей при самостоятельном ознакомлении с методами исследования операций и имитационного моделирования.

УДК 519.8:519.876.5

ISBN 978-5-906805-12-6

Оглавление

ПРЕДИСЛОВИЕ

Существуют различные мнения о происхождении Человека вообще и тем более о моменте, когда он получил право именоваться Homo sapiens и завоевать уважение своих прагматиков – соплеменников, мечтающих лишь об успехе в охоте на мамонта или хорошем урожае пшеницы. Возможно, что он раньше других научился посредством зарубок на камне или дереве предсказывать момент разлива Реки или место появления на небе Утренней звезды. Но, так или иначе, он пришел к осознанию понятия **Числа как абстракции** от конкретного количества пальцев, детей, полнолуний и т. п. Именно Число сопровождало цивилизацию, став основой науки, названной древними греками *математикой*. Эта наука, по крайней мере, долгие годы носила чисто прикладной характер, отвечая на вопросы «справедливого» распределения добычи и пахотных земель, взимания податей, строительства храмов, баллистики и др. В одних городах и странах перед носителями этой науки склонялись короли и султаны, в других презирали их за неспособность к «кулачному выяснению истины», в третьих избивали или сжигали как еретиков.

Тем не менее, математика выжила, создав универсальный язык мышления и общения представителей практически всех точных и естественных наук в попытках осознать причинно-следственные связи изучаемого ими явления и даже «поверить алгеброй гармонию» – построить так называемую **математическую модель**, то есть описать явление системой математических соотношений.

Всем известные закон всемирного тяготения $F = g \cdot m_1 m_2 / r^2$ или закон Ома $I = U / R$ представляют математические модели, позволяющие с достаточной надежностью дать **численный прогноз** итоговых оценок. Естественно, за такого рода законами стоят годы размышлений, реальных проб и ошибок незаурядных мыслителей.

Очевидно, что дешевле оценить устойчивость гидросооружения до начала строительства с помощью уравнений гидродинамики, температуру в пусковой шахте ракеты перед ее запуском с помощью уравнений теплопроводности и упругости, сочетать аэродинамическую трубу и уравнения газовой динамики при создании нового летательного аппарата.

Знаменитые книги Н. Винера «Кибернетика, или управление и связь в животном и машине» (1948 г.) и «Кибернетика и общество»

в сочетании с «Теорией информации» К. Шеннона и теорией автоматов Дж. Неймана создали плацдарм для вторжения математики не только в физику, но и во все естественные и технические науки. Не всегда это вторжение сопровождалось приветственными кликами – частично из-за скепсиса практиков, которым не повезло овладеть азами математики в школьные годы, частично из-за демагогичности предлагаемых моделей (существует обширная литература с моделями, либо содержащими лозунги «повысить, обеспечить, добиться…», либо не поддающимися численной реализации даже на современных суперкомпьютерах).

Многократно сложнее судьба математических моделей в сфере человеческой деятельности. Каждый второй знает, как и чему надо учить в школе и в вузе, каждый восьмой готов возглавить сборную страны по футболу, каждая политическая партия считает свой экономический курс *самым оптимальным* и готова вести страну по *градиенту* (уж больно слова красивые), мир экономических наук уже столетия критикует А. Смита, Д. Рикардо, К. Маркса, Дж. Кейнса и других, а поведение многих политиков и общностей, вопреки законам логики и историческому опыту, определяется наставлением королевы мужу, отправившемуся завоевывать чужую страну: «… получше их бей, а не то прослывешь пацифистом, и пряников сладких отнять у врага не забудь» (Б. Окуджава).

Однако существуют реально работающие полезные экономико-математические модели, отвечающие на вопросы «кто – кому – сколько?», «что и в каком объеме?» и другие. Не случайно, что многие математики последнего шестидесятилетия стали лауреатами Нобелевской премии по экономике.

Ниже мы остановимся на основных методах исследования операций, лежащих в основе экономико-математического моделирования, специфике моделирования в условиях случайности и на некоторых популярных имитационных моделях экономических систем и процессов.

Разумеется, от читателя потребуются хотя бы азбучные знания в области линейной алгебры, дифференциального и интегрального исчисления, теории вероятностей, математической статистики, алгоритмизации и языков программирования.

1. ВВЕДЕНИЕ В ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

1.1. Исследование операций и математическое моделирование

Термин «исследование операций» (*operation research*) возник в годы второй мировой войны как символ научного подхода к решению задач управления, в частности как «метод быстрого расчета программы поэтапного развертывания, подготовки и тылового снабжения» [1]. Сегодня исследование операций можно было бы определить как *совокупность методов поиска наилучших решений многообразия задач организационного управления при наличии тех или иных ограничений*.

Едва ли удастся четко определить круг таких задач, возникающих для человека в науке и технике, и методов их решения. Несомненно лишь одно – присутствие математического моделирования. Само по себе математическое моделирование в сфере человеческой деятельности появилось с момента возникновения алгебры. Иоганн Кеплер уже в 1615 году в статье «Новая стереометрия винных бочек» построил математическую модель экономически выгодного соотношения между геометрическими характеристиками упомянутой тары.

Непосредственное математическое моделирование начинается с попытки выяснить факторы, которые как-то характеризуют изучаемое явление (параметры явления) и допускают возможность количественной (численной) оценки. Затем необходимо выявить объективно существующие связи между этими факторами, представить их в виде математических соотношений (уравнений или неравенств).

Выбрав множество факторов (переменных величин), выявив соотношения между ними и ограничения на диапазон их значений, формулируем далее цель решаемой задачи – подбор среди всех допустимых значений переменных, которыми можно управлять, такого сочетания, при котором достигалось бы наилучшее (оптимальное) значение конкретной функции (целевой функции) от этих переменных. Существенно то, что цель должна быть единственной («за двумя зайцами не следует гнаться»), и, например, популярные в недавнем прошлом лозунги «*максимального* удовлетворения жизненных потребностей трудящихся при *минимальных* затратах» сами по себе звучат нелепо из-за противоположности объявленных целей.

Построенная модель явления представляет ценность, если она соответствует критерию практики – наблюдается согласие между теоретическими оценками и результатами наблюдений. Замечательно, если модель достаточно хорошо соответствует жизни при минимальном количестве учитываемых факторов. В большинстве же случаев для обеспечения близости теоретических оценок к реальности учитывают много значимых реальных факторов (тех, которыми можно управлять). Но рост сложности модели требует более сложного математического аппарата ее анализа и ведет к известному «проклятию размерности», делающему задачу выбора оптимальной политики практически неразрешимой.

Очевидно, что никакая математическая модель не может описать действительность с исчерпывающей полнотой (как утверждал известный Козьма Прутков, «никто не может объять необъятное»). Потому, согласно Р. Беллману [19], «… ученый должен идти прямой и узкой тропой между Западнями Переупрощения и Болотом Переусложнения».

Определенные сложности вносит в создаваемую модель и интерпретацию получаемых оценок учет факторов так называемой стохастической (вероятностной) природы, поскольку 100%-я предсказуемость получаемых оценок здесь вообще не достижима.

При исследовании физико-химических и даже ряда биологических процессов адекватность математической модели можно проверить постановкой эксперимента. В сфере же, где значима роль человеческого фактора, возникает множество проблем, с трудом поддающихся формализации из-за отсутствия однозначных воззрений на природу явлений и возможности проведения эксперимента.

Исследование операций, в зависимости от типа исследуемой математической модели и методов ее анализа, обычно разделяют на относительно самостоятельные дисциплины, такие как математическое программирование (центральное звено исследования операций), теория игр и статистических решений, сетевое планирование, теория расписаний, теория массового обслуживания, теория графов, вариационное исчисление и другие. Эта классификация достаточно условна, но общим для этих дисциплин является поиск наилучших решений на основе методов прикладной математики.

1.2. Они стояли у истоков исследования операций

В определенной мере создание математического программирования и его прикладной аспект связаны с выходом в 1939 г. работы выдающегося советского математика Леонида Витальевича Канторовича (1912 – 1986) «Математические методы в организации и планировании производства», где впервые была поставлена и решена задача линейного программирования. Но, как писал впоследствии Д. Данциг [1], «Л. Канторович добился многообещающих результатов, которыми в СССР пренебрегли». Его, наряду с единомышленниками (Г. Ш. Рубинштейн, В. А. Залгаллер, В. Л. Булавский, Д. Б. Юдин, Е. Г. Гольштейн и др.), обвиняли в стремлении «протащить буржуазные идеи в советскую науку», и лишь в 1959 г. Канторович стал известен на Западе благодаря публикации своей книги «Экономический расчет наилучшего использования ресурсов», написанной еще в 1942 г. [3, 4].

Т. К. Купманс, Д. Б. Данциг, Л. В. Канторович

В годы войны многие американские математики и экономисты были привлечены к поискам «оптимальных» решений в области экономического планирования, и к концу сороковых их интенсивный труд увенчался триумфальными итогами. Так в 1947 г. Джорджем Б. Данцигом (1914 – 2005) была выдвинута идея и предложен эффективный алгоритм *симплексного метода* для решения задач *линейного программирования* (терминология Данцига) [2]. Идея по достоинству была оценена Т. К. Купмансом[1] (1910 – 1985), в годы войны работавшим над транспортной моделью для снабжения войск на театре военных действий и сразу разглядевшим возможность

[1] Т. К. Купманс совместно с Л. В. Канторовичем были удостоены в 1975 г. Нобелевской премии по экономике.

применения этого метода для решения задач общего экономического планирования.

50-е годы ознаменовались фундаментальными результатами в области математического программирования и его приложений в экономике, ракетостроении и других сферах. В современных учебниках постоянно фигурируют имена талантливых математиков и экономистов того времени (Х. Хотеллинг, Г. Кун, А. Такер, К. Эрроу, Л. Гурвиц, Р. Гомори, Л. Форд, Д. Фалкерсон, С. Гасс, Т. Саати, Г. Вагнер и др.).

Джон фон Нейман

Нельзя не упомянуть Джона фон Неймана (1903 – 1957), одного из основателей кибернетики, участника Манхэттенского проекта (математически доказал осуществимость взрывного способа детонации атомной бомбы) и основоположника теории автоматов (выдвинул концепцию хранения команд компьютера в его внутренней памяти), что послужило огромным толчком к развитию вычислительной техники). В 1927 г. он сформулировал известную теорему о минимаксе – основополагающий элемент теории игр, а в 1944 г. вместе с Оскаром Моргенштерном подготовил к печати монографию «Теория игр и экономическое поведение» – фундамент для дальнейшего развития теории игр и статистических решений. Ему же принадлежит формулировка фундаментальной для математического программирования теоремы о двойственности (первое строгое ее доказательство было опубликовано впоследствии Такером, Куном и Джейлом).

Р. Э. Беллман (1920 – 1984)

В 1957 г. появляется монография выдающегося американского математика Ричарда Беллмана [7],

положившая начало одному из оригинальных методов исследования многошаговых процессов принятия решений – методу динамического программирования. Большой вклад в развитие методов оптимального управления динамическими системами внесла группа советских математиков во главе с Л. С. Понтрягиным.

Разумеется, некоторые результаты в области экономико-математического моделирования были получены и в предшествующие годы. Так в 1932 г. появилась знаменитая работа Василия Леонтьева, предложившего матричную структуру, которую он назвал «*межотраслевой моделью*» американской экономики типа «*вход – выход*». В дальнейшем она подверглась обобщению на многоальтернативность процессов производства продукта, приобрела фактор динамики и была использована в практике планирования экономики многих стран.

Истоки современной эконометрики, тесно связанной с исследованием операций, и теории массового обслуживания можно искать уже в трудах Якоба Бернулли (1654 – 1705) и Иоганна Бернулли (1744 – 1807) по теории вероятностей. Исключительный вклад в развитие этой области внесли П. Л. Чебышев, А. А. Марков, А. Эрланг, А. Я. Хинчин, С. Н. Бернштейн, А. Н. Колмогоров, Ежи Нейман, У. Феллер, У. С. Госсет (Стьюдент), Э. Пирсон и др. Бесспорный вклад в методы исследования операций (даже не задумываясь о приложениях в экономике) внесли творцы комбинаторной математики (Фибоначчи, Л. Эйлер, П. Ферма и др.) и создатели многообразия методов оптимизации.

Создание ЭВМ оказалось немаловажным стимулом развития численных методов оптимального планирования и управления. Увеличив вычислительные возможности человека, ЭВМ стимулировали появление методов Монте – Карло [27], «подающих надежду на спасение» не только при решении задач большой размерности, но и в так называемом имитационном моделировании сложных систем. Сегодня даже дилетант в состоянии решать многие задачи, над которыми бились тысячи исследователей в недавнем прошлом.

1.3. Математическое программирование и «проклятие размерности»

Не затрагивая пока других методов исследования операций, остановимся на задачах так называемого *математического программирования*.

Вообразите себя одним из руководителей крупного производства. Вы хотели бы осуществить мероприятия по увеличению прибыли, снижению затрат, повышению качества, обеспечению ритмичности и многие другие. Достижение всех этих целей одновременно – некорректная задача. Нужна некоторая глобальная цель, гармонично сочетающая локальные цели на основе выбранной системы уровней значимости или требующая поддержания всех целей на каком-то предельном уровне.

Очевидно, что достижение цели зависит от множества различных факторов (количества рабочих, уровня их квалификации, фондовооруженности, запасов сырья, спроса на создаваемый продукт и т. п.). Другими словами, цель является некоторой функцией от этих факторов.

Если вам известны (из статистических оценок, опыта или здравого смысла) функциональные связи между целью и факторами, которыми можно управлять, то возникает задача поиска сочетания значений производственных факторов, обеспечивающего оптимум для поставленной цели.

Задача **поиска экстремума (максимума или минимума) некоторой функции при наличии ограничений** на значения ее переменных и составляет **общую задачу математического программирования**[2].

Представляет ли какие-либо трудности ее решение?

Начнем с простой (?) задачи поиска максимума функции $F(x)$ при простейших условиях $A \le x \le B$ (при каком значении x, удовлетворяющем указанным условиям, $F(x)$ принимает самое большое значение?).

В предположении дифференцируемости $F(x)$ классическая математика предлагает взять производную от $F(x)$, приравнять ее к нулю, решить полученное уравнение (найти *критические точки*) и затем, например, выбрать максимальное из значений $F(x)$, вычисленных на концах заданного отрезка $[A, B]$ и в критических точках, попавших в него. Элементарно, если вы умеете решать уравнения, отличные от линейных и квадратных.

---

[2] Терминологически правильнее говорить о поиске супремума (*supremum*) или инфимума (*infimum*), т. е. наибольшего или наименьшего значения, поскольку понятия экстремума (*max, min*) ассоциируются с возможностью дифференцирования и обращением в нуль производных функции.

Если этот путь почему-то вас не устраивает, а на столе у вас стоит персональный компьютер, и вы хотя бы чуть-чуть в состоянии воспользоваться его вычислительными возможностями (догадываетесь хотя бы о том, что скрывается за буквосочетанием Excel, не говоря уже о возможностях MatLab или многочисленных систем программирования), то при желаемой точности искомого решения в 1 % доли заданного отрезка [A, B] достаточно разбить его на 100 частей, вычислить значения функции $F(x)$ в образовавшихся 101 точках и найти среди них максимальное.

Например, пытаясь найти максимальное значение функции $F(x) = x \cdot e^{-x} \ln(2 + x)$ для $x \in [0, 5]$, сооружаем и запускаем программу в среде MatLab, подобную приведенной ниже:

```
a=0; b=5; M=0; h=(b-a)/100;
for i=0 :100
    x=a+i*h; f=x*exp(-x)*log(1+x);
    if f>M  M=f; xopt=x end
end
xopt
M
```

Ответ {`xopt =1.6500 M =0.3088`} появится за доли секунды[3].

Для задачи с двумя переменными, состоящей в максимизации произвольной функции $F(x, y)$ при условиях $A \le x \le B$, $C \le y \le D$, попытка воспользоваться аппаратом классической математики будет успешной лишь в случае, когда есть уверенность, что максимум этой функции достигается внутри прямоугольника, определяющего множество допустимых точек (граничных точек здесь не две, а множество).

Если использовать численное решение, то при той же точности придется разбить интервалы по x и y на 100 частей и вычислять значения функции в 101 × 101 точках (время счета станет значительно, хотя и не слишком ощутимо, больше).

В реальной жизни управление зависит от десятков факторов. В случае наличия N переменных аналогичный подход к решению по-

[3] Поиск точки максимума вогнутой (минимума выпуклой) функции при точности 1 % интервала можно ускорить в 10 раз, а точность 10^{-6} интервала достичь всего лишь за 30 вычислений $F(x)$, если воспользоваться методом, основанным на числах Фибоначчи [7].

требует объема вычислений порядка 101^N. Если на вычисление одного значения тратить тысячную долю микросекунды (10^{-9} с), то при N = 8 время счета составит каких-то 125 дней, а при N = 10 – 3502 года.

Если вы надеетесь на то, что ниже будет изложен *простой* метод решения такого рода задач, то позвольте вас разочаровать. Универсального метода решения задач математического программирования нет, почему и приходится изобретать оригинальные методы для отдельных классов задач или искать приемлемые средства в руководствах по прикладной математике и библиотеках программ.

Так, для решения задач линейного программирования (задач с линейной целевой функцией и линейными же ограничениями) существует универсальный симплексный метод, имеющий множество модификаций и дающий решение задачи небольшой размерности за приемлемое время даже вручную (без использования компьютера). По крайней мере, компьютерное решение подобной задачи при N = 10 можно реализовать за секунды.

Увы, как только в задаче возникают нелинейности, для целенаправленности вышеуказанного поиска автор должен иметь некоторое представление о многообразии возможных методов решения своей задачи (известный извозчик, на которого надеялся небезызвестный Митрофанушка, не всегда дежурит у вашего подъезда).

Более того, в компьютерной рекламе ряда численных методов часто забывают предупреждать о том, что они при значительной размерности N или серьезной нелинейности часто выдают сообщения типа «за 512 итераций решение не достигнуто» или решение далеко от истинного из-за вычислительной погрешности (погрешность исходных данных + погрешность метода).

Ниже мы постараемся дать представление о классах задач, поддающихся решению известными методами, но нет гарантии, что для решения конкретной задачи вам не придется изобретать оригинальные приемы поиска оптимума.

2. ОСНОВЫ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Как известно из элементарной алгебры, выражение типа $a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n$, где a_i ($i = 1 \div n$) – некоторые константы и x_i – переменные величины, принято называть *линейным*. Так зависимость вида $y = A + Bx$ определяет величину y как линейную функцию от величины x и может быть изображена на плоскости в виде прямой линии (рис. 1). Аналогично в виде прямой на плоскости можно изобразить и линейное уравнение $\alpha x + \beta y = \gamma$ при произвольных значениях α, β, γ (α и $\beta \neq 0$ одновременно). Линейное неравенство $\alpha x + \beta y < \gamma$ (или $\alpha x + \beta y > \gamma$) определяет множество точек (полуплоскость), лежащее по какую-то сторону от соответствующей прямой линии (точки самой прямой исключены).

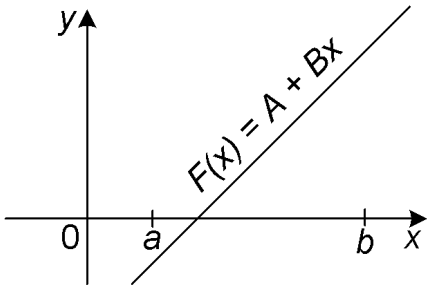


Рис. 1

В случае $n = 3$ геометрическим образом линейного уравнения $\alpha x + \beta y + \gamma z = \delta$ в трехмерном пространстве может служить плоскость, а линейному неравенству $\alpha x + \beta y + \gamma z \leq \delta$ соответствуют все точки полупространства, ограниченного соответствующей плоскостью (включительно). По аналогии для $n > 3$ используют термины *гиперплоскость* и *гиперпространство*.

Задача линейного программирования (ЛП) – частный случай задач математического программирования – состоит в отыскании максимума (минимума) линейной функции при наличии линейных ограничений на ее переменные. В случае такой функции одной переменной $F(x) = A + Bx$ в закрытом интервале $[a, b]$ решение элементарно – достаточно найти $F(a)$ и $F(b)$ и выбрать из них подходящее).

2.1. Линейная программа: случай двух переменных

Рассмотрим достаточно простой пример математического моделирования.

Пусть плановый прием в некотором вузе не превышает 5000 студентов, причем разрешен прием не более 4000 студентов своей страны и любого количества иностранных. Аудиторный фонд составляет 2800 мест. Педагогический персонал вуза – 400 человек.

По существующим нормативам для обучения 16 отечественных или 10 иностранных студентов требуется один преподаватель.

Статистика показывает, что посещаемость занятий студентами составляет соответственно 40 и 80 %. Ежегодно вуз получает дотацию 2 тыс. денежных единиц на каждого студента-соотечественника и в полтора раза большую сумму за каждого иностранного студента.

Предположив, что единственной целью вуза является максимизация денежных поступлений, попробуем выяснить наилучший план приема.

Обозначим искомую численность студентов соответственно через X и Y (очевидно, что $X \geq 0$ и $Y \geq 0$).

Ограничения на прием определяются требованиями $X + Y \leq 5000$ и $X \leq 4000$. Ограничение по аудиторному фонду с учетом не слишком высокой посещаемости занятий приводят к соблюдению условия $0{,}4\,X + 0{,}8\,Y \leq 2800$, а норматив численности студентов на одного педагога порождает условие $X / 16 + Y / 10 \leq 400$.

Поскольку итоговая денежная сумма составит $2\,X + 3\,Y$ тыс. денежных единиц, то задача сводится к минимизации функции

$$L(X, Y) = 2\,X + 3\,Y$$

при условиях

(1) $X / 16 + Y / 10 \leq 400$;
(2) $0{,}4\,X + 0{,}8\,Y \leq 2800$;
(3) $X + Y \leq 5000$;
(4) $X \leq 4000$;
(5) $X \geq 0$;
(6) $Y \geq 0$.

Другими словами, нам хочется найти значения X и Y, удовлетворяющие приведенным 6 условиям, при которых функция $L(X, Y)$ принимает самое большое значение.

Так как целевая функция $L(X, Y)$ и ограничения (1) – (6) линейны, мы имеем дело с задачей линейного программирования. Учитывая двумерность этой задачи, попытаемся решить ее графически.

Построив 6 прямых и выделив соответствующие полуплоскости, мы получаем **множество допустимых решений** (***планов***) в виде некоторого **выпуклого многоугольника** (рис. 2).

В какой же точке этого множества функция $L(X, Y)$ принимает самое большое значение (плоскость $Z = 2\,X + 3\,Y$ наиболее удалена по оси Z от координатной плоскости XY)?

Если взглянуть на рис. 3, можно прийти к выводу, что такая точка не может быть внутренней точкой многоугольника планов, и

наш поиск можно ограничить его вершинами или гранями.

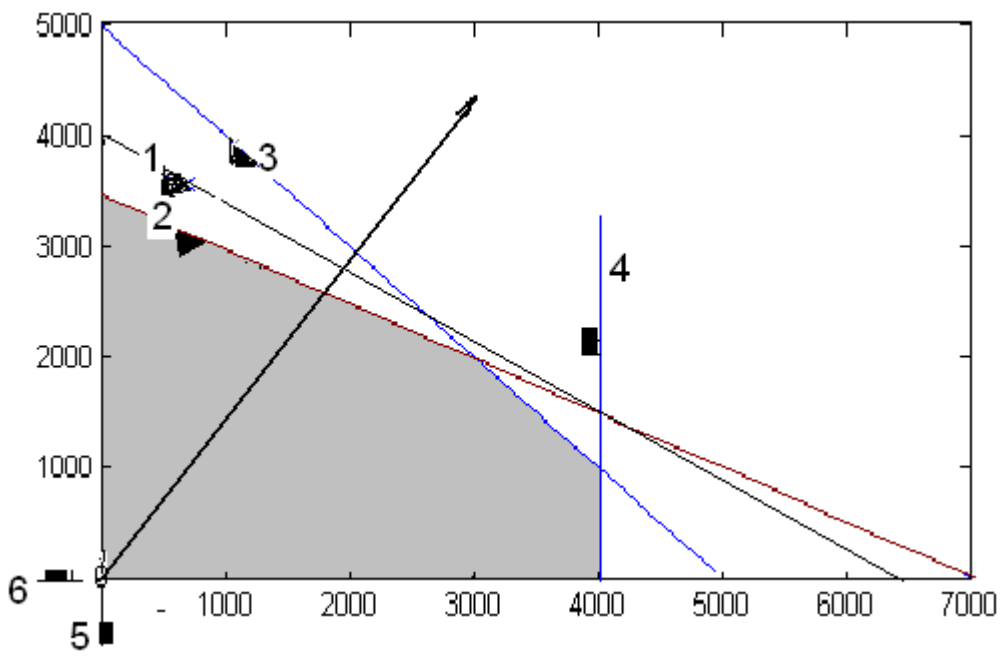


Рис. 2

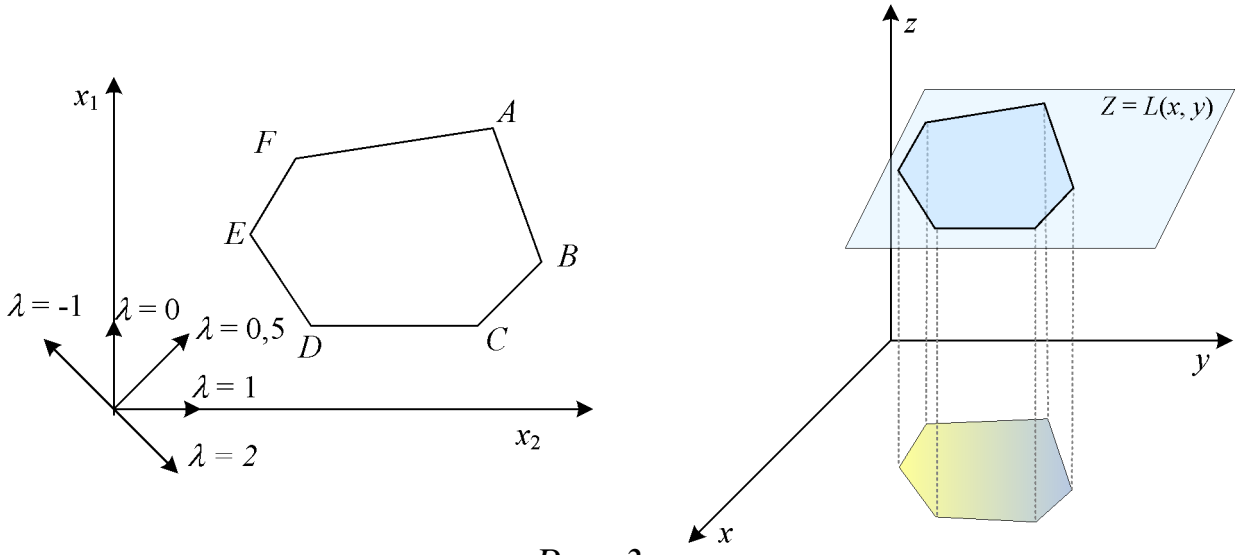


Рис. 3

Поскольку число вершин конечно (в нашем случае их 5), достаточно найти точки пересечения граничных прямых – координаты этих вершин (решить 5 систем линейных уравнений с двумя неизвестными) и вычислить соответствующие значения функции $L(X, Y)$.

Если вы хотите минимизировать затраты своей энергии на такой перебор, достаточно вспомнить понятие ***градиента функции в точке*** как вектора, составленного из частных производных функции, вычисленных в этой точке, и учесть, что *градиент указывает направление наибольшего возрастания функции в окрестности точки*.

Для нелинейных функций градиент меняется от точки к точке. Так, если $f(x, y) = x^2 + x\,y$, то grad $f(x) = \{2\,x + y, x\}$ в различных точках меняет свою ориентацию. В случае же линейной функции составляющие градиента совпадают с коэффициентами целевой функции, например grad$\{L(x, y) = 2\,x + 3\,y\} = \{2, 3\}$, т. е. градиент остается неизменным в любой точке плоскости. Опять-таки напрашивается вывод, что экстремумы линейной функции достигаются в вершинах множества планов (или на какой-то грани множества, если градиент перпендикулярен этой грани).

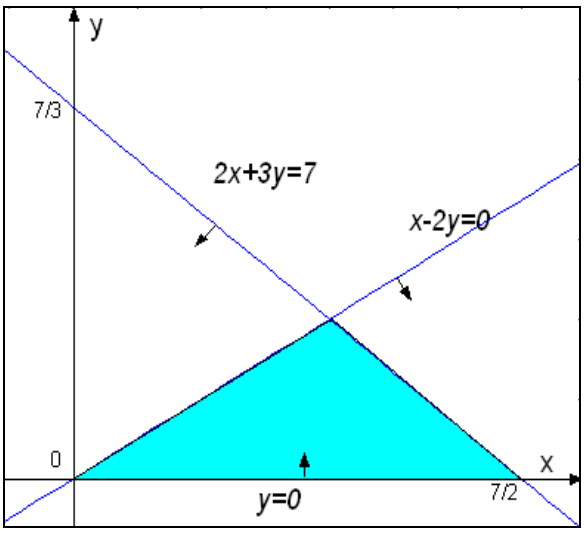


Рис. 4

$2\,x + 3\,y \le 7;\ x - 2\,y \ge 0;\ y \ge 0$

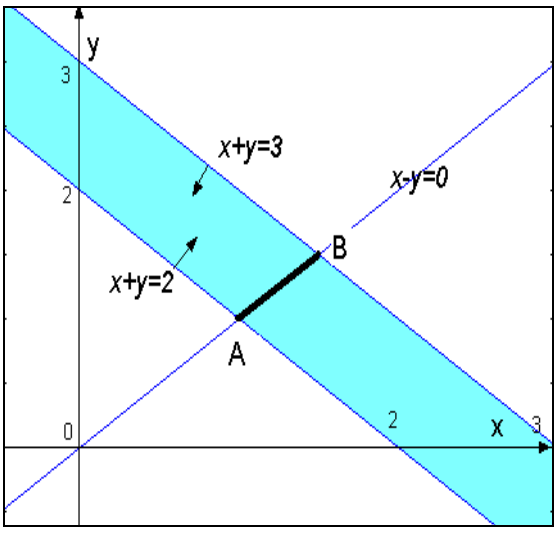


Рис. 5

$2 \le x + y \le 3;\ x - y = 0$

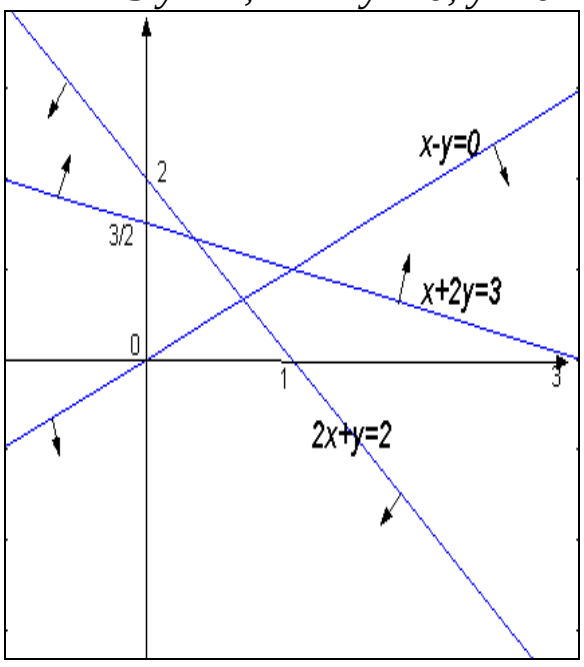


Рис. 6

$2\,x + y \le 2;\ x + 2\,y \ge 3;\ x - y \ge 0$

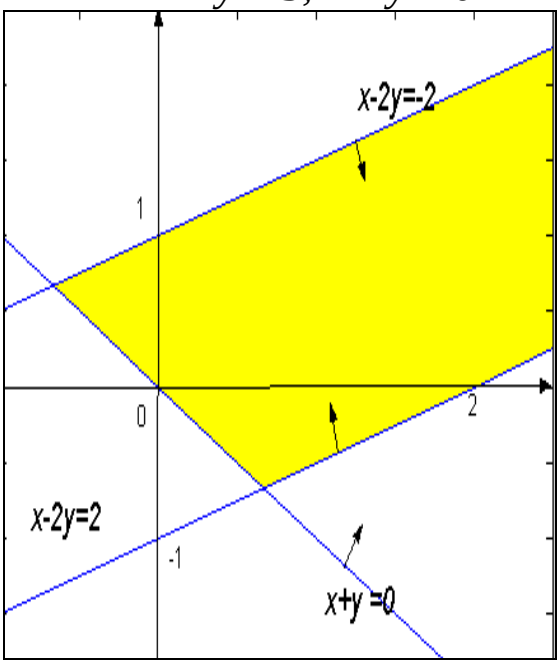


Рис. 7

$x - 2\,y \le 2;\ x - 2\,y \le 2;\ x + y \ge 0$

Для нашей задачи максимум явно достигается на пересечении прямых (2) и (3) – в точке с координатами (3000, 2000), т. е. оптимальный прием – 3 тыс. соотечественников и 2 тыс. иностранцев.

Рассматривая другие примеры систем линейных ограничений (рис. 4 – 7), мы делаем окончательный вывод, что ***для двухмерной задачи линейного программирования множество планов является выпуклым замкнутым многоугольником***. В частности, при наличии среди ограничений уравнений оно может оказаться отрезком (рис. 5), лучом или даже точкой. Ограничения могут оказаться противоречивыми (рис. 6).

Оптимальный план может не существовать (из-за противоречивости ограничений), оказаться единственным или представлять собой бесчисленное множество точек отрезка, служащего одной из граней множества планов (градиент перпендикулярен грани).

В случае неограниченного множества планов (рис. 7) может обнаружиться факт *неограниченности значений целевой функции по максимуму и* (*или*) *минимуму*.

Итак, в случае графического решения двумерной задачи ЛП достаточно отыскать множество планов (для этого надо уметь строить прямые линии, выбирать соответствующую полуплоскость и визуально выделять область, удовлетворяющую всем ограничениям), построить градиент, выбрать вершину – точку искомого экстремума и суметь решить систему двух линейных уравнений с двумя неизвестными (подстановками или методом Крамера).

В заключение заметим, что обнаруженные здесь свойства линейных программ переносятся и на случай трех переменных, где прямые превращаются в плоскости, полуплоскости – в полупространства, а многоугольник – в многогранник, но использовать графические приемы решения в этом случае будет почти всегда нереально.

2.2. Общие свойства линейных программ

Количество неизвестных величин, фигурирующих в постановке задачи, называют ее *размерностью*. Набор их значений, удовлетворяющих условиям задачи, называют *планом* (так программа производства – набор значений показателей, удовлетворяющий ограничениям по сырьевым, социальным и прочим факторам).

В роли неизвестных величин могут выступать, например, объемы выпуска продукции (обуви, приборов ночного видения или

микроскопов), распределение денежных средств на ликвидацию ветхого жилья и строительство медицинских учреждений, тираж планируемого издания сочинений М. Ю. Лермонтова и «заговоров от сглаза». Ограничения связаны с расходом материалов, наличием льгот на отдельные виды изданий и IQ современного читателя, рыночной конъюнктурой или личными предпочтениями, а целевая функция $L(X)$ может определять прибыль от произведенной продукции или объем неосвоенных средств.

Общая задача линейного программирования в канонической форме состоит в нахождении вектора[4] $X = (x_1, x_2, \ldots, x_n)$, обеспечивающего наибольшее (наименьшее) значение линейной функции

$$L(X) = \sum_{j=1}^{n} c_j x_j \qquad (1)$$

при условиях

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i = 1 \ldots m) \qquad (2)$$

$$x_j \geq 0 \qquad (j = 1 \ldots n) \qquad (3)$$

То же самое в развернутой форме: найти значения $x_1, x_2, \ldots, x_n$, при которых функция

$$L(X) = c_1x_1 + c_2x_2 + \ldots + c_nx_n$$

принимает наибольшее (наименьшее) значение при соблюдении условий

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1;$$
$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2;$$
$$\ldots.$$
$$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m;$$
$$x_1 \geq 0,\ x_2 \geq 0,\ \ldots,\ x_n \geq 0.$$

Остановимся на других компактных формах записи этой задачи.

Если обозначить через A_j вектор-столбец[5], составленный из коэффициентов при x_j в (2), а через B – вектор из элементов правой части системы, то (2) можно записать в векторной форме следующим образом:

---

[4] Под термином *вектор* в математике понимают последовательность величин (строка или столбец), не ассоциируя это понятие со «стрелкой».

[5] Здесь и далее под термином *вектор* мы понимаем вектор-столбец.

$$\sum_{j=1}^{n} A_j x_j = B. \qquad (4)$$

Если обозначить через X вектор неизвестных, через A – матрицу коэффициентов при неизвестных в (2) и через C – вектор коэффициентов линейной функции, задачу можно представить в матричном виде.

Максимизировать (минимизировать)

$$L(X) = C^T X$$

при условиях

$$A X = B, \quad X \geq 0$$

(здесь T – символ транспонирования).

В дальнейшем мы покажем, что любую линейную программу можно привести к такому виду.

Как мы уже отмечали ранее, вектор X, удовлетворяющий ограничениям задачи, называют *п л а н о м* и совокупность таких векторов – *м н о ж е с т в о м п л а н о в* .

По аналогии с линейным уравнением для случая двух или трех переменных линейное уравнение с $n \geq 4$ переменными можно интерпретировать как *гиперплоскость* в n-мерном пространстве и аналогичное неравенство как полупространство.

Нетрудно доказать [11], что m уравнений (2) и n неравенств (3) определяют замкнутое[6] множество планов как выпуклый многогранник и экстремумы (1) достигаются в его вершинах[7].

Множество планов ограничено $m + n$ гиперплоскостями и число вершин не превышает числа сочетаний из $m + n$ по n:

$$C_{m+n}^{n} = \frac{(m+n)!}{m!\,n!},$$

но может оказаться значительным (при $n = 10$ и m от 2 до 45).

В литературе по линейному программированию вместо термина *вершина* предпочитают использовать понятие опорного плана.

План называют о п о р н ы м , если он обращает в равенство

---

[6] Множество планов называют *замкнутым*, если его граница принадлежит этому множеству. Так множество $\{x + y \leq 3\}$ – замкнутое, а $\{x + y < 3\}$ – открытое.

[7] Точку X множества называют его *вершиной* (*крайней точкой*), если ее нельзя поместить внутри отрезка, соединяющего какие-то две точки множества, т. е. представить в виде $X = \lambda X_1 + (1 - \lambda) X_2$, где $0 < \lambda < 1$.

хотя бы n независимых ограничений (2) – (3). Поскольку *в вершине многогранника планов пересекаются хотя бы n граничных гиперплоскостей*, понятия опорного плана и вершины множества планов тождественны. Напрашивается вывод, что *оптимальный план всегда является опорным*.

Поскольку компоненты опорного плана должны обращать в равенство не менее n из имеющихся $m + n$ ограничений, то в (3) не более m ограничений будут выполняться в форме неравенств. Следовательно, *число положительных компонент опорного плана не превышает m*. Так, если решается задача при $n = 5$ и $m = 3$, то в ее оптимальном плане не более трех компонент положительны.

Опорный план, содержащий ровно m положительных компонент, называется невырожденным и в противном случае – *вырожденным* (m – число независимых ограничений в (2)).

Система m векторов A_j при положительных компонентах опорного плана называется ***б а з и с о м*** этого плана. Эта подсистема соответствующих столбцов матрицы A дает матрицу с ненулевым определителем (признак *линейной независимости векторов базиса*), поскольку в противном случае система уравнений единственного решения не имеет. Знание базиса автоматически определяет соответствующий опорный план.

Например, при ограничениях

$$\begin{pmatrix}3\\4\end{pmatrix}x_1+\begin{pmatrix}3\\5\end{pmatrix}x_2+\begin{pmatrix}9\\13\end{pmatrix}x_3+\begin{pmatrix}6\\8\end{pmatrix}x_4=\begin{pmatrix}9\\14\end{pmatrix},\ x_1,\ x_2,\ x_3,\ x_4\geq 0$$

за базис можно принять систему векторов A_1 и A_2 (определитель получаемой матрицы отличен от нуля), тогда в опорном плане $x_3 = 0$ и $x_4 = 0$, а из системы (разложения вектора B по векторам базиса)

$$\begin{pmatrix}3\\4\end{pmatrix}x_1+\begin{pmatrix}3\\5\end{pmatrix}x_2=\begin{pmatrix}9\\14\end{pmatrix}$$

получить неотрицательные $x_1 = 1$ и $x_2 = 2$.

Можно принять за базис и другие системы векторов, но не A_1 и A_4 (определитель получаемой матрицы обращается в нуль и решение системы уравнений не единственно).

2.3. Теоретические основы симплексного метода

Во избежание необходимости хаотического перебора предлагается *симплексный метод* [9, 11] – ***метод упорядоченного перебора опорных планов*** (упорядоченность обеспечивается монотонным изменением значения целевой функции при переходе к очеред-

ному плану).

Пусть стоит задача максимизации линейной функции

$$L(X) = \sum_{j=1}^{n} C_j x_j \tag{1}$$

при условиях

$$\sum_{j=1}^{n} A_j x_j = B, \tag{2}$$

$$x_j \geq 0, \quad j = 1, \ldots, n. \tag{3}$$

Предположим, что нам удалось найти опорный план X^0, в котором, например, первые m компонент отличны от нуля:

$$X^0 = (x_1^0, x_2^0, \ldots, x_m^0, 0, \ldots, 0), \tag{4}$$

соответствующий базис Б = ($A_1, A_2, \ldots, A_m$). Подставив в (2), имеем

$$\sum_{j=1}^{m} A_j x_j^0 = B. \tag{5}$$

Попытаемся выбрать другую систему базисных векторов с целью построения нового опорного плана, в котором k-я переменная ($k > m$) принимает ненулевое значение $\Theta > 0$:

$$X(\Theta) = (x_1(\Theta), x_2(\Theta), \ldots, x_m(\Theta), 0, \ldots, \Theta, \ldots, 0). \tag{6}$$

Подставив (6) в (2), получаем

$$\sum_{j=1}^{m} A_j x_j(\Theta) + A_k \Theta = B. \tag{7}$$

Разложим вектор A_k по векторам исходного базиса[8]

$$A_k = \sum_{j=1}^{m} Z_{jk} A_j. \tag{8}$$

Подставляя (8) в (7) с учетом (5), получаем

$$\sum_{j=1}^{m} A_j x_j(\Theta) + \Theta \sum_{j=1}^{m} A_j Z_{jk} = \sum_{j=1}^{m} A_j x_j^0, \tag{9}$$

откуда имеем

$$\sum_{j=1}^{m} A_j \left[x_j(\Theta) - x_j^0 + \Theta \cdot Z_{jk} \right] = 0. \tag{10}$$

---

[8] Для получения коэффициентов такого разложения придется решать систему m уравнений с m неизвестными, которая имеет единственное решение (еще раз напоминаем о линейной независимости базисных векторов и ненулевом определителе). Заметим, что в ситуации, когда базисные векторы являются единичными (образуют единичную матрицу), искомые коэффициенты совпадают с компонентами исходного вектора, поэтому в дальнейшем мы будем **«питать любовь» к единичному базису**.

Так как система уравнений (10) имеет единственное решение, то первые m компонент нового плана

$$x_j(\Theta) = x_j^0 - \Theta Z_{jk},\ j = 1, \ldots, m. \qquad (11)$$

Естественно потребовать неотрицательность компонент нового плана. Так как нарушение неотрицательности в (11) может возникнуть лишь при $Z_{jk} > 0$, то значение Θ нужно взять не превышающим наименьшего из отношений x_j^0 к положительным Z_{jk}.

Поскольку число положительных (базисных) компонент опорного плана должно оставаться равным m, то одну из первых m (ненулевых) компонент исходного плана обращаем в нуль выбором

$$\Theta = \min_{Z_{jk} > 0} \frac{x_j^0}{Z_{jk}}. \qquad (12)$$

Подставляя (11) в (1), имеем

$$L\{X(\Theta)\} = \sum_{j=1}^{n} C_j x_j(\Theta) = \sum_{j=1}^{m} C_j \left(x_j^0 - \Theta Z_{jk}\right) + C_k \Theta\ . \qquad (13)$$

Если обозначить

$$Z_k = \sum_{j=1}^{m} C_j Z_{jk},\ \Delta_k = Z_k - C_k, \qquad (14)$$

то (13) примет вид

$$L\{X(\Theta)\} = L(X^0) - \Theta \Delta_k. \qquad (15)$$

Отсюда напрашиваются следующие выводы.

Критерий 1. (критерий оптимальности). Если все $\Delta_k \geq 0$, то выбранный план для задачи максимизации оптимален. Для задачи на поиск минимума признак оптимальности – неположительность всех Δ_k.

Критерий 2. Если обнаруживается некоторое $\Delta_k < 0$ и хотя бы одно из $Z_{jk} > 0$, переход к новому плану увеличит значение целевой функции. В такой ситуации полагаем k-ю переменную равной Θ согласно (12) и преобразуем значения остальных переменных в соответствии с (11).

Критерий 3. Если найдется некоторое $\Delta_k < 0$, но все $Z_{jk} \leq 0$, то целевая функция не ограничена по максимуму (неограниченность по минимуму будет наблюдаться при $\Delta_k > 0$ и всех $Z_{jk} \leq 0$).

Это следует из того, что согласно (11) $x_j(\Theta) \geq 0$ при любом $\Theta > 0$ (в том числе при сколько угодно большом) и согласно (15)

можно неограниченно изменять значение целевой функции.

Предположение о том, что базисными (ненулевыми) являются первые m компонент плана, не является принципиальным и указание диапазона по j от 1 до m в (11) – (14) можно заменить на указание о принадлежности к базису «$j \in Б$».

Если все опорные планы задачи являются невырожденными (число положительных компонент равно m), то Θ отлично от нуля и переход к новому плану согласно (16) изменяет значение целевой функции, что из-за ограниченности количества опорных планов гарантирует достижение экстремума за конечное число шагов. При наличии вырожденных планов возврат к ранее рассмотренным планам возможен, но на практике такое *зацикливание* не возникало.

В сущности, симплексная процедура достаточно проста.

Руководствуясь школьным подходом к решению систем линейных уравнений, разрешите систему ограничений (2) относительно каких-то m переменных к виду

$$x_{i \in Б} = b_i^* + \sum_{j \notin Б} a_{ij}^*, \qquad (16)$$

но так чтобы значения b_i^* оставались неотрицательными. Если значения остальных переменных принять за нулевые, то выбранные равны b_i^*. Подставим выражения (16) в целевую функцию (1) и приведем подобные. Если среди коэффициентов при какой-то из оставшихся переменных найдется положительный (в задаче на поиск максимума), есть резон выбрать другой план – разрешить систему (16) относительно этой переменной, подставить в целевую функцию и т. д. Для минимизации затрат на такие преобразования можно прибегнуть к табличному представлению преобразований.

2.4. Прямой алгоритм симплексного метода

Пусть исходная задача приведена к канонической форме (далее мы покажем, как это делается) и начальный базис образует единичную матрицу. Тогда базисные компоненты опорного плана совпадают с правыми частями ограничений и коэффициенты Z_{jk} разложения вектора A_k по такому базису совпадают с компонентами этого вектора.

Для единообразия описания вычислительной процедуры далее будем пользоваться так называемыми симплексными таблицами вида

C баз	Базис плана	План X	C_1	C_2	...	C_m	C_{m+1}	...	C_k	...	C_n
			A_1	A_2	...	A_m	A_{m+1}	...	A_k	...	A_n
C_1	A_1	B_1	1	0	...	0	$Z_{1,m+1}$	...	Z_{1k}	...	Z_{1n}
C_2	A_2	B_2	0	1	...	0	$Z_{2,m+1}$	...	Z_{2k}	...	Z_{2n}
...	...	...	...	...	...	...	...	...	...	...	...
C_m	A_m	B_m	0	0	...	1	$Z_{m,m+1}$	...	Z_{mk}	...	Z_{mn}
Z_k		$L(X)$	Z_1	Z_2	...	Z_m	Z_{m+1}	...	Z_k	...	Z_n
Δ_k			Δ_1	Δ_2	...	Δ_m	Δ_{m+1}	...	Δ_k	...	Δ_n

В центральной части таблицы записываются коэффициенты при неизвестных в ограничениях, в столбце X – правая часть ограничений (базисные компоненты плана), в первой строке – коэффициенты целевой функции (линейной формы).

В первый столбец для удобства вычислений заносим коэффициенты целевой функции при базисных переменных (умножение его на столбец X с суммированием дает значение $L(X)$, аналогичное умножение его на столбец A_k даст Z_k). Последняя строка получается вычитанием из строки Z_k элементов первой строки.

Пусть стоит задача максимизации

$$L(X) = X_1 + 2X_2 + 2X_3 - X_5$$

при условиях

$$2X_1 + X_2 + X_3 + X_4 = 9;$$
$$X_1 + 2X_2 + X_3 + X_5 = 8;$$
$$X_1 + 3X_2 + 2X_3 + X_6 = 15;$$
$$X_k \geq 0,\ k = 1, \ldots, 6.$$

Здесь легко видеть наличие единичного базиса (A_4, A_5, A_6) и можно приступать к решению.

Переносим исходные данные в симплексную таблицу и получаем оценки для выбранного начального опорного плана.

C баз	Базис плана	План X	1	2	2	0	–1	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	9	2	1	1	1	0	0
–1	A_5	8	1	**2**	1	0	1	0
0	A_6	15	1	3	2	0	0	1
Z_k		$L(X)$	–1	–2	–1	0	–1	0
Δ_k		–8	–2	**–4**	–3	**0**	**0**	**0**

Так как имеются отрицательные значения Δ_k, то выбранный план $X = (0, 0, 0, 9, 8, 15)$ не оптимален. Резонно, например, перейти к другому опорному плану, где $X_2 \neq 0$ (ввести в базис вектор A_2).

Отыскав значение $\Theta = \min\left(\frac{9}{1}, \frac{8}{2}, \frac{15}{3}\right) = 4$, видим, что для сохранения неотрицательности компонент будущего плана из базиса нужно вывести вектор A_5. Соответственно, выражаем X_2 из второго уравнения (второе уравнение делим на коэффициент при X_2, т. е. на 2) и исключаем X_2 из остальных уравнений (из выбранного уравнения вычитаем разрешенное, умноженное на коэффициент при X_2).

C баз	Базис плана	План X	1	2	2	0	–1	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	5	3/2	0	1/2	1	–1/2	0
2	A_2	4	1/2	1	1/2	0	1/2	0
0	A_6	3	–1/2	0	**1/2**	0	–3/2	1
Z_k		$L(X)$	1	2	1	0	1	0
Δ_k		8	0	**0**	**–1**	**0**	2	**0**

Так как $\Delta_3 < 0$, выбранный план $X = (0, 4, 0, 5, 0, 3)$ не оптимален. Отыскав $\Theta = \min\left(\frac{5}{1/2}, \frac{4}{1/2}, \frac{3}{1/2}\right) = 6$ (третье уравнение), выражаем X_3 из третьего уравнения (делим его на коэффициент при X_3, т. е. на 1/2) и исключаем X_3 из остальных уравнений (вычитаем из них разрешенное, умноженное на коэффициент при X_3).

C баз	Базис плана	План X	1	2	2	0	–1	0
			A_1	A_2	A_3	A_4	A_5	A_6
0	A_4	2	**2**	0	0	1	1	–1
2	A_2	1	1	1	0	0	2	–1
2	A_3	6	–1	0	1	0	–3	2
Z_k		$L(X)$	0	2	2	0	–2	4
Δ_k		14	**–1**	**0**	**0**	**0**	–1	4

План $X = (0, 1, 6, 2, 0, 0)$ не оптимален. Из двух отрицательных значений Δ_k выберем $\Delta_1 < 0$. Отыскав $\Theta = \min(2 / 1, 1 / 1, –)$, видим, что из базиса можно удалить A_4 или A_2 (заметьте, что мы не рассматриваем отношение к отрицательной компоненте вектора).

C баз	Базис плана	План X	1	2	2	0	–1	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_1	1	1	0	0	1/2	1/2	–1/2
2	A_2	0	0	1	0	–1/2	**3/2**	–1/2
2	A_3	7	0	0	1	1/2	–5/2	3/2
Z_k		$L(X)$	1	2	2	1/2	–3/2	3/2
Δ_k		15	**0**	**0**	**0**	1/2	**–1/2**	3/2

Обратите внимание, что полученный план $X = (1, 0, 7, 0, 0, 0)$ вырожденный (одна из трех базисных переменных равна 0).

Т. к. $\Delta_5 < 0$ и $\Theta = 0$ (соответствует второму уравнению), перейдем к другому плану, вводя в базис A_5 вместо A_2. Заметьте, что здесь величина корректуры (см. (15)) $\Theta\Delta_5 = 0$ и значение целевой функции $L(X)$ останется неизменным.

C баз	Базис плана	План X	1	2	2	0	–1	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_1	1	1	–1/3	0	2/3	0	–1/3
–1	A_5	0	0	2/3	0	–1/3	1	–1/3
2	A_3	7	0	5/3	1	–1/3	0	2/3
Z_k		$L(X)$	1	7/3	2	1/3	–1	4/3
Δ_k		15	**0**	1/3	**0**	1/3	**0**	4/3

Так как все $\Delta_k \geq 0$, найденный план с компонентами (1, 0, 7, 0, 0, 0) оптимален и максимум значений $L(X)$ равен 15.

Полезно заметить, что здесь обращаются в нуль значения Δ_k только для базисных векторов и, следовательно, отсутствует возможность существования других оптимальных планов (с тем же значением целевой функции).

Рассмотрим другой пример, состоящий в максимизации целевой функции:

$$L(X) = 2X_1 - X_2$$

при условиях

$$-X_1 + X_2 + X_3 = 1;$$
$$3X_1 - 5X_2 + X_4 = 6;$$
$$X_k \geq 0,\ k = 1, \ldots, 4.$$

C баз	Базис плана	План X	2	– 1	0	0
			A_1	A_2	A_3	A_4
0	A_3	1	–1	1	1	0
0	A_4	6	**3**	–5	0	1
Z_k		$L(X)$	0	0	0	0
Δ_k		0	**–2**	1	**0**	**0**

C баз	Базис плана	План X	2	– 1	0	0
			A_1	A_2	A_3	A_4
0	A_3	3	0	**–2/3**	1	1/3
2	A_1	2	1	**–5/3**	0	1/3
Z_k		$L(X)$	2	–3/3	0	2/3
Δ_k		6	**0**	**–7/3**	**0**	2/3

Здесь после перехода от начального плана к очередному обнаруживаем $\Delta_2 < 0$ – признак неоптимальности найденного плана. Но все компоненты вектора A_2 неположительны и, соответственно, целевая функция не ограничена сверху (может принимать сколько угодно большие значения).

2.5. Приведение задачи к канонической форме

Прежде чем приступать к решению линейной программы симплексным методом, необходимо привести задачу к канонической форме (в приведенных выше примерах задачи были поставлены в канонической форме и с очевидностью отыскивался начальный базис и опорный план).

Рассмотрим более общий случай.

Если для переменной X_k отсутствует условие неотрицательности, ее заменяют разностью двух неотрицательных переменных

$$X_k = X_k^{'} - X_k^{''},\ X_k^{'} \geq 0,\ X_k^{''} \geq 0.$$

Если же $X_k \leq 0$, то производится замена $X_k = -X_k^{'},\ X_k^{'} \geq 0.$

Если некоторое из основных ограничений допускает неравенство, то вводят неотрицательную так называемую *ослабляющую* (*свободную*, *дополнительную*) переменную, уравновешивающую разность между левой и правой частями ограничения.

Например, ограничения

$$X_1 + 2\,X_2 + 3\,X_3 \geq 7;$$
$$X_1 - X_2 + X_3 \leq 2$$

преобразуются к виду

$$X_1 + 2\,X_2 + 3\,X_3 - X_4 = 7,\ X_4 \geq 0;$$
$$X_1 - X_2 + X_3 + X_5 = 2, \qquad X_5 \geq 0.$$

Если присутствует ограничение с отрицательной правой частью, его умножают на –1.

2.6. Выбор начального опорного плана и прямой алгоритм симплексного метода

Пусть задача приведена к канонической форме и компоненты вектора правой части неотрицательны. Если в системе векторов коэффициентов при переменных (матрице A) обнаруживается подсистема, образующая единичную подматрицу, то эти векторы образуют базис опорного плана и вектор правой части определяет базисные компоненты этого плана.

Если такой единичной подматрицы не обнаруживается, то либо придется перебирать все подсистемы из m уравнений с m неизвестными в надежде обнаружить неотрицательные решения, либо прибегнуть к методу *искусственного* базиса.

В последнем случае в ограничения добавляют неотрицатель-

ные так называемые ***искусственные переменные*** таким образом, чтобы возникла единичная подматрица коэффициентов, и эти переменные включают в целевую функцию с коэффициентом $+M$ для задачи минимизации или $-M$ при минимизации целевой функции, где $M > 0$ – сколько угодно большое число.

Полученная M-задача решается до получения оптимального плана.

Если в оптимальном плане М-задачи значения искусственных переменных равны нулю, то значения остальных компонент образуют оптимальный план исходной задачи.

Если в оптимальном плане М-задачи значение хотя бы одной из искусственных переменных отлично от нуля, то исходная задача не имеет ни одного плана (*ее ограничения противоречивы*).

Пусть стоит задача максимизации линейной формы (целевой функции)

$$L(X) = X_1 + 2\,X_2 - X_3$$

при условиях

$$2\,X_1 + X_2 - X_3 \le 7;$$
$$4\,X_1 - X_2 + 4\,X_3 \ge 6;$$
$$-X_1 + X_2 + X_3 \le 2;$$
$$X_k \ge 0,\ k = 1, 2, 3.$$

Приводим задачу к канонической форме, вводя три ослабляющих переменных, и получаем задачу в следующем виде.

Максимизировать

$$L(X) = X_1 + 2\,X_2 - X_3$$

при условиях

$$2\,X_1 + X_2 - X_3 + X_4 = 7;$$
$$4\,X_1 - X_2 + 4\,X_3 - X_5 = 6;$$
$$-X_1 + X_2 + X_3 + X_6 = 2;$$
$$X_k \ge 0,\ k = 1, \ldots, 6.$$

При попытке искать начальный единичный базис обнаруживаем лишь два единичных вектора A_4 и A_6. Вводим во второе уравнение искусственную переменную $X_7 \ge 0$ и включаем ее в $L(X)$ с коэффициентом $-M$, получая задачу:

максимизировать

$$L(X) = X_1 + 2\,X_2 - X_3 - M\,X_7$$

при условиях

$$2\,X_1 + X_2 - X_3 + X_4 = 7;$$

$$4X_1 - X_2 + 4X_3 - X_5 + X_7 = 6;$$
$$-X_1 + X_2 + X_3 + X_6 = 2;$$
$$X_k \ge 0,\ k = 1, \ldots, 7.$$

Решаем поставленную задачу прямым алгоритмом симплекс-метода (не забывайте, что $M > 0$ – очень большое число).

C баз	Базис плана	План X	1	2	–1	0	0	0	$-M$
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
0	A_4	7	2	1	– 1	1	0	0	0
$-M$	A_7	6	**4**	– 1	4	0	–1	0	1
0	A_6	2	–1	1	1	0	0	1	0
Z_k		–6M	$-4M$	M	$-4M$	0	M	0	$-M$
Δ_k			$\mathbf{-4M-1}$	$M-2$	$-4M-1$	**0**	M	**0**	**0**

==========

0	A_4	4	0	**3/2**	– 3	1	1/2	0	–1/2
1	A_1	3/2	**1**	– 1/4	1	0	–1/4	0	1/4
0	A_6	7/2	0	3/4	**2**	0	–1/4	1	1/4
Z_k		3/2	1	–1/4	1	0	–1/4	0	1/4
Δ_k			**0**	**–9/4**	2	**0**	–1/4	**0**	M+1/4

==========

2	A_2	8/3	0	1	– 2	2/3	1/3	0	–1/3
1	A_1	13/6	1	0	1/2	1/6	–1/6	0	1/6
0	A_6	3/2	0	0	**7/2**	–1/2	–1/2	1	1/2
Z_k		15/2	1	2	–7/2	3/2	1/2	0	–1/2
Δ_k			**0**	**0**	**–5/2**	3/2	1/2	**0**	M–1/2

==========

2	A_2	74/21	0	1	0	8/21	1/21	4/7	–1/21
1	A_1	41/21	1	0	0	5/21	–2/21	–1/7	2/21
–1	A_3	3/7	0	0	1	–1/7	–1/7	2/7	1/7
Z_k		60/7	1	2	–1	8/7	1/2	5/7	–1/7
Δ_k			**0**	**0**	**0**	8/7	1/2	**5/7**	M–1/7

Поскольку искусственная переменная обратилась в нуль, мы получили оптимальный план исходной задачи

$$X_{\text{opt}} = \left(\frac{41}{21},\ \frac{74}{21},\ \frac{3}{7}\right),\ L_{\max} = 60/7.$$

Рассмотрим еще один пример. Минимизировать

$$L(X) = 2X_1 - X_2 + X_3$$

при условиях

$$X_1 + X_2 + X_3 \le 3;$$
$$-3X_1 + X_2 - X_3 \ge 6;$$

$$X_k \geq 0,\ k = 1, 2, 3.$$

Приводим задачу к канонической форме вводом ослабляющих переменных X_4 и X_5 и для поиска начального базиса добавим искусственную переменную X_6.

C баз	Базис плана	План X	2 A_1	-1 A_2	1 A_3	0 A_4	0 A_5	M A_6
0	A_4	3	1	**1**	1	1	0	0
M	A_6	6	-3	1	-1	0	-1	1
Z_k		$6M$	$-3M$	M	$-M$	0	$-M$	M
Δ_k			$-3M-2$	$\boldsymbol{M+1}$	$-M-1$	**0**	$-M$	**0**

C баз	Базис плана	План X	2 A_1	-1 A_2	1 A_3	0 A_4	0 A_5	M A_6
-1	A_2	3	1	1	1	1	0	0
M	A_6	3	-4	0	-2	-1	-1	1
Z_k		$3M-3$	$-4M-1$	-1	$-2M-1$	$-M-1$	$-M$	M
Δ_k			$-4M-3$	**0**	$-2M-2$	$-M-1$	$-M$	**0**

Так как все $\Delta_k \leq 0$, найденный план оптимален. Однако значение искусственной переменной $X_6 = 3 > 0$, что свидетельствует о противоречивости ограничений поставленной задачи.

Заметим, что число «шагов» симплексного процесса имеет порядок $m \div 2m$, что существенно меньше возможного количества опорных планов, определяемого числом сочетаний из $m + n$ по n.

Замечание. *Если среди основных ограничений $L > 1$ условий представлены отношением $\geq$, то для уменьшения числа искусственных переменных после приведения к канонической форме можно выбрать уравнение с максимальной правой частью, а другие заменить результатом их вычитания из выбранного. В итоге получим $L - 1$ единичных векторов и потребность в единственной искусственной переменной.*

2.7. Двойственность в линейном программировании

Пусть стоит задача максимизации

$$L(X) = C^T X \tag{1}$$

при условиях

$$A X = B; \tag{2}$$

$$X \geq 0. \tag{3}$$

Задача минимизации

$$\tilde{L}(Y) = B^T Y \tag{4}$$

при условии

$$A^T Y \geq C \qquad (5)$$

называется *сопряженной* к задаче (1) – (3), и обе задачи образуют пару *двойственных* задач. Например,

Исходная задача:	Сопряженная задача:
максимизировать	минимизировать
$L(X) = X_1 + 3\,X_2 - X_3$	$\tilde{L}\,(Y) = 5\,Y_1 + 9\,Y_2$
при условиях	при условиях
$4\,X_1 - 7\,X_2 + X_3 = 5;$	$4\,Y_1 + 6\,Y_2 \geq 1;$
$6\,X_1 + 8\,X_2 - 2\,X_3 = 9;$	$-7\,Y_1 + 8\,Y_2 \geq 3;$
$X_1 \geq 0,\ X_2 \geq 0,\ X_3 \geq 0.$	$Y_1 - 2\,Y_2 \geq -1.$

Из (1) – (5) рядом элементарных преобразований можно получить и другую пару двойственных задач:

Минимизировать	Максимизировать
$L(X) = C^T X$	$\tilde{L}\,(Y) = B^T Y$
при	при
$A\,X = B;$	$A^T Y \leq C.$
$X \geq 0.$	

Путем приведения задачи к канонической форме и последующего использования вышеприведенных постановок можно получить так называемую *симметричную пару* двойственных задач:

Максимизировать	Минимизировать
$L(X) = C^T X$	$\tilde{L}\,(Y) = B^T Y$
при	при
$A\,X \leq B;$	$A^T Y \geq C;$
$X \geq 0.$	$Y \geq C.$

2.7.1. Первая теорема двойственности

Без особого труда можно доказать утверждение [11]:

Если одна из двойственных задач разрешима, то разрешима и сопряженная задача. При этом для оптимальных планов X^* и Y^* этих задач имеет место равенство значений целевых линейных функций

$$C^T X^* = B^T Y^*.$$

Если целевая линейная функция одной из задач не ограничена, то ограничения сопряженной задачи противоречивы. Если в одной из задач противоречивы ограничения, то в другой задаче либо не ограничена целевая функция, либо противоречивы ограничения.

2.7.2. Вторая теорема двойственности

Условия $X_j \geq 0$ и $\sum_{i=1}^{m} A_{ij} Y_i \geq C_j$ называют парой двойственных условий.

Без доказательства предлагаем следующую теорему (в дальнейшем мы покажем ее происхождение).

Если пара двойственных задач разрешима, то для их оптимальных планов в каждой паре двойственных условий если одно выполняется как строгое неравенство, то другое выполняется как равенство.

Ввиду особой значимости этой теоремы, рассмотрим серию примеров.

Пример. Примем за исходную задачу максимизации:

$$5 X_1 + 3 X_2$$

при условиях

$$X_1 + 2 X_2 \leq 4; \quad (1)$$
$$X_1 - X_2 \geq 2; \quad (2)$$
$$X_1 \geq 0; \quad (3)$$
$$X_2 \geq 0. \quad (4)$$

Приведением исходную задачу к канонической форме, получаем задачу максимизации:

$$5 X_1 + 3 X_2$$

при условиях

$$X_1 + 2 X_2 + X_3 = 4;$$
$$X_1 - X_2 - X_4 = 2;$$
$$X_1 \geq 0; \quad (3)$$
$$X_2 \geq 0; \quad (4)$$
$$X_3 \geq 0; \quad (1)$$
$$X_4 \geq 0. \quad (2)$$

Сопряженная задача состоит в минимизации целевой функции

$$4 Y_1 + 2 Y_2$$

при условиях

$$Y_1 + Y_2 \geq 5; \quad (3)$$
$$2 Y_1 - Y_2 \geq 3; \quad (4)$$
$$Y_1 \geq 0; \quad (1)$$
$$-Y_2 \geq 0. \quad (2)$$

Таким образом установлены следующие пары условий:

$$X_1 + 2\,X_2 \le 4,\ Y_1 \ge 0; \quad (1)$$

$$X_1 - X_2 \ge 2,\ -Y_2 \ge 0; \quad (2)$$

$$X_1 \ge 0,\ Y_1 + Y_2 \ge 5; \quad (3)$$

$$X_2 \ge 0,\ 2\,Y_1 - Y_2 \ge 3. \quad (4)$$

Если графически или симплексной процедурой удастся найти оптимальный план одной из задач, например, $X_{opt} = (4,\ 0)$, то оптимальный план сопряженной задачи отвечает условиям:

$$X_1 + 2\,X_2 = 4,\ Y_1 \ge 0; \quad (1)$$

$$X_1 - X_2 > 2,\ -Y_2 = 0; \quad (2)$$

$$X_1 > 0,\ Y_1 + Y_2 = 5; \quad (3)$$

$$X_2 = 0,\ 2\,Y_1 - Y_2 \ge 3. \quad (4)$$

Из пары равенств $-Y_2 = 0$ и $Y_1 + Y_2 = 5$ получаем решение $Y = (5,\ 0)$, удовлетворяющее всем остальным условиям и, следовательно, являющееся оптимальным планом сопряженной задачи. Подтверждением тому является и равенство значений целевых функций $5 \cdot 4 + 3 \cdot 0 = 4 \cdot 5 + 2 \cdot 0$.

Заметим, что если бы мы ошиблись в выборе X_{opt}, то получили бы противоречивые условия для Y_{opt}. Соответственно, вторую теорему двойственности можно использовать как основу проверки планов на оптимальность.

Обратите внимание на случай отсутствия требования неотрицательности, например, нет условия $X_1 \ge 0$. Здесь приходится прибегнуть к замене $X_1 = X_1' - 3\,X_1''$, $X_1' \ge 0$, $X_1'' \ge 0$. Соответственно, в сопряженной задаче возникает пара условий

$$Y_1 + Y_2 \ge 5;$$

$$-Y_1 - Y_2 \ge -5,$$

которую можно заменить равенством $Y_1 + Y_2 = 5$.

Для минимизации затрат времени на приведение условий задачи к каноническому виду можно учитывать два полезных замечания:

– если в исходной задаче нет условия неотрицательности на некоторую переменную, то построенное по коэффициентам при этой переменной ограничение сопряженной задачи выполняется равенством;

– если некоторое (основное) ограничение исходной задачи задано равенством, то двойственного условия сопряженной задачи не существует.

Если некоторая задача решается прямым алгоритмом симплексного метода, то *решение сопряженной задачи можно видеть в строке Z конечной симплексной таблицы в позициях, соответствующих начальному единичному базису*.

Нетрудно заметить, что разумнее из двух задач решать симплексным методом ту, которая после приведения к канонической форме содержит наименьшее число основных (типа $A\,X = B$) ограничений.

Постановка двойственных задач с необходимостью возникает при решении больших линейных программ со специфическими матрицами коэффициентов. Примерами такого рода являются так называемые транспортные, распределительные и другие задачи.

2.7.3. Экономическая интерпретация симметрической пары двойственных задач

Представьте себе, что вам удалось некоторую производственную задачу сформулировать как задачу линейного программирования. Например, решая задачу выпуска продукции при ограниченных ресурсах (видах сырья), вы ввели серию обозначений:

X_j – искомый объем производства j-го вида продукта ($j = 1 \div n$),

B_i – запас i-го вида сырья ($i = 1 \div m$),

A_{ij} – затраты i-го вида сырья на создание единицы j-го вида продукта,

C_j – стоимость производства единицы j-го вида продукта.

В результате некоторых размышлений вы пришли к задаче поиска значений X_j, удовлетворяющих условиям (расход сырья на производство продукции не превышает его запасов)

$$\sum_{j=1}^{n} A_{ij} X_j \le B_i \qquad (i = 1..m);$$

$$X_j \ge 0 \ (j = 1 \,..\, n)$$

и обеспечивающих максимум выручки от продажи продукции

$$L(X) = \sum_{j=1}^{n} C_j X_j .$$

Подставив конкретные данные (расценки на готовую продукцию, объем запасов сырья, расходные нормы и т. д.), в итоге получаем оптимальную производственную программу.

Событие радостное, но всякое полученное знание требует своего развития, заставляет задуматься, в частности, о разумности принятых расценок на продукцию, о соответствии их тем ценам, по которым вы приобретали сырье.

Обозначив через Y_i *объективную оценку стоимости единицы* i*-го сырья*, осознаем, что совокупная стоимость ресурсов, затраченных на производство единицы j-го продукта, должна быть не меньше объявленной стоимости, то есть

$$\sum_{i=1}^{m} A_{ij} Y_i \geq C_j \ , \ j = 1..n \quad ,$$

$$Y_i \geq 0, \ i = 1 .. m,$$

и общая стоимость затраченного сырья должна быть минимальной

$$\tilde{L}(Y) = \sum_{i=1}^{m} B_i Y_i \to \min \quad .$$

Высказанные соображения приводят к целесообразности решения так называемой *симметричной пары двойственных задач*:

максимизировать	минимизировать
$L(X) = C^T X$	$\tilde{L}(Y) = B^T Y$
при $A X \leq B;$	при $A^T Y \geq C;$
$X \geq 0.$	$Y \geq 0.$

Напомним, что в приведенной матричной записи C, B, X, Y выступают как векторы-столбцы (T – знак транспонирования).

Здесь первая теорема двойственности требует равенства стоимости затраченных ресурсов и объявленной стоимости произведенной продукции.

Что касается второй теоремы о соотношениях в парах двойственных условий для оптимальных планов, то обнаружение неравенства $\sum_{j=1}^{n} A_{ij} X_j < B_i$ при некотором i говорит о том, что i-е сырье не является лимитирующим и его объективная стоимость $Y_i = 0$.

Таким образом, в рассмотренной постановке *двойственная задача является математической формулировкой объективной оценки всех производственных факторов*.

2.7.4. Постоптимальный анализ и устойчивость решений

Получив оптимальный план (программу производства продукции на очередной период), вы не гарантированы от того, что завтра изменятся цены на сырье, часть какого-то сырья окажется похищенной или изменится технология производства (последнее встречается в жизни гораздо реже остальных случаев). Останется ли найденный вами план по своей структуре оптимальным или потребуется искать таковой заново?

Другими словами, обладает ли найденное оптимальное решение устойчивостью при изменении значений параметров задачи (исходных данных)? В каких диапазонах можно менять значения тех или иных параметров при сохранении оптимальности найденного плана?

Пусть нам удалось получить решение задачи максимизации функции $L(X) = C^T X$ при условиях $A X \le B$, $X \ge 0$, то есть найден оптимальный план и (самое существенное!) соответствующая ему базисная система векторов условий $Б_{\text{opt}}$. Обратите внимание на то, что если задача решалась каким-то алгоритмом симплексного метода, то в последней симплексной таблице на месте исходного единичного базиса получается матрица $Б_{\text{опт}}^{-1}$, обратная к матрице, составленной из векторов упомянутой системы[9].

Можно показать, что ненулевые (базисные) составляющие X_{opt} определяются как произведение $Б_{\text{опт}}^{-1} \cdot B$, а компоненты решения сопряженной задачи как $C_{\text{опт}}^T \cdot Б_{\text{опт}}^{-1}$, где $C_{\text{опт}}^T$ – коэффициенты целевой функции $L(X)$ при базисных переменных. В векторно-матричной записи высказанные замечания представимы в виде

$$X_{\text{opt}} = Б_{\text{опт}}^{-1} \cdot B,\ Y_{\text{опт}}^T = C_{\text{опт}}^T \cdot Б_{\text{опт}}^{-1}$$

Отыскав

$$L(X_{\text{опт}}) = C_{\text{опт}}^T\ X_{\text{опт}} = C_{\text{опт}}^T\ Б_{\text{опт}}^{-1}\ B,$$

---

[9] Квадратная матрица, обозначенная как A^{-1}, называется *обратной* по отношению к квадратной матрице A, если их произведение дает единичную матрицу $A^{-1}A=AA^{-1}=E$. Обратная матрица существует, если строки (или столбцы) исходной матрицы линейно независимы, то есть если определитель отличен от нуля. Здесь существование обратной матрицы несомненно, поскольку базис – система линейно независимых векторов.

$$\tilde{L}(Y_{\text{опт}}) = B^T\, Y_{\text{опт}} = B^T\,(C_{\text{опт}}^T\, Б_{\text{опт}}^{-1})^T = B^T\,(Б_{\text{опт}}^{-1})^T\, C_{\text{опт}} = (C_{\text{опт}}^T\, Б_{\text{опт}}^{-1}\, B)^T$$

и убедившись в равенстве правых частей этих выражений, можете лишний раз увидеть выполнение первой теоремы двойственности.

Например, решая задачу максимизации

$$L(X) = 3\,x_1 + 2\,x_2$$

при условиях

$$x_1 + 5\,x_2 \le 7;$$

$$5\,x_1 + x_2 \le 8;$$

$$x_1,\, x_2 \ge 0,$$

получаем последовательность симплексных таблиц, где матрица оптимального базиса состоит из векторов коэффициентов при x_2 и x_1:

C баз	Базисные переменные	План X	3	2	0	0
			x_1	x_2	x_3	x_4
0	x_3	7	1	5	1	0
0	x_4	8	5	1	0	1
z_k		0	0	0	0	0
Δ_k			–3	–2	·	·
==========						
0	x_3	27/5	0	24/5	1	–1/5
3	x_1	8/5	1	1/5	0	1/5
z_k		24/5	3	3/5	0	3/5
Δ_k			·	–7/5	·	3/5
==========						
2	x_2	**9/8**	0	1	**5/24**	**–1/24**
3	x_1	**11/8**	1	0	**–1/24**	**5/24**
z_k		**51/8**	3	2	**7/24**	**13/24**
Δ_k			·	·	7/24	13/24

Поскольку здесь базис оптимального плана образован векторами коэффициентов при x_2 и x_1, а начальный базис – векторами коэффициентов при x_3 и x_4, то

$$Б_{\text{опт}} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix},\; Б_{\text{опт}}^{-1} = \begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix}.$$

Можете убедиться, что оценки

$$Б_{\text{опт}}^{-1} \cdot B = \begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 9/8 \\ 11/8 \end{bmatrix} = \begin{bmatrix} X_2 \\ X_1 \end{bmatrix}_{\text{опт}},$$

$$Б_{\text{опт}}^{-1} = \begin{bmatrix} 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix} = \begin{bmatrix} \frac{7}{24} & \frac{13}{24} \end{bmatrix} = Y_{\text{опт}}$$

соответствуют выводам, фигурирующим в итоговой симплексной таблице.

Из приведенного примера видно, что знание базиса оптимального плана автоматически дает возможность найти компоненты этого плана и решение сопряженной задачи (не забывайте обращать внимание на порядок следования базисных переменных).

А что если значения b_i изменятся на ε_i? Останется ли базис оптимального плана неизменным и решение можно искать как $X_{\text{опт}} = Б_{\text{опт}}^{-1} \cdot B$ при любом B?

Очевидно, что это справедливо, если компоненты соответствующего оптимального плана останутся неотрицательными:

$$X_{\text{опт}} = Б_{\text{опт}}^{-1} \cdot (B + \varepsilon) \geq 0 .$$

Для нашего примера

$$X_{\text{опт}} = Б_{\text{опт}}^{-1} \cdot (B + \varepsilon) =$$

$$= \begin{bmatrix} \frac{5}{24} & -\frac{1}{24} \\ -\frac{1}{24} & \frac{5}{24} \end{bmatrix} \cdot \begin{bmatrix} 7 + \varepsilon_1 \\ 8 + \varepsilon_2 \end{bmatrix} = \begin{bmatrix} 9/8 \\ 11/8 \end{bmatrix} + \begin{bmatrix} \frac{5}{24}\varepsilon_1 - \frac{1}{24}\varepsilon_2 \\ -\frac{1}{24}\varepsilon_1 + \frac{5}{24}\varepsilon_2 \end{bmatrix} \geq 0,$$

то есть условие сохранности базиса определяется решением системы неравенств:

$$\begin{aligned} &\tfrac{9}{8} + \tfrac{5}{24}\varepsilon_1 - \tfrac{1}{24}\varepsilon_2 \geq 0; \quad && -5\varepsilon_1 + \varepsilon_2 \leq 27; \\ &\tfrac{11}{8} - \tfrac{1}{24}\varepsilon_1 + \tfrac{5}{24}\varepsilon_2 \geq 0 \equiv && \varepsilon_1 - 5\varepsilon_2 \leq 33. \end{aligned}$$

Если вам известны значения ожидаемых корректур ε_i, то достаточно проверить для них выполнение полученных неравенств. Если же ограничиться вариантами, где меняется единственное из значений b_i при сохранности остальных, то вынести суждение можно без особого труда.

Так, если принять $\varepsilon_2 = 0$, то получается система неравенств с одной переменной, решение которой доступно школьнику:

$$\begin{aligned} &\tfrac{9}{8} + \tfrac{5}{24}\varepsilon_1 \geq 0; \\ &\tfrac{11}{8} - \tfrac{1}{24}\varepsilon_1 \geq 0, \end{aligned}$$

отсюда имеем $-\frac{27}{5} \leq \varepsilon_1 \leq 33$, то есть при изменении ресурса b_1 в диапазоне $\left[7 - \frac{27}{5} = \frac{8}{5};\ \ 7 + 33 = 40\right]$ базис оптимального плана остается неизменным и для выяснения диапазона изменения компонент

оптимального плана достаточно искать $X_{\text{опт}} = Б_{\text{опт}}^{-1} \cdot B$ при предельных значениях b_1.

Аналогично можно сделать выводы в случае возможной корректуры коэффициентов целевой функции. Так при корректуре коэффициента при x_1 имеем

$$C_{\text{опт}}^{T} \cdot Б_{\text{опт}}^{-1} = \begin{bmatrix} 2 & ,3+\Delta \end{bmatrix} \cdot \begin{bmatrix} 5/24 & -1/24 \\ -1/24 & 5/24 \end{bmatrix} = \begin{bmatrix} \frac{7}{24} - \frac{\Delta}{24}; & \frac{13}{24} + \frac{5\Delta}{24} \end{bmatrix} \geq 0,$$

откуда допустимый диапазон корректуры $-13/5 \leq \Delta \leq 7$.

Знание $Б_{\text{опт}}^{-1}$ может облегчить анализ ситуации, когда появляются новые виды продукции со своими характеристиками, определяемые вектором A_k.

Не пересчитывая симплексную таблицу заново, можно получить соответствующий ее столбец умножением $Б_{\text{опт}}^{-1}$ на A_k.

2.8. Параметрическое линейное программирование

Параметрическое программирование связано с изучением задач, в которых целевая функция или ограничения зависят от одного или нескольких параметров.

Необходимость рассмотрения подобных задач обусловлена различными причинами, основная из которых связана с тем, что исходные данные для численного решения любой реальной задачи оптимизации практически всегда определяются приближенно или могут изменяться под влиянием каких-то факторов, что может существенно сказаться на оптимальности выбираемой программы (плана) действий. В роли таких факторов могут выступать время, температура, курс доллара по отношению к рублю, цена на сырье, удаленность от поставщиков и др. Соответственно, чтобы быть готовым к изменениям ситуации (исходных данных), при решении оптимизационной задачи разумно указывать не конкретные данные, а *диапазон их возможного изменения*.

С математической точки зрения параметрическое программирование выступает в качестве средства *анализа чувствительности решения к вариации исходных данных*, оценки *устойчивости* решения.

В отличие от *постоптимального анализа*, где мы пытаемся выяснить диапазоны вариации характеристик задачи, при которых структура найденного оптимального плана остается оптимальной,

здесь мы включаем некоторые параметры в математическую модель и строим оптимальные решения как функции от них.

Сразу оговоримся, что универсальных методов анализа устойчивости решений произвольной задачи математического программирования нет и в случае множественности параметров или нелинейных связях трудности такого анализа даже с помощью средств современной вычислительной техники исключительны.

Рассмотрим задачу параметрического линейного программирования, в которой **только коэффициенты целевой функции линейно зависят от некоторого единственного параметра λ**:
отыскать максимум (или минимум) функции

$$L(X, \lambda) = \sum_{j=1}^{n} (C_j + D_j \lambda) X_j$$

при условиях

$$\sum_{j=1}^{n} A_j x_j = B,\ X_j \geq 0,\ j = 1 \ldots n;$$

$$\lambda_1 \leq \lambda \leq \lambda_2.$$

Если обратиться к геометрической интерпретации задачи, то можно заметить, что градиент целевой функции зависит от параметра. Например, для целевой функции $L(X, \lambda) = \lambda X_1 + (1 - \lambda) X_2$ при различных значениях параметра λ градиент определяет различные направления роста функции.

Нетрудно видеть (рис. 3), что если при некотором значении λ максимум достигается в вершине А, то небольшая вариация этого значения несколько изменит направление градиента, но не изменит положение точки максимума. Отсюда напрашивается вывод, что некоторый план, оптимальный при $\lambda = \lambda_0$, оптимален и в окрестности λ_0, т. е. при $\alpha \leq \lambda \leq \beta$, где $\lambda_0 \in [\alpha, \beta]$.

Можно заметить, что при градиенте, перпендикулярном некоторой грани множества планов, имеем два разных оптимальных опорных плана с одним и тем же значением целевой функции.

Можно доказать, что *если целевая функция при* $\lambda = \lambda_0$ *не ограничена, она не ограничена при всех* λ*, больших или меньших* λ_0.

Процедура решения задач параметрического линейного программирования в случае зависимости от параметра коэффициентов целевой функции незначительно отличается от обычного симплексного метода (примеры решения подобных задач приведены ниже).

При **зависимости от параметра компонент вектора правых частей ограничений,** т. е. при поиске экстремума функции

$$L(X) = \sum_{j=1}^{n} C_j X_j$$

при условиях

$$\sum_{j=1}^{n} A_j X_j = B + \lambda D, X_j \geq 0, j = 1 \ldots n, \lambda_1 \leq \lambda \leq \lambda_2,$$

во избежание сложностей, связанных с требованием сохранения неотрицательности компонент плана при любых λ (сохранения неотрицательности правой части системы уравнений при всех ее тождественных преобразованиях), достаточно поставить и решить сопряженную задачу, воспользовавшись вышеупомянутым алгоритмом решения задач параметрического линейного программирования при зависимости от параметра коэффициентов целевой функции, и с помощью известных двойственных соотношений отыскать решение исходной задачи.

Пример 1. Рассмотрим задачу минимизации

$$L(X, \lambda) = \lambda X_1 - \lambda X_2 - X_3 + X_4$$

при условиях

$$3 X_1 - 3 X_2 - X_3 + X_4 \geq 5;$$
$$2 X_1 - 2 X_2 + X_3 - X_4 \leq 3;$$
$$X_k \geq 0, k = 1 .. 4; -\infty < \lambda < \infty.$$

Как обычно, приводим задачу к канонической форме и с использованием метода искусственного базиса отыскиваем начальный опорный план $X^0 = (0, 0, 0, 0, 0, 3, 5)$ с $L(X^0, \lambda) = 5M$.

C баз	Базис 1	План $X_{баз}$	λ	$-\lambda$	-1	1	0	0	M
			A_1	A_2	A_3	A_4	A_5	A_6	A_7
M	A_7	**5**	3	–3	–1	**1**	–1	**0**	**1**
0	A_6	**3**	2	–2	1	–1	0	**1**	**0**
Δ_k		$5M$	$3M-\lambda$	$-3M+\lambda$	$-M+1$	***M*–1**	$-M$	0	0

Так как определяющую роль здесь играет величина M, превышающая все величины задачи, не обращаем внимания на λ и, обнаружив $\Delta_4 >> 0$, вводим в базис A_4 вместо A_7 (переходим к следующему опорному плану):

C баз	Базис 2	План $X_{баз}$	λ	$-\lambda$	-1	1	0	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_4	**5**	3	–3	–1	**1**	–1	**0**
0	A_6	**8**	**5**	–5	0	**0**	–1	**1**
Δ_k		5	**3 –λ**	–3+λ	0	0	–1	0

Полученный опорный план $X^1 = (0, 0, 0, 5, 0, 8)$ с $L(X^1, \lambda) = 5$ будет оптимальным, если все значения Δ_k не положительны, т. е.

$$\begin{cases} \Delta_1 \equiv 3 - \lambda \le 0 \\ \Delta_2 \equiv -3 + \lambda \le 0 \end{cases} \qquad \begin{cases} \lambda \ge 3 \\ \lambda \le 3 \end{cases}.$$

Решаем систему двух линейных неравенств и обнаруживаем, что найденный *план X^1 оптимален лишь при* $\lambda = 3$.

Исследуем оставшиеся из заданного диапазона значения λ.

Пусть $\lambda > 3$. Тогда $\Delta_2 > 0$ и вектор A_2 подлежит вводу в базис, но в силу неположительности его компонент приходим к выводу, что *при* $\lambda > 3$ *целевая функция не ограничена снизу.*

Пусть $\lambda < 3$. Тогда $\Delta_1 > 0$ и в базис вводится вектор A_1:

C баз	Базис 3	План $X_{баз}$	λ	$-\lambda$	-1	1	0	0
			A_1	A_2	A_3	A_4	A_5	A_6
1	A_4	**1/5**	**0**	0	–1	**1**	–2/5	–3/5
λ	A_1	**8/5**	**1**	–1	0	0	–1/5	1/5
Δ_k		(8λ+1)/5	0	0	0	0	**– (λ+2)/5**	(λ–3)/5

Полученный опорный план является оптимальным, если все значения Δ_k неположительны, т. е.

$$\Delta_5 = -(\lambda + 2) / 5 \le 0 \mid \lambda \ge -2;$$
$$\Delta_6 = (\lambda - 3) / 5 \le 0 \mid \lambda \le 3.$$

Очевидно, что найденный *план* $X = (8/5, 0, 0, 1/5)$ с $L(X, \lambda) = (8\lambda + 1) / 5$ *оптимален при* $-2 \le \lambda \le 3$.

Пусть $\lambda < -2$. Тогда $\Delta_5 > 0$ и вектор A_5 подлежит вводу в базис, в силу неположительности его компонент приходим к выводу, что *при* $\lambda < -2$ *целевая функция не ограничена снизу*.

Таким образом, мы получили решение задачи:

$$L_{\min}(X, \lambda) = \begin{cases} \to -\infty, & \lambda < -2; \\ (8\lambda+1)/5, & -2 \le \lambda \le 3; \quad X_{\text{opt}} = (8/5,\ 0,\ 0,1/5) \\ 5, & \lambda = 3; \quad X_{\text{opt}} = (0,\ 0,\ 0,\ 5) \\ \to -\infty, & \lambda > 3. \end{cases}.$$

Пример 2. Рассмотрим задачу максимизации

$$L(X, \lambda) = X_1 - X_2 - 2\,X_3$$

при условиях

$$X_1 + X_2 + X_3 \le 3 + \lambda;$$
$$2\,X_1 - X_2 + X_3 \le 5 - \lambda;$$
$$X_k \ge 0,\ k = 1\ ..\ 3;\ -\infty < \lambda < \infty.$$

Поставим двойственную к ней задачу, имеющую вид минимизировать

$$L(Y, \lambda) = (3 + \lambda)\,Y_1 + (5 - \lambda)\,Y_2$$

при условиях

$$Y_1 + 2\,Y_2 \ge 1;$$
$$Y_1 - Y_2 \ge -1;$$
$$Y_1 + Y_2 \ge -2;$$
$$Y_1,\ Y_2 \ge 0;\ -\infty < \lambda < \infty.$$

Приводим двойственную задачу к канонической форме (умножив предварительно второе и третье неравенства на –1) и начинаем решение обычным симплексным методом. Заметьте, что указанное умножение тождественно смене знака у переменных x_2 и x_3 исходной задачи.

C баз	Базис 1	План $Y_{баз}$	3+λ A_1	5–λ A_2	0 A_3	0 A_4	0 A_5	M A_6
M	A_6	**1**	**1**	2	–1	0	0	1
0	A_4	**1**	–1	1	0	1	0	0
0	A_5	**2**	–1	–1	0	0	1	0
Δ_k		M	***M*–3–λ**	2*M*–5+λ	–*M*	0	0	0

C баз	Базис 2	План $Y_{баз}$	3+λ A_1	5–λ A_2	0 A_3	0 A_4	0 A_5	M A_6
3+λ	A_1	**1**	1	2	–1	0	0	1
0	A_4	**2**	0	3	–1	1	0	1
0	A_5	**3**	0	1	–1	0	1	1
Z_k		3+λ	3+λ	6+2λ	– (3+λ)	**0**	**0**	**3+λ**
Δ_k			0	1+3λ	– (3+λ)	0	0	–*M*+..

Найденный план $Y = (1, 0)$ оптимален, если $\Delta_2 = (1 + 3\,\lambda) \le 0$ и $\Delta_3 = -(3 + \lambda) \le 0$, т. е. *при* $-3 \le \lambda \le -1/3$ $Y_{opt} = (1, 0)$. В строке Z_k (в позициях 6, 4 и 5 в соответствии с начальным базисом) имеем решение прямой задачи: $X_{opt} = (3 + \lambda, 0, 0)$, $L(X_{opt}) = 3 + \lambda$.

Пусть $\lambda < -3$. Попытка ввода в базис вектора A_3 позволяет обнаружить, что в этом случае *целевая функция решаемой* (*двойственной*) *задачи не ограничена снизу и, следовательно, ограничения исходной задачи противоречивы*.

В случае $\lambda > -1/3$ имеем:

C баз	Базис 3	План $Y_{баз}$	3+λ A_1	5–λ A_2	0 A_3	0 A_4	0 A_5	M A_6
5–λ	A_2	**1/2**	1/2	1	–1/2	0	0	1/2
0	A_4	**1/2**	–3/2	0	1/2	1	0	–1/2
0	A_5	**5/2**	–1/2	0	–1/2	0	1	1/2
Z_k		(5–λ)/2	(5–λ)/2	5–λ	– (5–λ)/2	**0**	**0**	**(5–λ)/2**
Δ_k			– (3λ+1)/2	0	**– (5–λ)/2**	0	0	–M+...

Решив систему неравенств $\Delta_1 = -(3\lambda + 1) / 2 \le 0$, $\Delta_3 = -(5 - \lambda) / 2 \le 0$, обнаруживаем *при* $-1/3 \le \lambda \le 5$ $Y_{opt} = (0, 1/2)$, $X_{opt} = ((5 - \lambda) / 2, 0, 0)$, $L(X_{opt}) = (5 - \lambda) / 2$.

Продолжаем решение задачи при $\lambda > 5$. Получаем:

C баз	Базис 4	План $Y_{баз}$	3+λ A_1	5–λ A_2	0 A_3	0 A_4	0 A_5	M A_6
5–λ	A_2	**1**	–1	1	0	1	0	0
0	A_3	**1**	–3	0	1	2	0	–1
0	A_5	**3**	–2	0	0	1	1	0
Z_k		5–λ	– (5–λ)	5–λ	0	**5–λ**	**0**	**0**
Δ_k			–8	0	0	5–λ	0	–M

Видим, что *при* $\lambda \ge 5$ $Y_{opt} = (0, 1)$, $X_{opt} = (0, -5 + \lambda, 0)$, $L(X_{opt}) = 5 - \lambda$.

Интервал значений параметра λ исчерпан, выявлены четыре интервала устойчивости (рис. 8) оптимальных решений задачи.

Диапазон λ	Сопряженная задача	Исходная задача
$\lambda < -3$	$L(Y, \lambda) \to -\infty$	*ограничения противоречивы*
$-3 \le \lambda \le -1/3$	$Y_{opt} = (1, 0)$	$X_{opt} = (3 + \lambda, 0, 0)$, $L(X_{opt}) = 3 + \lambda$
$-1/3 \le \lambda \le 5$	$Y_{opt} = (0, 1/2)$	$X_{opt} = ((5 - \lambda) / 2, 0, 0)$, $L(X_{opt}) = (5 - \lambda) / 2$
$\lambda \ge 5$	$Y_{opt} = (0, 1)$	$X_{opt} = (0, -5 + \lambda, 0)$, $L(X_{opt}) = 5 - \lambda$

Увы, в случае *зависимости от параметра компонент матрицы ограничений* столь простого *универсального* подхода к решению не существует.

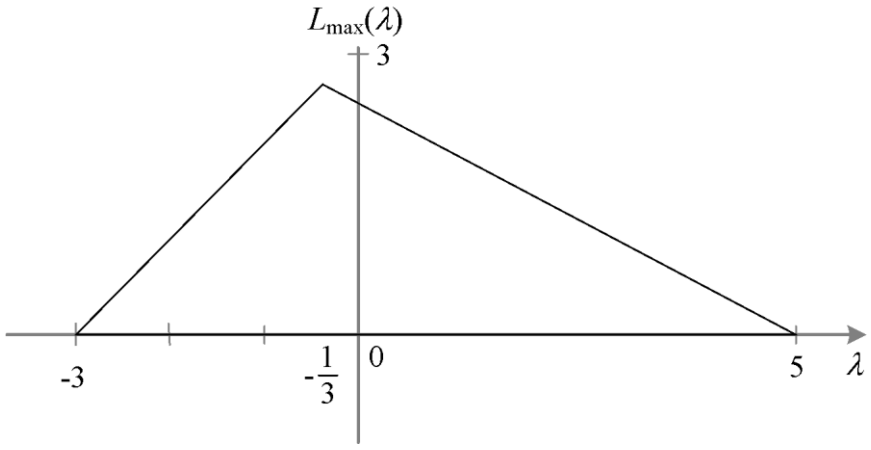


Рис. 8

Использованный здесь подход применим и в случае нелинейной зависимости от параметра (например, квадратичной), но здесь возникают системы нелинейных неравенств, решение которых часто представляет достаточно сложную, нестандартную задачу.

Иногда рассмотренные приемы проверки на оптимальность могут оказаться небесполезными в ситуации нескольких параметров, но решение систем линейных неравенств относительно нескольких переменных – не из тех задач, которые можно доверить даже самой умной машине.

3. ЦЕЛОЧИСЛЕННОЕ ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

3.1. Постановка задачи

Задача линейного целочисленного программирования отличается от общего случая линейной программы одним словом:
найти такое решение (*план*)

$$X = (x_1, x_2, \ldots, x_n),$$

при котором линейная функция

$$L(X) = \sum_{j=1}^{n} c_j x_j$$

принимает максимальное (*минимальное*) *значение при ограничениях*

$$\sum_{j=1}^{n} a_{ij} x_j = b_i,\ i = 1, 2, ..., m;$$

$$x_j \geq 0,\ j = 1, 2, ... n;$$

$$x_j - \textit{целые числа},\ j = 1, 2, ..., n.$$

Заметим, что последнее условие может накладываться не на все неизвестные, и тогда говорят о *частично целочисленных* задачах.

Круг подобных задач достаточно широк. При решении многих задач оптимального планирования, связанных с определением численности работников в структурных подразделениях предприятия, количества станков в заводском цехе, объемом производства или транспортировки дорогой крупногабаритной, неделимой продукции (подводных лодок, вертолетов, угольных комбайнов) и т. д., искомые величины могут быть только целочисленными. Пытаться округлять получаемые оценки в большую или меньшую сторону неразумно: или мы превысим свои ресурсные возможности, или не используем их в должной мере (излишки иногда можно употребить с пользой на что-нибудь еще).

Иногда встречаются задачи, где некая величина x_j может принимать лишь какие-то дискретные значения, например, +1, 0 или –1. Если провести замену $x_j = 1 \cdot z_{j1} + 0 \cdot z_{j2} - 1 \cdot z_{j3}$ и наложить требования $z_{j1} + z_{j2} + z_{j3} = 1$, $z_{j1}, z_{j2}, z_{j3} \geq 0$ – целые числа, то получим эквивалентную задачу целочисленного программирования. К подобным относятся, например, задачи с фиксированными размерами производимых или закупаемых партий каких-то продуктов.

Достаточно часто возникают задачи с так называемыми *булевыми переменными*, решениями которых являются суждения типа «да – нет». Если значению «да» сопоставить единицу, а «нет» – ноль, то добавлением условий $\{0 \le x_j \le 1, x_j - целое\}$ мы опять-таки получаем целочисленную программу. Линейность целевой функции и ограничений при этом сохраняется.

Встречаются задачи, где целевая функция содержит нелинейные элементы типа

$$C(x) = \begin{cases} A + B \cdot x, & \text{если } x > 0 \\ 0, & \text{если } x = 0 \end{cases}$$

(здесь B определяет непосредственные затраты на единицу производимой продукции, тогда как A – затраты на подготовку производства, отсутствующие в периоды производственного простоя).

Если ввести новую переменную $0 \le \delta \le 1$ и потребовать ее целочисленности, то заменой $C(x) = A \cdot \delta + B \cdot x$ и дополнительным условием $x \le M \cdot \delta$, где M – верхняя граница для x или попросту очень большое число, получаем целочисленную линейную программу с числом переменных на единицу большим.

Задачи целочисленного программирования часто называют задачами дискретного или диофантова[10] программирования.

3.2. Метод Гомори – метод последовательных отсечений

Сущность метода Гомори заключается в последовательном построении дополнительных ограничений, *отсекающих* найденные нецелочисленные решения (нецелочисленные оптимальные планы) задачи, но не отсекающих ни одного целочисленного плана. Мы пользуемся термином «отсекает» в смысле «делает неприемлемым», «делает недопустимым», «делает не удовлетворяющим новой системе условий».

Хотя идея такого подхода принадлежит Дж. Данцигу[11] и

---

[10] Диофант Александри́йский (III век н. э.) – древнегреческий математик. Одной из задач его труда «Арифметика» был поиск целочисленных решений уравнений.

[11] Джордж Бернард Данциг (1914 – 2005) – видный американский математик, предложил симплексный метод в современной форме и, вместе с Л. В. Канторовичем, является создателем *линейного программирования* (термин введен в 1949 г.).

Р. Гомори[12], достаточно эффективная ее реализация с гарантией конечности процесса отсечений справедливо называется методом Гомори.

Алгоритм решения задачи этим методом определяется следующими действиями.

1. Решаем поставленную задачу симплексным методом без учета условия целочисленности.

2. Если же среди компонент найденного оптимального решения есть нецелые, то к ограничениям задачи добавляем новое ограничение, обладающее следующими свойствами:

– *оно должно быть линейным*;

– *должно отсечь найденный оптимальный нецелочисленный план*;

– *не должно отсекать ни одного целочисленного плана.*

Р. Гомори

Эта идея впервые была представлена в форме дополнительного ограничения:

$$\sum_{j \notin \text{базису}} x_j \geq 1 \qquad (1)$$

(сумма небазисных компонент оптимального плана должна быть отлична от нуля, т. е. хотя бы одна из небазисных компонент должна быть ненулевой). В самом деле, оптимальный план с нулевыми значениями небазисных компонент этому условию не удовлетворяет, что подтверждает разумность отсечения этого плана от исходного множества.

К сожалению, для абсолютного большинства задач, где значения целочисленных переменных меняются в большом диапазоне, скорость сходимости процесса таких отсечений мала.

---

[12] Ральф Гомори (род. 1929). Окончил Кембридж, в 1954 г. получил докторскую степень по математике в Принстонском университете и с 1957 г. занялся там математическими методами исследования операций и создал первые алгоритмы отсечения планов (*cutting-plane algorithms*). В 1959 – 2007 гг. работал в IBM, возглавляя исследовательские работы.

Позднее Р. Гомори предлагает другой вариант – выбирать компоненту оптимального плана *с наибольшей дробной частью* и по соответствующей этой компоненте k-й строке симплексной таблицы (k-му уравнению)

$$B_k = \sum_{j=1}^{n} A_{kj} x_j$$

строить ограничение вида

$$f_k = \sum_{j \notin \text{базису}} f_{kj} x_j - S^*, \; S^* \geq 0, \qquad (2)$$

где $f_k = B_j - [B_j]$; $f_{kj} = A_{kj} - [A_{kj}]$; S^* – новая переменная; $[B_j]$, $[A_{kj}]$ – целые части соответствующих величин.

Обратите внимание на то, что [2,57] = 2, но [– 2,57] = – 3, и соответственно значения f_{kj} равны 0,57 и 0,43.

Так, если выбранное уравнение имеет вид

$$3{,}14 = -2{,}57 \cdot x_1 + x_2 + 3{,}12 \cdot x_3 - 4{,}12 \cdot x_4,$$

то получаем новое (дополнительное) ограничение

$$0{,}14 = 0{,}43 \cdot x_1 + 0{,}12 \cdot x_3 + 0{,}88 \cdot x_4 - S, \; S \geq 0,$$

которое присоединяем к имеющимся в симплексной таблице, тем самым получая расширенную задачу.

Поскольку для продолжения ее решения опять необходимо выбрать начальный опорный план, то для включения в число базисных рекомендуем ту переменную, для которой величина $\left|\Delta_j / f_{kj}\right|$ минимальна. Если для этой переменной величина θ (см. симплексный метод) достигается по дополнительной строке, то эту строку (уравнение) разрешаем относительно выбранной переменной и исключаем эту переменную из остальных уравнений (получен опорный план для расширенной задачи).

Если же величина θ не соответствует дополнительной строке, то приходится вводить искусственную переменную и надеяться на то, что в дальнейшем эта переменная покинет базис (этого не случится, если новое множество планов окажется пустым – новая система ограничений противоречива).

Упомянутые действия повторяются до получения целочисленного решения или установления неразрешимости задачи.

Существует доказательство конечности этого процесса, но оценить заранее число таких шагов невозможно. Все зависит от ко-

личества нецелочисленных планов задачи и ее размерности (хотя многие задачи даже малой размерности решаются весьма долго).

Необходимо сделать некоторые замечания.

1) Если для дробного x_j обнаружится целочисленность всех коэффициентов соответствующего уравнения, то задача не имеет целочисленного решения.

2) Если дополнительная переменная S^* вошла в число базисных (появились более жесткие условия, ранее введенное становится лишним), то соответствующие ей строку и столбец можно удалить. Соблюдение этого правила сохраняет размерность решаемой задачи в разумных пределах – число уравнений не превысит $m + n$.

Рассмотрим высказанные соображения на несложном примере (здесь мы ограничились двумя переменными, чтобы иметь возможность геометрической интерпретации).

Пример. Пусть для приобретения нового оборудования предприятие выделяет 19 денежных единиц. Оборудование должно быть размещено на площади, не превышающей 16 м^2. Предприятие может заказать оборудование двух видов: машины типа «А» стоимостью 2 денежных единицы, требующие производственную площадь 4 м^2 и обеспечивающие производительность за смену 8 т продукции, и машины типа «В» стоимостью 5 денежных единиц, занимающие площадь 1 м^2 и обеспечивающие производительность за смену 6 т продукции. Требуется составить оптимальный план приобретения оборудования, обеспечивающий максимальную общую производительность.

Обозначив через x_1, x_2 количество приобретаемых машин соответственно типа «А» и «В» и через L – их общую производительность, получаем математическую модель задачи:

$$\max L = 8\,x_1 + 6\,x_2$$

при ограничениях:

$$2\,x_1 + 5\,x_2 \le 19;$$
$$4\,x_1 + x_2 \le 16;$$
$$x_1 \ge 0,\ x_2 \ge 0;$$
$$x_1, x_2 - \textit{целые числа}.$$

Решаем задачу симплексным методом без учета целочисленности (предполагаем, что читатель освоился с технологией этого метода).

C_j	Б$_0$	X_1	8	6	0	0
			x_1	x_2	x_3	x_4
0	x_3	19	2	5	1	0
0	x_4	16	4	1	0	1
z_j		0	0	0	0	0
Δ_j			–8	–6	0	0

C_j	Б$_1$	X_1	8	6	0	0
			x_1	x_2	x_3	x_4
0	x_3	11	0	9/2	1	–1/2
8	x_1	4	1	1/4	0	1/4
z_j		32	8	2	0	2
Δ_j			0	–4	0	2

C_j	Б$_1$	X_1	8	6	0	0
			x_1	x_2	x_3	x_4
6	x_2	22/9	0	1	2/9	–1/9
8	x_1	61/18	1	0	–1/18	5/18
Δ_j		376/9	0	0	8/9	14/9

Получен оптимальный нецелочисленный план X_{opt} = (61/18; 22/9), $L_{\max}$ = 376/9.

Максимальной дробной частью обладает компонента плана x_2, относительно которой разрешено первое уравнение (4/9 > 7/18), дополнительное ограничение строим по первой строке, получая *первое ограничение Гомори*

$$4/9 = 2/9\, x_3 + 8/9\, x_4 - S_1,\ S_1 \geq 0$$

(подставьте найденный оптимальный план в это условие и убедитесь в его «отсечении»). Составленное уравнение дописываем к уже имеющимся в симплексной таблице:

C_j	Б$_1$	X_1	8	6	0	0	0
			x_1	x_2	x_3	x_4	S_1
6	x_2	22/9	0	1	2/9	–1/9	0
8	x_1	61/18	1	0	–1/18	5/18	0
		4/9	0	0	2/9	8/9	–1
Δ_j		376/9	0	0	8/9	14/9	

получая новую задачу линейного программирования, в которой 3 ограничения на 5 неотрицательных переменных. Для получения опорного плана этой задачи необходимо выбрать третью базисную переменную. Для этого определяем $\min \dfrac{\Delta_j}{f_{kj}} = \min\left(\dfrac{8/9}{2/9}; \dfrac{14/9}{8/9}\right) = 7/4$ и предлагаем для ввода в базис переменную x_4.

Отыскав величину $\theta = \min\left(-; \dfrac{61/18}{5/18}; \dfrac{4/9}{8/9}\right)$ и обнаружив ее соответствие дополнительному (третьему) уравнению, не прибегая к

искусственной переменной и лишнему преобразованию, выражаем x_4 из этого уравнения и получаем опорный план расширенной задачи. Найденный план оптимален, но нецелочисленен. Строим *второе ограничение Гомори* по первой строке (можно и по третьей):

$$1/2 = 1/4\ x_3 + 7/8\ S_1 - S_2,\ S_2 \geq 0.$$

C_j	Б$_2$	X_2	8	6	0	0	0	0
			x_1	x_2	x_3	x_4	S_1	S_2
6	x_2	5/2	0	1	1/4	0	–1/8	0
8	x_1	13/4	1	0	–1/8	0	5/16	0
0	x_4	1/2	0	0	1/4	1	–9/8	0
Δ_j		41	0	0	1/2	0	7/4	·
		1/2	0	0	1/4	0	7/8	–1

Определяя переменную, вводимую в базис, из совпадения $\min\left(\frac{1/2}{1/4};\frac{7/4}{7/8}\right)=2$ видим свободу выбора между x_3 и S_1. Выбрав S_1 и обнаружив соответствие $\theta=\min\left(-;\frac{13/4}{5/16};-;\frac{1/2}{7/8}\right)=4/7$ дополнительному ограничению, получаем

C_j	Б$_3$	X_3	8	6	0	0	0	0	0
			x_1	x_2	x_3	x_4	S_1	S_2	S_3
6	x_2	18/7	0	1	2/7	0	**0**	–1/7	0
8	x_1	43/14	1	0	–3/14	0	**0**	5/14	0
0	x_4	8/7	0	0	4/7	1	**0**	–9/7	0
0	$\boldsymbol{S_1}$	**4/7**	**0**	**0**	**0**	**0**	**1**	**–8/7**	**0**
Δ_j		40	0	0	0	0	0	2	·
		4/7	0	0	2/7	0	0	6/7	–1

Получаем новый оптимальный нецелочисленный план. Учитывая сделанное выше замечание 2, вычеркиваем строку и столбец, соответствующие переменной S_1.

В полученном плане максимальную дробную часть имеет компонента x_2, поэтому записываем по первой строке *третье ограничение Гомори*:

$$4/7 = 2/7\ x_3 + 6/7\ S_2 - S_3,\ S_3 \geq 0.$$

Отыскав $\min\left(\frac{0}{2/7};\frac{2}{6/7}\right)=0$, вводим в базис переменную x_3. Минимальное значение $\theta = 2$, что соответствует дополнительной строке, и после очередных симплексных преобразований получаем:

C_j	Б$_4$	X_4	8	6	0	0	0	0	0
			x_1	x_2	x_3	x_4	S_2	S_3	S_4
6	x_2	2	0	1	0	0	–1	1	0
8	x_1	7/2	1	0	0	0	1	–3/4	0
0	x_4	0	0	0	0	1	–3	2	0
0	x_3	2	0	0	1	0	3	–7/2	0
Δ_j		40	0	0	0	0	2	0	·
		1/2	0	0	0	0	0	1/4	–1

Обнаружив оптимальность, но нецелочисленность найденного плана, строим *четвертое ограничение Гомори*

$$1/2 = 1/4\ S_3 - S_4,\ S_4 \geq 0.$$

Намереваясь ввести в базис S_3, обнаруживаем, что $\theta = 0$ соответствует не дополнительному уравнению, а третьему. Как это не прискорбно, приходится прибегнуть к искусственной переменной (обозначаем ее x_5):

C_j	Б$_5$	X_5	8	6	0	0	0	0	0	$-M$
			x_1	x_2	x_3	x_4	S_2	S_3	S_4	x_5
6	x_2	2	0	1	0	0	–1	1	0	0
8	x_1	7/2	1	0	0	0	1	–3/4	0	0
0	x_4	0	0	0	0	1	–3	**2**	0	0
0	x_3	2	0	0	1	0	3	–7/2	0	0
$-M$	x_5	1/2	0	0	0	0	0	1/4	–1	1
Δ_j		40–M/2	0	0	0	0	2	–M/4	M	0

6	x_2	2	0	1	o	–1/2	1/2	**0**	0	0
8	x_1	7/2	1	0	0	3/8	–1/8	**0**	0	0
0	$\boldsymbol{S_3}$	**0**	**0**	**0**	**0**	**1/2**	**–3/2**	**1**	**0**	**0**
0	x_3	2	0	0	1	7/4	–9/4	**0**	0	0
$-M$	x_5	1/2	0	0	0	–1/8	**3/8**	**0**	–1	1
Δ_j		40–M/2	0	0	0	M/8	2–3M/8	**0**	M	0

C_j	Б$_6$	X_6	8	6	0	0	0	0	0
			x_1	x_2	x_3	x_4	$\boldsymbol{S_2}$	S_4	S_5
6	x_2	4/3	0	1	o	–1/3	**0**	4/3	0
8	x_1	11/3	1	0	0	1/3	**0**	–1/3	0
0	x_3	5	0	0	1	1	**0**	–6	0
0	$\boldsymbol{S_2}$	**4/3**	**0**	**0**	**0**	**–1/3**	**1**	**–8/3**	**0**
Δ_j		112/3	0	0	0	2/3	0	16/3	·
		2/3	0	0	0	1/3	0	2/3	–1

И опять получили нецелочисленный оптимальный план. На основе второго уравнения строим *пятое ограничение Гомори*

$$2/3 = 1/3\ x_4 + 2/3\ S_4 - S_5,\ S_5 \geq 0.$$

Выбрав для ввода в базис переменную x_4 и обнаружив соответствие $\theta = 2$ новому уравнению, имеем

C_j	Б7	X_7	8	6	0	0	0	0
			x_1	x_2	x_3	x_4	S_4	S_5
6	x_2	2	0	1	0	0	2	–1
8	x_1	3	1	0	0	0	–1	1
0	x_3	3	0	0	1	0	–8	3
0	x_4	2	0	0	0	1	2	–3
Δ_j		36	0	0	0	0	4	2

Оптимальный целочисленный план $X_7 = (\mathbf{3}, \mathbf{2}, 3, 2)$; $L_{\max} = 36$.

В чем же результат столь длительного решения?

Экономическая интерпретация: согласно полученному решению, предприятию необходимо закупить 3 машины типа «А» и 2 машины типа «В». При этом будет достигнута максимальная производительность работы оборудования, равная 36 т продукции за смену. Полученную экономию денежных средств в размере $x_3 = 3$ денежным единицам можно будет направить на какие-либо иные цели, например на премирование рабочих, которые будут заниматься отладкой полученного оборудования. На лишнюю площадь $x_4 = 2$ м^2 можно поставить ящик с цветами, если рабочие не страдают аллергией на них.

Геометрическую интерпретацию процедуры выполненных отсечений начинаем с построения множества планов (рис. 9). Точка 1 определяет оптимальный нецелочисленный план.

Из условий задачи имеем:

$$x_3 = 19 - 2\ x_1 - 5\ x_2;\ x_4 = 16 - 4\ x_1 - x_2.$$

Подставляем эти выражения в первое ограничение Гомори

$$4/9 = 2/9\ x_3 + 8/9\ x_4 - S_1,\ S_1 \geq 0$$

и после преобразований получаем $S_1 = 18 - 4\ x_1 - 2\ x_2 \geq 0$.

Полученное ограничение $\mathbf{4\ x_1 + 2\ x_2 \leq 18}$ отсекает от множества планов область, содержащую точку 1. Новый оптимальный нецелочисленный план – точка 2.

Подставляя выражения x_3 и S_1 во второе ограничение Гомори

$$1/2 = 1/4\ x_3 + 7/8\ S_1 - S_2,\ S_2 \geq 0,$$

получаем $S_2 = 20 - 4\,x_1 - 3\,x_2 \geq 0$ или $\mathbf{4\,x_1 + 3\,x_2 \leq 20}$. Это ограничение отсекает от множества планов область, содержащую точку 2. Новый оптимальный нецелочисленный план – точка 3.

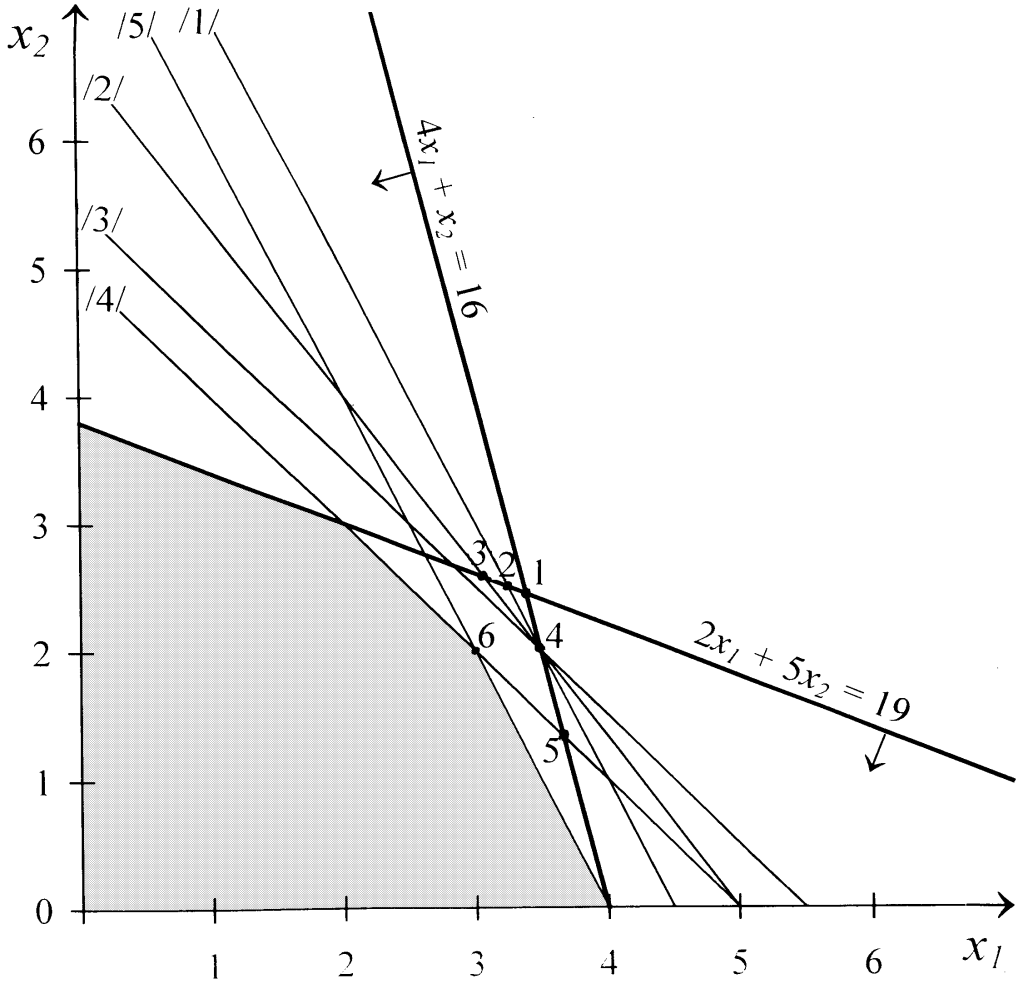


Рис. 9

Аналогичные подстановки в третье ограничение Гомори

$$4/7 = 2/7\,x_3 + 6/7\,S_2 - S_3,\ S_3 \geq 0$$

дают $S_3 = 22 - 4\,x_1 - 4\,x_2 \geq 0$. На рисунке видим, что условие $\mathbf{4\,x_1 + 4\,x_2 \leq 22}$ отсекает от множества планов область, содержащую точку 3. Новый оптимальный нецелочисленный план – точка 4.

Подстановки в четвертое ограничение Гомори $1/2 = 1/4\,S_3 - S_4$, $S_4 \geq 0$ дают $S_4 = 5 - x_1 - x_2 \geq 0$, и мы наблюдаем, как условие $\mathbf{x_1 + x_2 \leq 5}$ делает недопустимой точку 4. Новый оптимальный нецелочисленный план – точка 5.

Наконец, пятое ограничение Гомори $2/3 = 1/3\,x_4 + 2/3\,S_4 - S_5$, $S_5 \geq 0$ преобразуется к эквивалентному виду $S_5 = 8 - 2\,x_1 - x_2 \geq 0$, и мы наблюдаем, как условие $\mathbf{2\,x_1 + x_2 \leq 8}$ отсекает от множества пла-

нов область, содержащую точку 5.

В итоге устанавливаем оптимальность и целочисленность плана – точки 6 с координатами (3, 2).

3.3. Метод ветвей и границ

Этот метод с красивым названием идеологически гораздо проще метода Гомори, хотя и использует идею отсечения оптимального нецелочисленного плана.

Получив нецелочисленный оптимальный план задачи, в котором какая-то составляющая x_k оказалась нецелочисленной ($x_k = A$), мы ставим **две задачи** на основе ограничений решенной задачи и дополнительных условий $x_k \le [A]$ и $x_k \ge [A] + 1$ соответственно. Например, если в найденном оптимальном плане $x_3 = 4{,}7$, то в новых задачах будут условия $x_3 \le 4$ и $x_3 \ge 5$ (недопустимость найденного плана в будущем очевидна). Каждую из полученных задач решаем до получения оптимального плана и нового разветвления.

В общем случае процесс ветвления продолжается до обнаружения противоречивости ограничений или получения целочисленных решений, из которых остается выбрать наилучшее.

Очевидно, что компьютерная реализация этого метода едва ли доставит удовольствие программисту.

Для иллюстрации метода возьмем пример из предыдущего раздела:

$$\max L = 8\,x_1 + 6\,x_2$$

при ограничениях:

$$2\,x_1 + 5\,x_2 \le 19;$$
$$4\,x_1 + x_2 \le 16;$$
$$x_1 \ge 0,\ x_2 \ge 0 - \textit{целые числа}.$$

Решая эту задачу симплексным методом, мы получаем оптимальный план $X_{\text{opt}} = (61/18;\ 22/9)$, $L_{\max} = 376/9 = 41{,}78$.

Если вам не по душе нецелочисленность x_1, ставим две задачи:

A1	**A2**
$\max L = 8\,x_1 + 6\,x_2$	$\max L = 8\,x_1 + 6\,x_2$
при ограничениях:	при ограничениях:
$2\,x_1 + 5\,x_2 \le 19;$	$2\,x_1 + 5\,x_2 \le 19;$
$4\,x_1 + x_2 \le 16;$	$4\,x_1 + x_2 \le 16;$
$x_1 \le 3;$	$x_1 \ge 4;$
$x_1 \ge 0,\ x_2 \ge 0 - \textit{целые числа}.$	$x_1 \ge 0,\ x_2 \ge 0 - \textit{целые числа}.$

Для задачи **A2** обнаруживается целочисленный оптимальный (повезло!) план (4, 0), для которого $L_{max} = 32$.

Задача **A1** решается до получения оптимального плана (3, 13/5), и появляются две ветви:

A11	**A12**
max $L = 8\,x_1 + 6\,x_2$	max $L = 8\,x_1 + 6\,x_2$
при ограничениях:	при ограничениях:
$2\,x_1 + 5x_2 \le 19;$	$2\,x_1 + 5x_2 \le 19;$
$4\,x_1 + x_2 \le 16;$	$4\,x_1 + x_2 \le 16;$
$x_1 \le 3;$	$x_1 \le 3;$
$x_2 \le 2;$	$x_2 \ge 3;$
$x_1 \ge 0, x_2 \ge 0$ – *целые числа*.	$x_1 \ge 0, x_2 \ge 0$ – *целые числа*.

На приведенном рисунке (рис. 10) точка 1 определяет решение исходной задачи, точки 3 и 2 соответственно – решение задач **A1** и **A2** ($L_{max} = 32$), точки 4 и 5 – решение задач **A11** ($L_{max} = 36$) и **A12** ($L_{max} = 34$). Сравнивая конечные решения по ветвям, видим оптимальный план (3, 2).

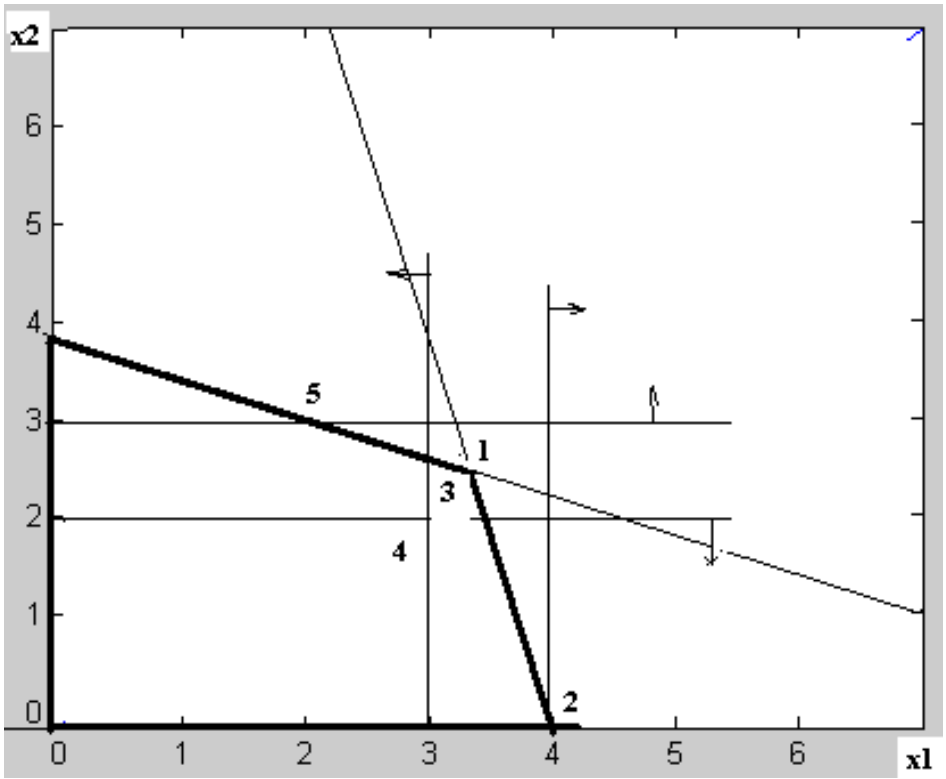


Рис. 10

В задачах небольшой размерности или при малом числе вариантов значений переменных метод вполне эффективен и часто используется в приложениях. Существуют многочисленные его про-

граммные реализации, но не всегда реклама соответствует факту – решение сколько-нибудь серьезной задачи завершается, как правило, сообщением о том, что после *N*-го количества шагов результат не достигнут.

Заметим, что существует много задач, где требование целочисленности накладывается не на все переменные. В этом случае ограничение Гомори чуть-чуть модифицируется

$$f_k = \sum_{j \not\in \text{базису}} f_{kj} x_j - S^*, S^* \geq 0, \qquad (3)$$

где

$$f_{kj} = \begin{cases} A_{kj},\ j \notin Y,\ A_{kj} \geq 0 \\ \dfrac{-f_k}{(1-f_k)A_{kj}},\ j \notin Y,\ A_{kj} < 0 \\ f_{kj},\ j \in Y,\ f_{kj} \leq f_k \\ \dfrac{f_k}{(1-f_k)(1-f_{kj})},\ j \in Y,\ f_{kj} > f_k \end{cases}$$

(здесь A_{kj} – коэффициенты выбранного уравнения, Y – множество целочисленных переменных). Модификация метода ветвей и границ не нуждается в комментариях.

4. ЗАДАЧИ ТРАНСПОРТНОГО ТИПА

Сам термин «транспортная задача» уже говорит о необходимости решения проблем, возникающих при организации перемещения тех или иных продуктов с помощью транспортных средств (транспортировка угля из угольных центров Кузбасса по существующей сети железных дорог к потребителям, доставка ночной выпечки от хлебокомбината до булочных города, перекачка нефти, газа или воды по трубопроводам от источников к потребителям и др.).

Решение подобных задач может преследовать самые разные цели. Это может быть желание проехать *кратчайшим путем* на автомашине от «солнечного Магадана» до «Гренадской волости в Испании» или в *кратчайший срок* доставить отопительные батареи из Новокузнецка в Приморье, куда «неожиданно» в декабре пришла зима. Это может быть поставка грузов с *минимальными затратами* на эту процедуру или организация полетов бомбардировщиков от аэродромов для *наилучшего подавления* каких-либо целей. Это может быть выбор для назначения на вакантные должности среди множества претендентов с тем, чтобы достичь *наибольшего общего эффекта* таких назначений (разумеется, мы сможем решить эту задачу лишь при наличии независимой экспертизы соответствия каждого претендента каждой из должностей).

В таких условиях возникает естественный вопрос: **«от кого, кому и сколько?»**, дополненный естественным желанием, чтобы найденный план отвечал одному из двух критериев – минимуму денежных затрат или возможности реализации в кратчайшие сроки, что плохо сочетается (редко сочетается «дешево и быстро», ибо за скорость надо доплачивать, а при плохой организации получается подчас «и медленно, и дорого»).

Как мы увидим ниже, некоторые задачи, например поиска кратчайшего или самого длинного пути между некоторыми пунктами сети хороших автодорог по какому-либо критерию, решаются достаточно просто методом динамического программирования.

Несмотря на внешнюю простоту постановки многих задач транспортировки привычных грузов, их решение часто исключительно сложно. Так значительные сложности возникают при наличии каких-то промежуточных пунктов в транспортной сети или ограничений на «объемы» поставок, но, тем не менее, для подобных задач существуют эффективные методы решения.

Несоизмеримо сложнее решать так называемые многопродуктовые задачи, где в одной автомашине везут ящики с пивом, мешки с мукой и садовый инструмент, причем для разных получателей.

4.1. Классическая транспортная задача

4.1.1. Постановка задачи и свойства решений

Пусть имеется m поставщиков некоторого продукта в количествах A_i (i = 1 .. m) и n его потребителей (рис. 11) в количествах B_j (j = 1 .. n). Пусть при этом соблюдается условие баланса

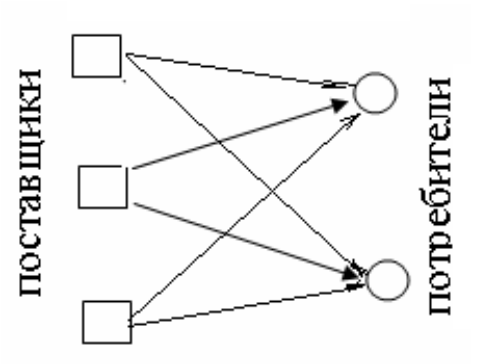


Рис. 11

$$\sum_{i=1}^{m} A_i = \sum_{j=1}^{n} B_j \ . \qquad (1)$$

К тому же поставки происходят напрямую без промежуточных пунктов (как показано на рисунке).

Известны стоимость C_{ij} перевозки единицы продукта от i-го поставщика к j-му потребителю. Если обозначить через X_{ij} соответствующие объемы перевозок, естественно попытаться найти план перевозок, минимизирующий суммарную их стоимость

$$L(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij} \qquad (2)$$

при условиях

$$\sum_{j=1}^{n} X_{ij} = A_i \quad (i = 1 .. m); \qquad (3)$$

$$\sum_{i=1}^{m} X_{i\,j} = B_j \quad (j = 1 .. n); \qquad (4)$$

$$X_{ij} \geq 0 \quad (i = 1 .. m;\ j = 1 .. n). \qquad (5)$$

Полученная задача представляет частный случай общей линейной программы с $m + n$ ограничениями на $m \times n$ неотрицательных переменных. Однако даже при m = 4 и n = 6 едва ли вас соблазнит перспектива решения задачи ЛП с 10 ограничениями и 24 переменными прямым алгоритмом симплексного метода.

Убедимся, что эта задача не из тех, решение которых требует гениальности, но едва ли элементарна для непосвященного.

Отметим некоторые «приятные» особенности этой задачи, выделяющие ее среди произвольных линейных программ.

1. *Ограничения транспортной задачи непротиворечивы*.

В подтверждение этому достаточно взять набор значений

$$X_{ij} = \frac{A_i B_j}{\sum_{i=1}^{m} A_i} \quad (i = 1 .. m; j = 1 .. n) \qquad (6)$$

и подстановкой в (3) – (5) с учетом (1) убедиться, что он является планом (удовлетворяет ограничениям).

2. *Целевая функция транспортной задачи ограничена как сверху, так и снизу*, что является следствием ограниченности множества планов

$$0 \le X_{ij} \le \min(A, B_j).$$

Из указанных истин следует, что *транспортная задача имеет оптимальный план*.

3. Поскольку ограничения (3) – (4) с учетом (1) линейно зависимы (достаточно по отдельности сложить равенства (3) и (4) и следствием получить (1)), то одним из них можно пренебречь. Следовательно, *опорный план транспортной задачи имеет не более* $m + n - 1$ *положительных компонент* (если их число равно $m + n - 1$, то опорный план называется *невырожденным*).

4. Приятной *особенностью опорных планов транспортной задачи является их целочисленность при целых значениях* A_i *и* B_j (на всех этапах симплексного преобразования коэффициенты при неизвестных равны 1, –1 или 0).

5. Если условие баланса (1) нарушено, то можно ввести фиктивного (воображаемого) $(m + 1)$-го поставщика или фиктивного $(n + 1)$-го потребителя с объемом поставки (потребления) соответственно $\sum_{j=1}^{n} B_j - \sum_{i=1}^{m} A_i > 0$ или $\sum_{i=1}^{m} A_i - \sum_{j=1}^{n} B_j > 0$ с нулевыми стоимостями перевозок.

Если обозначить через U_i $(i = 1 .. m)$ и V_j $(j = 1 .. n)$ двойственные переменные, то сопряженная задача будет состоять в максимизации

$$L(U, V) = \sum_{i=1}^{m} A_i U_i + \sum_{j=1}^{n} B_j V_j \qquad (7)$$

при условиях

$$U_i + V_j \le C_{ij}\ (i = 1\ ..\ m,\ j = 1\ ..\ n). \qquad (8)$$

В этом вы можете убедиться, если поставите задачу, сопряженную задаче минимизации ($m = 2$, $n = 3$):

$$C_{11}X_{11} + C_{12}X_{12} + C_{13}X_{13} + C_{21}X_{21} + C_{22}X_{22} + C_{23}X_{23}$$

при

$$\begin{array}{rrrrrrr|l}
X_{11} + & X_{12} + & X_{13} & & & & = A_1 & U_1 \\
 & & & X_{21} + & X_{22} + & X_{23} = & A_2 & U_2 \\
X_{11} + & & & X_{21} & & & = B_1 & V_1 \\
 & X_{12} + & & & X_{22} & & = B_2 & V_2 \\
 & & X_{13} + & & & X_{23} = & B_3 & V_3 \\
X_{11}, & X_{12}, & X_{13}, & X_{21}, & X_{22}, & X_{23} & \ge 0 &
\end{array}$$

Заметим, что многие задачи можно описать в терминах классической транспортной задачи или ее аналога с требованием максимума «затрат». В последнем случае сопряженная задача состоит в минимизации (7) при условиях $U_i + V_j \ge C_{ij}$ ($i = 1\ ..\ m,\ j = 1\ ..\ n$).

4.1.2. Выбор начального опорного плана

Идея всех методов поиска начального опорного плана заключается в установке максимального возможного объема перевозки по маршруту, т. е. минимума из предложения и спроса $X_{ij} = \min(A_i, B_j)$.

Самым простейшим из этих методов является «*метод северо-западного угла*», отправной точкой для которого является левый верхний угол таблицы.

Пусть $m = 3$, $A = (17, 33, 20)$, $n = 5$, $B = (14, 14, 14, 14, 14)$. Так как баланс здесь соблюдается, ищем начальный опорный план в табличной форме (для удобства перенесем в оцифровку таблицы значения A и B). Имеем $X_{11} = \min(17, 14) = 14$, $X_{21} = X_{31} = 0$. Продолжаем поиск с того же угла незаполненной части таблицы:

$X_{12} = \min(17 - 14, 14) = 3$, $X_{13} = X_{14} = X_{15} = 0$,

$X_{22} = \min(33, 14 - 3) = 11$, $X_{32} = 0$,

$X_{23} = \min(33 - 11, 14) = 14$, $X_{33} = 0$,

$X_{24} = \min(33 - 11 - 14, 14) = 8$, $X_{25} = 0$, $X_{34} = 14 - 8 = 6$,

$X_{35} = 20 - 6 = 14$.

$X^0 =$

14	3	∘	∘	∘	17
∘	11	14	8	∘	33
∘	∘	∘	6	14	20
14	14	14	14	14	$B \setminus A$

Можно придумать и другие модификации этого метода. В частности, *метод минимального элемента матрицы стоимостей* определяется правилом первоочередного выбора перевозки на самом «дешевом» маршруте.

Пусть значения A и B те же, что и в рассмотренном примере, а стоимости перевозок определяются матрицей C. Так как минимальная стоимость соответствует C_{22}, начинаем поиск компонент опорного плана с элемента $X_{22} = \min(33, 14) = 14$, $X_{12} = X_{32} = 0$, затем

$X_{13} = \min(17, 14) = 14$, $X_{23} = X_3 = 0$,

$X_{34} = \min(20, 14) = 14$, $X_{14} = X_{24} = 0$,

$X_{35} = \min(20 - 14, 14) = 6$, $X_{31} = 0$,

$X_{21} = \min(33 - 14, 14) = 14$, $X_{11} = 0$

и, наконец, $X_{15} = 17 - 14 = 3$, $X_{25} = 33 - 14 - 14 = 5$.

$$C = \begin{array}{|ccccc|} \hline 6 & 4 & 2 & 4 & 6 \\ 5 & 1 & 6 & 8 & 5 \\ 8 & 5 & 7 & 2 & 3 \\ \hline \end{array} \qquad X^0 = \begin{array}{|ccccc|c} \hline \circ & \circ & 14 & \circ & 3 & 17 \\ 14 & 14 & \circ & \circ & 5 & 33 \\ \circ & \circ & \circ & 14 & 6 & 20 \\ \hline 14 & 14 & 14 & 14 & 14 & B \setminus A \end{array}$$

Какой из найденных планов ближе к оптимальному? Ответить на этот вопрос можно сравнением суммарных стоимостей перевозок. Какой из методов поиска опорного плана лучше? Однозначного ответа нет: при машинной реализации метод северо-западного угла проще по своему алгоритму и имеет ряд преимуществ, связанных с проблемой вырожденности. При ручном решении в большинстве случаев предпочтение следует отдать методу минимального элемента матрицы.

Обратите внимание на то, что оба найденных плана являются невырожденными (содержат $m + n - 1 = 3 + 5 - 1 = 7$ положительных компонент).

Чтобы базисные нулевые компоненты вырожденного плана не терялись среди небазисных, в дальнейшем небазисные компоненты плана будем выделять «точками» вместо нулей.

При компьютерной реализации вырожденность устраняется довольно просто: увеличивают все A_i на ε (неотрицательное число, много меньшее значений A_i и B_j) и одно из значений B_j – на $m \times \varepsilon$, а по окончании решения задачи «округляют» компоненты оптимального плана.

При ручном решении при выборе начального плана достаточно следить, чтобы не возникало «замкнутой цепочки по базису» (это

требование возникает из условия независимости системы базисных векторов). Так из приведенных ниже вариантов опорного плана приемлем только четвертый.

0	**0**	1	
	1		
			1
1	**0**		

		1	
	1		**0**
	0		**1**
1			0

		1	
	1		
0			**1**
1		0	**0**

		1	
	1	0	
0			1
1	0		

Очевидно, что в случае транспортной задачи максимизации поиск начального плана идет точно также, но предпочтение отдается максимальным «стоимостям».

4.1.3. Метод Д. Данцига последовательного улучшения плана

Данный метод представляет компактную форму обычной симплексной процедуры.

Пусть найден некоторый начальный опорный план. Согласно второй теореме двойственности, этот план будет оптимальным, если для его базисных компонент $X_{ij} > 0$ условия сопряженной задачи выполняются в виде $U_i + V_j = C_{ij}$ и для остальных компонент – в виде $U_i + V_j \le C_{ij}$ (соответственно для транспортной задачи максимизации $U_i + V_j \ge C_{ij}$).

Поэтому $m + n - 1$ базисным компонентам выбранного плана сопоставляется система $m + n - 1$ уравнений $U_i + V_j = C_{ij}$ с $m + n$ неизвестными, которая решается с точностью до константы (берем, например, $U_1 = 0$). Если все $\Delta_{ij} = U_i + V_j - C_{ij} \le 0$, то выбранный план оптимален.

Если же нашлось $\Delta_{kp} > 0$, то план не оптимален и подвергается перестройке к плану с ненулевым значением соответствующей компоненты.

Полагаем $X_{kp} = \theta > 0$ и ищем так называемую минимизирующую цепочку по базису так, чтобы новый набор значений X_{ij} удовлетворял требованиям предложения и спроса. Затем выбираем максимальное допустимое θ, сохраняя неотрицательность компонент нового плана.

Пример. Для предложенных условий отыскиваем начальный опорный план по минимальной стоимости:

$C =$

1	2	3	4	8
2	6	7	9	10
1	2	5	8	2
9	3	4	4	$B\backslash A$

$X^0 =$

8	∘	∘	∘	8
∘	2	4	4	10
1	1	∘	∘	2
9	3	4	4	$B\backslash A$

Для проверки этого невырожденного плана на оптимальность строим систему 6 уравнений с 7 неизвестными

$$U_1 + V_1 = 1,\ U_2 + V_2 = 6,\ U_2 + V_3 = 7,$$
$$U_2 + V_4 = 9,\ U_3 + V_1 = 1,\ U_3 + V_2 = 2,$$

которую решаем при $U_1 = 0$, и затем находим $\Delta_{ij} = U_i + V_j - C_{ij}$.

Построение этой системы и ее решение удобнее выполнять в табличной форме (значения правых частей этой системы заносятся в таблицу, отыскиваются решения и их суммы).

$U_i + V_j =$

U|V	1	2	3	5
0	**1**	2	3	5
4	5	**6**	**7**	**9**
0	**1**	**2**	3	5

$\Delta_{ij} =$

∘	0	0	1
3	∘	∘	∘
∘	∘	–2	–3

Так как $\Delta_{21} > 0$, то план X^0 не оптимален. Перестраиваем его, полагая $X_{21} = \theta$:

$X^1 =$

8	∘	∘	∘
θ	2–θ	4	4
1–θ	1+θ	∘	∘

$\theta = 1$

$=$

8	∘	∘	∘
1	1	4	4
∘	2	∘	∘

Проверяем этот план на оптимальность:

$U_i + V_j =$

U\V	1	5	6	8
0	**1**	5	6	8
1	2	**6**	**7**	**9**
–3	–2	**2**	3	5

$\Delta_{ij} =$

∘	3	3	4
∘	∘	∘	∘
–3	∘	–2	–3

Так как $\Delta_{14} > 0$, то план X^1 не оптимален. Перестраиваем его, полагая $X_{14} = \theta$:

$X^2 =$

8–θ	∘	∘	θ
1+θ	1	4	4–θ
∘	2	∘	∘

$\theta = 4$

$=$

4	∘	∘	4
5	1	4	∘
∘	2	∘	∘

Проверка плана X^2 опровергает его оптимальность из-за $\Delta_{12} > 0$.

$U_i + V_j =$

U\ V	1	5	6	4
0	**1**	5	6	**4**
1	**2**	**6**	**7**	5
–3	–2	**2**	3	1

$\Delta_{ij} =$

∘	3	3	∘
∘	∘	∘	–4
–3	∘	–2	–7

Перестраиваем его, полагая $X_{12} = \theta$:

$X^3 =$

4–θ	θ	∘	4
5+θ	1–θ	4	∘
∘	2	∘	∘

$\theta = 1$ =

3	1	∘	4
6	∘	4	∘
∘	2	∘	∘

Оптимальность X^3 опровергается фактом $\Delta_{14} > 0$.

$U_i + V_j =$

U\ V	1	2	6	4
0	**1**	**2**	6	**4**
1	**2**	3	**7**	5
0	1	**2**	6	4

$\Delta_{ij} =$

∘	∘	3	∘
∘	–3	∘	– 4
0	∘	1	–4

Перестраиваем его, полагая $X_{13} = \theta$:

$X^4 =$

3–θ	1	θ	4
6+θ	∘	4–θ	∘
∘	2	∘	∘

$\theta = 3$ =

∘	1	3	4
9	∘	1	∘
∘	2	∘	∘

Проверяем этот план на оптимальность:

$U_i + V_j =$

U\ V	–2	2	3	4
0	–2	**2**	**3**	**4**
4	**2**	6	**7**	8
0	–2	**2**	3	4

$\Delta_{ij} =$

–3	∘	∘	∘
∘	0	∘	–1
–3	∘	–2	–4

Так как теперь все $\Delta_{ij} \le 0$, то план X^4 оптимален, но существуют и другие оптимальные планы (для небазисного $X_{22} = 0$ имеем $\Delta_{22} = 0$, что дает возможность получения другого плана с тем же значением целевой функции), например,

$X^5 =$

∘	1–θ	3+θ	4
9	θ	1–θ	∘
∘	2	∘	∘

$\theta = 1$ =

∘	0	4	4
9	1	∘	∘
∘	2	∘	∘

Обратите, кстати, внимание на вырожденность плана X^3.

4.1.4. Задача о назначении персонала

Пусть имеется m категорий претендентов в количестве A_i ($i = 1 .. m$) и n групп вакантных должностей по B_j ($j = 1 .. n$) в каждой. Известны оценки $C_{i\,j}$ использования претендента i-й категории на должности из j-й группы.

Задача поиска распределения с максимальной суммарной эффективностью приводит к задаче, отличающейся от транспортной лишь требованием максимизации целевой функции.

Возьмем для примера $m = 3$, $n = 3$, $A = (6, 3, 1)$, $B = (2, 2, 2)$ и приведенную ниже матрицу эффективностей C.

Так как условие баланса нарушено за счет избытка претендентов, вводим фиктивную группу должностей с нулевой эффективностью, после чего ищем начальный опорный план по правилу предпочтения максимального значения из элементов C.

$$C = \begin{array}{|ccc|}\hline 4 & 6 & 1 \\ 5 & 6 & 3 \\ 9 & 8 & 8 \\ \hline\end{array} \quad C' = \begin{array}{|cccc|}\hline 4 & 6 & 1 & 0 \\ 5 & 6 & 3 & 0 \\ 9 & 8 & 8 & 0 \\ \hline\end{array} \quad X^0 = \begin{array}{cccc|c} \circ & 2 & 0 & 4 & 6 \\ 1 & \circ & 2 & \circ & 3 \\ 1 & \circ & \circ & \circ & 1 \\ \hline 2 & 2 & 2 & 4 & B\backslash A \end{array}$$

Проверка на оптимальность дает

$$U_i + V_j = \begin{array}{|c|cccc|}\hline \text{U\V} & 0 & 6 & 1 & 0 \\ \hline 0 & 0 & \mathbf{6} & \mathbf{1} & \mathbf{0} \\ 2 & \mathbf{2} & 8 & \mathbf{3} & 2 \\ 9 & \mathbf{9} & 15 & 10 & 9 \\ \hline\end{array} \qquad \Delta_{i\,j} = \begin{array}{|cccc|}\hline -4 & \circ & \circ & \circ \\ \circ & 2 & \circ & -1 \\ \circ & 7 & 2 & 1 \\ \hline\end{array}$$

Так как $\Delta_{11} < 0$, то план X^0 не оптимален. Перестраиваем его, полагая $X_{11} = \theta$:

$$X^1 = \begin{array}{|ccc|}\hline \theta & 2 & 0-\theta & 4 \\ 1-\theta & \circ & 2+\theta & \circ \\ 1 & \circ & \circ & \circ \\ \hline\end{array} \quad \theta = 0 \Rightarrow \quad \begin{array}{|ccc|}\hline 0 & 2 & \circ & 4 \\ 1 & \circ & 2 & \circ \\ 1 & \circ & \circ & \circ \\ \hline\end{array}$$

Проверка на оптимальность дает:

$$U_i + V_j = \begin{array}{|c|cccc|}\hline \text{U\V} & 4 & 6 & 2 & 0 \\ \hline 0 & \mathbf{4} & \mathbf{6} & 2 & \mathbf{0} \\ 1 & \mathbf{5} & 7 & \mathbf{3} & 1 \\ 5 & \mathbf{1} & 11 & 7 & 5 \\ \hline\end{array} \qquad \Delta_{i\,j} = \begin{array}{|cccc|}\hline \circ & \circ & 1 & \circ \\ \circ & 1 & \circ & 1 \\ \circ & 3 & -1 & 5 \\ \hline\end{array}$$

Так как $\Delta_{31} < 0$, перестраиваем план, полагая $X_{31}=\theta$:

$$X^2 = \begin{array}{|cccc|}\hline 0 & 2 & \circ & 4 \\ 1+\theta & \circ & 2-\theta & \circ \\ 1-\theta & \circ & \theta & \circ \\ \hline\end{array} \quad \theta=1 \quad = \quad \begin{array}{|cccc|}\hline 0 & 2 & \circ & 4 \\ 2 & \circ & 1 & \circ \\ \circ & \circ & 1 & \circ \\ \hline\end{array}$$

И снова проверка плана на оптимальность:

$$U_i + V_j = \begin{array}{|c|cccc|}\hline \text{U\V} & 4 & 5 & 2 & 0 \\ \hline 0 & \mathbf{4} & \mathbf{6} & 2 & \mathbf{0} \\ 1 & \mathbf{5} & 6 & \mathbf{3} & 1 \\ 6 & 10 & 11 & \mathbf{8} & 6 \\ \hline\end{array} \qquad \Delta_{ij} = \begin{array}{|cccc|}\hline \circ & \circ & 1 & \circ \\ \circ & 0 & \circ & 0 \\ 1 & 3 & \circ & 6 \\ \hline\end{array}$$

Все $\Delta_{ij} \geq 0$, план X^2 оптимален (есть и другие оптимальные планы).

Нами рассмотрена классическая однопродуктовая транспортная задача с прямыми связями между «поставщиками» и «потребителями» и методика ее решения. На практике же чаще приходится иметь дело с транспортировкой в сети с промежуточными узлами и ограничениями на пропускные способности коммуникаций. Такого рода *задача в сетевой постановке* решается более сложными методами, но и эти методы (в частности, известный *венгерский метод*) базируются на соотношениях двойственности.

Транспортная задача с критерием оптимальности не по затратам, а по времени выполнения комплекса перевозок решается опять-таки посредством постановки сопряженной задачи с привлечением алгоритма поиска максимального потока в транспортной сети (этот же алгоритм используется и в венгерском методе).

Обе упомянутые задачи будут нами рассмотрены ниже.

4.2. Распределительные задачи

Указанная группа задач выступает своеобразным обобщением транспортной задачи и возникает при закреплении ресурсов по видам работ, при закреплении транспортных средств за маршрутами, при составлении топливно-энергетических балансов, при планировании военных операций с максимальным поражением целей, при распределении заказов между фирмами и т. п.

Рассмотрим одну из такого рода задач.

Пусть имеется m типов автомашин в количествах A_i (i = 1 .. m), которые должны быть закреплены за n маршрутами с объемами перевозок, равными B_j (j = 1 .. n). Известно, что расходы на эксплуата-

цию и объем перевозок одной машины i-го типа на j-м маршруте равны соответственно C_{ij} и L_{ij}. Требуется найти распределение с минимальными эксплуатационными затратами.

Поставленная задача сводится к минимизации

$$L(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij} \tag{1}$$

при условиях

$$\sum_{j=1}^{n} X_{ij} \le A_i \ (i = 1 .. m), \tag{2}$$

$$\sum_{i=1}^{m} L_{ij} X_{ij} = B_j \ (j = 1 .. n), \tag{3}$$

$$X_{ij} \ge 0 \ (i = 1 .. m, j = 1 .. n). \tag{4}$$

Для приведения к канонической форме вводим в (2) ослабляющие переменные (эквивалентно добавлению фиктивного $(n + 1)$-го маршрута с нулевыми значениями $C_{i\ n+1}$). Сопоставим этому маршруту значения $L_{i\ n+1} = 1$ и добавим $(n + 1)$-е ограничение в (3) с каким-то значением B_{n+1}, требуя минимизации

$$L(X) = \sum_{i=1}^{m} \sum_{j=1}^{n+1} C_{ij} X_{ij} \tag{5}$$

при условиях

$$\sum_{j=1}^{n+1} X_{ij} = A_i \ \ (i = 1 .. m), \tag{6}$$

$$\sum_{i=1}^{m} L_{ij} X_{ij} = B_j \ \ (j = 1 .. n + 1), \tag{7}$$

$$X_{ij} \ge 0 \ (i = 1 .. m, j = 1 .. n + 1). \tag{8}$$

Сопряженная задача состоит в максимизации

$$\tilde{L}(U,V) = \sum_{i=1}^{m} A_i U_i + \sum_{j=1}^{n+1} B_j V_j \tag{9}$$

при условиях

$$U_i + L_{ij}\ V_j \le C_{i\,j} \text{ при всех } i, j. \tag{10}$$

Рассмотрим алгоритм У. Х. Малкова, использующий идею А. Фергюсона и Дж. Данцига.

Решение начинается с построения начального опорного плана

задачи (5) – (8) с использованием правила предпочтения, например, максимальной производительности на единицу затрат

$$R_{ij} = \frac{L_{ij}}{C_{ij}} \ . \qquad (11)$$

Если для транспортной задачи использовалось $X_{i\,j} = \min(A_i, B_j)$, то здесь этот прием заменяется на

$$X_{ij} = \min(A_i, B_j / L_{ij}). \qquad (12)$$

Число базисных компонент в найденном плане равно $m + n$. Если план оказывается вырожденным, то к числу базисных можно отнести и нулевые значения (не допуская замкнутой *вне фиктивного маршрута* цепочки по базису). Для автоматического устранения вырожденности используют ранее упомянутый ε-метод. После получения опорного плана производится его проверка на оптимальность: согласно второй теореме двойственности базисным X_{ij} сопоставляется система $m + n$ уравнений с $m + n + 1$ неизвестными $U_i + L_{i\,j}\,V_j = C_{ij}$, которая решается с точностью до константы (например, $V_{n+1} = 0$) и осуществляется проверка выполнения (10).

При обнаружении $\Delta_{kp} = U_k + L_{kp}\,V_p - C_{kp} > 0$ для соответствующего элемента X_{kp} отыскиваем так называемую «минимизирующую цепочку» с одним или двумя выходами на фиктивный столбец. Величины коррекции ΔX_{ij} для компонент нового плана полагаем равными $\gamma\ \theta_{ij}$.

По выделенной цепочке строим систему уравнений для сохранения баланса по строкам и столбцам относительно величин коррекции компонент нового плана (для строк – $\theta_{kp} + \theta_{kj} = 0$, для столбцов – $L_{kp}\,\theta_{kp} + L_{ip}\,\theta_{ip} = 0$). Решения полученной однородной системы находим с точностью до постоянного множителя. Приняв $\theta_{kp} = 1$, ищем остальные θ_{ij} и для неотрицательности компонент нового плана берем

$$\gamma = \min_{\theta_{ij} < 0} \frac{X_{ij}}{|\theta_{ij}|} \ . \qquad (13)$$

Очевидно, что при таком выборе соответствующая компонента плана обратится в нуль и выведется из базиса.

Пример. Пусть $m = 4$, $n = 5$, $A = (10, 50, 15, 30)$, $B = (60, 175, 400, 100, 100)$ и матрицы стоимостей и производитель-

ностей (с добавлением фиктивного маршрута) имеют вид

$$C = \left|\begin{array}{ccccc||c} 1 & 1 & 4 & 5 & 2 & 0 \\ 10 & 1 & 5 & 2 & 4 & 0 \\ 10 & 2 & 2 & 5 & 1 & 0 \\ 5 & 4 & 10 & 5 & 5 & 0 \end{array}\right| \qquad L = \left|\begin{array}{ccccc||c} 5 & 10 & 20 & 15 & 20 & 1 \\ 2 & 5 & 5 & 10 & 4 & 1 \\ 30 & 8 & 20 & 25 & 5 & 1 \\ 15 & 20 & 10 & 20 & 10 & 1 \end{array}\right|$$

Для поиска начального опорного плана определим значения R_{ij} согласно (11) и затем руководствуемся (12):

$$R = \begin{array}{|ccccc|} \hline 1 & 5 & 5 & 3 & 10 \\ 1/5 & 5 & 1 & 5 & 1 \\ 3 & 4 & 10 & 5 & 5 \\ 3 & 5 & 1 & 4 & 2 \\ \hline \end{array} \quad X^0 = \begin{array}{|ccccc|c|c|} \hline & 10 & & & & & 10 \\ & 15 & 20 & 10 & & 5 & 50 \\ & & 5 & & & & 15 \\ 4 & & & & 10 & 16 & 30 \\ \hline 60 & 175 & 400 & 100 & 100 & ... & B|A \\ \hline \end{array}$$

$X_{12} = \min(10, 175/10) = 10$, $X_{11} = X_{13} = X_{14} = X_{15} = X_{16} = 0$,
$X_{33} = \min(15, 400/20) = 15$, $X_{31} = X_{32} = X_{34} = X_{35} = X_{36} = 0$,
$X_{22} = \min(50, (175 - 10 \cdot 10) / 5) = 15$, $X_{42} = 0$,
$X_{24} = \min(50 - 15, 100 / 10) = 10$, $X_{44} = 0$,
$X_{41} = \min(30, 60 / 15) = 4$, $X_{21} = 0$,
$X_{45} = \min(30 - 4, 100 / 10) = 10$, $X_{25} = 0$,
$X_{23} = \min(50 - 15 - 10, (400 - 15 \cdot 20) / 5) = 20$, $X_{43} = 0$,
$X_{26} = 50 - 15 - 20 - 10 = 5$, $X_{46} = 30 - 4 - 10 = 16$.

Строим систему уравнений

$U_1 + 10\ V_2 = 1$, $U_2 + 10\ V_4 = 2$, $U_4 + 15\ V_1 = 5$, $U_2 + 5\ V_2 = 1$,
$U_2 + V_6 = 0$, $U_4 + 10\ V_5 = 5$, $U_2 + 5\ V_3 = 5$, $U_3 + 20\ V_3 = 2$, $U_4 + V_6 = 0$

и, получив решение этой системы при $V_6 = 0$ (уже не так просто, как для рассмотренной ранее транспортной задачи)

$U_1 = -1$, $U_2 = 0$, $U_3 = -18$, $U_4 = 0$,
$V_1 = 1/3$, $V_2 = 1/5$, $V_3 = 1$, $V_4 = 1/5$, $V_5 = 1/2$,

находим $\Delta_{i\,j} = U_i + L_{ij}\ V_j - C_{i\,j}$:

$$U_i + L_{ij}\ V_j = \begin{array}{|ccccc|c|} \hline 2/3 & \mathbf{1} & 19 & 2 & 9 & -1 \\ 2/3 & \mathbf{1} & \mathbf{5} & \mathbf{2} & 2 & \mathbf{0} \\ -8 & -16 & \mathbf{2} & -13 & -15 & -18 \\ \mathbf{5} & 4 & 10 & 4 & \mathbf{5} & \mathbf{0} \\ \hline \end{array} \quad \Delta_{ij} = \begin{array}{|ccccc|c|} \hline -1/3 & ° & \mathbf{15} & -3 & 7 & -1 \\ -9{,}3 & ° & ° & ° & -2 & ° \\ -18 & -18 & ° & -18 & -16 & -18 \\ ° & 0 & 0 & -1 & ° & ° \\ \hline \end{array}$$

Обнаружив положительные значения Δ_{ij}, переходим к другому плану, где, например, компонента X_{13} отлична от нуля (точнее, входит в число базисных переменных).

Согласно обнаруженной «цепочке», элементы которой помечены *, строим систему уравнений:

$X^1 =$

	10*	+*			
	15*	20*	10		5*
		15			
4				10	16

$$\theta_{13} + \theta_{12} = 0;$$
$$20\,\theta_{13} + 5\,\theta_{23} = 0;$$
$$10\,\theta_{12} + 5\,\theta_{22} = 0;$$
$$\theta_{22} + \theta_{23} + \theta_{26} = 0.$$

Решив эту систему при $\theta = 1$, имеем $\theta_{12} = -1$, $\theta_{22} = 2$, $\theta_{23} = -4$, $\theta_{26} = 2$.

Отыскав согласно (13)

$X^1 =$

	5	5			
	25		10		15
		15			
4				10	16

$$\gamma = \min\left(\frac{10}{1}, \frac{20}{5}\right) = 5,$$

получаем умножением на θ_{ij} значения корректур для компонент плана:

$\Delta X_{13} = 5$, $\Delta X_{12} = -5$, $\Delta X_{22} = 10$, $\Delta X_{23} = -20$, $\Delta X_{26} = 10$. Отсюда находим компоненты плана X^1.

Вновь строим систему уравнений:

$U_1 + 10\,V_2 = 1$, $U_1 + 20\,V_3 = 4$, $U_2 + 5\,V_2 = 1$, $U_2 + 10\,V_4 = 2$,
$U_2 + V_6 = 0$, $U_3 + 20\,V_3 = 2$, $U_4 + 15\,V_1 = 5$, $U_4 + 10\,V_5 = 5$, $U_4 + V_6 = 0$
и, отыскав решение этой системы при $V_6 = 0$
$U_1 = -1$, $U_2 = 0$, $U_3 = -3$, $U_4 = 0$,
$V_1 = 1/3$, $V_2 = 1/5$, $V_3 = 1/4$, $V_4 = 1/5$, $V_5 = 1/2$,
имеем

$U_i + L_{ij}\,V_j =$

2/3	**1**	**4**	2	9	–1
2/3	**1**	5	**2**	2	**0**
7	–1,4	**2**	2	–0,5	–3
5	4	2,5	4	**5**	**0**

$\Delta_{ij} =$

–1/3	°	°	–3	**7**	–1
–9,3	°	0	°	–2	°
–3	–3,4	°	–3	–2,5	–3
°	0	0	–1	°	°

Обнаружив $\Delta_{15} > 0$, переходим к плану, где $X_{15} \neq 0$. Согласно найденной «цепочке», строим очередную систему уравнений:

$X^2 =$

	5*	5		+*	
	25*		10		15*
		15			
4				10*	16*

$$\theta_{15} + \theta_{12} = 0;$$
$$20\,\theta_{15} + 10\,\theta_{45} = 0;$$
$$10\,\theta_{12} + 5\,\theta_{22} = 0;$$
$$\theta_{22} + \theta_{26} = 0,\ \theta_{45} + \theta_{46} = 0.$$

Решив эту систему при $\theta_{15} = 1$, имеем

$\theta_{12} = -1$, $\theta_{45} = -2$, $\theta_{22} = 2$, $\theta_{26} = -2$, $\theta_{46} = 2$.

Отыскав $\gamma = \min(\frac{5}{1}, \frac{10}{2}, \frac{10}{2}) = 5$, получаем умножением на θ_{ij} значения корректур для компонент плана:

$\Delta X_{15} = 5$, $\Delta X_{12} = -5$, $\Delta X_{22} = 10$, $\Delta X_{45} = -10$, $\Delta X_{26} = -10$, $\Delta X_{46} = 10$,

откуда получаем план X^2 (здесь мы оставили в базисе $X_{12} = 0$, а не X_{45}, так как при одинаковой производительности $R_{12} > R_{45}$).

Вновь строим систему уравнений:

$U_1 + 10\ V_2 = 1$, $U_2 + 5\ V_2 = 1$, $U_3 + 20\ V_3 = 2$,

$U_1 + 20\ V_2 = 4$, $U_2 + 10\ V_4 = 2$, $U_4 + 15\ V_1 = 5$,

$U_1 + 20\ V_5 = 2$, $U_3 + 1\ V_6 = 0$, $U_4 + V_6 = 0$.

Получив решение этой системы при $V_6 = 0$:

$U_1 = -1$, $U_2 = 0$, $U_3 = -3$, $U_4 = 0$,

$V_1 = 1/3$, $V_2 = 1/5$, $V_3 = 1/4$, $V_4 = 1/5$, $V_5 = 3/20$, имеем

$U_i + L_{ij}\ V_j =$

2/3	**1**	**4**	2	**2**	–1
2/3	**1**	5/4	**2**	3/5	**0**
7	–1,4	**2**	2	–9/4	–3
5	4	5/2	4	3/2	**0**

$\Delta_{ij} =$

–1/3	°	°	–3	°	–1
–9,3	°	–3,7	°	–3,4	°
–3	–3,4	°	–3	–3,2	–3
°	**0**	–7,5	–1	–3,5	°

Поскольку все $\Delta_{ij} \geq 0$, план X^2 оптимален. Однако наличие $\Delta_{42} = 0$ для небазисной компоненты плана определяет возможность существования и других оптимальных планов.

Очевидно, что при сохранении идеологии метода решения распределительная задача, в отличие от транспортной, не гарантирована от противоречивости условий и тем более от получения нецелочисленного оптимального плана даже при целочисленных исходных показателях.

4.3. Задачи на транспортных сетях

Рассмотрим некоторые задачи, решение которых базируется на понятии максимального потока в транспортной сети.

4.3.1. Задача о максимальном потоке

Рассмотрим транспортную сеть, где выделены пункты 0 (вход, источник) и n (выход, сток) и каждой дуге (отрезку), связывающей пункты i и j, сопоставлено число $d_{ij} > 0$, называемое *пропускной способностью* дуги. Эта величина характеризует максимальное допустимое количество вещества (воды, газа, самолетов, вагонов и др.),

которое может проходить по дуге в единицу времени. Количество вещества, реально проходящего по дуге от i до j, называем потоком по дуге (i, j) и обозначаем через X_{ij}.

Очевидно, что

$$0 \le X_{ij} \le d_{ij} \text{ для всех } i, j. \qquad (1)$$

Если учесть, что все вещество, вошедшее в промежуточный пункт сети, должно полностью выйти из него, получаем

$$\sum_i X_{ij} = \sum_k X_{jk}, \quad j \ne 0, n. \qquad (2)$$

Из естественного требования равенства суммарного потока на входе и на выходе имеем

$$\sum_j X_{0j} = \sum_i X_{in} = Z \ . \qquad (3)$$

Величину Z называем *величиной потока* в сети и ставим задачу максимизации Z при условиях (1) – (3). Решить задачу можно и симплексным методом, но едва ли эта возможность осуществима для сколько-нибудь реальной сети.

В случае так называемых $(0, n)$ – *плоских сетей*, т. е. сетей, которые можно изобразить на плоскости по одну сторону от линии, соединяющей вершины 0 и n, без пересечения дуг вне вершин (наша сеть относится к таковым), можно предложить простой наглядный алгоритм решения.

Для иллюстрации возьмем ориентированную сеть, приведенную на рис. 12 (числа на дугах-стрелках – пропускные способности). Берем самый «верхний» (по принципу левостороннего движения) путь от вершины 0 к вершине 7 [0 – 1 – 5 – 7], находим минимальную пропускную способность составляющих его дуг (рис. 13), равную 5, и уменьшаем пропускные способности дуг на эту величину.

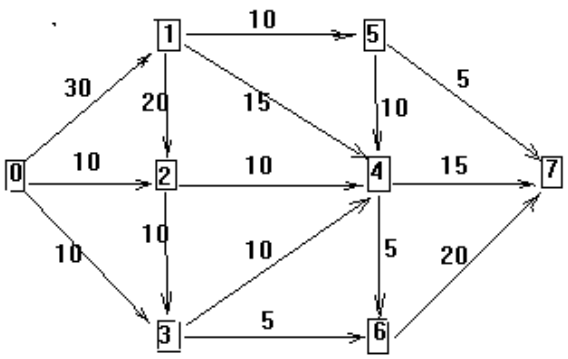


Рис. 12

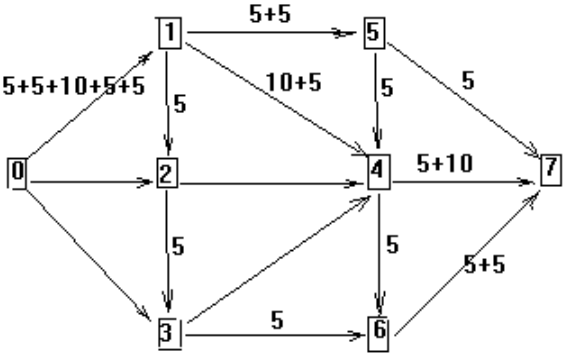


Рис. 13

Последующий поиск обнаруживает потоки по «верхним» путям: [0 – 1 – 5 – 4 – 7], [0 – 1 – 4 – 7], [0 – 1 – 4 – 6 – 7], [0 – 1 – 2 – 3 – 6 – 7] с пропускными способностями (рис. 13) соответственно 5, 10, 5, 5. В итоге сеть оказалась разорванной, максимальный поток равен 30.

Поскольку для больших сетей такой метод неприемлем, рассмотрим универсальный алгоритм поиска в матричной форме.

Строим матрицу D_0, куда заносим значения d_{ij} (для неориентированной дуги $d_{ij} = d_{ji}$), затем повторяем процесс поиска некоторого пути и коррекции потока на этом пути, используя процесс отмечаний.

Метим символом * нулевые строку и столбец матрицы (вход сети). В строке 0 отыскиваем $d_{0j} > 0$ и метим соответствующие столбцы индексами $\mu_j = 0$, $V_j = d_{0j}$ и переносим метки столбцов на строки. Затем берем i-ю отмеченную строку, ищем в ней непомеченный столбец с $d_{ij} > 0$, которому сопоставляем метки-индексы $\mu_j = i$, $V_j = \min(V_j, d_{ij})$. Метки столбцов переносим на строки и этот процесс продолжаем до тех пор, пока не будет отмечен n-й столбец.

Затем «обратным ходом» по индексам выясняем путь, приведший к n-й вершине, и уменьшаем пропускные способности дуг пути (элементы матрицы) на V_n, увеличивая симметричные элементы на эту же величину. Такая процедура продолжается до тех пор, пока отмечание n-й вершины не станет невозможным.

Максимальный поток может быть найден вычитанием из исходной матрицы D_0 получаемой после вышеприведенной корректуры матрицы пропускных способностей $X = \max(D_0 – D_k, 0)$.

Для рассмотренного примера строим матрицу D_0. Из строки 0 метим вершины 1, 2 и 3 (строки-столбцы) индексами $\mu = 0$ и V, равными 30, 10 и 10. Из меченой строки 1 метим вершины 4 и 5 индексами $\mu = 1$, $V_4 = \min(30, 15) = 15$, $V_5 = \min(30, 10) = 10$. Из строки 3 метим вершину 6 и, наконец, из строки 4 – вершину 7.

Обратным ходом по μ обнаруживаем путь: к вершине 7 от 4, к 4 от 1, к 1 от 0; корректируем элементы D_0 на величину $V_7 = 15$.

Очередной шаг дает путь [0 – 1 – 5 – 7] с потоком 5.

$D_0 =$

	*	0/30	0/10	0/10	1/15	1/10	3/5	4/15	
	0	1	2	3	4	5	6	7	
0		30	10	10					*
1			20		15	10			0/30
2				10	10				0/10
3					10		5		0/10
4							5	15	1/15
5					10			5	1/10
6								20	3/5
7									4/15

$D_1 =$

	*	/15	0/10	0/10	2/10	1/10	3/5	5/5	
	0	1	2	3	4	5	6	7	
0		15	10	10					*
1	15		20		0	10			0/15
2				10	10				0/10
3					10		5		0/10
4		15					5	0	2/10
5					10			5	1/10
6								20	3/5
7					15				5/5

Аналогичные действия приводят к матрице D_4, где дальнейшее отмечание невозможно. Отсюда получаем матрицу $X_{\max}$ максимального потока.

$D_4 =$

	*	0/10	0/5	0/5	2/5	1/5			
	0	1	2	3	4	5	6		
0		10	5	5				7	*
1	20		20		0	5			0/10
2	5			10	5				0/10
3	5				10		0		0/5
4		15	5				0	0	2/5
5	5				10			0	1/5
6				5	5			10	
7					15	5	10		

Этот алгоритм не накладывает никаких ограничений на транспортную сеть (любая ориентация, не обязательно плоская) и элементарно реализуется в программном виде.

$X_{\max} =$

0	1	2	3	4	5	6	7
0	20	5	5				
1		0		15			
2			0	5			
3				0		5	
4						5	15
5				0			5
6							10
7							

4.3.2. Обобщенная задача о максимальном потоке

В отличие от описанной выше задачи здесь предполагается полная ориентированность сети (поток по дуге только в одном направлении), и ограничения на пропускную способность не только сверху, но и снизу.

Такая обобщенная задача состоит в максимизации

$$\sum_j X_{0j} = \sum_i X_{in} = Z$$

при условиях $\sum_i X_{ij} = \sum_k X_{jk}$, $j = 1..n$; $0 \le b_{ij} \le X_{ij} \le d_{ij}$ для всех i, j.

Если в простой задаче о максимальном потоке всегда можно найти допустимый (например, нулевой) поток, то здесь такого потока может не существовать. Примером такой ситуации может служить задача, где при некотором j $\sum_i d_{ij} < \sum_k b_{jk}$.

Для поиска допустимого потока расширяем сеть добавлением вершин –1 (псевдовход) и $n + 1$ (псевдовыход). Псевдовход соединяем со всеми вершинами, кроме 0, дугами, пропускная способность которых равна сумме нижних границ по дугам, входящим в соответствующую вершину. Из всех вершин, кроме n, проводим в вершину $n + 1$ дуги, пропускная способность которых равна сумме нижних границ пропускных способностей дуг, выходящих из соответствующей вершины. Кроме того, соединяем вершину n с вершиной 0 дугой с неограниченной пропускной способностью. Если нам удастся в такой расширенной сети найти максимальный поток, «насыщающий» псевдовыход, тем самым мы докажем существование допустимого потока.

В матричном виде алгоритм реализуется в два этапа.

На первом этапе строится матрица $D - B$, окаймляемая строками и столбцами -1 и $n + 1$. Элементы с индексами $(-1, j)$ полагаем равными $\sum_i b_{ij}$, элементы $(i, n + 1)$ равными $\sum_j b_{ij}$ и элемент $(n, 0)$ – произвольно большому числу. К полученной матрице применяем алгоритм поиска максимального потока.

Если найденный максимальный поток не является насыщающим, то задача неразрешима. В противном случае отбрасываем дополнительные строки и столбцы, исключаем связь $(n, 0)$; если $d_{0n} > 0$, соответствующий элемент матрицы берем равным $d_{0n} - b_{0n}$ и 0 в противном случае. К полученной матрице вновь применяем алгоритм поиска максимального потока.

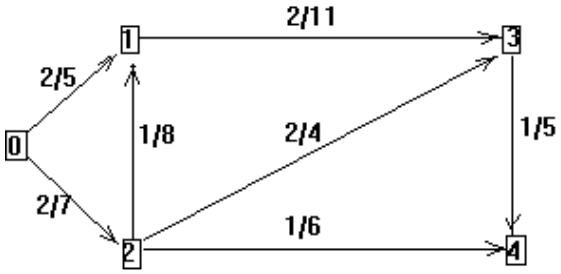


Рис. 14

Пример. Рассмотрим приведенную здесь сеть (рис. 14), где числа на дугах – ограничения пропускных способностей снизу и сверху.

Строим исходные и расширенную матрицы:

$B =$

	0	1	2	3	4
0		2	2		
1				2	
2		1		2	1
3					1
4					

$D =$

	0	1	2	3	4
0		5	7		
1				11	
2		8		4	6
3					5
4					

$T_0 =$

	−1	0	1	2	3	4	5
−1			3	2	4	2	
0			3	5			4
1					9		2
2			7		2	5	4
3						4	1
4		∞					
5							

Не приводя здесь ряд промежуточных матриц, найдем очевидные потоки с пропускной способностью r: $[-1 - 1 - 5]$, $r = 2$; $[-1 - 2 - 5]$, $r = 2$; $[-1 - 3 - 5]$, $r = 1$; $[-1 - 4 - 0 - 5]$, $r = 2$ и в результате обычной коррекции получаем матрицу T_4, из которой в процессе следующих отмечаний находим пути: $[-1 - 1 - 3 - 4 - 0 - 5]$, $r = 1$; $[-1 - 3 - 4 - 0 - 5]$, $r = 1$; $[-1 - 3 - 4 - 0 - 2 - 5]$, $r = 2$ и получаем матрицу T_7, показывающую исчерпание возможностей псевдовхода и псевдовыхода.

$T_4 =$

	−1	0	1	2	3	4	5
−1			1	0	3	0	
0			3	5		2	2
1	2				9		0
2	2		7		2	5	2
3	1					4	0
4	2	∞					
5		2	2	2	1		

$T_7 =$

	−1	0	1	2	3	4	5
−1			0	0	0	0	
0			3	3		6	0
1	3				8		0
2	2	2	7		2	5	0
3	4		1			0	0
4	2	∞			4		
5		4	2	4	1		

$T_8 =$

	0	1	2	3	4
0		3	3		
1				8	
2	2	7		2	5
3		1			0
4				4	

$T_9 =$

	0	1	2	3	4
0		3	0		
1				8	
2	5	7		2	2
3		1			0
4			3	4	

$X =$

	0	1	2	3	4
0		2	7		
1				3	
2		1		2	4
3					5
4					

Проведя усечение T_7 к виду T_8 и обнаружив путь $[0 - 2 - 4]$ с $r = 3$, получаем матрицу T_9, в которой дальнейшее отмечание невозможно.

Отыскав матрицу $D - T_9$ и ограничиваясь лишь положительными значениями, получаем матрицу X максимального потока.

4.3.3. Венгерский метод и транспортные задачи

В случае классической транспортной задачи в отличие от ранее рассмотренного метода Данцига, где выбирался опорный план исходной задачи и проверялся на оптимальность посредством соотношений сопряженной задачи, здесь[13] выбирается какой-нибудь план сопряженной задачи , например

$$U_i = \min_j C_{ij},\ i = 1 \,..\, m,$$

$$V_j = \min_i [C_{ij} - U_i],\ j = 1 \,..\, n,$$

строится матрица $T = ||C_{ij} - U_i - V_j|| \geq 0$ и, в соответствии с $T_{ij} = 0$ (по соображениям второй теоремы двойственности), строится допустимая сеть, где ищется максимальный поток (решение исходной задачи). Если он не исчерпывает возможности поставщиков, берем другой опорный план сопряженной задачи (другую допустимую

[13] Идея этого метода высказана еще в 1931 году венгерским математиком Ф. Эгервари. Эта забытая работа была обнаружена в 1953 году американским математиком Г. Куном, который развил эту идею и назвал созданный им метод венгерским.

сеть) и т. д. Технологию этого метода с иллюстративным примером читатель может найти в [11].

Этот метод позволяет решать не только классическую транспортную задачу, но и транспортную задачу в сетевой постановке, где кроме «поставщиков» и «потребителей» присутствуют промежуточные пункты и даже ограничения на пропускные способности отдельных дуг сети.

Рассмотрим более общую задачу на примере нижеприведенной сети (рис. 15), где числа на дугах определяют стоимость единичной перевозки, вершины 1 и 2 – пункты производства продукта в количествах 20 и 10 единиц, вершины 6 и 7 – пункты потребления в количествах 15 и 15 единиц. К тому же на маршрутах 2 – 5 и 3 – 6 пропускная способность не превышает соответственно 5 и 7 единиц продукта.

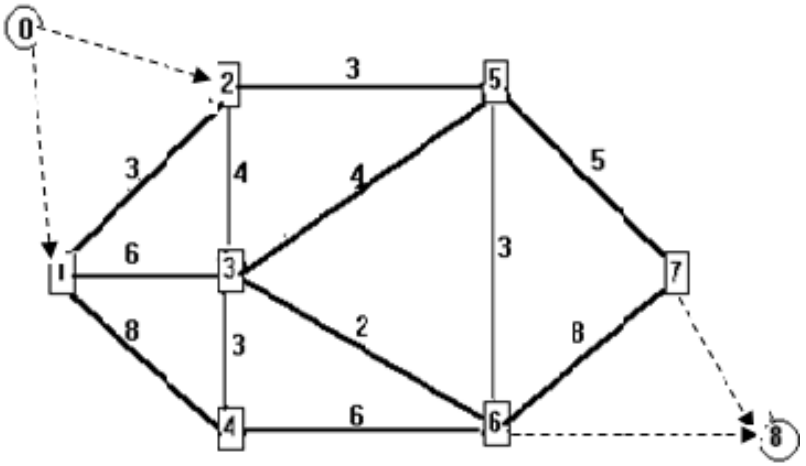


Рис. 15

Воплощая основную идею метода, введем фиктивные вход 0 и выход 8. Соединим вход 0 с вершинами 1 и 2 дугами с пропускной способностью, равной объемам производства, и вершины 6 и 7 с выходом 8 – дугами с пропускной способностью, равной объемам потребления. Стоимости перевозок на этих дугах берем нулевыми.

В полученной сети ищем максимальный поток от входа к выходу, обеспечивающий минимум стоимости перевозок

$$L(X) = \sum_{(i\ j)} C_{ij}\ X_{ij} \qquad (1)$$

при условиях

$$\sum_j X_{0\,j} = A, \qquad (2)$$

$$\sum_i X_{i\,n} = A, \qquad (3)$$

$$\sum_i X_{ij} = \sum_k X_{jk},\ j \neq 0, n \qquad (4)$$

$$0 \le X_{ij} \le d_{ij} \text{ для всех } i, j \qquad (5)$$

и соблюдении баланса

$$A = \sum a_i = \sum b_j\ . \qquad (6)$$

Один из возможных алгоритмов состоит в следующем.

1. Отмечаем вход 0 некоторым значком (множество отмеченных вершин в дальнейшем будем обозначать через M).

2. Отыскиваем неотмеченные вершины (j), в которые ведут ненасыщенные дуги с нулевыми стоимостями перевозок из вершин множества M, т. е. дуги с характеристиками $C_{Mj} = 0$, $X_{Mj} < d_{Mj}$. При отсутствии таковых переходим к п. 3 и при наличии – к п. 4.

3. Ищем среди неотмеченных вершину, из которой идет дуга во множество M с нулевой стоимостью и ненулевой перевозкой, и переходим к п. 4. При отсутствии таковой – к п. 7.

4. Выбранную вершину включаем во множество M и, если выход остался непомеченным, переходим к п. 2.

5. По меткам, сопоставленным отмеченным вершинам (откуда и сколько), ищем путь от входа к выходу и его пропускную способность и корректируем суммарный поток и пропускные способности.

6. Если пропускные способности дуг, исходящих из вершины входа, не исчерпаны, переходим к п. 1.

7. Выясняем наличие дуг, для которых $C_{Mj} > 0$ при $j \notin M$ или $C_{iM} < 0$ при $i \in M$. Если таковых нет, задача неразрешима.

8. Среди модулей найденных C_{Mj} и C_{iM} отыскиваем минимальное значение δ и корректируем матрицу стоимостей, вычитая δ из стоимостей на дугах, ведущих из M, и добавляя к стоимостям на дугах, ведущих в M. Возвращаемся к п. 2.

Покажем реализацию алгоритма в матричной форме для вышеприведенной сети (рис. 15).

$C =$

	0	1	2	3	4	5	6	7	8	
0		0	0							*
1			3	6	8					*
2		3		4		1				*
3		6	4		3	5	2			
4		8		3			6			
5			1	5			3	5		
6				2	6	3		8	0	
7						5	8		0	
8										
	*	*	*							

$D =$

	0	1	2	3	4	5	6	7	8
0		20	10						
1			∞	∞	∞				
2	∞		∞		5				
3	∞	∞		∞	∞	7			
4	∞		∞			∞			
5		∞	∞			∞	∞		
6			∞	∞	∞		∞	15	
7					∞	∞		15	
8									

Выбрав начальным поток $X^0 = 0$, задаем начальную метку:
$\mu_0 = -1$, $V_0 = 20 + 10 = 30$ (общий объем производства); и с учетом $C_{01} = 0$, $X_{01} < D_{01}$ и $C_{02} = 0$, $X_{02} < D_{02}$:
$\mu_1 = 0$, $V_1 = \min(V_0, D_{01}) = 20$; $\mu_2 = 0$, $V_2 = \min(V_0, D_{02}) = 10$.

Так как в отмеченных строках нет нулевых стоимостей C_{Mj}, $j \notin M$, то и дальнейшее отмечание невозможно. Найдя

$$\delta = \min_{\substack{i=0,1,2 \\ j\neq 0,1,2}} |C_{ij}| = \min(6, 8, 4, 1) = 1,$$

вычитаем δ из отмеченных строк матрицы C, добавляя к отмеченным столбцам (получаем матрицу C_1) и, продолжая процедуру отмечаний с учетом C_1 и D, можем отметить в дополнение к отмеченным ранее вершину 5 $\mu_5 = 2$, $V_5 = \min(V_2, D_{25}) = 5$;

$C_1 =$

	0	1	2	3	4	5	6	7	8	
0		0	0							
1			3	5	7					*
2		3		3		0				*
3		7	5		3	5	2			
4		9		3			6			
5			2	5			3	5		*
6				2	6	3		8	0	
7						5	8		0	
8										
		*	*			*				

$C_2 =$

	0	1	2	3	4	5	6	7	8
0		0	0						
1			3	2	4				
2		3		0		0			
3		10	8		3	8	2		
4		12		3			6		
5			2	2			0	2	
6				2	6	6		8	0
7						8	8		0
8									

после чего опять-таки приходится расширять сеть путем коррекции матрицы C_1 на величину

$$\delta = \min_{\substack{i=0,1,2,5 \\ j\neq 0,1,2,5}} |C_{ij}| = \min(5, 7, 3, 5, 3, 5) = 3,$$

получая матрицу C_2. Продолжив отмечания, имеем:
$\mu_3 = 2$, $V_3 = \min(V_2, D_{23}) = 10$; $\mu_6 = 5$, $V_6 = \min(V_5, D_{56}) = 5$; $\mu_8 = 6$, $V_8 = \min(V_6, D_{68}) = 5$.

Обратным ходом по меткам выясняем путь $[0 \to 2 \to 5 \to 6 \to 8]$ с пропускной способностью $V_8 = 5$ и корректируем матрицу D, уменьшая ее элементы, соответствующие дугам пути в прямом направлении, и увеличивая симметричные на $V_8 = 5$.

X^1=4	0	1	2	3	4	5	6	7	8
0		0	5						
1			0	0	0				
2		0		0		5			
3		0	0		0	0	0		
4		0		0			0		
5			0	0			5	0	
6				0	0	0		0	5
7						0	0		0
8									

D_1 = 4	0	1	2	3	4	5	6	7	8
0		20	5						
1			∞	∞	∞				
2	5	∞		∞		0			
3		∞	∞		∞	∞	7		
4		∞		∞			∞		
5			∞	∞			∞	∞	
6				∞	∞	∞		∞	10
7						∞	∞		15
8							5		

Повторяя процедуру отмечаний (на основе C_2, D_1 и X^1), имеем: $\mu_0 = -1$, $V_0 = 20 + 5 = 25$; $\mu_1 = 0$, $V_1 = \min(V_0, D_{01}) = 20$; $\mu_2 = 0$, $V_2 = \min(V_0, D_{02}) = 5$; $\mu_3 = 2$, $V_3 = \min(V_2, D_{23}) = 5$ (вершину 5 отметить не удается, так как $D_{25} = 0$). Соответственно $\delta = \min(4, 3, 7, 2) = 2$ (значение $C_{25} = 0$ не учитываем, так как соответствующей дуги на этом этапе решения нет). После вычитания из отмеченных строк и добавления к отмеченным столбцам получаем матрицу C_3, и, продолжая процесс отмечаний, имеем: $\mu_6 = 3$, $V_6 = \min(V_3, D_{36}) = 5$; $\mu_8 = 6$, $V_8 = \min(V_6, D_{68}) = 5$, откуда имеем путь $[0 \to 2 \to 3 \to 6 \to 8]$ с пропускной способностью 5.

C_3 = 4	0	1	2	3	4	5	6	7	8
0		0	0						
1			3	2	2				
2		3		0		0			
3		10	8		1	6	0		
4		14		5			6		
5			4	4			0	2	
6				4	6	6		8	0
7						8	8		0
8									

Выполнив обычную коррекцию матрицы пропускных способностей и потока, получаем D_2 и X_2.

D_2 = 4	0	1	2	3	4	5	6	7	8
0		20	0						
1			∞	∞	∞				
2	10	∞		∞		0			
3		∞	∞		∞	∞	2		
4		∞		∞			∞		
5			∞	∞			∞	∞	
6				∞	∞	∞		∞	5
7						∞	∞		15
8							10		

X_2 = 4	0	1	2	3	4	5	6	7	8
0		0	1 0						
1			0	0	0				
2		0		5		5			
3		0	0		0	0	5		
4		0		0			0		
5			0	0			5	0	
6				0	0	0		0	10
7						0	0		0
8									

Продолжая аналогичные действия, обнаруживаем путь $[0 \to 1 \to 3 \to 6 \to 8]$ с величиной потока 2, $[0 \to 1 \to 4 \to 6 \to 8]$ с потоком 3, путь $[0 \to 1 \to 3 \to 5 \to 7 \to 8]$ с потоком 15.

На завершающем этапе получается матрица C_7 и матрицы D_5 и X_5, свидетельствующие о том, что возможности поставщиков исчерпаны и потребности удовлетворены. Граф максимального потока представлен на рис. 16.

Заметим, что матрицу X_5 можно получать и по окончании счета, вычитая D_5 из исходной матрицы D_0.

C_7=

	0	1	2	3	4	5	6	7	8
0		0	0						
1			0	0	0				
2		6		1		0			
3		12	7		1	0	0		
4		16		5			0		
5			9	10			0	0	
6				10	12	6		6	0
7						10	10		0
8									

D_5 =

	0	1	2	3	4	5	6	7	8
0		0	0						
1	20		∞	∞	∞				
2	10	∞		∞		0			
3		∞	∞		∞	∞	0		
4		∞		∞			∞		
5			∞	∞			∞	∞	
6				∞	∞	∞		∞	0
7						∞	∞		0
8							15	15	

X_5 =

	0	1	2	3	4	5	6	7	8
0		20	10						
1			0	20	0				
2		0		5		5			
3		0	0		0	18	7		
4		0		0			0		
5			0	0			8	15	
6				0	0	0		0	15
7						0	0		15
8									

Нет сомнений том, что мало-мальски приличную транспортную задачу вручную никто не будет решать и естественно в этой ситуации прибегнуть к помощи компьютера (если у вас есть возможность стать обладателем соответствующего программного средства).

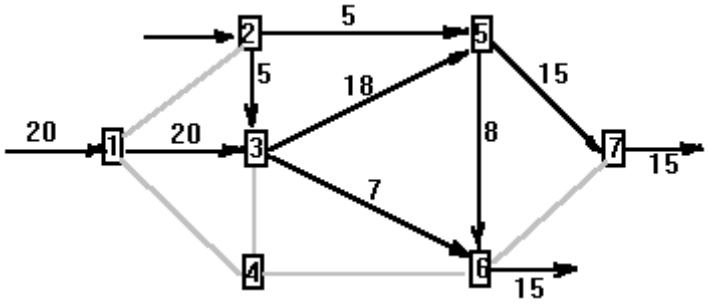


Рис. 16

4.3.4. Транспортная задача по критерию времени

В рассматриваемой задаче критерием качества организации перевозок являются не суммарные затраты, а время реализации перевозок (подобные проблемы возникают при транспортировке скоропортящихся грузов, при переброске сил быстрого реагирования и т. д.).

Пусть имеется m поставщиков продукта в количествах A_i ($i = 1 .. m$) и n потребителей в количествах B_j ($j = 1 .. n$), причем соблюдается баланс предложения и спроса. Известно *время* t_{ij} доставки груза от i-го поставщика к j-му потребителю, *не зависящее от объема перевозки*.

Хотелось бы среди всех допустимых планов перевозок $X = \{X_{ij}\}$ найти план, оптимальный по времени. Очевидно, что время, необходимое для реализации плана, совпадает с наибольшим временем его отдельных перевозок и оптимальное время перевозок равно

$$T_{\text{opt}} = \min_{X} \max_{X_{ij}>0} t_{ij} \qquad (1)$$

при условиях

$$\sum_{j=1}^{n} X_{ij} = A_i,\ i = 1 .. m; \qquad (2)$$

$$\sum_{i=1}^{m} X_{ij} = B_j,\ j = 1 .. n; \qquad (3)$$

$$X_{ij} \geq 0,\ i = 1 .. m,\ j = 1 .. n; \qquad (4)$$

$$\sum_{i=1}^{m} A_i = \sum_{j=1}^{n} B_j . \qquad (5)$$

Алгоритм решения базируется на идеях венгерского метода для классической транспортной задачи и элементарном здравом смысле.

Предварительно выбирается минимальное значение t_{ij} и строится допустимая сеть S по значениям t_{ij}, не превышающим выбранного. В этой сети отыскивается максимальный поток. Если этот поток отвечает условиям задачи, то выбранное время оптимально. В противном случае выбирается очередное наименьшее время, сеть расширяется и в ней вновь ищется максимальный поток.

Очевидно, что для выбора начального времени t^* имеет смысл учесть возможность вывоза от каждого поставщика и доступа к каждому потребителю и отталкиваться от максимального среди минимальных времен в строках и столбцах матрицы T.

Что касается поиска максимального потока в сети с прямыми поставками, то есть возможность оперировать с матрицами размерности m на n, используя приведенные ниже приемы.

Поэтому рассмотрим компактную схему поиска максимального потока, позволяющую работать с матрицами размерности $m \times n$.

Пусть найден какой-нибудь поток $X \geq 0$ в допустимой сети S, удовлетворяющий естественным условиям:

$$\sum_{j=1}^{n} X_{i\,j} \leq a_i \ (i = 1 \,..\, m); \ \sum_{i=1}^{m} X_{i\,j} \leq b_j \ (j = 1 \,..\, n);$$

(поиск начального приближения для потока можно осуществлять любым методом, например, любой вариацией метода северо-западного угла).

Для строк i, в которых

$$\sum_{j=1}^{n} X_{i\,j} < a_i, \tag{6}$$

сопоставим метки

$$\mu_i = 0, \ \upsilon_i = a_i - \sum_{j=1}^{n} X_{i\,j} \ . \tag{7}$$

Выбираем отмеченные строки (например, i-ю) и отмечаем неотмеченные столбцы j такие, что дуга $(i\,j) \in S$, метками

$$\lambda_j = i, \ \omega_j = \upsilon_i. \tag{8}$$

Затем берем отмеченные столбцы (например, j-й) и неотмеченным строкам i, в которых $X_{i\,j} > 0$, сопоставляем метки

$$\mu_i = j, \ \upsilon_i = \min_i (\omega_j, X_{i\,j}). \tag{9}$$

Повторяем процесс отмечания столбцов и строк до тех пор, пока не будет отмечен «ненасыщенный» столбец j^*, для которого

$$\sum_{i=1}^{m} X_{i\,j^*} < b_{j^*}. \tag{10}$$

Отыскиваем величину

$$\Theta = \min\left(\omega_{j^*}, b_{j^*} - \sum_{i=1}^{m} X_{i\,j^*}\right), \qquad (11)$$

определяющую величину потока в найденном пути, поочередно добавляем и вычитаем Θ из значений X_{ij} в цепочке

$$(i_0\ j^*)\ (i_0\ j_1)\ (i_1\ j_1)\ (i_1\ j_2)\ (i_2\ j_2)\ \ldots (i_{k-1}\ j_k)\ (i_k\ j_k),$$

где

$$i_0 = l_{j^*},\ j_1 = m_{i_0},\ i_1 = l_{j_1},\ j_2 = m_{i_1},\ i_2 = l_{j_2},\ \ldots,\ i_k = l_{j_k}\ (m_{i_k} = 0),$$

и вновь продолжаем процесс отмечаний. Если не удастся отметить ни один из ненасыщенных столбцов, перестраиваем сеть, расширяя ее за счет увеличения времени.

Пример. Пусть задача определена данными, приведенными в таблице.

$T =$

1	13	17	18	18
2	18	10	10	10
16	1	4	12	12
11	9	13	7	*B**A*

Минимальные значения t_{ij} в строках равны 1, 2, 1 и в столбцах 1, 1, 4, 10. Выбираем $t^* = 10$, строим вспомогательную сеть по $t_{ij} \le t^*$ и отыскиваем в ней начальное приближение для потока.

Так как найденный поток X_0 не является насыщающим (исчерпывающим возможности поставки и потребления), пытаемся его улучшить с использованием процесса отмечаний венгерского метода $\mu_1 = 0$, $\upsilon_1 = 18 - 11 = 7$; $\lambda_1 = 1$, $\omega_1 = \upsilon_1 = 7$.

$X_0 =$

11				18
0		10	0	10
	9	3		12
11	9	13	7	*B**A*

Дальнейшее отмечание невозможно: приходится расширить сеть, взяв $t^* = 12$ (появится возможность перевозки на маршруте 1 – 4, поток X_0').

$X_0' =$

11				18
0		10	0	10
	9	3	0	12
11	9	13	7	*B**A*

Очевидно, что это расширение ничего нового не даст; берем $t^* = 13$ (поток X_0'').

$X_0'' =$

11	0			18
0		10	0	10
	9	3	0	12
11	9	13	7	*B**A*

Отталкиваясь от ранее выбранного потока, пытаемся его улучшить. Продолжая процесс отмечаний, имеем

$\lambda_2 = 1$, $\omega_2 = \upsilon_1 = 7$; $\mu_3 = 2$,

$\upsilon_3 = \min(X_{32}, \omega_2) = 7$; $\lambda_2 = 3$, $\omega_4 = \upsilon_3 = 7$.

$X_1 =$

11	7			18
0		10	0	10
	2	3	7	12
11	9	13	7	*B**A*

Так как отмечен ненасыщенный столбец, отыскиваем минимизирующую цепочку [X_{34} X_{32} X_{12}] и корректируем ее на величину $\theta = \min(\omega_4, B_4 - 0) = 7$. В итоге мы получаем насыщающий поток и можем утверждать, что минимальное время транспортировки составляет 13 единиц.

4.3.5. Замечания

Исторически первые опыты в сфере исследования операций связаны с задачами транспортировки грузов задолго до появления самого термина. Пристальное внимание к этой сфере было привлечено в США в годы Второй мировой войны в связи с организацией снабжения войск союзников на театре военных действий.

Не обошли вниманием эту тематику и советские математики, уже к началу 60-х разработавшие множество оригинальных методов решения сетевых задач. Так в опубликованной в 1962 г. монографии И. Я. Бирмана «Транспортная задача линейного программирования» рассматривались методы потенциалов и дифференциальных рент, задачи оптимальных перевозок угля по Сибири и Дальнему Востоку, распределения объектов стройиндустрии Казахстана и другие.

Принципиальную роль в методологии решения сетевых задач сыграли фундаментальные работы Л. Форда и Д. Фалкерсона [12] с их алгоритмами, базирующимися на поиске максимального потока.

Фантастически обширную литературу собрала задача о коммивояжере (кратчайшим путем обойти все вершины транспортной сети, но только по одному разу, и вернуться в исходную вершину) – до сих пор предлагаются эвристические алгоритмы ее решения – от метода ветвей и границ до генетических алгоритмов.

Давно стали популярными простые, но требующие помощи компьютеров методы поиска кратчайших или самых длинных путей в транспортной сети (далее мы коснемся одного из этих методов, базирующегося на идеях Р. Беллмана).

Заслуживает упоминания возникающая иногда при синтезе цепей задача поиска кратчайших связывающих сетей, реализуемая достаточно простым алгоритмом Р. К. Прима [13].

Широкий круг комбинаторных задач, связанных с прогулками по вершинам и дугам (ребрам) сети, решается с привлечением аппарата теории графов (среди множества пособий по теории графов начинающему исследователю рекомендуется обратиться к содержательной, немногословной, доступной непрофессионалу книге

К. Бержа [15]). Примерами подобных задач служат различные вариации задачи коммивояжера, поиска эйлеровых цепей и циклов (обход всех ребер по одному разу с возвратом или без такового), поиска гамильтоновых путей (обход вершин в ориентированной сети), проектирования транспортных маршрутов без пересечения дорог вне узловых пунктов, разработки структуры тайной организации и многие другие.

Увы, существует множество сетевых задач, не укладывающихся в рамки упомянутых выше. Так в приложениях возникают задачи о *многопродуктовых потоках* в сети с ограниченными пропускными способностями, где сложно установить саму возможность транспортировки или минимизировать затраты даже при линейных связях между затратами и объемами перевозок.

5. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

5.1. Специфика нелинейных программ и методы их решения

Выше мы рассмотрели основные подходы к решению задач линейного программирования, где экстремум целевой функции достигается в вершинах многогранника планов и соответственно возникает идея упорядоченного перебора опорных планов (вершин множества планов) исходной или сопряженной задачи.

Однако в реальной жизни взаимосвязи между ее характеристиками, как правило, нелинейны и близки к линейным лишь в ограниченных условиях. Для нелинейных же программ простой метод решения, подобный симплексному, отсутствует по ряду причин.

Во-первых, множество планов может оказаться невыпуклым или иметь бесконечное количество «вершин» (рис. 17). Хуже того, оно может оказаться несвязным (рис. 18).

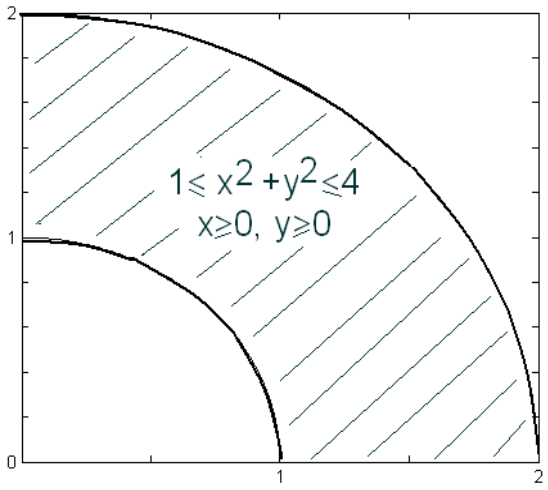


Рис. 17

Во-вторых, искомые экстремумы могут достигаться как на границе, так и внутри множества планов.

В-третьих, в нелинейных программах может возникнуть проблема поиска **глобального** экстремума среди множества **локальных** (рис. 19).

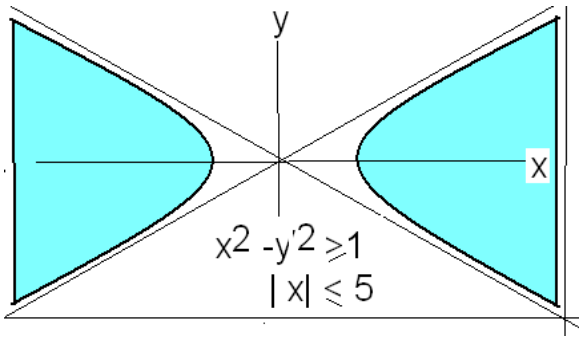


Рис. 18

Как мы показали ранее, использование аппарата производных или прямое табулирование целевой функции на множестве планов не решают проблему в случае более трех переменных. Поэтому каждая нелинейная программа требует индивидуального подхода, учитывающего ее специфику.

Большинство методов нелинейного программирования можно подразделить на следующие основные классы.

1. *Градиентные методы*, в основе которых лежит свойство градиента функции в точке (вектора частных производных, вычисленного в точке) как указателя направления наибольшего роста функции в окрестности этой точки.

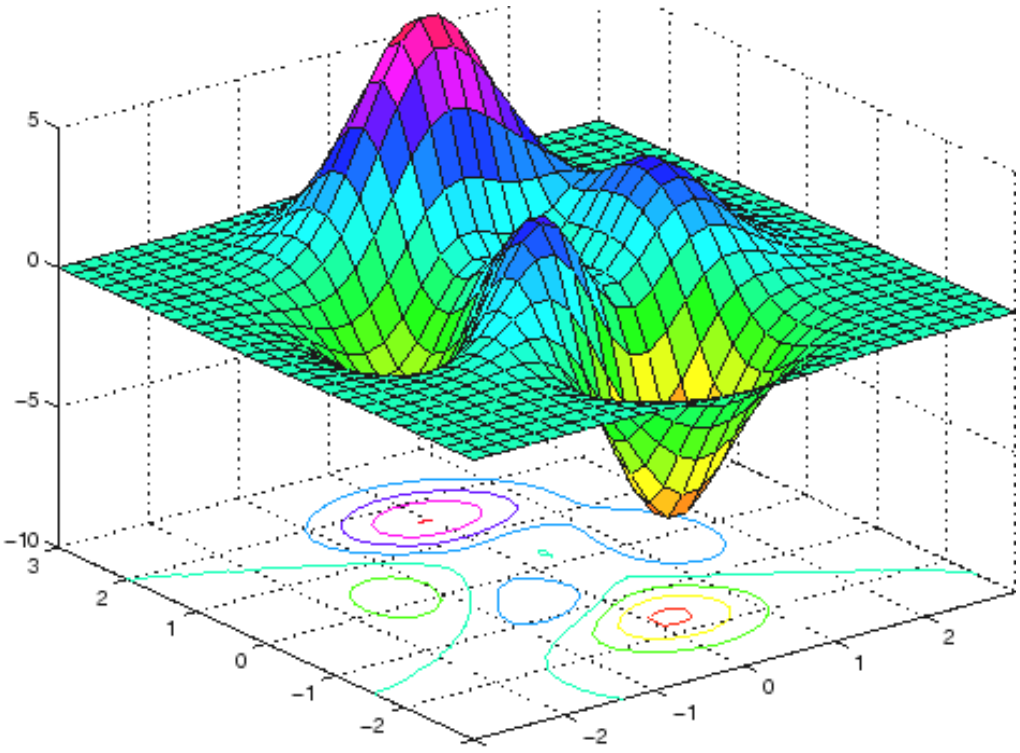


Рис. 19

Так для удовлетворения желания найти точку минимума $F(X)$ при отсутствии ограничений одна из простейших разновидностей градиентных методов – *метод наискорейшего спуска* предлагает выбрать некоторую точку (план) X_k и начальный шаг H_k, вычислить градиент функции в выбранной точке $\text{grad}\, F(X_k)$ и осуществить переход в направлении, обратном градиенту, с выбранным шагом. Если значение функции в новой точке меньше предыдущего, то новая точка принимается за исходную и повторяется аналогичное действие. При попадании в точку с бóльшим значением шаг уменьшается (например, вдвое) и переход повторяется от предыдущей точки. Такие действия продолжаются до получения достаточно малого шага.

Существуют и более эффективные переходы по градиенту, связанные с выбором различного шага по разным координатам или с автоматическим определением шага (при каждом переходе решается задача поиска экстремума функции в заданном направлении). Но гарантии достичь именно глобальный экстремум нет (при разных начальных данных для многоэкстремальных функций получаем разные решения).

Градиентные методы для решения задач с ограничениями, где при смещениях по градиенту приходится сталкиваться с опасно-

стью «выскочить» за пределы допустимого множества решений, существенно усложняются (модифицированный метод Ньютона, методы возможных направлений Зойтендейка, сопряженных градиентов, проектируемых градиентов Розена и др.).

Существует обширная литература по численному анализу, где значительное внимание уделяется градиентным и другим итерационным методам, но тем не менее решение нелинейных задач оптимизации при наличии ограничений почти всегда затруднительно.

2. *Методы Монте – Карло*. Здесь отыскивается n-мерный параллелепипед $\{a_i \le x_i \le b_i,\ i = 1, 2, \ldots, n\}$, включающий в себя множество планов, и затем моделируются координаты N случайных точек (рис. 20) с равномерным законом распределения в параллелепипеде (практически во всех программных средах предусмотрено наличие таких датчиков псевдослучайных чисел в интервале (0,1)). Для точек, попавших во множество планов, вычисляются значения функции и запоминается точка текущего экстремума. Для достижения бóльшей точности затем берется параллелепипед меньших размеров с центром в найденной точке и в нем вновь моделируются N случайных точек. Процесс такого *стохастического моделирования* заканчивается при малых размерах параллелепипеда.

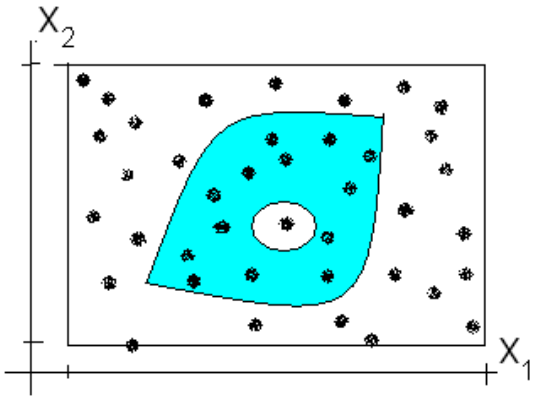


Рис. 20

Для отыскания приемлемого значения N можно воспользоваться законом больших чисел и прийти к оценке $N \sim 1 / (4\,\varepsilon^2\,\delta)$, где ε – допустимая абсолютная погрешность, δ – вероятность ошибки. Если задаться погрешностью 0,001 и вероятностью ошибки 0,01, то получаем «фантастическое» значение $N = 25 \cdot 10^6$! Однако методы Монте – Карло имеют то преимущество над моделированием на детерминированной сетке, что точность получаемых оценок имеет порядок $1 / \sqrt{N}$ и не зависит от размерности задачи (при $n > 3$ их эффективность несомненна). Естественно, эти методы никто не применяет при ручном счете, но они просты для программной реализации и часто используются при поиске начального приближения для градиентных методов.

3. *Методы динамического программирования*, сводящие мно-

гомерную задачу оптимизации к последовательности задач меньшей размерности. Их применение особенно успешно в случае *сепарабельных функций*, т. е. функций, представимых суммой функций одной переменной $F(X_1, X_2, \ldots, X_n) = f_1(X_1) + f_2(X_2) + \ldots + f_n(X_n)$.

4. *Методы выпуклого программирования*, реализующие поиск минимума выпуклой (максимума вогнутой) функции на выпуклом множестве планов. Если множество планов – выпуклый многогранник, то эти методы допускают использование симплексного метода. В основе этих методов лежат так называемые двойственные оценки и теорема Куна – Такера (понятие о двойственности в линейном программировании – частный случай).

5.2. Дробно-линейное программирование

Пусть стоит задача максимизации дробно-линейной функции

$$Q(X) = \frac{C_0 + \sum_{j=1}^{n} C_j X_j}{D_0 + \sum_{j=1}^{n} D_j X_j} \tag{1}$$

при линейных ограничениях

$$\sum_{j=1}^{n} A_{ij} X_j = B_i \ (i = 1 .. m); \tag{2}$$

$$X_j \geq 0 \ (j = 1 .. n). \tag{3}$$

Предположим, что знаменатель в (1) положителен при всех X, удовлетворяющих (2) – (3). Тогда при вводе обозначений

$$D_0 + \sum_{j=1}^{n} D_j X_j = \frac{1}{R} \ (R > 0), \ Z_j = R \cdot X_j, \ j = 1 .. n \tag{4}$$

задача сведется к линейной программе максимизации

$$Q(Z, R) = C_0 R + \sum_{j=1}^{n} C_j Z_j \tag{5}$$

при условиях

$$D_0 R + \sum_{j=1}^{n} D_j Z_j = 1; \tag{6}$$

$$-B_i R + \sum_{j=1}^{n} A_{ij} Z_j = 0 \ , \ i = 1 .. m; \tag{7}$$

$$R > 0,\ Z_j \geq 0,\ (j = 1\ ..\ n). \qquad (8)$$

Так задача максимизации дробно-линейной функции

$$\frac{-3 + 2\,X_1 + 4\,X_2 - 5\,X_3}{5 + 3\,X_1 - X_2}$$

при условиях

$$X_1 - X_2 \geq 0;$$

$$5\,X_1 + 3\,X_2 + 10\,X_3 \leq 15;$$

$$X_1,\ X_2,\ X_3 \geq 0.$$

с учетом (4) преобразуется в задачу максимизации линейной функции

$$-3\,R + 2\,Z_1 + 4\,Z_2 - 5\,Z_3$$

при условиях (линейных ограничениях)

$$5\,R + 3\,Z_1 - Z_2 = 1;$$

$$Z_1 - Z_2 \geq 0;$$

$$-15\,R + 5\,Z_1 + 3\,Z_2 + 10\,Z_3 \leq 0;$$

$$R > 0,\ Z_1,\ Z_2,\ Z_3 \geq 0.$$

C баз	Базис	План Z	–3	2	4	–5	0	0	$-M$	$-M$
			R	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7
$-M$	Z_6	1	5	3	–1	0	0	0	1	0
$-M$	Z_7	0	0	1	–1	0	–1	0	0	1
0	Z_5	0	–15	5	3	10	0	1	0	0
Δ_k		$-M$	$-5M$	$-3M$	$2M$	5	M	0	0	0

C баз	Базис	План Z	–3	2	4	–5	0	0	$-M$
			R	Z_1	Z_2	Z_3	Z_4	Z_5	Z_7
–3	R	1/5	1	3/5	–1/5	0	0	0	0
$-M$	Z_7	0	0	1	–1	0	–1	0	1
0	Z_5	3	0	14	0	10	0	1	0
Δ_k		–3/5	0	$-M$	M+	5	M	0	0

C баз	Базис	План Z	–3	2	4	–5	0	0
			R	Z_1	Z_2	Z_3	Z_4	Z_5
–3	R	1/5	1	0	2/5	0	3/5	0
2	Z_1	0	0	1	–1	0	–1	0
0	Z_5	3	0	0	14	10	14	1
Δ_k		–3/5	0	0	–7,2	5	–3,8	0

C баз	Базис	План Z	–3	2	4	–5	0	0
			R	Z_1	Z_2	Z_3	Z_4	Z_5
–3	R	4/35	1	0	0	–2/7	1/5	–1/35
2	Z_1	3/14	0	1	0	5/7	0	1/14
4	Z_2	3/14	0	0	1	5/7	1	1/14
Δ_k		–3/5	33/35	0	0	0	1/7	17/5

Отсюда $Z_1 = Z_2 = 3/14$, $Z_3 = 0$, $R = 4/35$ и соответственно $X_{\text{opt}} = (15/8, 15/8, 0)$, $\max Q(X) = 33/35$.

Следует заметить, что дробно-линейные программы представляют достаточно большой интерес для приложений, например, некорректная задача, преследующая взаимно противоположные цели типа максимизации прибыли при минимальных капиталовложениях, становится корректной, если поставить целью максимум прибыли на единицу затрат или минимум затрат на единицу прибыли.

5.3. Метод множителей Лагранжа

Приводимый здесь метод был предложен еще в 1797 г. Ж. Лагранжем для задачи поиска экстремумов функции $F(X)$ при условиях $f_i(X) = 0$ ($i = 1 .. m$).

Функция $\Phi(X, \lambda) = F(X) + \sum_{i=1}^{m} \lambda_i f_i(X)$ называется *функцией Лагранжа* и коэффициенты λ_i – *множителями Лагранжа*.

Можно доказать, что *необходимым условием существования экстремума исходной задачи является обращение в нуль всех частных производных функции Лагранжа*.

Так, при поиске экстремальных значений функции $F(X_1, X_2) = X_1 + X_2$ при единственном условии $X_1^2 + X_2^2 = 1$ строится функция Лагранжа

$$\Phi(X, \lambda) = X_1 + X_2 + \lambda (X_1^2 + X_2^2 - 1).$$

Затем строим систему уравнений

$$\frac{\partial \Phi}{\partial X_1} = 1 + 2\lambda X_1 = 0, \quad \frac{\partial \Phi}{\partial X_2} = 1 + 2\lambda X_2 = 0, \quad \frac{\partial \Phi}{\partial \lambda} = X_1^2 + X_2^2 - 1 = 0,$$

решение которой дает $\lambda = \pm 1/\sqrt{2}$, $X_1 = X_2 = -\lambda$ и экстремальные значения целевой функции $-\sqrt{2}$ и $\sqrt{2}$.

При поиске параметров цилиндрической бочки, обеспечивающих максимум ее объема $V = \pi R^2 H$ при фиксированной площади ее

оболочки $S = 2\,\pi\,R^2 + 2\,\pi\,R\,H$, берем функцию Лагранжа в виде $\Phi(R, H, \lambda) = \pi\,R^2\,H + \lambda\cdot[2\,\pi\,R^2 + 2\,\pi\,R\,H - S]$ и приходим к системе

$$\frac{\partial\Phi}{\partial R} = 2\pi RH + \lambda\cdot 2\pi(2R+H) = 0;\ \frac{\partial\Phi}{\partial H} = \pi R^2 + \lambda\cdot 2\pi R = 0;$$

$$\frac{\partial\Phi}{\partial\lambda} = 2\,\pi R^2 + 2\pi RH - \mathrm{S} = 0,$$

решение которой дает $\lambda = -R/2$ и искомые оценки $H = 2\,R$, $R = \sqrt{S/6\pi}$.

Для определения типа найденного экстремума (*правило Сильвестра*) можно построить матрицу вторых производных $F(X)$, вычисленных в экстремальной точке, и определить знаки главных ее миноров. Если все они положительны, то найден минимум, если они чередуются, начиная с минуса, то найден максимум.

Простота метода множителей Лагранжа является кажущейся, поскольку он обычно приводит к нелинейной системе уравнений и не гарантирует тип отыскиваемого экстремума, кроме глобальных дает и множество локальных экстремумов. Тем не менее, он полезен как база генерации идей и методов нелинейного программирования.

5.4. Теорема Куна – Такера

Пусть стоит задача минимизации $F(X)$ при условиях:

$$f_i(X) \le 0\ (i = 1\ ..\ m); \qquad (1)$$

$$X \ge 0, \qquad (2)$$

где X – n-мерный вектор, $F(X)$ и $f_i(X)$ – выпуклые функции, что обеспечивает выпуклость множества планов и единственность искомого минимума (рис. 21). Напомним, что мы считаем функцию выпуклой в некоторой точке, если главные миноры матрицы вторых производных положительны.

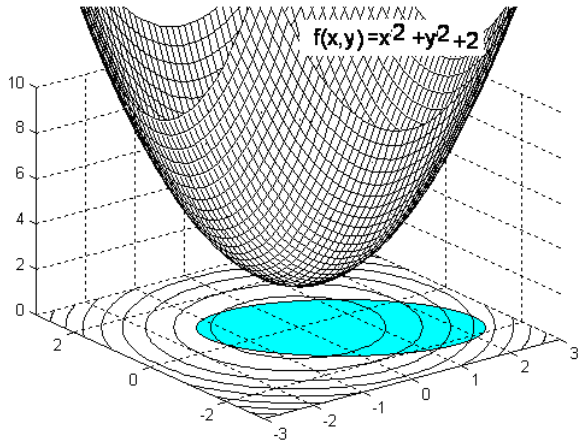


Рис. 21

Введем функцию Лагранжа

$$\Phi(X,\lambda) = F(X) + \sum_{i=1}^{m}\lambda_i\, f_i(X). \qquad (3)$$

Теорема Куна – Такера утверждает, что *вектор* $X^* \geq 0$ *является решением поставленной задачи тогда и только тогда, когда существует вектор* $\lambda^* \geq 0$ *такой, что* при всех $X \geq 0$, $\lambda \geq 0$

$$\Phi(X^*, \lambda) \leq \Phi(X^*, \lambda^*) \leq \Phi(X, \lambda^*). \qquad (4)$$

Так как функция Лагранжа в точке (X^*, λ^*) принимает минимальное значение по X и максимум по λ (рис. 22), эта точка называется *седловой* и теорему называют *теоремой о седловой точке* или *теоремой о минимаксе*[14].

Достаточность условий этой теоремы доказывается сравнительно просто. Доказательство их необходимости предполагает выполнение *условий регулярности*, т. е. существования хотя бы одной допустимой точки X, где $f_i(X) < 0$ при всех i.

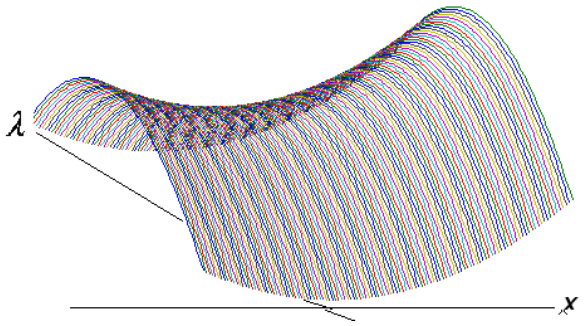


Рис. 22

Если $F(X)$ и $f_i(X)$ дифференцируемы, то условия теоремы эквивалентны «локальным» условиям, утверждающим, что в точке (X^*, λ^*)

$$\frac{\partial \Phi}{\partial X} \geq 0, \; X^T \frac{\partial \Phi}{\partial X} = 0, \; X \geq 0, \qquad (5)$$

$$\frac{\partial \Phi}{\partial \lambda} \leq 0, \; \lambda^T \frac{\partial \Phi}{\partial \lambda} = 0, \; \lambda \geq 0. \qquad (6)$$

Остановимся на частных случаях задачи. Если в поставленной задаче отсутствуют требования неотрицательности (2), то заменой X разностью двух неотрицательных векторов можно показать, что условие (5) упрощается к виду $\frac{\partial \Phi}{\partial X} = 0$ (если нет требования неотрицательности для одной из компонент X, то обратится в нуль производная $\Phi(X)$ по соответствующей переменной).

Если $f_i(X) = 0$ при каком-то i, то заменой этого равенства системой неравенств $f_i(X) \leq 0$, $-f_i(X) \leq 0$ обнаруживаем, что в (6) соот-

[14] Работа над нелинейным программированием и теорией двойственности была начата Алом Такером и его студентами Х. Куном и Д. Джейлом в 1948 г., но, по утверждению Д. Данцига, создателем теоремы о двойственности является Джон фон Нейман, а Такера и его коллег ценят как авторов первого строгого доказательства.

ветствующая производная обращается в нуль и исчезает условие неотрицательности по λ_i, если $f_i(X) = 0$ при всех i, то (6) упрощается к виду $\frac{\partial \Phi}{\partial \lambda} = 0$.

Как предельный случай условий Куна – Такера могут быть построены двойственные задачи линейного программирования. Так при минимизации $C^T X$ при условиях $A X \geq B, X \geq 0$ функция Лагранжа имеет вид

$$\Phi(X, \lambda) = C^T X + \lambda^T (A X - B)$$

и условия Куна – Такера

$$\frac{\partial \Phi}{\partial \lambda} = C - A^T \lambda \geq 0, \; X^T \frac{\partial \Phi}{\partial X} = X^T (C - A^T \lambda) = 0, \; X \geq 0,$$

$$\frac{\partial \Phi}{\partial \lambda} = B - AX \leq 0, \; \lambda^T \frac{\partial \Phi}{\partial \lambda} = \lambda^T (B - AX) = 0, \; \lambda \geq 0.$$

Отсюда с учетом $X^T C = C^T X$, $\lambda^T B = B^T \lambda$, $X^T A^T \lambda = \lambda^T A X$ получаем, что в седловой точке достигается минимум по X и максимум по λ, причем $C^T X = B^T \lambda$ и $A X \geq B$, $A^T \lambda \leq C$, $X \geq 0$, $\lambda \geq 0$.

5.5. Квадратичное программирование и метод Вулфа – Фрэнка

Рассмотрим задачу минимизации квадратичной функции n переменных

$$F(X) = C^T X + X^T D X \tag{1}$$

при линейных ограничениях

$$A X \leq B; X \geq 0, \tag{2}$$

где A – матрица размерности $m \times n$; C, X – n-мерные векторы; B – m-мерный вектор; D – положительно определенная n-мерная квадратная матрица[15].

Так, например, целевая функция

$F(X) = 7 X_1 - 3 X_2 + 22 X_1^2 - 8 X_1 X_2 + X_2^2 - 10 X_1 X_3 + 12 X_2 X_3 + 40 X_3^2$

представится в виде (1), где

$$C = \begin{Vmatrix} 7 \\ -3 \\ 0 \end{Vmatrix}, \; D = \begin{Vmatrix} 22 & -4 & -5 \\ -4 & 1 & 6 \\ -5 & 6 & 40 \end{Vmatrix}.$$

Положительная определенность D и линейность ограничений (выпуклое множество планов) позволяют использовать теорему Куна – Такера.

---

[15] Матрица называется положительно определенной, если положительны ее главные миноры.

Функция Лагранжа здесь имеет вид

$$\Phi(X, \lambda) = C^T X + X^T DX + \lambda^T (A X - B), \quad (3)$$

и условия Куна – Такера приводятся к форме:

$$\frac{\partial \Phi}{\partial X} = C + 2 D X + A^T \lambda \geq 0, \; X^T \frac{\partial \Phi}{\partial X} = 0, \; X \geq 0; \quad (4)$$

$$\frac{\partial \Phi}{\partial X} = A X - B \leq 0, \; \lambda^T \frac{\partial \Phi}{\partial X} = 0, \; \lambda \geq 0. \quad (5)$$

Если ввести векторы ослабляющих переменных, то (4) – (5) примут вид

$$C + 2 D X + A^T \lambda - V = 0, \; X^T V = 0, \; X \geq 0, \; V \geq 0;$$

$$A X - B + Y = 0, \; -\lambda^T Y = 0, \; \lambda \geq 0, \; Y \geq 0.$$

С учетом неотрицательности переменных можно поставить эквивалентную задачу *минимизации* (*до нуля*) функции

$$g(X, \lambda, Y, V) = X^T V + \lambda^T Y \quad (6)$$

при условиях

$$\begin{gathered} 2 D X + A^T \lambda - V = -C; \\ A X + Y = B; \\ X, \lambda, Y, V \geq 0. \end{gathered} \quad (7)$$

Обозначив совокупный вектор переменных как

$$W = (X^T, \lambda^T, Y^T, V^T)^T, \quad (8)$$

(7) – (9) можно записать в виде:

$$R W = S, \; W \geq 0, \quad (9)$$

где

$$R = \begin{Vmatrix} 2D & A^T & 0 & -E \\ A & 0 & E & 0 \end{Vmatrix}, \; S = \begin{Vmatrix} -C \\ B \end{Vmatrix}. \quad (10)$$

Таким образом, метод Вулфа – Фрэнка сводит решение задачи к форме, допускающей применение симплексной процедуры.

Здесь находится некоторый опорный план W^0. Если $g(W^0) = 0$, то этот план оптимален. В противном случае отыскивается градиент

$$\text{grad}\, g(W^0) = (V^T, Y^T, \lambda^T, X^T) \quad (11)$$

и его компоненты на один шаг симплексного преобразования (перехода к другому опорному плану) принимаются за коэффициенты «линейной целевой функции». После выбора нового опорного плана выполняются вышеописанные рассуждения.

Пример. Минимизировать

$$F(X_1, X_2) = -4 X_1 - 6 X_2 + X_1^2 + 3 X_2^2$$

при условиях

$$2X_1 + X_2 \le 4;$$

$$X_1, X_2 \ge 0.$$

Ставим задачу минимизации до нуля функции

$$g(X_1, X_2, \lambda, Y, V_1, V_2) = X_1 V_1 + X_2 V_2 + \lambda Y$$

при условиях (9), где

$$R = \begin{array}{|cccccc|}\hline 2 & 0 & 2 & 0 & -1 & 0 \\ 0 & 6 & 1 & 0 & 0 & -1 \\ \hline 2 & 1 & 0 & 1 & 0 & 0 \\ \hline \end{array} \qquad S = \begin{array}{|c|}\hline 4 \\ 6 \\ \hline 4 \\ \hline \end{array}$$

Ищем начальный опорный план методом искусственного базиса (не обращая внимания на коэффициенты, меньшие M):

C баз	Базис	План W						M u_1	M u_2
			X_1	X_2	λ	Y	V_1	V_2		
M	u_1	4	2	0	2	0	-1	0	1	0
M	u_2	6	0	6	1	0	0	-1	0	1
	Y	4	2	1	0	1	0	0	0	0
$\Delta_k = g(W)$		$10M$	$2M$	$6M$	$3M$	0	$-M$	$-M$	0	0

C баз	Базис	План W						M u_1
			X_1	X_2	λ	Y	V_1	V_2	
M	λ	2	1	0	1	0	$-1/2$	0	1/2
	u_2	4	-1	6	0	0	1/2	-1	$-1/2$
	Y	4	2	1	0	1	0	0	0
$\Delta_k = g(W)$		M	$-M$	$6M$	0	0	$M/2$	$-M$	0

Изгнав искусственные переменные, получаем начальный опорный план $X_1 = X_2 = V_2 = 0$, $V_1 = 8$, $Y = 4$, $\lambda = 6$, для которого $g(W) = 0 \times 8 + 0 \times 0 + 6 \times 8 = 48 > 0$.

Находим градиент grad $g(W) = \{8, 0, 4, 6, 0, 0\}$ и берем его компоненты в качестве коэффициентов «линейной целевой функции» (нормали к поверхности, определяемой функцией $g(W)$, в выбранной точке).

C баз	Базис	План W	8	0	4	6	0	0
			X_1	X_2	λ	Y	V_1	V_2
4	λ	6	0	6	1	0	0	-1
0	V_1	8	-2	12	0	0	1	-2
6	Y	4	2	1	0	1	0	0
$\Delta_k = g(W)$		48	4	30	0	0	0	-4

C баз	Базис	План W	12	0	0	6	2	0
			X_1	X_2	λ	Y	V_1	V_2
0	λ	6	0	6	1	0	0	–1
2	V_1	12	0	13	0	1	1	–2
12	X_1	2	1	1/2	0	1/2	0	0
$\Delta_k = g(W)$		48	0	32	0	2	0	–2

C баз	Базис	План W	0	0	0	7/13	20/13	12/13
			X_1	X_2	λ	Y	V_1	V_2
0	λ	**7/13**	0	0	1	–6/13	–6/13	–1/13
0	X_2	**12/13**	0	1	0	1/13	1/13	–2/13
0	X_1	**20/13**	1	0	0	6/13	–1/26	1/13
$\Delta_k = g(W)$		0						

Получен оптимальный план X = (20/13, 12/13) – точка, лежащая на границе множества планов.

Если отвлечься от ограничений, то минимум рассматриваемой функции $F(X_1, X_2) = -4\,X_1 - 6\,X_2 + X_1^2 + 3\,X_2^2$ достигается в точке (2, 1) за пределами множества планов.

На приведенном ниже рис. 23, полученном с помощью процедуры `contour` в среде MatLab, представлены линии уровня функции $F(X_1, X_2)$ = const и наглядно видно взаимное расположение точек минимума с учетом ограничений и без такового.

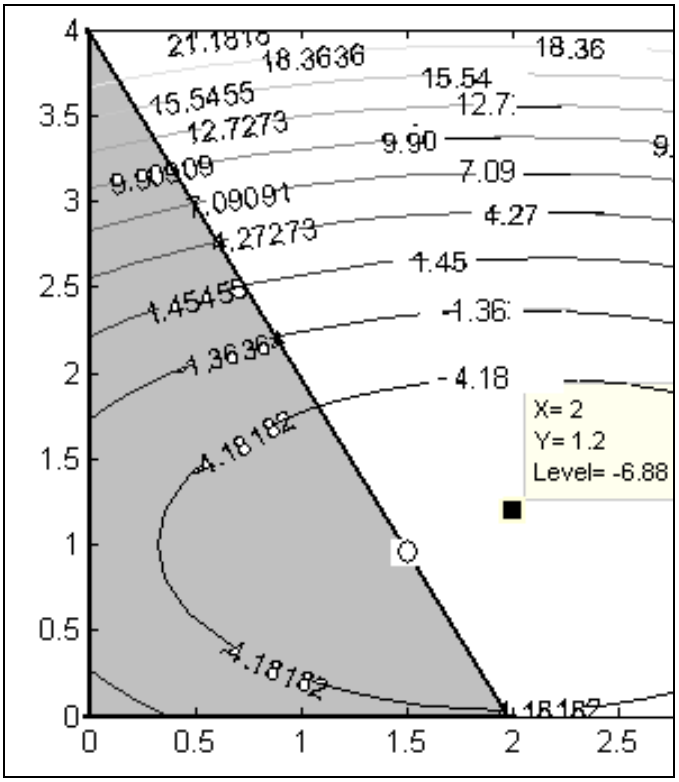


Рис. 23

Заметим, что при ограничениях $A\,X = B$, $X \geq 0$ исчезнет переменная Y, снимутся условия неотрицательности на λ и потребуется минимизировать $X^T V$ при условиях:

$$2\,D\,X + A^T\lambda - V = -C;$$

$$A\,X = B;$$

$$X, V \geq 0.$$

При других отклонениях постановки задачи от выбранного здесь стандарта (1) – (2) можно элементарными приемами прийти к нему (вместо максимизации $F(X)$ искать минимум $-F(X)$, условие $A\,X \geq B$ заменить на $B - A\,X \leq 0$ и т. п.).

6. ВВЕДЕНИЕ В ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

6.1. Многошаговые процессы принятия решений

В начале шахматной партии игрок делает ход, определяемый некоторым выбором из 20 возможных, следующий его ход определяется ответным ходом партнера и возникшей совокупностью возможных выборов и т. д. Как найти последовательность выборов, приводящую к достижению успеха? Перебор всех вариантов, которые могут возникнуть в процессе игры, невозможен даже для суперкомпьютера.

Жизнь человечества полна примерами принятия решений, оптимальных с точки зрения текущего момента, но с плачевным результатом через какое-то время.

«Хозяин» лесохозяйственного предприятия, ставящий целью получение максимума прибыли в ближайшем столетии, едва ли будет руководствоваться на каждом этапе деятельности примитивным принципом «руби побольше» (очень скоро рубить будет нечего).

Автоматическое управление ракетой-перехватчиком на очередном участке ее траектории не повторяет управляющих характеристик предыдущего типа, хотя конечная цель остается неизменной.

Единственно разумно требовать от человека поэтапного (пошагового) принятия решений, не забывая о необходимости достижения некоторой привлекательной конечной цели.

Не возникает сомнений в разумности использования для управления сложными системами методов математического моделирования, которое подчас дает решения, неожиданные не только для начинающего, но и для опытного управляющего.

Исчерпывающее математическое моделирование действительности невозможно, ибо, как писал Козьма Прутков, невозможно «объять необъятное». Как правило, отсутствует глубокое знание законов развития природы и общества, отсутствует достоверная информация об изучаемой системе, вмешивается субъективная оценка последствий решения и др.

Естественно, что, создавая математическую модель управления сложной системой, разработчик не может ограничиться рассмотрением только «верхушки айсберга», но и не должен погрязнуть в деталях непринципиального характера. Последствия переупрощения очевидны, учет же большого количества факторов при-

водит к известному «проклятию размерности», делающему задачу выбора «оптимальной» политики практически неразрешимой.

Для иллюстрации математических проблем, возникающих при исследовании многошаговых процессов принятия решений, рассмотрим следующую идеализированную задачу [7].

6.2. Многошаговый процесс распределения однородного ресурса

Пусть имеется X денежных единиц, часть которых Y используется для вложений в сферу A, а оставшаяся часть $X - Y$ – в сферу B. В течение некоторого периода эти вложения дают доход, определяемый значениями некоторых функций $g(Y)$ и $h(X - Y)$ соответственно, который идет на удовлетворение каких-либо личных потребностей. По истечении периода вложения изымаются $\alpha Y + \beta (X - Y)$ денежных единиц (как правило, $\alpha, \beta < 1$), которые в очередном периоде используются для вложений в те же сферы.

Требуется найти политику разделения денежного ресурса, дающую максимум дохода за N периодов.

В случае одношагового процесса (при $N = 1$) доход, зависящий от начального ресурса X и выбора доли $Y \in [0, X]$, равен

$$R_1(X, Y) = g(Y) + h(X - Y).$$

Столь же очевидно, что максимальный доход в одношаговом процессе при начальном ресурсе X равен

$$\max_{0 \le Y \le X} [g(Y) + h(X - Y)]$$

(с такой ситуацией столкнемся, например, достигнув последнего периода, когда достаточно принимать решение лишь на шаг вперед).

В случае двух периодов ($N = 2$) суммарный доход зависит от начального ресурса X, выборов в первом и втором периодах Y и Y_1:

$$R_2(X, Y, Y_1) = [g(Y) + h(X - Y)] + [g(Y_1) + h(X_1 - Y_1)],$$

где

$$X_1 = \alpha Y + \beta (X - Y);\ 0 \le Y \le X;\ 0 \le Y_1 \le X_1.$$

Аналогично при любом $N > 1$ доход за N периодов складывается из дохода в первом периоде и суммарного дохода в *последующих* $N - 1$ периодах

$$R_N(X, Y, Y_1, Y_2, \ldots) = g(Y) + h(X - Y) + \sum_{k=1}^{N-1} \left[g(Y_k) + h(X_k - Y_k) \right],$$

где

$$X_1 = \alpha\, Y + \beta\,(X - Y);\ X_k = \alpha\, Y_{k-1} + \beta\,(X_{k-1} - Y_{k-1}),\ k = 2\ ..\ N - 1;$$

$$0 \le Y \le X,\ 0 \le Y_1 \le X_1,\ 0 \le Y_2 \le X_2,\ \ldots\ .$$

Следовательно, задача поиска максимального дохода в N-шаговом процессе при начальном денежном ресурсе X сводится к максимизации функции N неизвестных Y, Y_1, Y_2, … при указанных ограничениях, т. е. к задаче математического программирования.

Если функции g и h линейны, то имеем дело с задачей линейного программирования и нет принципиальных преград для ее решения при *конкретном* значении X. Если же они нелинейны, то при $N > 2$ возникает уже упоминавшееся *проклятие размерности*.

А как быть в случае, когда величина исходного ресурса заранее точно неизвестна и лишь имеется информация о диапазоне возможных ее значений? Ведь здесь даже при линейных функциях придется либо решать очень много линейных программ, либо решать достаточно сложную, нестандартную параметрическую линейную программу.

6.3. Принцип оптимальности и рекуррентные соотношения

Один из путей изучения многошаговых процессов связан с использованием интуитивного принципа оптимальности, сформулированного в 1957 году выдающимся американским математиком Р. Беллманом[16] в книге «Динамическое программирование» [7] и определяющего фундаментальное свойство оптимальной стратегии (политики, поведения).

«*Оптимальное поведение обладает тем свойством, что, каковы бы ни были первоначальное состояние и решение в начальный момент, последующие решения должны составлять оптимальное поведение относительно состояния, полученного в результате начального решения*».

---

[16] Ричард Эрнст Беллман (1920 – 1984) – американский математик, один из ведущих специалистов в области прикладной математики и вычислительной техники, автор 619 статей и 39 книг. Получил многочисленные результаты, связанные с применением *динамического программирования* в разных областях математики (вариационное исчисление, автоматическое регулирование, теория аппроксимации, исследование операций и др.).

Это *рекурсивное* определение можно интерпретировать самым примитивным высказыванием: если вы намерены добиться наилучшего эффекта вашей многолетней деятельности, то на любом ее этапе, независимо от того, в каких состояниях вы оказывались ранее и какие действия предпринимали, действуйте оптимально (с точки зрения конечной цели), насколько вам позволяет то состояние, в которое вы загнали себя предыдущими действиями. Если на любом этапе своей целенаправленной деятельности руководствоваться стремлением к достижению некоторой высшей цели, то вся последовательность действий будет оптимальна. Заметьте, что принцип оптимальности определяет особенности политики, направленной не на получение сиюминутной выгоды, а на достижение некоторой удаленной цели.

Р. Беллман

Обратимся к поставленной выше задаче поэтапного распределения ресурса, где оптимальность политики определяется достижением максимума суммарного дохода. Обозначим через $F_k(Z)$ *суммарный доход в k-шаговом процессе при начальном ресурсе Z и использовании оптимальной политики* (максимальный суммарный доход за k шагов при начальном ресурсе Z). Как мы уже отмечали ранее, доход в k-шаговом процессе можно представить суммой дохода на первом шаге и дохода на последующих $k-1$ шагах (из всех возможных представлений нам нравится именно такое!):

$$R_N(X, Y, Y_1, Y_2, \ldots) = g(Y) + h(X - Y) + \sum_{k=1}^{N-1}\left[g(Y_k) + h(X_k - Y_k)\right].$$

Если после первого шага вспомнить о принципе оптимальности и необходимости дальнейшей оптимальной политики, то доход в k-шаговом процессе сложится из *дохода на первом шаге* $g(Y) + h(X - Y)$ и *максимального дохода на последующих* $k-1$ *шагах*, которые начнутся уже при денежном ресурсе $\alpha Y + \beta (X - Y)$:

$$\left[g(Y) + h(X - Y) + F_{k-1}(\alpha Y + \beta(X - Y))\right]$$

(здесь мы уже предполагаем оптимальность последующих выборов $Y_1, Y_2, \ldots$).

Если подобрать величину Y (выбор на первом шаге при ресурсе X) так, чтобы эта сумма была максимальна, то мы получим оценку максимального дохода в k-шаговом процессе с начальным ресурсом X (!):

$$F_k(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X - Y) + F_{k-1}(\alpha Y + \beta(X - Y)) \right] \qquad (1)$$

(такое соотношение справедливо при любом $k \ge 2$).

Если учесть, что максимальный доход в одношаговом процессе при начальном денежном ресурсе X равен

$$F_1(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X - Y) \right], \qquad (2)$$

напрашивается последовательность решения поставленной задачи: найти функцию $F_1(X)$ и затем на основе *рекуррентных* соотношений (1) – функций $F_2(X)$, …, $F_N(X)$.

В дополнение к введенным ранее обозначениям $F_k(Z)$ будем обозначать через $Y_k(Z)$ – *оптимальный выбор на первом шаге k-шагового процесса с начальным ресурсом Z*.

Кстати, если в условиях поставленной задачи предположить, что доходы, получаемые на шагах процесса, также используются для вложений в упомянутые сферы, то при $k \ge 2$ (1) преобразуется к виду

$$F_k(X) = \max_{0 \le Y \le X} F_{k-1}\left[g(Y) + h(X - Y) + (\alpha Y + \beta(X - Y)) \right].$$

6.4. Структура решения

Как уже было сказано, решение задачи для N-шагового процесса разделения ресурса начинаем поиском функции $F_1(X)$ – максимального дохода в одношаговом процессе при начальном денежном ресурсе X и соответствующем значении Y, обеспечивающих максимум, т. е. функции $Y_1(X)$ – оптимального выбора для одношагового процесса с начальным ресурсом X. Тем самым мы узнаем, как следует действовать, если длительность предстоящего процесса составит только один шаг и начальный ресурс для него окажется равным X.

На следующем этапе решения, обладая информацией о функции F_1 при произвольном аргументе, выясняем, каков же окажется максимальный доход в случае двухшагового процесса при произвольном начальном ресурсе X:

$$F_2(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X - Y) + F_1(\alpha Y + \beta(X - Y)) \right]$$

и $Y_2(X)$ – оптимальный выбор на первом шаге двухшагового процесса при начальном ресурсе X.

Затем аналогично можно получить оценки функций максимального дохода $[F_3(X), \ldots, F_N(X)]$ и поведений на первом шаге процесса соответствующей длительности $[Y_3(X), \ldots, Y_N(X)]$. Появляется возможность отыскания оптимальной политики при конкретном начальном ресурсе $X = Z$.

Оптимальный выбор на первом шаге N-шагового процесса при этом ресурсе равен $\hat{y}_1 = Y_N(Z)$, в результате чего к началу второго шага начальный ресурс будет равен $Z_1 = \alpha\, \hat{y}_1 + \beta\, (Z - \hat{y}_1)$. Следовательно, оптимальный выбор на втором шаге совпадет с оптимальным выбором на первом шаге оставшегося $(N-1)$-шагового процесса, т. е. $\hat{y}_2 = Y_{N-1}(Z_1)$, и к началу третьего шага ресурс станет равным $Z_2 = \alpha\, \hat{y}_2 + \beta\, (Z_1 - \hat{y}_2)$. Оптимальный выбор на третьем шаге совпадет с оптимальным выбором на первом шаге оставшегося $(N-2)$-шагового процесса, т. е. $Y_{N-2}(Z_2)$ и т. д. Оптимальный выбор на последнем шаге будет определяться значением функции $Y_1(Z_{N-1})$ оптимального выбора для одношагового процесса.

6.5. Простейший случай: выпуклые и линейные функции

Пусть функции $g(X)$ и $h(X)$ являются выпуклыми (выпуклыми вниз, $g''(X) \geq 0$ и $h''(X) \geq 0$).

При поиске $F_1(X) = \max\limits_{0 \leq Y \leq X} \left[g(Y) + h(X - Y) \right]$ имеем дело с максимумом суммы выпуклых функций, которая является выпуклой функцией. Очевидно, что максимум выпуклой функции в интервале достигается на одном из концов интервала, и если допустить, что $g(0) = h(0) = 0$, то $F_1(X) = \max\{h(X), g(X)\}$ и $Y_1(X)$ равно 0 или X соответственно.

Так как $F_1(X)$ выпукла, то при поиске

$$F_2(X) = \max_{0 \leq Y \leq X} \left[g(Y) + h(X - Y) + F_1(\alpha Y + \beta(X - Y)) \right]$$

мы вновь ищем максимум суммы выпуклых функций

$$F_2(X) = \max\left\{ h(X) + F_1(\beta X),\ g(X) + F_1(\alpha X) \right\}$$

и оптимальный выбор на первом шаге двухшагового процесса с начальным ресурсом X, т. е. $Y_2(X)$, равен 0 или X.

По индукции можно показать, что при любом $n > 1$

$$F_n(X) = \max\left\{ h(X) + F_{n-1}(\beta X),\ g(X) + F_{n-1}(\alpha X) \right\}$$

и $Y_n(X)$ равно 0 или X соответственно.

Пример. Пусть $g(X) = 9 \cdot X$, $h(X) = 5 \cdot X$, $\alpha = 0{,}4$, $\beta = 0{,}75$, $N = 4$. Так как функции линейны (частный случай выпуклости), то при $X > 0$ $F_1(X) = \max\{5\,X, 9\,X\} = 9\,X$; $Y_1(X) = X$.

Далее $F_2(X) = \max\{5\,X + F_1(0{,}75\,X), 9\,X + F_1(0{,}4\,X)\} =$

$= \max(5\,X + 9 \cdot 0{,}75 \cdot X, 9 \cdot X + 9 \cdot 0{,}4 \cdot X) =$

$= \max(11{,}7\,X; 12{,}6\,X) = 12{,}6\,X$; $Y_2(X) = X$;

$F_3(X) = \max\{5\,X + F_2(0{,}75\,X), 9\,X + F_2(0{,}4X)\} =$

$= \max(5\,X + 12{,}6 \cdot 0{,}75 \cdot X, 9\,X + 12{,}6 \cdot 0{,}4 \cdot X) =$

$= \max(14{,}45\,X; 14{,}04\,X) = 14{,}45\,X$; $Y_3(X) = 0$;

$F_4(X) = \max\{5\,X + F_3(0{,}75\,X), 9\,X + F_3(0{,}4X)\} =$

$= \max(5\,X + 14{,}45 \cdot 0{,}75 \cdot X, 9\,X + 14{,}45 \cdot 0{,}4 \cdot X) =$

$= \max(15{,}8375\,X; 14{,}78\,X) = 15{,}8375\,X$; $Y_4(X) = 0$.

Обратите внимание на то, что мы нашли не значения функций $F_k(X)$, $Y_k(X)$, а сами функции, что позволит нам выяснить оптимальную политику при любом начальном ресурсе Z.

Величина дохода в заданном 4-шаговом процессе равна $F_4(Z) = 15{,}8375\,Z$. Оптимальный выбор на первом шаге $\{\hat{y}_1\}$ совпадает с $Y_4(Z) = 0$, т. е. весь денежный ресурс вкладывается в оборудование типа B. Оптимальный выбор на втором шаге $\{\hat{y}_2\}$ совпадает с оптимальным выбором на первом шаге оставшегося трехшагового процесса с изменившимся начальным ресурсом

$$Z_1 = 0{,}4\,\hat{y}_1 + 0{,}75\,(Z - \hat{y}_1) = 0{,}75\,Z,$$

т. е. с $Y_3(Z_1) = 0$ (на втором шаге политика та же, что и на первом). Оптимальный выбор на третьем шаге $\{\hat{y}_3\}$ совпадает с оптимальным выбором на первом шаге оставшегося двухшагового процесса с начальным ресурсом

$$Z_2 = 0{,}4\,\hat{y}_2 + 0{,}75\,(Z_1 - \hat{y}_2) = 0{,}5625\,Z,$$

т. е. с $Y_2(Z_2) = Z_2$ (весь ресурс в сферу A). Наконец, на последнем (четвертом) шаге оптимальный выбор $\{\hat{y}_4\}$ совпадает с оптимальным выбором для оставшегося одношагового процесса с начальным ресурсом

$$Z_3 = 0{,}4\,\hat{y}_3 + 0{,}75\,(Z_2 - \hat{y}_3) = 0{,}225\,Z,$$

т. е. с $_1(Z_3) = Z_3$ (весь ресурс в сферу A).

Скептически настроенный читатель заметит, что рассмотренный пример можно было свести к элементарной задаче линейного программирования с 4 неизвестными:

максимизировать

$$L(Z) = 9\,(Y_1 + Y_2 + Y_3 + Y_4) + 5\,(Z - Y_1 + Z_1 - Y_2 + Z_2 - Y_3 + Z_3 - Y_4)$$

при условиях

$$0 \le Y_1 \le Z;\ 0 \le Y_2 \le Z_1;\ 0 \le Y_3 \le Z_2;\ 0 \le Y_4 \le Z_3;$$

$$Z_1 = 0{,}4\,Y_1 + 0{,}75\,(Z - Y_1);\ Z_2 = 0{,}4\,Y_2 + 0{,}75\,(Z_1 - Y_2);$$

$$Z_3 = 0{,}4\,Y_3 + 0{,}75\,(Z_2 - Y_3).$$

Но решение этой задачи, например, симплексным методом придется выполнять для конкретного значения Z, тогда как решение с помощью рекуррентных соотношений дает готовую политику для любого Z.

6.6. Эффективность метода динамического программирования

Из приведенных выше замечаний и решений можно сделать некоторые выводы относительно метода динамического программирования, базирующегося на принципе оптимальности и реализуемого, в частности, в форме метода рекуррентных соотношений.

1. *Метод динамического программирования позволяет свести N-мерную задачу оптимизации к совокупности задач меньшей размерности* (очевидно, что легче решить 20 одномерных задач оптимизации, чем одну 20-мерную).

2. *Метод ориентирован на решение не конкретной задачи, а целого класса подобных задач.*

3. Как будет видно из дальнейшего рассмотрения численных реализаций, *появление дополнительных ограничений подчас облегчает решение задачи за счет уменьшения объема перебора вариантов.*

Идеология, использованная при рассмотрении задачи разделения ресурса, может быть обобщена на случай любого многошагового процесса принятия решений.

Ниже мы рассмотрим несколько достаточно простых задач [7, 8] для иллюстрации метода динамического программирования (построение рекуррентных соотношений и тем самым сведение многомерной задачи оптимизации к последовательности задач меньшей размерности; аналитическое решение рекуррентных соотношений и выявление структуры полученного решения).

6.7. Задача складирования однородного продукта

Пусть имеется склад вместимости B с начальным запасом V некоторого продукта, цены на который подвержены сезонным изменениям, но мы тем не менее обладаем надежным прогнозом.

В начале i-го сезона часть сохраненного продукта Y_i можно продать по цене P_i и в конце сезона закупить X_i этого продукта по цене C_i. Образовавшийся на складе запас хранится до следующего сезона. Хотелось бы найти политику продажи-покупки, максимизирующую суммарный доход за N сезонов.

Если пренебречь затратами на хранение и сопутствующими факторами, задачу можно свести к максимизации целевой функции

$$\sum_{i=1}^{N} [\, P_i\, Y_i - C_i\, X_i\,]$$

при условиях

$$0 \le Y_1 \le V,\ X_1 \ge 0,\ V_1 = V - Y_1 + X_1 \le B;$$
$$0 \le Y_2 \le V_1,\ X_2 \ge 0,\ V_2 = V_1 - Y_2 + X_2 \le B;$$
$$0 \le Y_3 \le V_2,\ X_3 \ge 0,\ V_3 = V_2 - Y_3 + X_3 \le B,$$

т. е. к задаче линейного программирования с $2N$ неизвестными и $4N$ ограничениями, которую можно решить симплексным методом при фиксированных значениях N и V.

Рассмотрим задачу с других позиций, введя обозначения:
k – число предстоящих сезонов;
$F_k(V)$ – максимальный доход за k сезонов при начальном запасе V;
P_k, C_k – цены продажи-покупки в первом из очередных k сезонов;
Y, X – объемы продажи-покупки в этом сезоне;
$Y_k(V)$, $X_k(V)$ – оптимальные объемы продажи-покупки в этом сезоне (на первом шаге k-шагового процесса с начальным запасом V).

Для процесса длительностью в один этап (с таковым имеем дело, когда до конца процесса останется один шаг), если к его началу запас составит V единиц,

$$F_1(V) = \max\{P_1\, Y - C_1\, X\}, \qquad (1)$$

где область максимизации определяется условиями $0 \le Y \le V$; $X \ge 0$.

Очевидно, что максимум здесь достигается при $Y = V$ и $X = 0$ (естественно при завершении деятельности по купле-продаже продать весь запас и ничего не покупать), т. е.

$$F_1(V) = P_1\, V;\ Y_1(V) = V;\ X_1(V) = 0. \qquad (1а)$$

Обратимся к случаю k-шагового процесса при $k > 1$, повторив традиционные рассуждения. Доход в k-шаговом процессе складывается из дохода на первом шаге $\{P_k Y - C_k X\}$ и дохода на оставшихся $k - 1$ шагах. Если мы руководствуемся принципом оптимальности: независимо от начального состояния (запаса V) и начального поведения (объемов Y, X продажи-покупки на первом шаге), дальнейшая политика должна быть оптимальной, исходя из возникающего состояния (запаса $V - Y + X$). Тогда доход на оставшихся $k - 1$ шагах будет равен $F_{k-1}(V - Y + X)$ и с учетом желания действовать оптимально с первого же шага

$$F_k(V) = \max\{P_k Y - C_k X + F_{k-1}(V - Y + X)\},\ k = 2\ ..\ N, \qquad (2)$$

где область максимизации определяется условиями:

$$0 \le Y \le V;\ X \ge 0;\ V - Y + X \le B.$$

Легко видеть, что множество планов (область максимизации) при любом k – многоугольник (рис. 24) с координатами вершин, линейно зависящими от V.

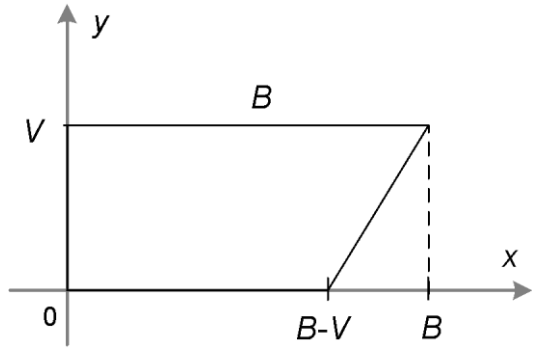


Рис. 24

Так как функция $F_1(V) = P_1 V$ линейна по V, то при поиске $F_2(V)$ мы сталкиваемся с максимизацией функции, линейной по X и Y, над указанным множеством планов, т. е. с задачей линейного программирования с экстремумами в вершинах множества. Можно показать, что аналогичное явление наблюдается при любом $k > 1$. Соответственно (2) упрощается к виду:

$$F_k(V) = \begin{cases} F_{k-1}(V); & Y_k(V) = 0;\ \ X_k(V) = 0 \\ P_k V + F_{k-1}(0); & Y_k(V) = V;\ \ X_k(V) = 0 \\ -C_k(B - V) + F_k(B); & Y_k(V) = 0;\ \ X_k(V) = B - V \\ P_k V - C_k B + F_{k-1}(B); & Y_k(V) = V;\ \ X_k(V) = B \end{cases}$$

Так исходная задача линейного программирования с $2N$ неизвестными сведена к N элементарным линейным программам с двумя неизвестными. Более того, здесь мы имеем решение при любых V и B.

Пример. Пусть $N = 5$ и

Сезон	1	2	3	4	5
Продажная цена (P)	7	6	4	4	8
Закупочная цена (C)	6	7	5	3	5

Решение задачи дает:

$$F_1(V) = 8\,V;\ Y_1(V) = V;\ X_1(V) = 0;$$

$$F_2(V) = \max\begin{cases} F_1(V) = 8V \\ 4V + F_1(0) = 4V \\ -3(B-V) + F_1(B) = 3V + 5B \\ 4V - 3B + F_1(B) = 4V + 5B \end{cases} = 4V + 5B;\ Y_2(V) = V;\ X_2(V) = B;$$

$$F_3(V) = \max\begin{cases} F_2(V) = 4V + 5B \\ 4V + F_2(0) = 4V + 5B \\ -5(B-V) + F_2(B) = 5V + 4B \\ 4V - 5B + F_2(B) = 4V + 4B \end{cases} = 4V + 5B;\ Y_3(V) = 0 \vee V;\ X_3(V) = 0;$$

$$F_4(V) = \max\begin{cases} F_3(V) = 4V + 5B \\ 6V + F_3(0) = 6V + 5B \\ -7(B-V) + F_3(B) = 7V + 2B \\ 6V - 7B + F_3(B) = 6V + 2B \end{cases} = 6V + 5B;\ Y_4(V) = V;\ X_4(V) = 0;$$

$$F_5(V) = \max\begin{cases} F_4(V) = 6V + 5B \\ 7V + F_4(0) = 7V + 5B \\ -6(B-V) + F_4(B) = 6V + 5B \\ 7V - 6B + F_4(B) = 7V + 5B \end{cases} = 7V + 5B;\ Y_5(V) = V;\ X_5(V) = 0 \vee B.$$

Отсюда видим, что искомый максимум дохода в 5-шаговом процессе при начальном запасе V равен $F_5(V) = 7\,V + 5\,B$ и *оптимальная политика по шагам* определяется следующим набором действий.

1. Объем продажи = $Y_5(V) = V$ – весь запас.
   Объем закупок = $X_5(V) = 0$ или B – ничего или полный склад.
2. Объем продажи = $Y_4(0$ или $B) = 0$ или B – весь запас, если он есть.
   Объем закупок = $X_4(0$ или $B) = 0$ – ничего не покупать.
3. Объем продажи = $Y_3(0) = 0$ – можно все продать, но нечего.
   Объем закупок = $X_3(0) = 0$ – ничего не покупать.
4. Объем продажи = $Y_2(0) = 0$ – продать запас, но его нет.
   Объем закупок = $X_2(0) = B$ – закупить полный склад.

5. Объем продажи = $Y_1(B) = B$ – продать весь запас.

Объем закупок $= X_1(V)$ – ничего не покупать.

Идеология рассмотренного решения не претерпит изменений, если появятся дополнительные ограничения на объемы продаж-покупок или дополнительные затраты, пропорциональные объемам.

Читатель может сопоставить необходимые затраты энергии на решение этой задачи методом линейного программирования и использованным здесь методом.

6.8. Численное решение рекуррентных соотношений

Большинство рассматривавшихся ранее примеров допускали аналитическое решение за счет выпуклости или линейности исследуемых функций. В случае функций более сложной природы решение рекуррентных соотношений приходится производить численно.

Обратимся к поставленной ранее задаче разделения ресурса:

$$F_k(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X - Y) + F_{k-1}(\alpha Y + \beta(X - Y)) \right], \; k = 2 \; .. \; N;$$

$$F_1(X) = \max_{0 \le Y \le X} \left[g(Y) + h(X - Y) \right].$$

Поскольку знание значения $F_{k-1}(X)$ лишь при одном конкретном значении X не дает возможности поиска $F_k(X)$ при том же аргументе, приходится выбирать сетку M значений X с каким-то шагом в интервале от нуля до некоторого предельного значения $X = Z$, соответствующего максимально возможному начальному ресурсу, и отыскиваем $F_1(X)$ и $Y_1(X)$ в узлах этой сетки, решая с этой целью M одномерных задач оптимизации.

Приступив к поиску значений $F_2(X)$ и $Y_2(X)$ в тех же узлах, мы сталкиваемся с тем, что значения $\alpha\ Y + \beta\ (X - Y)$ не совпадают с узлами сетки и приходится прибегнуть к приближенной оценке значений $F_1(\alpha\ Y + \beta\ (X - Y))$ путем интерполяции.

Напомним, что если задана таблица значений $R(T)$ на равномерной сетке с шагом h, для нахождения $R(t)$ при $T_i \le t \le T_{i+1}$ можно прибегнуть к линейной *интерполяции*

$$R(t) = R(T_i) + (t - T_i)\ \frac{R(T_{i+1}) - R(T_i)}{h}$$

или интерполяции более высоких порядков.

Аналогично отыскиваются $F_3(X)$, $F_4(X)$ и т. д.

В результате получается таблица:

X	$F_1(X)$	$Y_1(X)$	$F_2(X)$	$Y_2(\mathrm{X})$	$F_3(\mathrm{X})$	$Y_3(\mathrm{X})$	…	$F_N(\mathrm{X})$	$Y_N(\mathrm{X})$
0	$F_1(0)$	$Y_1(0)$	$F_2(0)$	$Y_2(0)$	$F_3(0)$	$Y_3(0)$	…	$F_N(0)$	$Y_N(0)$
h	$F_1(h)$	$Y_1(h)$	$F_2(h)$	$Y_2(h)$	$F_3(h)$	$Y_3(h)$	…	$F_N(h)$	$F_N(h)$
$2h$	$F_1(2h)$	$Y_1(2h)$	$F_2(2h)$	$Y_2(2h)$	$F_3(2h)$	$Y_3(2h)$	…	$F_N(2h)$	$Y_N(2h)$
…	…	…	…	…	…	…	…	…	…
Z	$F_1(Z)$	$Y_1(Z)$	$F_2(Z)$	$Y_2(Z)$	$F_3(Z)$	$Y_3(Z)$	…	$F_N(Z)$	$Y_N(Z)$

Для поиска решения при $X = C$ в пределах сетки достаточно провести «обратный ход»:

– оптимальный выбор на первом шаге N-шагового процесса $\hat{y}_1 = Y_N(C)$;

– оптимальный выбор на втором шаге $\hat{y}_2 = Y_{N-1}(C_1)$, где $C_1 = \alpha\,\hat{y}_1 + \beta\,(C - \hat{y}_1)$;

– оптимальный выбор на третьем шаге $\hat{y}_3 = Y_{N-2}(C_2)$, где $C_2 = \alpha\,\hat{y}_2 + \beta\,(C_1 - \hat{y}_2)$ …

Пример. Пусть

$$F_k(X) = \min_{0 \le Y \le X} \left\{ \frac{4}{Y} + \frac{9}{X - Y} + F_{k-1}(X - Y) \right\}, k = 2, 3, \ldots, k = 2, 3, \ldots,$$

$$F_1(X) = \min_{0 \le Y \le X} \left\{ \frac{4}{Y} + \frac{9}{X - Y} \right\}$$

и требуется найти решение при $X = 10$.

Выберем сетку значений X от 1 до 10 с шагом 1 (можно взять и меньший шаг). Для каждого значения X находим значения минимизируемой функции, перебирая значения Y в пределах от 0 до X с более мелким шагом, и сохраняем наименьшее.

В данном случае можно заметить, что здесь функции выпуклые и искомый минимум единственный. Потому можно воспользоваться аппаратом производных и обнаружить, что минимум функции для $F_1(X)$ достигается при Y, удовлетворяющих условию $9\,Y^2 - 4\,(X - Y)^2 = 0$, т. е. при $Y = 0{,}4\,X$, и $F_1(X) = {}^{25}\!/_{X}$.

Аналогично находим $F_2(X)$ и $F_3(X)$ перебором допустимых значений Y или использованием аппарата производных (возникающие уравнения для производных легко решаются аналитически) и получаем таблицу:

X	$F_1(X)$	$Y_1(X)$	$F_2(X)$	$Y_2(X)$	$F_3(X)$	$Y_3(X)$
1	25,0	0,4	61,3	0,26	107,8	0,19
2	12,5	0,8	30,6	0,51	53,9	0,38
3	8,3	1,2	20,4	0,77	35,9	0,58
4	6,3	1,6	15,3	1,02	26,9	0,77
5	5,0	2,0	12,3	1,28	21,6	0,96
6	4,2	2,4	10,2	1,54	17,9	1,15
7	3,6	2,8	8,8	1,79	15,4	1,34
8	3,1	3,2	7,7	2,06	13,5	1,54
9	2,8	3,6	6,8	2,30	12,0	1,73
10	2,5	4,0	6,1	2,56	10,8	1,92

Отсюда $F_3(10) = 10{,}8$; выбор на первом шаге $\bar{y}_1 = Y_3(10) = 1{,}92$; выбор на втором шаге

$\bar{y}_2 = Y_2(10 - 1{,}92) = Y_2(8{,}08) = 2{,}06 + (2{,}30 - 2{,}06)\ 0{,}08 = 2{,}08$; выбор на третьем шаге $\bar{y}_3 = Y_1(8{,}08 - 2{,}08) = Y_1(6) = 2{,}4$.

6.9. Примеры постановки и решения задач динамического программирования

Здесь мы рассмотрим ряд задач, решение которых поможет читателю убедиться в глубине усвоения метода рекуррентных соотношений и, может быть, откроет некоторые «изюминки» (решения некоторых из этих задач можно найти в [8, 11]).

1. *Задача надежности многокомпонентных схем*. При конструировании сложной аппаратуры одной из основных является задача обеспечения надежности, которая решается иногда путем дублирования компонент. Например, при построении последовательной схемы из N ступеней надежность (вероятность безотказной работы) равна произведению вероятностей безотказной работы ее ступеней.

Если j-я ступень содержит $1 + M$ компонент-дублеров и $\Phi_j(M_j)$ – вероятность ее безотказной работы, то надежность всей схемы $P(N)$ равна произведению значений $\Phi_j(M_j)$ при $j = 1, 2, \ldots, N$.

Очевидно, что большое количество дублеров ведет к росту стоимости, объема и появлению дополнительных ошибок. Если учитывать только фактор стоимости, то возникают ограничения

$$\sum_{j=1}^{N} C_j M_j \le S,\ M \ge 0 \text{ целые}, j = 1, 2, ..., N,$$

где S – заданная предельная стоимость дублирования схемы; C_j – стоимость одного дублера в j-й ступени.

Если вообразить N-шаговый процесс, на первом шаге которого выясняется число дублеров для N-й ступени, на втором – для $(N-1)$-й и т. д. и обозначить через $F_k(S)$ максимум надежности k-компонентной схемы, то из принципа оптимальности получим систему рекуррентных соотношений вида:

$$F_k(S) = \max_X \left\{ \Phi_k(X)\, F_{k-1}(S - C_k X) \right\}, \; k = 2, \ldots, N;$$

$$F_1(S) = \max\{\Phi_1(X)\},$$

где X – число дублеров k-й ступени, $X = 0, 1, 2, \ldots, [S / C_k]$.

Очевидно, что требование целочисленности существенно упростит численную процедуру решения задачи.

2. *Задача планирования производственной линии* (*задача Джонсона*). Рассмотрим ситуацию последовательной обработки на двух машинах N различных деталей, если известно время A_i и B_i обработки i-й детали на соответствующих машинах. Очевидно, что первая машина будет загружена полностью, но вторая может периодически оказываться в состоянии простоя. Как найти порядок обработки, минимизирующий время ее простоя и тем самым общее время обработки.

Обозначив через X_i простой в ожидании i-й детали, имеем

$$X_1 = A_1;$$

$$X_1 + X_2 = \max(A_1 + A_2 - B_1, A_1);$$

$$X_1 + X_2 + X_3 = \max(A_1 + A_2 + A_3 - B_1 - B_2, A_1 + A_2 - B_1, A_1); \ldots$$

$$\sum_{i=1}^{N} X_i = \max_{1 \le k \le N} \left\{ \sum_{i=1}^{k} A_i - \sum_{i=1}^{k-1} B_i \right\}.$$

Читатель, обратившийся к [20], может увидеть, как с помощью рекуррентных соотношений показано условие необходимости перестановки в паре деталей (i, j)

$$\min(A_j, B_i) < \min(A_i, B_j).$$

Соответственно среди всех значений A_i и B_i ищем наименьшее. Если оно совпадает с некоторым A_i, то i-ю деталь ставим на обработку первой, если совпадает с некоторым B_i – последней. Эту процедуру повторяем для всех остальных деталей.

Пусть, например, время обработки задано таблицей

i	1	2	3	4	5	6	7	8
A_i	4	4	30	6	2	9	13	9
B_i	5	1	4	30	3	13	9	9

В итоге указанного упорядочения получаем оптимальную перестановку:

i	5	1	4	8	6	7	3	2
A_i	2	4	6	9	9	13	30	4
B_i	3	5	30	9	13	9	4	1

Время простоя второй машины при первичном порядке равно
max(4, 4 + 4 – 5, 4 + 4 + 30 – 5 – 1, 4 + 4 + 30 + 6 – 5 – 1 – 4,
4 + 4 + 30 + 6 + 2 – 5 – 1 – 4 – 30,
4 + 4 + 30 + 6 + 2 + 9 – 5 – 1 – 4 – 30 – 3,
4 + 4 + 30 + 6 + 2 + 9 + 13 – 5 – 1 – 4 – 30 – 3 – 13,
4 + 4 + 30 + 6 + 2 + 9+ 13 – 5 – 1 – 4 – 30 – 3 – 13) =
= max(4, 3, 32, 34, 6, 12, 12, 12) = 34.

Простой при оптимальной перестановке составит
max(2, 2 + 4 – 3, 2 + 4 + 6 – 3 – 5, 2 + 4 + 6 + 9 – 3 – 5 – 30,
2 + 4 + 6 + 9 + 9 – 3 – 5 – 30 – 9,
2 + 4 + 6 + 9 + 9 + 13 – 3 – 5 – 30 – 9 – 13,
2 + 4 + 6 + 9 + 9 + 13 + 30 – 3 – 5– 30 – 9 – 13 – 9,
2 + 4 + 6 + 9 + 9 + 13 + 30 + 4 – 3 – 5 – 30 – 9 – 13 – 9 – 4) =
= max(2, 3, 4, –17, –17, –17, 4, 4) = 4.

Существует и другой оптимальный порядок обработки, связанный с неоднозначностью установки детали 8.

К сожалению, простого решения задачи Джонсона для случая последовательной обработки на $L > 2$ машинах нет.

3. *Задача о загрузке корабля*. Имеется корабль грузоподъемностью W, который может быть загружен неделимыми предметами N типов. Если обозначить через W_k и C_k вес и ценность предмета k-го типа и через X_k число таких предметов, можно поставить задачу сбора груза максимальной ценности в виде:

максимизировать $\sum_{k=1}^{N} C_k X_k$ при условиях $\sum_{k=1}^{N} W_k X_k \le W$, $X_k \ge 0$ – целые при всех k.

Возникшую задачу целочисленного линейного программирования можно решать методом ветвей и границ или методом Гомори,

получая решение для конкретного W. Появление дополнительных условий приведет к усложнению процесса решения.

Вообразим искусственный N-шаговый процесс, на первом шаге которого планируется загрузка предметов N-го типа (на втором $(N-1)$-го, затем $(N-2)$-го и на последнем – первого типа). Тогда, обозначив через $F_k(W)$ суммарную ценность загрузки в k-шаговом процессе при заданной грузоподъемности W и при использовании оптимальной политики, на основе принципа оптимальности сводим задачу к системе рекуррентных соотношений

$$F_k(W) = \max[C_k X + F_{k-1}(W - W_k X)],\ k = 2, 3, \ldots, N;$$

$$F_1(W) = \max[C_1 X],$$

где область максимизации определяется целыми значениями X в диапазоне от нуля до целой части отношения W / W_n. Обозначаем через $X_k(W)$ – оптимальное количество предметов, загружаемое на первом шаге k-шагового процесса.

Пример. Решим поставленную задачу при $N = 3$, $C_1 = 8$, $C_2 = 7$, $C_3 = 4$, $W = 10$, $W_1 = 4$, $W_2 = 3$, $W_3 = 2$.

Возьмем сетку значений W от 0 до 10. Очевидно, что $F_1(W)$ равно C_1, умноженному на максимально возможное X, равное целой части отношения W к W_1.

При поиске $F_2(W)$ перебираем значения X в диапазоне от 0 до целой части W / W_2 и выбираем среди значений максимизируемой функции наибольшее. Так при $W = 10$ перебираем $X = 0, 1, 2, 3$ и

$F_2(10) = \max[0 + F_1(10), 7 + F_1(7), 14 + F_1(4), 21 + F_1(1)] =$

$= \max[16, 15, \mathbf{22}, 21] = 22$; $X_2(10) = 2$.

Аналогично отыскиваем значения $F_3(W)$ и $X_3(W)$.

X	$F_1(W)$	$X_1(W)$	$F_2(W)$	$X_2(W)$	$F_3(W)$	$X_3(W)$
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	4	1
3	0	0	7	1	7	0
4	8	1	8	0	8	0,2
5	8	1	8	0	11	1
6	8	1	14	2	14	0
7	8	1	15	1	15	0,2
8	16	2	16	0	18	1
9	16	2	21	3	21	0
10	16	2	22	2	22	0,2

Полученная таблица позволяет найти оптимальную политику загрузки при любом W, не превышающем 10.

Так при $W = 8$ оптимальное количество предметов третьего типа (загружаемых на первом шаге 3-шагового процесса) равно $X_3(8) = 1$, предметов второго типа (загружаемых на первом шаге оставшегося двухшагового процесса) равно $X_2(8 - 2 \times 1) = X_2(6) = 1$, предметов первого типа (загружаемых на последнем шаге трехшагового процесса) равно $X_1(6 - 3 \times 2) = X_1(0) = 0$.

При W=10 возникают два варианта оптимальной загрузки:
1) $X_3(10) = 0$; $X_2(10 - 2 \times 0) = X_2(10) = 2$; $X_1(10 - 3 \times 2) = X_1(4) = 1$;
2) $X_3(10) = 2$; $X_2(10 - 2 \times 2) = X_2(6) = 2$; $X_1(6 - 3 \times 2) = X_1(0) = 0$.

Приведенный пример еще раз показывает, что решение методом динамического программирования обеспечивает получение политики *при любом значении* W. Более того, появление дополнительных ограничений (число предметов такого-то типа не менее или не более указанного значения) упрощает решение за счет уменьшения объема перебора (решение методами целочисленного программирования стало бы более трудоемким).

4. *Задача о замене оборудования*. Имеется машина возраста t, которая может в течение сезона обеспечить прибыль в размере $R(t)$, причем в каждом сезоне имеется возможность ее замены новой машиной с затратами $C(t)$. Требуется найти политику замены, которая обеспечивает максимальную прибыль за N сезонов.

Естественно, что функция $R(t)$ является убывающей по t из-за возрастания эксплуатационных расходов и к какому-то моменту времени оказывается более выгодным произвести модернизацию (замену) оборудования, чем продолжать эксплуатировать старое.

Если обозначить $F_N(t)$ максимальную прибыль за N сезонов от использования машины с начальным возрастом t, то получаем рекуррентные соотношения:

$$F_N(t) = \max\begin{cases} -C(t) + R(0) + F_{N-1}(1) \\ R(t) + F_{N-1}(t+1) \end{cases}, N > 1; \; F_1(t) = \max\begin{cases} -C(t) + R(0) \\ R(t) \end{cases}$$

(первый выбор соответствует замене, второй – отказу от замены).

Решение полученной системы осложняется наличием факта расширения сетки: знание значений $F_{N-1}(t)$ при $t \le T$ не дает возможности найти $F_N(T)$. Поэтому разумно принять значения $F_N(t) = 0$ при любом N и $t > T_{\max}$.

Такой подход позволяет воспользоваться обычным вычисли-

тельным алгоритмом на сетке значений t от 0 до T.

Можно построить варианты рекуррентных соотношений, где прибыль и затраты на замену являются функциями от времени (момента изготовления машины).

5. *Задача планирования развития отрасли*. Пусть в N регионах действуют или могут быть созданы предприятия по выпуску некоторой продукции и требуется довести ее выпуск до B единиц.

Введем следующие обозначения:

D_j – существующая мощность j-го предприятия;

E_j – его максимальная возможная мощность;

$M_j = E_j - D_j$ – максимально возможный прирост мощности;

$\Phi_j(X)$ – годовые затраты на приращение X мощности;

$Б = B - \Sigma D_j$ – недобор мощности.

Соответственно можно поставить задачу минимизации функции

$$F(X) = \sum_{j=1}^{N} \Phi_j(X_j)$$

при условиях $\sum_{j=1}^{N} X_j = Б$, $0 \le X_j \le M_j$ ($j = 1 .. N$).

Заметим, что задача неразрешима, если $\Sigma M_j < Б$.

Вообразим N-шаговый процесс принятия решений, на первом шаге которого принимается суждение о N-м предприятии, на втором – $(N - 1)$-м и т. д.

Обозначим через $F_k(t)$ минимальные затраты на создание дополнительной мощности t на k предприятиях. Очевидно, что достаточно ограничиться значениями t от 0 до $\min(Б, \Sigma M_j)$.

При $k > 1$ $F_k(t) = \min[\Phi_k(X) + F_{k-1}(t - X)]$, $F_1(t) = \Phi_1(t)$ (область минимизации определяется условиями $0 \le X \le \min(M_k, t)$).

Пусть планируемый выпуск $B = 100$, число предприятий $N = 4$, существующие мощности $D = (0, 20, 0, 0)$, предельные значения мощности $E = (20, 60, 50, 30)$. Требуемый прирост мощности предприятий $Б = 80$ и возможности прироста по предприятиям $M = (20, 40, 50, 30)$. Известны затраты на приращение мощности X:

X	0	10	20	30	40	50
$\Phi_1(X)$	0	10	13			
$\Phi_2(X)$	0	6	11	16	20	
$\Phi_3(X)$	0	9	18	26	32	38
$\Phi_4(X)$	0	8	16	19		

При $k = 1$ достаточно взять сетку значений t от 0 до 20 и $F_1(t) = \Phi_1(t)$, $X_1(t = (0, 10, 13))$.

При $k = 2$ берем сетку значений t от 0 до 20 + 40 и $F_2(t) = \min[\Phi_2(X) + F_1(t - X)]$ при $0 \le X \le \min(t, 40)$:

t	X=0	X=10	X=20	X=30	X=40	$F_2(t)$	$X_2(t)$
0	0+ 0					0	0
10	0+10	6+ 0				6	10
20	0+13	6+10	11+ 0			11	20
30		6+13	11+10	16+ 0		16	30
40			11+13	16+10	20 + 0	20	40
50				16+13	20+10	29	30
60					20+13	33	40

При $k = 3$ берем сетку значений t от 0 до $80 < 20 + 40 + 50$ и $F_3(t) = \min[\Phi_3(X) + F_2(t - X)]$ при $0 \le X \le \min(t, 50)$:

t	X =0	X=10	X=20	X=30	X=40	X=50	$F_3(t)$	$X_3(t)$
0	0+ 0						0	0
10	0+ 6	9+ 0					6	0
20	0+11	9+6	18+0				11	0
30	0+16	9+11	18+6	26+0			16	0
40	0+20	9+16	18+11	26+6	32+0		20	0
50	0+29	9+20	18+16	26+11	32+6	38+0	29	0,10
60	0+33	9+29	18+20	26+16	32+11	38+ 6	33	0
70		9+30	18+29	26+20	32+16	38+11	42	10
80			18+30	26+29	32+20	38+16	51	20

При $k = 4$ берем t от 0 до $80 < 20 + 40 + 50 + 0$ и $0 \le X \le \min(t, 30)$:

t	X =0	X=10	X=20	X=30	$F_4(t)$	$X_4(t)$
0	0+ 0				0	0
10	0+ 6	8+ 0			6	0
20	0+11	8+ 6	16+ 0		11	0
30	0+16	8+11	16+ 6	19+ 0	16	0
40	0+20	8+16	16+11	19+ 6	20	0
50	0+29	8+20	16+16	19+11	28	10
60	0+33	8+29	16+20	19+16	33	0
70	0+42	8+33	16+29	19+20	39	10
80	0+51	8+42	16+33	19+29	48	30

Отсюда мы получаем, что для прироста мощности на 80 единиц требуются затраты $F_4(80) = 48$ путем прироста по предприятиям следующим образом:

– предприятие 4 – на $X_4(80) = 30$;

– предприятие 3 – на $X_3(50) = 0$ или 10;

– предприятие 2 – на $X_2(50) = 30$ или $X_2(40) = 40$;

– предприятие 1 – на $X_1(20) = 20$ или $X_1(0) = 0$ единиц.

6. *В мире животных*. Медведь-арендатор имеет X кг моркови и Y кг морковных семян. Часть Z моркови он может обменять у Лисы на $R(Z)$ кг меда. В очередном сезоне оставшаяся морковь может быть использована для выращивания семян, а имеющиеся семена – для выращивания моркови.

Допустим, что наш Медведь работает в зоне устойчивого земледелия и урожайность является известной величиной. Более того, Лиса отличается честностью в торговых операциях.

Убедитесь, что поиск максимума суммарного объема полученного меда за N сезонов сводится к рекуррентным соотношениям

$$F_k(X, Y) = \max_{0 \le Z \le X} [R(Z) + F_k(\alpha(Y), \beta\,(X - Z))],\ k = 1\ ..\ N,$$

где $\alpha(Y)$; $\beta\,(X–Z)$ – урожай моркови и семян.

7. *Оптимальное разбиение*. Требуется разбить некоторое число C на N слагаемых так, чтобы их произведение было максимально. Осознайте, что задачу можно свести к рекуррентным соотношениям следующего вида:

$$F_k(C) = \max_{0 \le X \le C} \{X\, F_{k-1}(C - X)\},\ k = 1\ ..\ N;\ F_0(C) = 1.$$

Убедитесь, что оптимальное решение

$$X_k(C) = C / k;\ F_k(C) = (C / k)^k.$$

Рассмотрите аналогичную задачу максимизации

$$F(C) = \prod_{i=1}^{n} x_i^{a_i} \text{ при условиях } \sum_{i=1}^{n} x_i = C,\ X_i \ge 0,\ i = 1\ ..\ n.$$

Многочисленные примеры постановки и численного решения задач динамического программирования читатель может обнаружить в [7, 8, 11].

7. БЕСКОНЕЧНОШАГОВЫЕ ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЙ

7.1. Бесконечношаговая аппроксимация и функциональные уравнения

Обратимся к рекуррентным соотношениям, построенным ранее для задачи разделения денежного ресурса:

$$F_k(X) = \max_{0 \le Y \le X} \{g(Y) + h(X - Y) + F_{k-1}[\alpha\, Y + \beta\, (X - Y)]\},\ k = 2\ ..\ N;$$

$$F_1(X) = \max_{0 \le Y \le X} \{g(Y) + h(X - Y)\}.$$

Как показано выше, при численном их решении объем хранимой информации имеет порядок $2\ M \times N$, где M – число узлов таблиц. Если даже не сохранять предшествующие таблицы $F_k(X)$ (таблицы поведений $Y_k(X)$ нуждаются в сохранении для последующего нахождения оптимальной политики по шагам), то объем хранимой информации сведется к $M \times (N + 1)$. Само собой, что при больших N значительны как емкость сохраняемой информации, так и затраты времени вычислений.

Если $\alpha, \beta < 1$, $g(0) = h(0) = 0$, то значения $\alpha\, Y + \beta\, (X - Y) < X$ с ростом N стремятся к нулю, т. е. при больших N значения $F_N(X)$ и $F_{N+1}(X)$ оказываются достаточно близкими. Другими словами, процесс стабилизируется. Напрашивается мысль о замене процесса большой длительности бесконечношаговым и переходе от рекуррентных соотношений к *функциональному уравнению* (в отличие от обычных уравнений, решением функциональных являются функции, а не скалярные величины):

$$F(X) = \max_{0 \le Y \le X} \{g(Y) + h(X - Y) + F[\alpha\, Y + \beta\, (X - Y)]\},$$

где $F(X)$ – максимальный доход в бесконечношаговом процессе с начальным ресурсом X (оптимальный выбор на первом шаге такого процесса будем обозначать через $Y(X)$).

Разумеется, такой переход должен сопровождаться уверенностью в существовании и единственности решения, что на практике почти нереально. Соответственно обычно пытаются решить уравнение и убедиться в наличии или отсутствии решения (вопрос о единственности решения остается открытым).

7.2. Методы решения функциональных уравнений

Для решения функциональных уравнений применяют два основных метода: *приближение в пространстве функций* и *приближение в пространстве поведений*.

Возьмем функциональное уравнение общего вида

$$F(X) = \max_{y} R[y, F(X)].$$

Приближение в пространстве функций начинается с выбора некоторой функции $F_0(X)$ с последующим использованием итерационного процесса (процесса последовательных приближений)

$$F_n(X) = \max_{y} R[y, F_{n-1}(X)], n = 1, 2, \ldots,$$

который продолжается, пока при некотором n не обнаружится близость $F_n(X)$ и $F_{n\text{-}1}(X)$ при всех X. Тогда функция $F_n(X)$ принимается за $F(X)$.

Приближение в пространстве поведений начинается с выбора поведения $y_0(X)$ и поиска функции $F_0(X)$, являющейся решением уравнения $F(X) = R[y_0(X), F(X)]$. Отыскивая $\max_{y} R[y, F_0(X)]$, получаем улучшенное поведение $y_1(X)$, для которого из уравнения $F(X) = R[y_1(X), F(X)]$ находим функцию $F_1(X)$. Продолжаем аналогичные действия до тех пор, пока очередные приближения поведений не окажутся достаточно близкими.

Оба метода требуют сохранять лишь очередное и предыдущее значения приближений для искомых функций и поведений, т. е. емкость хранимых таблиц имеет порядок не более 3 M.

Рассмотрим постановку и решение интересных задач прикладного характера приближением среди функций.

7.3. Задача о кратчайшем пути в транспортной сети

Пусть задана транспортная сеть с известными длинами дуг L_{ij}, связывающих смежные вершины, в которой отмечены вершина с номером 0 (вход) и вершина N (выход). Требуется найти путь кратчайшей длины от входа к выходу.

Если обозначить через F_i длину кратчайшего пути от i-й вершины до N-й, то $F_N = 0$ и для остальных вершин по принципу оптимальности $F_i = \min_{(i\ j)} [\, L_{ij} + F_j \,]$, $i = 0, 1, \ldots, N - 1$ (из какой бы вер-

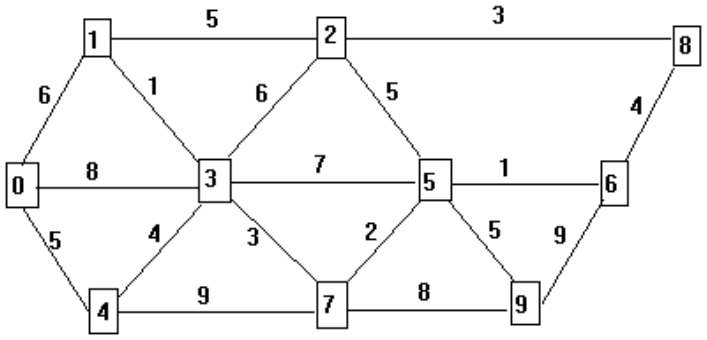
шины i мы не исходили и в какую бы вершину j не перешли, дальнейший путь должен быть кратчайшим).

Воспользуемся приближением в пространстве функций, приняв за начальную функцию, равную нулю при $i = N$ и бесконечности (очень большому числу) при остальных i.

Осуществляем итерационный процесс

$$F_i^{(k)} = \min_{(i\ j)}[\ L_{i\ j} + F_j^{(k-1)}],\ i \neq N,\ k = 1, 2, \ldots; F_N^{(k)} = 0,\ k = 1, 2, \ldots$$

до совпадения очередных приближений. Если к тому же запоминать для каждого i индекс последующей вершины j, обеспечивающей минимум, то можно будет найти искомый кратчайший путь. Рассмотрим пример (рис. 25).



$L=$

L_{ij}	0	1	2	3	4	5	6	7	8	9
0		6		8	5					
1	6		5	1						
2		5		6		5			3	
3	8	1	6		4	7		3		
4	5			4				9		
5			5	7			1	2		5
6						1			4	9
7				3	9	2				8
8			3				4			
9						5	9	8		

Рис. 25

Задав начальные приближения и перебирая вершины в порядке роста индексов, получаем очередные приближения; например,

$$F_6^{(1)} = \min(1 + F_5^{(1)}, 4 + F_8^{(0)}, 9 + F_9^{(0)}) = \min(1 + 5, 4 + \infty, 9 + 0) = 6,\ j = 5; \ldots$$

$$F_2^{(2)} = \min(3 + F_1^{(2)}, 6 + F_3^{(1)}, 5 + F_5^{(1)}, 3 + F_8^{(1)}) = \min(\infty, \infty, 5 + 5, 3 + 10) = 10, \ldots$$

i	F^0	F^1	j^1	F^2	j^2	F^3	j^3	F^4	j^4	F^5	j^5
0	∞	∞	4	∞	4	18	3	17	3	17	3
1	∞	∞	3	∞	3	11	3	11	3	11	3
2	∞	∞	8	10	5	10	5	10	5	10	5
3	∞	∞	1	10	7	10	7	10	7	10	7
4	∞	∞	3	14	3	14	3	14	3	14	3
5	∞	5	9	5	9	5	9	5	9	5	9
6	∞	6	5	6	5	6	5	6	5	6	5
7	∞	7	5	7	5	7	5	7	5	7	5
8	∞	10	6	10	6	10	6	10	6	10	6
9	0	0	-	0	-	0	-	0	-	0	-

Из полученной таблицы по значениям j[5] легко выяснить кратчайший путь [0 – 1 – 3 – 7 – 5 – 9] с длиной $F_0 = 17$. Аналогично можно найти кратчайшие пути от любой вершины до 9-й.

Ускорения сходимости процесса итераций можно добиться, если при поиске k-го приближения ссылаться не на $(k-1)$-е приближение, а на последнюю из полученных оценок:

$$F_i^{(k)} = \min_{(i\ j)}[\ L_{i\ j} + F_j^{(\mathrm{s})}]\ ,\ i \neq N\ ,\ \mathrm{s} = \begin{cases} k \\ k-1 \end{cases},\ k = 1, 2, \ldots$$

7.4. Задача о критическом пути в сетевом графике

Во многих прикладных задачах возникает необходимость поиска пути максимальной длины (критического пути) от входа к выходу в ориентированной транспортной сети, не содержащей контуров (т. е. возможности возврата к пройденным вершинам).

Если обозначить через F_i длину критического пути от i-й вершины до выхода N, то задача сводится к системе функциональных уравнений

$$F_i = \max_{(i\ j)}[\ L_{i\ j} + F_j\ ],\ i = 0, 1\ , \ldots, N-1;\ F_N = 0,$$

решение которой можно, как и в задаче о кратчайшем пути, вести приближением в пространстве функций (приняв начальные приближения для всех вершин, кроме N-й, равными нулю). Если предварительно проранжировать вершины сети, процесс итераций можно осуществить за один шаг (существенно при больших N).

Суть ранжировки поясним на примере нижеприведенной сети (рис. 26).

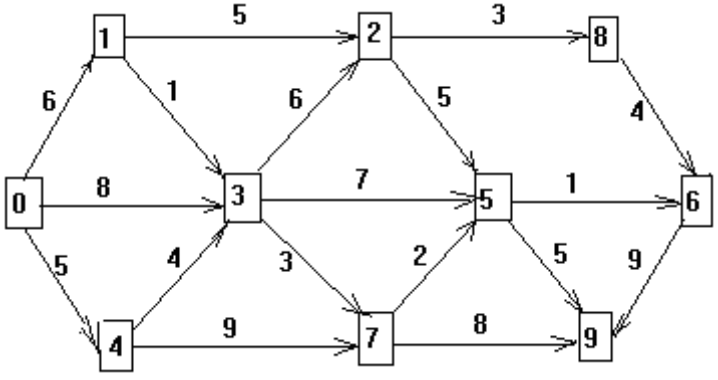


Рис. 26

Отнесем к рангу 0 вершину входа 0. К рангу 1 отнесем вершины, в которые ведут дуги только из вершины ранга 0, т. е. вершины 1 и 4. Ранг 2 присвоим вершинам, в которые ведут дуги только из вершин меньшего ранга, т. е. вершине 3. К рангу 3 отнесутся вершины 2 и 7, к рангу 4 – вершины 5 и 8, к рангу 5 – вершина 6, к рангу 6 – вершина 9.

Перебирая вершины в порядке убывания ранга, получаем значения

$F_9 = 0;$	$F_7=\max[2+F_5,8+F_9] = 12, j_7 = 5;$
$F_6 = 9 + F_9 = 9, j_6 = 9;$	$F_3=\max[6+F_2,7+F_5,3+F_7] =22, j_3 =2;$
$F_5 =\max[5+F_9,1+F_6] = 10, j_5 = 6;$	$F_1=\max[5+F_2,1+F_3] = 23, j_1 = 3;$
$F_8 = 4 + F_6 = 13, j_8 = 6;$	$F_4=\max[4+F_3,9+F_7] = 26, j_4 = 3;$
$F_2 =\max[5+F_5, 3+F_8] = 16, j_2 = 8;$	$F_0=\max[6+F_1,8+F_3,5+F_4]=31, j_0 =4$

и критический путь [0 – 4 – 3 –2 – 8 – 6 – 9] с длиной 31.

7.5. Выбор критерия оптимальности

Многие реальные процессы управления являются частью некоторых бесконечных процессов (по крайней мере, длительных) – работа электростанции или предприятия большой химии предполагается в течение десятилетий. Если их предыстория обычно не представляет интереса для последующего планирования (мы оставляем в стороне пользу опыта), то учет достаточно далекого будущего необходим, ибо за счет оптимизации лишь на фиксированный интервал времени мы можем принять решение, ведущее наших потомков к критической ситуации.

Рассмотрение бесконечношаговых процессов базируется на *гипотезе стационарности*, т. е. постоянства или известной закономерности изменений характеристик среды или хотя бы постоянства их вероятностных оценок.

Замена рекуррентных соотношений функциональными уравнениями обеспечивает поиск универсальной политики для любого шага процесса, не говоря об оправданности такого подхода с вычислительной точки зрения. Разумеется, едва ли возможно надежно учесть далекое будущее. Многие характеристики процесса изменяются под влиянием неучтенных воздействий. Соответственно, найденная универсальная политика перестает быть оптимальной через некоторое время, но правомерна на первых шагах процесса.

На этом и строится так называемое «*скользящее планирование*», где оптимальная политика для бесконечношагового процесса используется на конечном числе шагов и периодически подвергается коррекции (пересчету) с учетом возникающих изменений параметров.

При конечном плановом периоде не возникает проблем с выбором критерия эффективности и мы спокойно говорим о максимальном доходе, минимальных издержках и т. п.

В бесконечношаговом процессе эти критерии терпят крах, т. к. при *любой* реальной политике эти величины стремятся к бесконечности.

Рассмотрим для примера таблицу значений эффектов по шагам некоторого процесса при разных политиках *Q*:

Q	шаг 1	шаг 2	шаг 3	шаг 4	шаг 5	шаг 6	…
1	3	2	1	3	2	1	…
2	3	1	3	1	3	1	…
3	1	6	–1	1	6	–1	…
4	2	2	2	2	2	2	…
5	1	3	1	3	1	3	…
6	2	2	1	2	2	1	…
7	4	0	0	4	0	0	…

Очевидно обнаруживается, что политика 6 при любой длительности процесса дает худший эффект в сравнении, например, с политикой 1. Отдать предпочтение какой-либо из других политик затруднительно.

Обычно для бесконечношаговых процессов используются три критерия эффективности:

– *средний эффект* за отрезок времени (СЭ);

– *интегральный дисконтированный эффект* (ИДЭ);

– *эквивалентный средний эффект* (ЭСЭ).

Критерий СЭ интуитивно наиболее прост и определяется отношением суммы эффектов за несколько шагов к числу шагов. Для приведенного примера можно найти оценки СЭ по первым 6 шагам и попытаться сделать грубый прогноз на большую длительность.

Q	шаг 1	шаг 2	шаг 3	шаг 4	шаг 5	шаг 6	…	прогноз
1	3	5/2	2	9/4	11/5	2	…	2 +
2	3	2	7/3	2	11/5	2	…	2 +
3	1	7/2	2	7/4	13/5	2	…	2 ±
4	2	2	2	2	2	2	…	2
5	1	2	5/3	2	9/5	2	…	2 –
7	4	2	4/3	2	8/5	4/3	…	< 2

Если мы не ошиблись в прогнозе, с позиций СЭ политики 1 и 2 предпочтительнее, а политика 7 хуже других.

Критерий ИДЭ состоит в вычислении интегрального показателя эффекта, дисконтированного (приведенного) к начальному мо-

менту времени. Если значения эффекта по шагам процесса равны $[R_1, R_2, ..., R_n, ...]$, то

$$\text{ИДЭ} = R_1 + \alpha R_2 + \alpha^2 R_3 + ... + \alpha^{n-1} R_n + ...,$$

где $0 \le \alpha < 1$ называют *коэффициентом дисконтирования* (*приведения, скидки*).

Если максимальное из значений R_i ($i = 1, 2, ...$) меньше некоторого R, то

$$\text{ИДЭ} < R (1 + \alpha + \alpha^2 + \alpha^3 + ...) = R / (1 - \alpha),$$

т. е. при $\alpha < 1$ величина ИДЭ всегда ограничена и тем самым дает возможность сравнения политик.

В основе этого критерия лежит утверждение об уменьшении полезности денег (денежная сумма сегодня полезнее той же денежной суммы завтра). В самом деле, если фирма обладает денежной единицей и имеет возможность вложений под P % годовых, то через период в k лет она будет обладать $S = (1 + P / 100)^k$ денежными единицами, т. е. обладание $1 / S$ денежной единицы сегодня эквивалентно обладанию полной денежной единицей через период. В такой ситуации коэффициент дисконтирования равен

$$\alpha = \left| 1 \Big/ 1 + \frac{\text{p}}{100} \right|^k$$

и убывает с ростом процентной ставки.

Сравним вышеприведенные политики с позиций ИДЭ. Имеем

$\text{ИДЭ}(1) = 3 + 2\alpha + \alpha^2 + 3\alpha^3 + 2\alpha^4 + \alpha^5 + ... =$

$= (3 + 2\alpha + \alpha^2)(1 + \alpha + \alpha^2 + \alpha^3 + ...) = (3 + 2\alpha + \alpha^2) / (1 - \alpha^3).$

Аналогично получаем

$\text{ИДЭ}(2) = (3 + \alpha) / (1 - \alpha^2)$, $\text{ИДЭ}(3) = (1 + 6\alpha - \alpha^2) / (1 - \alpha^3)$,

$\text{ИДЭ}(4) = 2 / (1 - \alpha)$, $\text{ИДЭ}(5) = (1 + 3\alpha) / (1 - \alpha^2)$,

$\text{ИДЭ}(6) = 4 / (1 - \alpha^3)$.

Сопоставляя полученные значения, обнаруживаем

$\text{ИДЭ}(1) \ge \text{ИДЭ}(2)$ при всех α (при $\alpha = 0$ равенство),

$\text{ИДЭ}(1) \ge \text{ИДЭ}(3)$ при всех α (при $\alpha = 1$ равенство),

$\text{ИДЭ}(1) \ge \text{ИДЭ}(4)$ при всех α (при $\alpha = 1$ равенство),

$\text{ИДЭ}(1) \ge \text{ИДЭ}(5)$ при всех α (при $\alpha = 1$ равенство),

$\text{ИДЭ}(1) > \text{ИДЭ}(7)$ при $\alpha > 0{,}414$ (!).

Очевидно, что политика 1 имеет преимущество над политиками 2, 3, 4, 5 и превосходит политику 7 при $\alpha > 0{,}414$, что соответ-

ствует $P < 141$ %. При бóльших P политика 7 имеет преимущество за счет бóльшей начальной денежной суммы.

Критерий ЭСЭ связывает подходы с позиций ИДЭ и СЭ [22]:
$ЭСЭ(\alpha) = (1 - \alpha) \times ИДЭ(\alpha)$; $СЭ = \lim\limits_{\alpha \to 1} ЭСЭ(\alpha)$.

Для рассмотренного примера имеем
$ЭСЭ(1) = (3 + 2\alpha + \alpha^2) / (1 + \alpha + \alpha^2)$, $СЭ(1) = 2$;
$ЭСЭ(2) = (3 + \alpha) / (1 + \alpha)$, $СЭ(2) = 2$;
$ЭСЭ(3) = (1 + 6\alpha - \alpha^2) / (1 + \alpha + \alpha^2)$, $СЭ(3) = 2$;
$ЭСЭ(4) = 2$, $СЭ(4) = 2$;
$ЭСЭ(5) = (1 + 3\alpha^2) / (1 + \alpha)$, $СЭ(5) = 2$;
$ЭСЭ(7) = 4 / (1 + \alpha + \alpha^2)$, $СЭ(7) = 4 / 3$.

Появившаяся возможность точной оценки СЭ обнаруживает, что с позиций СЭ политика 7 хуже остальных политик.

7.6. Управление запасами: конечношаговый процесс

Рассмотрим задачу разработки программы выпуска некоторого изделия на плановый период из N отрезков времени в предположении наличия точного прогноза спроса на эту продукцию.

Пусть продукция, создаваемая в течение отрезка времени t, может быть использована для покрытия спроса на этом отрезке и имеется возможность сохранения остатка до начала очередного временного периода.

Требуется разработать такую программу, при которой общая сумма затрат на производство и хранение минимальна при условии полного и своевременного удовлетворения спроса [18].

Обозначим через k число временных интервалов до конца процесса, через X_k – объем производства в первом из предстоящих k периодов (при $N = 5$ индекс $k = 5$ соответствует первому шагу, $k = 4$ – второму и т. д.), через Y_k – уровень запаса на конец этого периода, через $C_k(X_k, Y_k)$ – функцию затрат на производство и хранение в этом периоде. Аналогично через S_k обозначим спрос на продукцию, R_k – максимальный объем производства и T_k – максимальную емкость склада.

Если обозначить через $F_k(Y)$ минимальные затраты в k-шаговом процессе с начальным запасом Y, а через $X_k(Y)$ – оптимальный объем производства на k-м шаге такого процесса, то возникает система рекуррентных соотношений

$$F_k(Y) = \min[C_k(X_k, Y + X_k - S_k) + F_{k-1}(Y + X_k - S_k)],\ k = 2\ ..\ N,$$

где область минимизации определяется условиями

$$0 \le X_k \le R_k,\ 0 \le Y + X_k - S_k \le T_k.$$

В случае $k = 1$ напрашивается очевидная политика: производить лишь минимум продукции, необходимый для покрытия спроса, т. е.

$$F_1(Y) = C_1(S_1 - Y, 0),\ X_1(Y) = S_1 - Y \text{ при } Y \le S_1,$$

$$F_1(Y) = C_1(0, Y - S_1),\ X_1(Y) = 0 \text{ при } Y \ge S_1.$$

Рассмотрим пример 6-шагового процесса с характеристиками производственной мощности $R_k = 5$, емкости склада $T_k = 4$ и спроса $S_k = 3$, неизменными на всех шагах процесса. Возьмем неизменную функцию затрат $C_k(X, Y) = C(X) + H(Y)$, где $H(Y) = H \times Y$ – стоимость хранения Y единиц продукции, пропорциональная объему хранения.

Функцию производственных затрат возьмем в виде:

$$C(X) = \begin{cases} A + B\,X\ , & X > 0 \\ 0\ , & X = 0 \end{cases}$$

(если в текущем интервале времени производится продукция, то кроме непосредственных затрат, пропорциональных объему, присутствуют косвенные затраты на содержание контролеров, слесарей-ремонтников и др.). В этом случае в зависимости от стоимостных показателей H, A, B может возникнуть резон усиленной работы некоторое время с целью, чтобы затем какое-то время иметь передышку и жить созданными запасами.

Пусть $A = 13$, $B = 2$, $H = 1$. Примем диапазон целочисленных значений Y от 0 до 4 и с учетом

$$\text{при } Y \le 3\ F_1(Y) = 13 + 2\,(3 - Y),\ X_1(Y) = 3 - Y;$$

$$\text{при } Y \ge 3\ F_1(Y) = 1\,(Y - 3),\ X_1(Y) = 0;$$

имеем

Y	0	1	2	3	4
$F_1(Y)$	19	17	15	0	1
$X_1(Y)$	3	2	1	0	0

При $k > 1$

$$F_k(Y) = \min\left[\begin{cases} 13 + 2\,X\ , & X > 0 \\ 0\ , & X = 0 \end{cases} + 1\,(Y + X - 3) + F_{k-1}(Y + X - 3)\right],$$

где область минимизации определена требованиями

$$0 \le X \le 5,\ 0 \le Y + X - 3 \le 4.$$

При $k = 2$ имеем расчетную таблицу значений

Y	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$F_2(Y)$	$X_2(Y)$
0				19+0+19	21+1+17	23+2+15	38	3
1			17+0+19	19+1+17	21+2+15	23+3+0	26	5
2		15+0+19	17+1+17	19+2+15	21+3+0	23+4+1	24	4
3	0+0+19	15+1+17	17+2+15	19+3+0	21+4+1		19	0
4	0+1+17	15+2+15	17+3+0	19+4+1			18	0

При $k = 3$

Y	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$F_3(Y)$	$X_3(Y)$
0				19+0+38	21+1+26	23+2+24	48	4
1			17+0+38	19+1+26	21+2+24	23+3+19	45	5
2		15+0+38	17+1+26	19+2+24	21+3+19	23+4+18	43	4
3	0+0+38	15+1+26	17+2+24	19+3+19	21+4+18		38	0
4	0+1+26	15+2+24	17+3+19	19+4+18			27	0

и т. д., получая в итоге сводную таблицу $F_k(Y)$, $X_k(Y)$:

Y	$F_1(Y)$	$X_1(Y)$	$F_2(Y)$	$X_2(Y)$	$F_3(Y)$	$X_3(Y)$	$F_4(Y)$	$X_4(Y)$	$F_5(Y)$	$X_5(Y)$	$F_6(Y)$	$X_6(Y)$
0	19	3	38	3	48	4	67	3,4	79	5	96	4
1	17	2	26	5	45	5	64	5	74	5	93	5
2	15	1	24	4	43	4	54	5	72	4	91	4
3	0	0	19	0	38	0	48	0	67	0	79	0
4	1	0	18	0	27	0	46	0	65	0	75	0

Если уровень запаса на начало процесса равен 0, то оптимальные объемы производства по периодам от начала процесса равны

$X_1 = X_6(0) = 4$, $X_2 = X_5(0 + 4 – 3) = 5$, $X_3 = X_4(1 + 5 – 3) = 0$,

$X_4 = X_3(3 + 0 – 3) = 4$, $X_5 = X_2(0 + 4 – 3) = 5$, $X_6 = X_1(1 + 5 – 3) = 0$.

Здесь наблюдается двукратное повторение последовательности объемов производства – *производственный цикл* [4 – 5 – 0] и средние затраты 96 / 6 = 16. Сохранится ли наблюдаемое явление для большей длительности? Вычислительный эксперимент дает положительный ответ при $N = 9$ и 12, но для $N = 15$ уже обнаруживается цикл [5 – 5 – 0 – 5 – 0] и средние затраты составляют 15,8 < 16. А для длительности 52 или 365? При наличии программного средства это нетрудно выяснить, но может быть разумнее перейти от большой (!?) длительности к бесконечной и не тратить зря время и деньги?

P. S. Если принять функцию производственных затрат пропорциональной объему, то можно убедиться, что минимум затрат обеспечится требованием производить лишь необходимое для покрытия спроса на каждом шаге и не тратить деньги на создание запасов.

7.7. Управление запасами: бесконечношаговый процесс

При неизменных во времени спросе и других характеристиках рассматриваемый процесс проявляет свойства стационарности – стремления некоторых его итоговых характеристик к постоянству.

Если обозначить через $F(Y)$ величину интегрального дисконтированного эффекта (ИДЭ) при начальном запасе Y и использовании оптимальной политики, возникает функциональное уравнение

$$F(Y) = \min[C(X, Y + X - S) + \alpha F(Y + X - S)],$$

где α – коэффициент дисконтирования и область минимизации определяется условиями $0 \le X \le R$, $0 \le Y + X - S \le T$.

Поставленную задачу при дискретных значениях спроса и объема производства по соображениям компактности записи можно интерпретировать как систему переходов между состояниями (уровнями запаса).

Если обозначить Y через i, $Y + X - S$ через j, $C(X, Y + X - S)$ через C_{ij}, $F(Y)$ через $F(i)$, то C_{ij} можно понимать как затраты на переход из состояния i в состояние j, а функциональное уравнение можно переписать в следующем виде:

$$F(i) = \min_j [C_{ij} + F(j)],\ i = 0, 1, \ldots, T. \qquad (*)$$

Обозначим через g средние затраты (СЭ) и через F_i – составляющие затрат, определяемые начальным состоянием (запасом).

Тогда можно принять

$$F(i) = F_i + g / (1 - \alpha),$$

и при $\alpha = 1$ приведенные функциональные уравнения (*) примут вид

$$F_i + g = \min_j [C_{ij} + F_j],\ i = 0, 1, \ldots, T.$$

Решение этих уравнений осуществляем приближением в поведениях.

Обратимся к примеру, рассмотренному в 7.6, и соответствующую систему управления запасами будем интерпретировать как систему с пятью состояниями – уровнями запаса ($i = 0, 1, 2, 3, 4$).

Для удобства последующей работы заблаговременно рассчитаем таблицу значений C_{ij} ($j = i + X - 3$), где учитываются затраты на производство, равные $13 + 2 \cdot X$ при $X > 0$ и нулю при $X = 0$, и затраты на хранение остатков, равные $1 \cdot j$ (X не превышает 5).

C_{ij} =

$i \setminus j$	0	1	2	3	4
0	19+0	21+1	23+2		
1	17+0	19+1	21+2	23+3	
2	15+0	17+1	19+2	21+3	23+4
3	0+0	15+1	17+2	19+3	21+4
4		0+1	15+2	17+3	19+4

Берем за начальное поведение вполне разумную политику производства, предлагающую минимальный объем выпуска, лишь достаточный для покрытия текущего спроса.

Тогда для $i = 0$ берем $X = 3$, для $i = 1$ – $X = 2$, для $i = 2$ – $X = 1$, для $i = 3$ и $i = 4$ – $X = 0$: начальное поведение определяется системой переходов $(0 \to 0)$, $(1 \to 0)$, $(2 \to 0)$, $(3 \to 0)$, $(4 \to 1)$.

Для этого поведения возникает система 5 уравнений с 6 неизвестными, разрешимая с точностью до константы (приняв некоторое F_i, например F_0, равным какой-то константе, другие F_i определяем с точностью до константы, тогда как значение g не зависит от этого выбора):

$$F_0 + g = 19 + F_0,\ F_0 = 0,\ \boldsymbol{g = 19},$$
$$F_1 + g = 17 + F_0,\ F_1 = -2,$$
$$F_2 + g = 15 + F_0,\ F_2 = -4,$$
$$F_3 + g = 0 + F_0,\ F_3 = -19,$$
$$F_4 + g = 1 + F_1,\ F_4 = -20.$$

Обнаружив, что выбранная политика обеспечивает средние затраты, равные 19, попытаемся найти улучшенное поведение:

$i = 0$: $\min[C_{0j} + F_j] = \min[19 + 0,\ 22 - 2,\ 25 - 4,\ -\ ,\ -\ ]$ при $j = 0$;

$i = 1$: $\min[C_{1j} + F_j] = \min[17 + 0,\ 20 - 2,\ 23 - 4,\ 26 - 19,\ -\ ]$ при $j = 3$;

$i = 2$: $\min[C_{2j} + F_j] = \min[15 + 0,\ 18 - 2,\ 21 - 4,\ 24 - 19,\ 27 - 20]$ при $j = 3$;

$i = 3$: $\min[C_{3j} + F\ ] = \min[0 + 0,\ 16 - 2,\ 19 - 4,\ 22 - 19,\ 25 - 20]$ при $j = 0$;

$i = 4$: $\min[C_{4j} + F_j] = \min[-\ ,\ 1 - 2,\ 17 - 4,\ 20 - 19,\ 23 - 20]$ при $j = 1$.

Найденное улучшенное поведение определяет систему переходов: $(0 \to 0)$, $(1 \to 3)$, $(2 \to 3)$, $(3 \to 0)$, $(4 \to 1)$.

Строим и решаем соответствующую систему уравнений

$$F_0 + g = 19 + F_0,\ F_0 = 0,\ \boldsymbol{g = 19},$$
$$F_1 + g = 26 + F_3,\ F_1 = -12,$$
$$F_2 + g = 24 + F_3,\ F_2 = -14,$$
$$F_3 + g = 0 + F_0,\ F_3 = -19,$$
$$F_4 + g = 1 + F_1,\ F_4 = -30.$$

Видим, что и эта политика обеспечит средние затраты, равные 19.

Аналогичные попытки улучшения дают поведения:

$(0 \to 1), (1 \to 3), (2 \to 3), (3 \to 4), (4 \to 1)$, $g = 17{,}3$;

$(0 \to 2), (1 \to 2), (2 \to 4), (3 \to 0), (4 \to 1)$, $g = 17$;

$(0 \to 2), (1 \to 3), (2 \to 3), (3 \to 0), (4 \to 1)$, $g = 16{,}3$;

$(0 \to 1), (1 \to 3), (2 \to 4), (3 \to 0), (4 \to 1)$, $g = 16$;

$(0 \to 2), (1 \to 3), (2 \to 4), (3 \to 0), (4 \to 1)$, $g = 15{,}8$;

$(0 \to 2), (1 \to 3), (2 \to 4), (3 \to 0), (4 \to 1)$.

Поскольку два очередных приближения в поведениях совпали, то можно утверждать, что оптимальная политика переходов между состояниями имеет вид, приведенный на рис. 27, и оптимальный производственный цикл определяется последовательностью объемов производства [5 – 5 – 0 –5 – 0]. В частности, при скользящем планировании найденную политику можно принять за отправную.

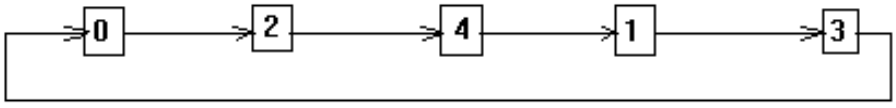


Рис. 27

7.8. Бесконечношаговый процесс замены оборудования

Обратимся к задаче, рассмотренной в предшествующей главе для условий стационарного режима:

$$F_n(t) = \max \begin{bmatrix} -C(t) + R(t) + F_{n-1}(1) \\ R(t) + F_{n-1}(t+1) \end{bmatrix}.$$

Если рассмотреть бесконечношаговый процесс и обозначить через $F(t)$ максимальный дисконтированный эффект при начальном возрасте t, то

$$F(t) = \max \begin{bmatrix} -C(t) + R(0) + \alpha F(1) \\ R(t) + \alpha F(t+1) \end{bmatrix}.$$

Предполагая оптимальным возрастом замены $t = T$, имеем

$F(0) = R(0) + \alpha F(1)$; $F(1) = R(1) + \alpha F(2)$; …,

$F(T) = -C(T) + R(0) + \alpha F(1)$,

откуда последовательными подстановками получаем

$$F(0) = R(0) + \frac{\alpha}{1-\alpha^T}\{R(1) + \alpha\, R(2) + \alpha^2\, R(3) + \ldots + \alpha^{T-1}\,[R(0) - C(T)]\}.$$

Достаточно последовательно табулировать значения функции $F(0)$ при $T = 0, 1, 2, \ldots$ до тех пор, пока сохраняется возрастание этих значений. Предельное значение T соответствует оптимальному возрасту замены машины на новую.

Можно рассмотреть и другие подходы к организации замены оборудования [18].

8. СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЙ

8.1. Специфика выбора критерия оптимальности

До сих пор мы рассматривали оптимизационные модели, где предполагалась полная *детерминированность* исходных данных, т. е. значения спроса, дохода, затрат и прочих характеристик определялись однозначно. Однако часто значения параметров задачи являются функциями от многих факторов, неучтенных в математической модели, и с позиций решающего задачу выступают в роли случайных величин, для которых известны лишь параметры соответствующего распределения вероятностей или другие вероятностные оценки.

Естественно, что в таких условиях мы не можем говорить о точных значениях прибыли, затрат и т. п., а лишь об ожидаемых значениях этих величин или о вероятностях того, что указанные величины принимают значения в некотором заданном диапазоне.

Если в случае детерминированного процесса обнаруживается одна или несколько оптимальных стратегий, представляющих вполне определенную последовательность управляющих воздействий (такие стратегии называют чистыми), то для стохастического процесса оптимальная стратегия часто представляет совокупность таких воздействий, смешанных в некоторых пропорциях (соответственно такие стратегии и называют смешанными).

Не претендуя на исчерпывающее изложение этой необъятной тематики, рассмотрим несколько примеров многошаговых стохастических процессов принятия решений.

8.2. Управление запасами в условиях неопределенности

Рассмотрим уже упоминавшуюся ранее задачу управления запасами для случая, когда величина спроса является случайной. Тогда запланированный в начале очередного периода объем производства может оказаться заниженным по сравнению с возникающим спросом и для покрытия неудовлетворенного спроса приходится нести дополнительные расходы (штрафы, срочные заказы и т. п.).

Рассмотрим случай, когда спрос определяется дискретной случайной величиной.

Если при детерминированном спросе задача управления запасами сводилась к рекуррентным соотношениям вида

$$F_k(Y) = \min\{C_k(X) + H_k(Y + X - S_k) + F_{k-1}(Y + X - S_k)\},$$

то в ситуации, когда S_k принимает значения S_{k1}, S_{k2}, ..., S_{km} с вероятностями P_{k1}, P_{k2}, ..., P_{km} придется определить $F_k(Y)$ как минимальные *ожидаемые* затраты производства и хранения в k-шаговом процессе с начальным запасом Y.

Учитывая, что для дискретной случайной величины математическое ожидание равно сумме произведений всех ее возможных значений на соответствующие вероятности, получаем рекуррентные соотношения следующего вида:

$$F_k(Y) = \min\{C_k(X) + \sum_{i=1}^{m}[H_k(Y + X - S_{ki}) + F_{k-1}(Y + X - S_{ki})]\, P_{ki}\}.$$

Область минимизации в условиях обязательного покрытия любого возможного спроса должна учитывать то, что спрос должен удовлетворяться в любом случае $Y + X \geq \max\limits_{ki}\{S_{ki}\}$, но остаток $Y + X - \min\limits_{ki}\{S_{ki}\}$ не превышает емкости склада.

Пусть, например, максимальный объем производства равен 5, емкость склада равна 4, для всех k $H_k = 1$, $C_k(X) = 13 + 2X$ при $X > 0$ и равна 0 при $X = 0$, спрос может быть равен 2 или 4 с равными вероятностями.

Область минимизации определяется условиями $4 \leq Y + X \leq 6$ и $0 \leq X \leq 5$.

Отыскивая

$F_1(Y) = \min\{C(X) + 0{,}5\cdot(Y + X - 2) + 0{,}5\cdot(Y + X - 4)\} =$

$= \min\{C(X) + Y + X - 3\}$, получаем

Y	0	1	2	3	4
$F_1(Y)$	22	20	18	16	1
$X_1(Y)$	4	3	2	1	0

Аналогично находим

$F_2(Y) = \min\{C(X) + (Y + X - 3) + 0{,}5\, F_1(Y + X - 2) + 0{,}5\, F_1(Y + X - 4)\}$:

Y	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$F_2(Y)$	$X_2(Y)$
0					22+20	25+18	42	4
1				20+20	23+18	26+9,5	35,5	5
2			18+20	21+18	24+9,5		33,5	4
3		16+20	19+18	22+9,5			31,5	3
4	1+20	17+18	20+9,5				21	0

$F_3(Y) = \min\{C(X) + (Y + X - 3) + 0{,}5\, F_2(Y + X - 2) + 0{,}5\, F_2(Y + X - 4)\}$:

Y	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$F_3(Y)$	$X_3(Y)$
0					22+37,75	25+33,5	58,5	5
1				20+37,75	23+33,5	26+27,25	53,25	5
2			18+37,75	21+33,5	24+27,25		51,25	4
3		16+37,75	19+33,5	22+27,25			49,25	3
4	1+37,75	17+33,5	20+27,25				38,75	0

Если в случае детерминированного процесса со спросом 3 для трехшагового процесса с нулевым начальным запасом мы получали оптимальную политику производства по шагам (4, 5, 0), здесь объемы производства несколько выше и равны (5, 4, 1).

Если обратиться к рассмотрению соответствующего бесконечношагового процесса, то функциональное уравнение с учетом дисконтирования имеет вид

$$F(Y) = \min\{C(X) + \sum_{i=1}^{m}[H(Y + X - S) + \alpha\, F(Y + X - S_i)]\, P_i\}.$$

Не останавливаясь на проблеме его решения, отметим, что если в рассмотренном ранее примере детерминированного процесса оптимальные режимы производства в зависимости от начального запаса были равны (5, 5, 5, 0, 0), то решение этой системы приближением в поведениях дает значения (5, 5, 4, 3, 0).

Предположим, что спрос является случайной величиной из диапазона от 0 до ∞ и известны функции (для простоты считаем их независимыми от номера сезона):

$P(S)$ – плотность распределения величины спроса S;

$K(X)$ – затраты на сезонный объем производства X;

$R(V)$ – дополнительные затраты на компенсацию неудовлетворенного спроса в объеме V.

Напомним, что в случае непрерывной случайной величины S математическое ожидание некоторой функции $F(S)$ равно значению интеграла:

$$M[F(S)] = \int_{(S)} F(S)P(S)dS.$$

Обозначим через $F_k(Y)$ математическое ожидание издержек в k-шаговом процессе при начальном запасе Y и при использовании оптимальной политики.

Очевидно, что затраты на первом шаге процесса при начальном запасе Y равны сумме затрат на производство X единиц продукции, т. е. значению $K(X)$, и ожидаемых затрат на покрытие неудовлетворенного спроса. Так как при $S \le Y + X$ спрос удовлетворяется полностью, то последние равны

$$\int_{Y+X}^{\infty} R(S - Y - X)P(S)dS .$$

Последующий $(k-1)$-шаговый процесс начнется с начальным запасом $Y + X - S$ при S в диапазоне от 0 до $Y + X$ и нулевым начальным запасом при S, превысившим $Y + X$.

Тогда в соответствии с принципом оптимальности задача минимизации затрат в процессе конечной длительности сводится к рекуррентным соотношениям вида [7, 8]:

$$F_k(Y) = \min_X \{ K(X) + \int_{Y+X}^{\infty} R(S - Y - X)P(S)dS +$$

$$+ \int_0^{Y+X} F_{k-1}(Y + X - S)P(S)dS + F_{k-1}(0) \int_{Y+X}^{\infty} P(S)dS \}, \; k > 1;$$

$$F_1(Y) = \min_X \{ K(X) + \int_{Y+X}^{\infty} R(S - Y - X)P(S)dS \}.$$

Очевидно, что аналитическое решение для рекуррентных соотношений и функциональных уравнений получить трудно даже в случае простейших исходных функций. Тем не менее, численное решение, нереальное в XX веке, на современной компьютерной базе вполне осуществимо при разумной длительности процесса и не слишком высокой требовательности к точности оценок.

8.3. Марковские процессы принятия решений

Пусть некоторая система (техническая, биологическая, информационная) в любой фиксированный момент t может находиться в одном из n состояний и переходить из этого состояния в любое другое. Если вероятность $P_t(i, j)$ перехода в момент t из i-го состояния в j-е не зависит от предыстории системы, такая система называется *марковской*[17]. Если управление некоторым процессом идет без уче-

[17] По имени выдающегося русского математика Андрея Андреевича Маркова (1856 – 1922).

та накапливаемого опыта, то система управления может считаться таковой.

Обозначив через $X_t(i)$ ожидаемую вероятность того, что в момент t система находится в i-м состоянии, находим ожидаемую вероятность нахождения системы в любом состоянии в любой последующий момент:

$$X_{t+1}(j) = \sum_{i=1}^{n} P_t(i, j)\, X_t(i),\ j = 1 \,..\, n.$$

Так, если вероятности перехода не зависят от t, определяются матрицей

$$P = \begin{Vmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{Vmatrix}$$

и в начальный момент времени система находится в состоянии 1, т. е. $X_0(1) = 1$, $X_0(2) = 0$, то:

$X_1(1) = 1 / 2\ X_0(1) + 1 / 3\ X_0(2) = 0{,}5, \quad X_1(2) = 0{,}5,$

$X_2(1) = 1 / 2\ X_1(1) + 1 / 3\ X_1(2) = 0{,}4167, X_2(2) = 0{,}5833,$

$X_3(1) = 1 / 2\ X_2(1) + 1 / 3\ X_2(2) = 0{,}4028, X_3(2) = 0{,}5972,$

$X_4(1) = 1 / 2\ X_3(1) + 1 / 3\ X_3(2) = 0{,}4005, X_4(2) = 0{,}5995$ т. д.

Заметим, что при вероятностях, не зависящих от времени, система обладает свойством *стационарности*, т. е. функция $X_t(j)$ при $t \to \infty$ асимптотически сходится к функции $X(j)$, удовлетворяющей уравнениям:

$$X(j) = \sum_{i=1}^{n} P(i, j)\, X(i),\ j = 1 \,..\, n.$$

Для нашего примера

$X(1) = 1 / 2\ X(1) + 1 / 3\ X(2)$, $X(2) = 1 / 2\ X(1) + 2 / 3\ X(2)$, откуда с учетом $X(1) + X(2) = 1$ получаем $X(1) = 0{,}4$, $X(2) = 0{,}6$.

Предположим, что в каждый момент времени выбор вероятностей перехода зависит от некоторой политики (выбора) q и переход сопровождается получением некоторого благоприятного эффекта $R_{ij}(q)$. Обозначим через $F_k(i)$ *ожидаемый эффект* функционирования системы, находившейся в начальный момент в i-м состоянии, за k периодов при использовании оптимальной политики. Руководствуясь принципом оптимальности, требующим независимо от начального состояния i и от начального выбора q далее действовать

оптимально, т. е. гарантировать максимум ожидаемого эффекта в последующем процессе, приходим к рекуррентным соотношениям следующего вида:

$$F_k(i) = \max_q \sum_{j=1}^{n} P_{ij}(q)[\, R_{ij}(q) + F_{k-1}(j)\,],\ i = 1 \,..\, n,\ k \geq 1; \qquad (1)$$

$$F_0(i) = 0,\ i = 1 \,..\, n.$$

Для процессов большой длительности использование приведенных соотношений требует существенных затрат времени даже при машинной реализации процесса вычислений. Если учесть, что при независимости значений вероятностей и эффектов от времени процесс обладает свойством стационарности, то в предположении *регулярности* (возможности прямого или опосредованного перехода из любого состояния в любое) полагаем для больших k

$$F_k(i) = F_i + k\,G, \qquad (2)$$

где G – средний эффект за период и F_i – составляющая суммарного эффекта, определяемая начальным состоянием. Подставляя (2) в (1) с учетом $\sum_{j=1}^{n} P_{ij}(q) = 1$, имеем систему функциональных уравнений

$$F_i + G = \max_q \sum_{j=1}^{n} P_{ij}(q)[\, R_{ij}(q) + F_j\,],\ i = 1 \,..\, n, \qquad (3)$$

которую можно решать приближением в поведениях.

Эту систему можно получить, если записать уравнение для бесконечношагового процесса с учетом дисконтирования, положить величину дисконтированного эффекта равной $F_i + G / (1 - \alpha)$ и принять $\alpha = 1$.

Для иллюстрации марковского процесса принятия решений рассмотрим ставшую ныне классической «задачу о такси» [21].

Таксист обслуживает окрестности трех городов (три возможных состояния) и может руководствоваться одним из трех выборов: ездить по городу в поисках случайного пассажира, ждать вызова по радио или поехать на стоянку и стать там в очередь.

Для каждого города-состояния (i) и каждого выбора (q) известны вероятности поездки в тот или иной город и соответствующие доходы, сведенные в следующей таблице:

Город	Выбор	Вероятности перехода			Значения дохода		
i	q	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
1	1	1/2	1/4	1/4	10	4	8
	2	1/16	3/4	3/16	8	2	4
	3	1/4	1/8	5/8	4	6	4
2	1	1/2	0	1/2	14	0	18
	2	1/16	7/8	1/16	8	16	8
3	1	1/4	1/2	1/4	10	2	8
	2	1/8	3/4	1/8	6	4	2
	3	3/4	1/16	3/16	4	0	8

Возьмем за начальное поведение $q_0 = (1, 1, 1)$, т. е. во всех городах придерживаться первого выбора. Для выбранного поведения строим систему n уравнений с $n + 1$ неизвестными

$$F_i + G = \sum_{j=1}^{n} P_{i\,j}(q_0)[\,R_{i\,j}(q_0) + F_j\,],\ i = 1 \,..\, n,$$

разрешимую с точностью до константы. Для нашего примера:

$F_1 + G = 1 / 2\ [10 + F_1] + 1 / 4\ [4 + F_2] + 1 / 4\ [8 + F_3]$,

$F_2 + G = 1 / 2\ [14 + F_1] + 1 / 2\ [18 + F_3]$,

$F_3 + G = 1 / 4\ [10 + F_1] + 1 / 2\ [2 + F_2] + 1 / 4\ [8 + F_3]$.

Полагая, например, $F_3 = 0$, получаем $F_1 = 0{,}95$, $F_2 = 7{,}16$ и $G = 9{,}32$; выбранная политика дает средний доход за поездку, равный 9,32.

Вычисляем

$$T_i(q) = \sum_{j=1}^{n} P_{i\,j}(q)[\,R_{i\,j}(q) + F_j\,]$$

при всех i и q и найденных значениях F_i:

$T_1(1) = 1 / 2\ [10 + 0{,}95] + 1 / 4\ [4 + 7{,}16\ ] + 1 / 4\ [8 + 0]$,

$T_1(2) = 1 / 16\ [8 + 0{,}95] + 3 / 4\ [2 + 7{,}16\ ] + 3 / 16\ [8 + 0]$,

$T_1(3) = 1 / 4\ [4 + 0{,}95] + 1 / 8\ [6 + 7{,}16] + 5 / 8\ [4 + 0]$,

$T_2(1) = 1 / 2\ [14 + 0{,}95] + 1 / 2\ [18 + 0]$,

$T_2(2) = 1 / 16\ [8 + 0{,}95] + 7 / 8\ [16 + 7{,}16] + 1 / 16\ [8 + 0]$,

$T_3(1) = 1 / 4\ [10 + 0{,}95] + 1 / 2\ [2 + 7{,}16] + 1 / 4\ [8 + 0]$,

$T_3(2) = 1 / 8\ [6 + 0{,}95] + 3 / 4\ [4 + 7{,}16] + 1 / 8\ [2 + 0]$,

$T_3(3) = 3 / 4\ [4 + 0{,}95] + 1 / 16\ [0 + 7{,}16] + 3 / 16\ [8 + 0]$.

Выбирая максимальное из значений $T_i(q)$ по q, получаем улучшенное поведение $q = (1, 2, 2)$. Строим и решаем систему уравнений:
$F_1 + G = 1 / 2\ [10 + F_1] + 1 / 4\ [4 + F_2] + 1 / 4\ [8 + F_3]$,
$F_2 + G = 1 / 16\ [8 + F_1] + 7 / 8\ [16 + F_2] + 1 / 16\ [8 + F_3]$,
$F_3 + G = 1 / 8\ [6 + F_1] + 3 / 4\ [4 + F_2] + 1 / 8\ [2 + F_3]$,
получая $F_3 = 0$, $F_2 = –3{,}88$, $F_1 = 12{,}85$, $G = 13{,}15$.

Попытка дальнейшего улучшения дает политику $q = (2, 2, 2)$, для которой $F_3 = 0$, $F_2 = –1{,}18$, $F_1 = 12{,}86$, $G = 13{,}34$. Очередная попытка улучшения приводит к той же политике, откуда напрашивается вывод о том, что оптимальная политика состоит в использовании второго выбора во всех городах со средним ожидаемым доходом за одну поездку, равным 13,34.

Величины F_k сами по себе не имеют значения, характеризуют относительный вклад начального состояния в общий ожидаемый эффект: в нашем случае при $F_3 = 0$ факт $F_2 < 0$, $F_1 > 0$ свидетельствует о предпочтительности исходить из города 1 и нежелательности в начальный момент оказаться в городе 2.

8.4. Примеры марковских процессов принятия решений

8.4.1. Задача о рекламе

Фирма рекламирует свою продукцию с помощью радио, телевидения и газет. Недельные затраты на рекламу оцениваются соответственно в 200, 900 и 300 денежных единиц. Фирма оценивает недельный сбыт тремя состояниями: удовлетворительным (3), хорошим (2) и отличным (1). Известны недельные доходы при разных объемах сбыта и способах рекламы

способ	радио			телевидение			газеты		
состояние	1	2	3	1	2	3	1	2	3
1	400	520	600	1000	1300	1600	400	530	710
2	300	400	700	800	1000	1700	350	450	800
3	200	250	500	600	700	1100	250	400	650

и переходные вероятности при различных видах рекламы

способ	радио			телевидение			газеты		
состояние	1	2	3	1	2	3	1	2	3
1	0,4	0,5	0,1	0,7	0,2	0,1	0,2	0,5	0,3
2	0,1	0,7	0,2	0,3	0,6	0,1	0	0,7	0,3
3	0,1	0,3	0,6	0,1	0,2	0,7	0	0,2	0,8

Убедитесь, что поиск способа рекламы, максимизирующего ожидаемый суммарный доход за определенный срок или средний недельный доход в процессе большой длительности, полностью аналогичен 8.3.

8.4.2. Задача ремонта оборудования

Станок находится в одном из трех состояний: хорошем (1), удовлетворительном (2) или плохом (3) и дает доход от выпуска продукции, равный 250, 200 и 50 денежных единиц. Есть возможность использовать обычный или капитальный ремонты или замену на новый станок со следующими затратами:

состояние	обычный ремонт	капитальный ремонт	замена на новый
1	10	15	30
2	50	60	100
3	150	180	200

Известны вероятности перехода при разных ремонтах

состояние	обычный ремонт			капитальный ремонт		
	1	2	3	1	2	3
1	0,8	0,2	0	0,9	0,1	0
2	0,1	0,6	0,3	0,5	0,4	0,1
3	0	0,1	0,9	0	0,7	0,3

Вероятность того, что новый станок будет находиться в соответствующем состоянии, равна 0,8, 0,15 и 0,05.

8.4.3. Простейшие задачи об очередях

Перед человеком стоит очередь из N претендентов на обслуживание. Известна полезность R от выстаивания очереди и вероятность P того, что в единицу времени будет обслужен один человек. За каждую единицу времени ожидания человек терпит убыток C. Стоит ли занимать очередь?

Убедитесь, что $F_k = \max[-C + P\ F_{k-1} + (1 - P)\ F_k, 0]$, $k = 1\ ..\ N$, $F_0 = R$, где F_k ожидаемый доход, получаемый от выстаивания очереди из k человек при оптимальной политике.

Покажите, что эти соотношения можно привести к виду $F_k = \max[(P\ F_{k-1} - C) / (1 - P), 0]$.

Поясните структуру оптимальной политики.

9. ЭЛЕМЕНТЫ ТЕОРИИ ИГР И СТАТИСТИЧЕСКИХ РЕШЕНИЙ

О сущности многих явлений природы и человеческого мышления наши представления ничтожны. Какой-либо детерминизм представлений – база для принимаемых решений – часто исчезает и подменяется неопределенностью. «Аннушка уже разлила подсолнечное масло… и Берлиоза выкинуло на рельсы» (кто мог предугадать эту причинно-следственную связь, кроме Воланда). Человек постоянно рискует при принятии решений, сталкиваясь с «дурной случайностью», природа которой для него остается тайной.

Тем не менее, параметры многих систем хотя и носят случайный характер, но допускают наличие вероятностных оценок и анализ с помощью аппарата теории вероятностей и математической статистики (выше мы уже рассматривали стохастические многошаговые процессы принятия решений).

Примечательно, что принимаемые рекомендации здесь не носят категорического характера, а звучат как «все будет хорошо с вероятностью 0,95», «велика вероятность того, что ваши ожидаемые издержки не превысят …», «вы достигнете максимума ожидаемого успеха, если два дня в неделю на вас будет коричневый костюм» (ответ на вопрос «в какие дни?» остается за кадром).

Наука о выработке таких суждений оформилась как самостоятельная математическая дисциплина «Теория игр и статистических решений» в 40-е годы XX столетия в работе Джона фон-Неймана и О. Моргенштерна [23].

9.1. Основные понятия теории игр

Теория игр занимается изучением так называемых конфликтных ситуаций, где сталкиваются интересы индивидов, партий, государств и т. п.

Как утверждал Г. Лейбниц, «...и игры заслуживают изучения; и если какой-нибудь проницательный математик посвятит себя их изучению, то получит много важных результатов, ибо нигде человек не показывает столько изобретательности, как в игре».

Во избежание терминологической путаницы следует различать понятие *игры* как совокупности правил и индивидуальных *партий* (реализации) игры: двусмысленная фраза «я играю в шахматы» мо-

жет означать знакомство с правилами этой популярной игры или факт пребывания за шахматной доской.

Важно различать понятия *выбор* и *ход*: игра состоит из последовательности ходов, тогда как партия – из последовательности выборов (мы часто путаем эти понятия – известный Остап Бендер, «сделавший ход e2–e4», фактически сделал соответствующий выбор из 20 допустимых. Самыми примитивными являются одноходовые игры, где всякая партия состоит из одного хода (например, дети показывают друг другу пальцы: если общее число показанных пальцев четное, выигрывает первый, в противном случае – второй).

Систему правил, однозначно определяющую выбор игрока в зависимости от сложившейся ситуации и позволяющую оценивать достижимые эффекты, называют *стратегией*. В одноходовых играх стратегия достаточно проста («всегда иди домой с цветами», «в половине случаев бери с собой зонтик», «вступай в коалицию с соседом слева» и т. п.), в многоходовых играх, где очередной выбор зависит от результата предыдущих, стратегия может оказаться много сложнее.

Каждая фиксированная стратегия игрока, где любой ситуации сопоставлен однозначно конкретный выбор, называется *чистой*. В реальности чаще используются так называемые *смешанные стратегии*, где какие-то выборы используются с некоторыми частотами.

Не претендуя на полноту, остановимся на классификации игр.

Интересы участников игры (игроков) могут оказаться несовпадающими и даже противоположными. В последнем случае игра называется *антагонистической*.

В игре могут участвовать два или более «игроков». Случай игры с одним участником (пасьянс, управление физическим объектом и т. д.), в сущности, является игрой двух лиц, где вторым участником выступает природа (судьба, рок, провидение).

Антагонистическую игру, где выигрыш одного коллектива равен проигрышу другого, называют *игрой с нулевой суммой*.

Игроки могут в игре выступать каждый за себя или объединяться в группы. В последнем случае игра называется *коалиционной*. Подобные игры представляют исключительный интерес для приложений, но результаты исследований в этой сфере достаточно скромны.

Игры, в которых игроки осведомлены о состоянии своем и партнеров, а также о прошлом поведении участников игры, относятся к категории игр *с полной информацией* (типичные примеры – шахматы, крестики-нолики и т. п.). Большинство же игр протекает в условиях неполной информации, где сведения о состоянии партнеров исчерпываются лишь вероятностными характеристиками (домино, карточные игры, игры против «природы»).

Если множество выборов определяется неким интервалом значений (угол атаки от 0 до 60°, случайное значение из (0, 1) и т. п.), то говорят о *непрерывных играх*. В противном случае можно говорить об игре *на дискретном множестве выборов*.

Простейшими среди игр являются одноходовые *матричные игры двух лиц с нулевой суммой*.

В такой игре игрок 1 имеет m и игрок 2 – n выборов. Если игрок 1 делает свой i-й выбор, а игрок 2 – свой j-й выбор, то выигрыш игрока 1 (проигрыш игрока 2) равен R_{ij}. Такая игра называется *матричной* и матрица $R = [R_{ij},\ i = 1\ ..\ m,\ j = 1\ ..\ n]$ называется *матрицей выигрышей* (*платежной матрицей*).

Выигрыш игрока 1 (проигрыш игрока 2) при оптимальной политике обоих игроков принято называть *ценой игры*.

При ведении игры разумный игрок с уважением относится к поведению партнера, ориентируется на его оптимальную политику и наказывает его за отступления от таковой.

Рассуждения игрока 1. Если Я запланирую использование i-го выбора, мой догадливый партнер для минимизации моего выигрыша сделает тот из своих выборов, который даст мне лишь $\min R_{ij}$. Соответственно, Я должен использовать тот выбор, который *гарантирует* мне выигрыш, не меньший

$$V_1 = \max_{i=1..m} \min_{j=1..n} R_{ij}\,.$$

Партнер, рассуждая аналогично, приходит к выводу о гарантированном проигрыше, не превышающем

$$V_2 = \min_{j=1..n} \max_{i=1..m} R_{ij}\,.$$

Если в матрице выигрышей существует элемент $R_{kl} = V_1 = V_2$, говорят о наличии оптимальной политики «в пространстве чистых

стратегий» и оптимальными выборами для игроков соответственно являются выборы k и l. Пару (k, l) называют *седловой* точкой и значение вышеуказанного элемента – ценой игры.

Пример 1. Пусть игра определяется матрицей

$$R = \begin{Vmatrix} 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 3 & 4 & 6 \\ 6 & 6 & 8 & 10 \end{Vmatrix}, \; V_1 = \max \begin{Vmatrix} 2 \\ 4 \\ 3 \\ 6 \end{Vmatrix} = 6, \; V_2 = \min[6, 6, 8, 10] = 6.$$

Цена игры равна 6, седловые точки (4, 1) и (4, 2), т. е. для игрока 1 – оптимален выбор четвертый, для игрока 2 равнозначны первый и второй. Здесь четвертая строка R *доминирует* над другими (ее элементы больше элементов других строк) – выбор 4 выгоднее других выборов при любой политике противника.

Обнаружение доминирования позволяет уменьшить размеры изучаемой матрицы исключением «невыгодных» строк и столбцов.

Пример 2. Пусть игра определяется матрицей

$$R = \begin{Vmatrix} 5 & 4 & 3 & 7 & 6 \\ 6 & 2 & 4 & 3 & 5 \\ 2 & 7 & 3 & 5 & 4 \\ 7 & 3 & 4 & 4 & 4 \end{Vmatrix}, \; V_1 = \max \begin{Vmatrix} 3 \\ 2 \\ 2 \\ 3 \end{Vmatrix} = 3, \; V_2 = \min[7, 7, 4, 7, 6] = 4.$$

Здесь $V_1 = 3$ определяет *гарантированный выигрыш* игрока 1 (он выиграет, по крайней мере, эту величину, если будет пользоваться первым или четвертым выбором). Значение $V_2 = 4$ определяет *гарантированный проигрыш* игрока 2 (можно гарантировать, что его проигрыш не превысит 4, если он будет пользоваться третьим из своих выборов).

Здесь равенство $V_1 = V_2$ не выполняется; *оптимальной чистой стратегии для игроков нет* и цена игры $V_1 \le V \le V_2$.

При отсутствии седловой точки среди чистых стратегий приходится искать таковую среди смешанных.

Если игрок 1 прибегает к своему выбору i с вероятностью P_i, а игрок 2 – к своему выбору j с вероятностью Q_j, то *ожидаемый выигрыш* игрока 1 (проигрыш игрока 2) равен

$$\sum_{i=1}^{m} \sum_{j=1}^{n} R_{ij} \, P_i \, Q_j = P^T R \, Q.$$

Основная теорема теории игр (*теорема Джона фон Неймана*) утверждает, что любая матричная игра с нулевой суммой всегда имеет седловую точку, т. е. существуют векторы P и Q такие, что

$$\max_P \min_Q P^T R\ Q = \min_Q \max_P P^T R\ Q = V$$

(V – цена игры, ожидаемый выигрыш – проигрыш при оптимальной политике партнеров).

9.2. Матричные игры и линейное программирование

Очевидно, что при отступлении игрока 1 от оптимальной политики при оптимальных действиях игрока 2 выигрыш игрока 1 будет меньше цены игры, а аналогичное поведение игрока 2 даст превышение его проигрыша над ценой игры:

$$P^T R\, Q_{\text{opt}} \le V = P_{\text{opt}}^T R\, Q_{\text{opt}} \le P_{\text{opt}}^T R\, Q.$$

Игрок 1 хочет сделать цену игры как можно большей и подобрать значения P_i так, чтобы ожидаемый выигрыш при любых выборах игрока 2 был больше цены игры. Аналогично игрок 2 хочет уменьшить гарантированный проигрыш и при любом выборе игрока 1 подобрать Q_j так, чтобы проигрыш был меньше цены игры.

Отсюда возникают две задачи:

максимизировать V	минимизировать V
при условиях	при условиях
$\sum_{i=1}^{m} R_{i\,j} P_i \ge V$, $j = 1 .. n$;	$\sum_{j=1}^{n} R_{i\,j} Q_j \le V$, $i = 1 .. m$;
$\sum_{i=1}^{m} P_i = 1$;	$\sum_{j=1}^{n} Q_j = 1$;
$P_i \ge 0$, $i = 1 .. m$.	$Q_j \ge 0$, $j = 1 .. n$.

Легко увидеть, что эти задачи образуют пару двойственных задач линейного программирования и решение матричной игры сводится к решению пары двойственных линейных программ.

Обратим внимание на то, что при увеличении элементов матрицы R на любую константу C цена игры увеличится на C и это изменение не окажет влияния на искомые вероятности выборов (так можно добиться, например, положительности элементов матрицы и, следовательно, цены игры).

В предположении $V > 0$ проведем замену переменных

$$X_i = P_i / V,\ Y_j = Q_j / V.$$

Соответственно, поставленные задачи можно преобразовать к задачам с меньшим числом переменных:

минимизировать $\sum_{i=1}^{m} X_i$	максимизировать $\sum_{j=1}^{n} Y_j$
при условиях	при условиях
$\sum_{i=1}^{m} R_{i\,j} X_i \geq 1,\ j = 1 \,..\, n;$	$\sum_{j=1}^{n} R_{i\,j} Y_j \leq 1,\ i = 1 \,..\, m;$
$X_i \geq 0,\ i = 1 \,..\, m.$	$Y_j \geq 0,\ j = 1 \,..\, n.$

Например, для игры с матрицей $\begin{Vmatrix} 1 & 2 & 3 \\ 4 & 0 & 1 \\ 2 & 3 & 0 \end{Vmatrix}$ возникают задачи:

максимизировать	минимизировать
$Y_1 + Y_2 + Y_3$	$X_1 + X_2 + X_3$
при	при
$Y_1 + 2\,Y_2 + 3Y_3 \leq 1;$	$X_1 + 4\,X_2 + 2\,X_3 \geq 1;$
$4\,Y_1 + Y_3 \leq 1;$	$2\,X_1 + 3\,X_3 \geq 1;$
$2\,Y_1 + 3\,Y_2 \leq 1;$	$3\,X_1 + X_2 \geq 1;$
$Y_1, Y_2, Y_3 \geq 0.$	$X_1, X_2, X_3 \geq 0.$

Решение этих задач симплексным методом дает оптимальные значения $X = \{11 / 37, 4 / 37, 5 / 37\}$, $Y = \{8 / 37, 7 / 37, 5 / 37\}$ и экстремумы целевых функций, равные 20 / 37.

Отсюда $V = 1 / \sum X_i = 1 / \sum Y_j = 37 / 20$, $P = \{11 / 20, 4 / 20, 5 / 20\}$, $Q = \{8 / 20, 7 / 20, 5 / 20\}$.

Другими словами, при многократной реализации игры игрокам рекомендуется использовать свои выборы с соответствующими вероятностями (отступление от этого требования чревато неприятностями).

А как поступить в реальности с этими вероятностями?

Если бы вероятности оказались равными 0,25, 0,5 и 0,25, то достаточно бросить монету: выпадет «решка» – делай выбор 2, в противном случае брось монету еще раз, при выпадении «решки» делай выбор 1 и при «орле» – выбор 3 (или наоборот). В нашем случае можно включить компьютер, войти в любую знакомую программную среду и обратиться к *датчику случайных чисел равномер-*

ного распределения в (0, 1)[18]. Если полученное число меньше 0,25 – делай выбор 1, при числе из интервала от 0,25 до 0,75 – выбор 2 и числе большем 0,75 – выбор 3.

9.3. Итеративный метод решения матричных игр

При больших размерах платежной матрицы приведенное выше решение довольно трудоемкое и чревато большой вычислительной погрешностью − традиционная беда всех так называемых «точных методов».

Здесь мы рассмотрим *итеративный метод Брауна – Робинсон* (разумеется, мы не помышляем о решении вручную), допускающий простую программную реализацию. Выполняем многократную реализацию игры на основе знания предыстории с последовательным совершенствованием стратегий.

Для примера возьмем задачу, которую мы только что решили.

Пусть игрок 1 случайно сделал выбор 1 с ожидаемыми выигрышами 1, 2, 3. Противник, стремясь минимизировать свой проигрыш, прибегнет к выбору 1 с ожиданием проигрыша 1, 4, 2. Игрок 1 в стремлении максимизировать свой выигрыш прибегнет к выбору 2, что даст ему надежду на суммарный выигрыш (1 + 4, 2 + 0, 3 + 1). Но тогда его противник найдет среди этих значений меньшее и прибегнет к выбору 2 с ожидаемым суммарным проигрышем (1 + 2, 4 + 0, 2 + 3) и т. д.

Шаг	Выбор i	Суммарный выигрыш			Выбор j	Суммарный проигрыш		
1	1	1	2	3	1	1	4	2
2	2	5	2	4	2	3	4	5
3	3	7	5	4	3	6	5	5
4	1	8	7	7	2	8	5	8
5	1	9	9	10	1	9	9	10
6	3	11	12	10	3	12	10	10
7	1	12	14	13	1	13	14	12
8	2	16	14	13	3	16	15	12
9	1	17	16	16	2	18	15	15
10	1	18	18	19	1	19	19	17

[18] Вообще-то эти числа, получаемые по определенному правилу, но подчиняющиеся определенным статистическим критериям, называют *псевдослучайными*.

Этот процесс реализуется достаточно большое число раз (см. в 5.1 характеристику методов Монте – Карло) с последующим поиском частоты использования выборов и усреднением значений выигрышей – проигрышей.

В результате 10 выборов для игрока 1 частоты составили 0,6, 0,2, 0,2; для игрока 2 – 0,4, 0,3 и 0,3; оценка цены игры находится в диапазоне от 1,7 до 1,9.

9.4. Многошаговые игры

Предыдущее рассмотрение игр проводилось в предположении, что игра является одноходовой и реализация игры может осуществляться большое число раз.

Не замахиваясь на игры, подобные шахматам, рассмотрим так называемую *игру на выживание* – многоходовую игру с ограниченными ресурсами, где политика игроков (последовательность выборов) зависит от результата предыдущих выборов и от длительности игры [8].

Пусть общий начальный ресурс игроков $A + B = C$ и игра продолжается до разорения одного из игроков. Обозначим через $F(A)$ ожидаемую вероятность выживания (шансы не разориться) игрока 1 при его начальном ресурсе A и оптимальной политике обоих игроков.

Тогда

$$F(A) = \max_P \min_Q \sum_{i=1}^{m} \sum_{j=1}^{n} P_i\, Q_j F(A + R_{i\,j}) =$$

$$= \min_Q \max_P \sum_{i=1}^{m} \sum_{j=1}^{n} P_i\, Q_j F(A + R_{i\,j}),$$

$$F(A \le 0) = 0,\ F(A \ge C) = 1.$$

Если игра не обладает чистыми оптимальными стратегиями, то оптимальные значения вероятностей использования выборов соответствуют внутренним точкам множества планов ($0 < P < 1,\ 0 < Q < 1$) и напрашивается мысль для поиска оптимальных P, Q прибегнуть к аппарату производных.

Пример. Рассмотрим игру на выживание с матрицей $\begin{Vmatrix} 2 & -1 \\ -2 & 1 \end{Vmatrix}$ при полном капитале игроков $C = 4$.

Обозначив вероятности соответствующих выборов игроков через P, $1 - P$, Q, $1 - Q$, имеем $F(A \le 0) = 0$, $F(A \ge 4) = 1$,

$$F(1) = \max_P \min_Q [P\,Q\,F(3) + P\,(1 - Q)\,F(0) + (1 - P)\,Q\,F(-1) +$$

$$+ (1 - P)\,(1 - Q)\,F(2)] = \max_P \min_Q [P\,Q\,F(3) + (1 - P)\,(1 - Q)\,F(2)],$$

$$F(2) = \max_P \min_Q [P\,Q\,F(4) + P\,(1 - Q)\,F(1) + (1 - P)\,Q\,F(0) +$$

$$+ (1 - P)\,(1 - Q)\,F(3)] =$$

$$= \max_P \min_Q [P\,Q + P\,(1 - Q)\,F(1) + (1 - P)\,(1 - Q)\,F(3)],$$

$$F(3) = \max_P \min_Q [P\,Q\,F(5) + P\,(1 - Q)\,F(2) + (1 - P)\,Q\,F(1) +$$

$$+ (1 - P)\,(1 - Q)\,F(4)] =$$

$$= \max_P \min_Q [P\,Q + P\,(1 - Q)\,F(2) + (1 - P)\,Q\,F(1) + (1 - P)\,(1 - Q)].$$

Отыскивая частные производные, строим системы уравнений для поиска оптимальных значений $P(A)$, $Q(A)$:

1) $Q\,F(3) - (1 - Q)\,F(2) = 0$, $P\,F(3) - (1 - P)\,F(2) = 0$;

2) $Q + (1 - Q)\,F(1) - (1 - Q)\,F(3) = 0$, $P - P\,F(1) - (1 - P)\,F(3) = 0$;

3) $Q + (1 - Q)\,F(2) - Q\,F(1) - 1 - Q) = 0$, $P - P\,F(2) + (1 - P)\,F(1) - (1 - P) = 0$.

Решение приведенных систем дает

$$P(1) = Q(1) = \frac{F(2)}{F(2)+F(3)};\ P(2) = \frac{F(3)}{1-F(1)+F(3)}\ ,\ Q(2) = \frac{F(3)-F(1)}{1-F(1)+F(3)};$$

$$P(3) = \frac{1-F(1)}{2-F(1)-F(2)}\ ,\ Q(3) = \frac{1-F(2)}{2-F(1)-F(2)}.$$

Подставляя полученное в исходные выражения функций, имеем опять-таки нелинейную систему относительно $F(1)$, $F(2)$, $F(3)$

$$F(1) = \frac{F(2)\,F(3)}{F(2)+F(3)};\ F(2) = \frac{F(3)}{1-F(1)+F(3)};\ F(3) = \frac{1-F(1)\,F(2)}{2-F(1)-F(2)}.$$

Решая эту систему, имеем $F(1) = 0{,}3$, $F(2) = 0{,}5$, $F(3) = 0{,}7$ и $P(1) = 0{,}41$, $P(2) = 0{,}5$, $P(3) = 0{,}59$, $Q(1) = 0{,}41$, $Q(2) = 0{,}3$, $Q(3) = 0{,}41$.

А если m, $n >> 2$? В этой ситуации трудности решения нелинейных систем очевидны.

Другим примером многошаговых игр могут служить *игры погони*. В качестве простейшего примера можно привести задачу для

двух игроков, расположившихся на прямой на расстоянии d. На каждом шаге игры игроки могут *одновременно* смещаться влево или вправо при полной информации о позиции друг друга. После очередного шага игрок 2 уплачивает игроку 1 величину $G(S)$, где S – расстояние между ними. С вероятностью $A(d)$ игра может быть продолжена и с вероятностью $1 - A(d)$ окончена.

Существенно больший интерес может представить игра погони на плоскости или в пространстве, где устанавливается принципиальная возможность поимки одного игрока другим или отыскивается траектория, минимизирующая время поимки. Эти игры относятся к так называемым *непрерывным многошаговым играм*.

9.5. Статистические решения: основные понятия

Выбор наилучших решений в условиях полной и неполной информации – одно из основных занятий людей. Принятие управленческих решений в условиях неполной или неточной информации сопряжено с неизбежным риском понести немалые убытки, причинить вред здоровью или вместо Сочи оказаться в «солнечном» Магадане в случае принятия ошибочного решения.

Когда мы знакомились с азбучными истинами теории игр, то предполагали, что участниками игры являются люди, способные принимать разумные решения. Другая ситуация возникает в так называемых *играх против природы*, где человек, разумный по предположению, противостоит непознанному им явлению. Человек обычно сетует на судьбу и не благодарит ее, когда «повезло». Едва ли стоит связывать с природой (судьбой, высшими силами и т. п.) априорную злонамеренность или предрасположенность по отношению к человеку, хотя человек и делает подчас многое, чтобы вывести окружающую среду из состояния равновесия.

Теория статистических решений может быть истолкована как теория поиска оптимального недетерминированного поведения в условиях неопределенности. Современная концепция статистического решения выдвинута А. Вальдом. В рамках этой концепции поведение считается оптимальным, если оно *минимизирует риск в последовательных экспериментах*, т. е. математическое ожидание убытков статистического эксперимента. В такой постановке любая задача статистических решений может рассматриваться как игра двух лиц, в которой одним из игроков является «природа».

Если быть более точным в терминологии, то следует различать *ситуацию риска*, когда имеется статистическая информация о подобных решениях и существует возможность оценить вероятности, связанные с последствиями принятия решения, и *ситуацию неопределенности*, когда нет возможности объективно оценить указанные вероятности и остается прибегнуть к экспертным оценкам или надеяться на собственное озарение.

Формализованная (*математизированная*) постановка задачи выбора решения в условиях неопределенности (риска) сводится к следующему.

Пусть задан некоторый вектор $S = (S_1, S_2, \ldots, S_n)$, описывающий n состояний внешней среды, и вектор $X = (X_1, X_2, \ldots, X_m)$, описывающий m допустимых решений. Требуется найти вектор $X^* = (0, 0, \ldots, 0, X_i, 0, \ldots, 0)$, который обеспечивает оптимум некоторой *функции полезности* $W(X, S)$ по некоторому критерию K. Функция полезности сопоставляет каждому состоянию S внешней среды и каждому предлагаемому решению X значение так называемой полезности (дохода, прибыли, эффективности и т. п.). Информацию об указанной функции можно представлять матрицей размерности $m \times n$ с элементами $W_{ij} = F(X_i, S_j)$, где F – *решающее правило*.

Формирование решающего правила во многом предопределяет конечный результат расчетов (в случае его ошибочности едва ли принимаемое решение окажется наилучшим) и возможно лишь при достаточно четкой экономической постановке задачи.

Так, арендуя зал на m мест при заранее неизвестном числе n посетителей, предприниматель стоит перед выбором: потерять возможную прибыль, если посетителей окажется больше чем мест в зале, или зря потратить деньги на аренду просторного помещения при малом числе посетителей. Если стоимость билета равна k и затраты на аренду равны $a(m)$, то функция полезности $W(m, n) = k \cdot \min(m, n) - a(m)$ и следует рассмотреть ее значения для возможных значений m и n.

Планируя выпуск новой продукции, надо *заблаговременно* закупить станки. Возможна поставка до 50 станков (комплект поставки – 10), минимальный объем поставок – 20 станков. Производительность одного станка составляет 2 изделия в год, каждое из которых приносит доход в размере 21,9 тыс. руб. Оптовая цена станка составляет 4,775 тыс. руб., его содержание – 3,6 тыс. руб. Затраты

на подготовку производства составляют 25,5 тыс. руб. и не зависят от числа станков. Спрос на изделия прогнозируется в диапазоне от 20 до 100 штук.

Соответственно, вектор решений об объеме поставок $X = (20, 30, 40, 50)$, состояние же спроса можно описать вектором $S = (20, 40, 60, 80, 100)$ (едва ли резонно брать меньший шаг).

Если решающее правило сформулировать как «доход – издержки», то матрица полезности:

$$W(X, S) = 21{,}9 \min(2X, S) - 23{,}6\ X - 25{,}5 - 4{,}775\ X$$

(последнее вычитаемое может быть модифицировано, если закупаемые станки будут эксплуатироваться не один год).

Можно привести множество подобных примеров принятия решения в аналогичной ситуации.

Если *в ситуации риска* имеющиеся статистические данные позволяют оценить вероятность $P(S)$ того или иного состояния внешней среды, то можно найти математическое ожидание полезности и сделать выбор X_i, обеспечивающий его максимум:

$$W = \max_{i=1..m} \sum_{j=1}^{n} W_{ij} P_j . \qquad (1)$$

В ситуации неопределенности многообразие критериев (подходов к выбору наилучшего решения) несколько больше.

Критерий Лапласа. При незнании вероятности возникновения того или иного состояния внешней среды нет оснований полагать, что какое-то из них возникает чаще других (*принцип недостаточного основания*), тогда принимают *равные значения* $P_j = 1 / n$, находят *средний эффект* для каждого из рассматриваемых вариантов решения, выбирая тот, для которого средний эффект максимален:

$$W = \max_{i=1..m} \frac{1}{n} \sum_{j=1}^{n} W_{ij} . \qquad (2)$$

Критерий Вальда (критерий наибольшей осторожности, или пессимистический критерий). *Для каждого* варианта решения X_i рассматривается *самый худший отклик среды* (наименьшее из W_{ij}) и среди них отыскивается гарантированный максимальный эффект:

$$W = \max_{i=1..m} \min_{j=1..n} W_{ij} . \qquad (3)$$

Критерий Гурвица. Ориентация на самый худший исход является своеобразной перестраховкой, однако опрометчиво выбирать и излишне оптимистичную политику. Критерий Гурвица[19] предлагает некоторый компромисс:

$$W = \max_{i=1..m}\left[\alpha \max_{j=1..n} W_{ij} + (1-\alpha)\min_{j=1..n} W_{ij}\right], \qquad (4)$$

где параметр $0 \le \alpha \le 1$ выступает как *коэффициент оптимизма*.

При $\alpha = 0$ критерий Гурвица превращается в критерий Вальда, при $\alpha = 0{,}5$ шансы на успех и неудачу мы предполагаем равновероятными, при $\alpha = 0{,}8$ мы более радужно расцениваем свои шансы на успех. Выбор же $\alpha = 1$ вызывает определенные сомнения в трезвом подходе к решаемой проблеме. В некоторых сферах человеческой деятельности, например при оценке сроков выполнения работ, предпочитают выбор $\alpha = 0{,}4$ (выводы делайте сами).

Особое место занимает **критерий Сэвиджа**. При выборе решения по этому критерию:

1) матрице полезности сопоставляется новая матрица – *матрица сожалений* с элементами $D_{ij} = W_{ij} - \max_i (W_{ij})$, которые отражают убытки от ошибочного действия, т. е. выгоду, упущенную в результате принятия i-го решения в j-м состоянии;

2) для матрицы D выбирается решение по пессимистическому критерию Вальда, дающее наименьшее значение максимального сожаления:

$$W = \max_{i=1..m} \min_{j=1..n} D_{ij} \qquad (5)$$

(минимум упущенной выгоды при принятии данного решения).

Вполне логично, что *различные критерии приводят к различным выводам относительно наилучшего решения*, при этом они не

---

[19] Леонид (Леон) Гурвиц (1917 – 2008) – выходец из России, удостоенный в 2007 г. Нобелевской премии за создание основ теории механизмов распределения. Как отмечал Нобелевский комитет, теория, созданная им и развитая затем нобелевскими лауреатами Эриком Маскиным и Роджером Майерсоном, помогла «выявить эффективные торговые механизмы, схемы регулирования и процедуры голосования», а также значительно расширила знания об особенностях оптимального распределения экономических ресурсов.

категоричны и сопровождаются комментарием «если…».

Возможность выбора критерия дает свободу лицам, принимающим экономические решения (если они, конечно, располагают достаточной информацией для постановки подобной задачи). Любой критерий должен согласовываться с намерениями решающего задачу и соответствовать его характеру, знаниям и убеждениям.

Еще один пример постановки и решения задачи.

Некая фирма, идя навстречу пожеланиям сельхозпроизводителей, которые нуждаются в хранении зерна, решила построить элеватор и эксплуатировать его в течение 5 лет (за эти годы хотелось бы рассчитаться за беспроцентные кредиты на строительство и получить заслуживающую внимания выгоду). Имеются типовые проекты элеватора мощностью на 20, 30, 40, 50 и 60 тыс. ц зерна.

Посевные площади сельхозрайона составляют 1430 га. Строительство элеватора мощностью 20 тыс. ц обойдется в 300 тыс. денежных единиц (д. е.) и эти затраты возрастают на 10 % с ростом мощности элеватора на 10 тыс. ц. Согласование проекта с властями, реклама элеватора, строительство подъездных путей и вспомогательных сооружений обойдется в 185 тыс. д. е. Затраты на эксплуатацию элеватора мощностью 20 тыс. ц составляют 10 тыс. д. е. и убывают на 10 % при увеличении мощности на 10 тыс. ц. За хранение зерна на счет элеватора вносится плата в размере 10 д. е. за 1 ц. Урожайность в данном районе – от 14 до 20 ц с 1 га. Практика показывает, что зерновой запас элеватора расходуется за год и возобновляется при новом урожае. *Какой элеватор выгоднее построить*?

Если бы мы точно знали, каким будет урожай в эти пять лет, решение задачи можно было бы поручить добросовестному школьнику. Но в реальности, построив большой хороший элеватор, оснащенный автоматикой, и затратив значительную сумму, мы можем столкнуться с малым урожаем и, соответственно, с малым доходом от хранения. С другой стороны, построенный малый элеватор может не вместить большой урожай и будет упущена возможная выгода.

Примем типовые проекты элеваторов за *вектор допустимых решений*:

$$X = \{x_i\} = (20, 30, 40, 50, 60),\ i = 1 .. 5,$$

оценки урожайности в данном районе (здесь можно взять и другую сетку значений) примем за *вектор состояний внешней среды*:

$$S = \{s_j\} = (14, 15, 16, 17, 18, 19, 20),\ j = 1 .. 7,$$

и попытаемся построить *матрицу полезности* (эффективности принятия i-го решения в случае j-й урожайности).

Затраты на сооружение элеватора и инфраструктуру составляют 300000 + 30000 (x_i – 20) + 185000. Эти деньги кредиторам придется возвратить равными долями в течение пяти лет. Ежегодные затраты на эксплуатацию элеватора равны 10000 – 1000 (x_i – 20).

Что касается доходной части, то здесь приходится учесть, что деньги мы получаем от реального урожая, который оказалось возможным поместить в элеваторе. Реальный урожай определяется величиной 1430 S_j ц. Возможности элеватора – 1000 x_i ц. Следовательно, наибольший сохраняемый объем зерна не превысит минимального из этих значений и плата за его хранение составит $10 \times \min(1430\ S_j, 1000\ x_i)$.

Если просуммировать доходы – затраты за пять лет и разделить на 5, то мы получаем средний доход в год (принимаем его за функцию полезности) в виде матрицы с элементами

$W_{ij} = 10 \times \min(1430\ S_j, x_i\ 1000) –$

$–[60000 + 6000\ (x_i – 20) + 37000] – [10000 – 100\ (x_i – 20)].$

Выполнив несложные расчеты, заполним матрицу $\{W_{ij}\}$:

	$S_1 = 14$	$S_2 = 15$	$S_3 = 16$	$S_4 = 17$	$S_5 = 18$	$S_6 = 19$	$S_7 = 20$
$x_1 = 20$	**93000**	93000	93000	93000	93000	93000	93000
$x_2 = 30$	88200	**102500**	**116800**	**131100**	**145400**	**159700**	**174000**
$x_3 = 40$	83200	97500	111800	126100	140400	154700	169000
$x_4 = 50$	78200	92500	106800	121100	135400	149700	164000
$x_5 = 60$	73200	87500	101800	116100	130400	144700	159000

Полагая шансы на ту или иную урожайность равновероятными, находим средние значения полезности $W_i(L) = \frac{1}{n}\sum_{j=1}^{n} W_{ij}$ для каждого из вариантов решения, например,

$W_2 = (88200 + 102500 + 116800 + 131100 + 145400 + 159700 +$

$+ 174000) / 7 = 131100$, и по критерию Лапласа устанавливаем оптимальность выбора проекта мощностью 30 тыс. ц с ожидаемой прибылью 131,1 тыс. д. е.

Если выбирать самый худший вариант по величине прибыли для каждой альтернативы (наименьшие значения $W_{ij}(B)$ полезности

в строках матрицы *W*), то можно из таких самых плохих оценок эффекта наших возможных выборов выбрать наилучший. Таким образом, по критерию Вальда обнаруживаем, что следует построить элеватор мощностью 20 тыс. ц и оправдываться, что даже в самом худшем случае максимум возможной прибыли гарантирован на уровне 93 тыс. д. е.

Обратившись к оценкам *по критерию Гурвица* при трех различных уровнях оптимизма (α = 0,2; 0,5; 0,8), обнаруживаем целесообразность выбора проекта элеватора мощностью 30 тыс. ц с ожидаемой прибылью соответственно 105360, 131100, 156840 д. е.

	$W_i(L)$	$W_i(B)$	α = 0,2	α = 0,5	α = 0,8
x_1 = 20	93000	**93000**	93000	93000	93000
x_2 = 30	**131100**	88200	**105360**	**131100**	**156840**
x_3 = 40	126100	83200	100360	126100	151840
x_4 = 50	121100	78200	95360	121100	146840
x_5 = 60	116100	73200	90360	116100	141840

При подходе с позиций *критерия Сэвиджа* (упущенных возможностей и последующего сожаления об этом) строим матрицу сожалений *D*, вычитая из столбцов матрицы полезности наибольшие значения, и применяем к ней пессимистический критерий Вальда, дающий наименьшее значение максимального сожаления.

	S_1 = 14	S_2 = 15	S_3 = 16	S_4 = 17	S_5 = 18	S_6 = 19	S_7 = 20	min
x_1=20	0	–9500	–23800	–38100	–52400	–66700	–81000	–81000
x_2=30	–4800	0	0	0	0	0	0	**–4800**
x_3=40	–9800	–5000	–5000	–5000	–5000	–5000	–5000	–9800
x_4=50	–14800	–10000	–10000	–10000	–10000	–10000	–10000	–14800
x_5=60	–19800	–15000	–15000	–15000	–15000	–15000	–15000	–19800

Для нашего примера по этому критерию оптимален проект элеватора мощностью 30 тыс. ц (прибегая к этому выбору, мы рискуем потерять прибыль до 4800 д. е.).

Таким образом, практически по всем критериям отдается предпочтение проекту 30 тыс. ц, и лишь глубокий пессимист во взглядах на ожидаемый урожай отдаст предпочтение проекту 20 тыс. ц с гарантией ожидаемой прибыли лишь в 93 тыс. д. е. и значительных упущенных возможностей. Остальные проекты рассматривать явно нецелесообразно.

10. ВВЕДЕНИЕ В МОДЕЛИРОВАНИЕ СИСТЕМ

Выше мы уже затрагивали принципы, лежащие в основе математического моделирования, проблемы и методы решения задач, которые в свое время было принято называть задачами исследования операций (*operation research*). Сегодня методология математического моделирования бурно развивается, охватывая все новые сферы – от разработки больших технических систем и управления ими до анализа сложнейших экономических и социальных процессов. Вычислительные эксперименты с моделями объектов позволяют, опираясь на мощь современных вычислительных методов и технических средств, изучать объекты в достаточной полноте, недоступной чисто теоретическим подходам.

В 50 – 60-е годы прошлого столетия массовое распространение упомянутых методов тормозилось скромными возможностями вычислительной техники и программисты экономили каждую ячейку памяти и каждую секунду, превращая программирование в искусство, но сохранялась надежда. Однако, несмотря на фантастический рост вычислительных возможностей компьютеров, и сегодня удовлетворить потребности моделирования *больших* систем с помощью ставших уже классическими численных методов нереально.

Ниже мы рассмотрим современные формальные подходы к определению терминов, целей и задач моделирования, создающие возможность определенной автоматизации построения систем моделирования.

10.1. Модели систем и системы моделирования

Как уже говорилось выше, *моделирование* – процесс изучения системы путем ее замены более удобным для экспериментального исследования объектом (моделью), сохраняющим существенные черты оригинала.

В настоящее время формируется комплексная методология научных исследований – *математическое моделирование* и *вычислительный эксперимент*. Ее сущность состоит в замене исходного объекта его математической моделью и исследовании этой модели с помощью средств вычислительной техники.

Основу математического моделирования составляет триада «*модель – алгоритм – программа*».

Математические модели реальных исследуемых процессов становятся все более сложными. Если четверть века назад, например, экономико-математические модели ограничивались использованием аппарата линейной алгебры и традиционной математической статистики, то необходимость учета нелинейности поведения экономических характеристик заставила обратиться даже к более сложному аппарату вплоть до систем уравнений с частными производными, которыми до сих пор пользовались физики при моделировании прогноза погоды или летательных аппаратов.

Вычислительный эксперимент носит междисциплинарный характер. В совместных исследованиях участвуют специалисты в прикладной области, прикладной и вычислительной математике, по прикладному и системному программному обеспечению.

На первом этапе вычислительного эксперимента выбирается (или строится) модель исследуемого объекта, отражающая в математической форме важнейшие его свойства – законы, которым он подчиняется, связи, присущие составляющим его частям, и т. д. Математическая модель (ее основные фрагменты) исследуется традиционными средствами прикладной математики для получения предварительных знаний об объекте. Математическое моделирование базируется на четкой формулировке основных понятий и предположений, анализе адекватности используемых моделей, контроле точности вычислительных алгоритмов, квалифицированной обработке и анализе результатов расчетов.

Второй этап связан с разработкой вычислительного алгоритма для реализации модели на компьютере. Вычислительные алгоритмы не должны искажать основные свойства модели и, следовательно, исходного объекта, они должны быть адаптируемыми к особенностям решаемых задач и используемых вычислительных средств.

На третьем этапе создается программное обеспечение для реализации модели и алгоритма на компьютере. Программный продукт должен учитывать важнейшую специфику математического моделирования, связанную с многовариантностью расчетов.

Опираясь на триаду «модель – алгоритм – программа», исследователь получает в руки универсальный, гибкий и недорогой инструмент, который вначале тестируется на решении содержательного набора пробных задач. После этого проводится широкомасштабное исследование математической модели для получения необхо-

димых качественных и количественных свойств и характеристик исследуемого объекта [29].

Многие моделирующие системы, идеологически разработанные в 1970 – 1980-х гг., претерпели эволюцию вместе с компьютерной техникой и операционными системами (например, GPSS – General Purpose Simulation System) и эффективно используются в настоящее время на новых компьютерных платформах. Кроме того, в конце 1990-х гг. появились принципиально новые моделирующие системы. В последние два десятилетия XX века системы моделирования стали использоваться более активно, хотя каждая из них имела множество недостатков. Среди наиболее известных систем можно выделить GASP-IV (структурированный язык программирования, похожий на Фортран, набор методов моделирования, датчиков случайных чисел, требуется высокая математическая подготовка), SIMULA-67 (аналогичен GASP-IV по возможностям), GPSS-V (низкое быстродействие, мощный инструментарий, отсутствие возможности включать непрерывные динамические компоненты в модель), SLAM-II (чрезвычайно сложна в освоении) [30].

К концу XX века существенно выросло многообразие систем моделирования самого разного назначения, появляются специализированные системы для различных предметных областей. Наибольшее распространение получили следующие системы: Process Charter (ориентирована на дискретное моделирование, проста в освоении, однако имеет ограниченный набор средств моделирования), PowerSim (непрерывные модели, множество встроенных функций, сложная специальная система обозначений), IThink (те же возможности, что и PowerSim, большая подборка готовых элементов моделей для неподготовленных пользователей, малое число функций), Extend+BPR (гибкий и мощный пакет, реализован на платформе Macintosh), ReThink (аналогичен Extend+BPR), Pilgrim (отечественная разработка, высокое быстродействие, возможность встраивать блоки с помощью стандартного языка C++), СИМПАС (СИстема-Моделирования-на-ПАСкале, сложна в моделировании из-за использования языка общего назначения, имеет множество специальных процедур и функций, ориентирована на моделирование информационных систем, компьютерных сетей) [30].

Успех или неудача экспериментов с моделями сложных систем существенным образом зависит от набора аппаратно-программных

средств, представляемых разработчику или пользователю-исследователю машинной модели. В настоящее время существует большое количество языков имитационного моделирования – специальных языков программирования имитационных моделей на ЭВМ – и перед разработчиком машинной модели возникает проблема выбора языка, наиболее эффективного для целей моделирования конкретной системы [31].

10.2. Модели систем и их свойства

Все то, на что направлена человеческая деятельность, называется *объектом* (лат. *objectum* – предмет). Одна из целей деятельности человека – упорядочение получения и обработки информации об объектах, которые существуют вне нашего сознания и взаимодействуют между собой и внешней средой.

В научных исследованиях большую роль играют гипотезы, т. е. предсказания (предположения), основывающиеся на ограниченном количестве опытных данных. Большое значение при формулировании и проверке гипотез имеет метод *аналогии* – аналогии с проверенными на практике научными положениями, поиска сходства объектов.

Гипотезы и аналогии, отражающие реальный, объективно существующий мир, должны обладать наглядностью или сводиться к удобным для исследования логическим схемам. Такие логические схемы, упрощающие рассуждения и логические построения или позволяющие проводить эксперименты, уточняющие природу явлений, называются *моделями*. Другими словами, *модель* (лат. *modulus* – мера) – это *объект-заместитель объекта-оригинала, обеспечивающий изучение некоторых свойств оригинала*.

Совокупность сведений об исследуемой системе и условиях, при которых необходимо провести исследование, называют *описанием*. Оно может быть представлено схемами, текстами, формулами, таблицами экспериментальных данных, характеризующих предполагаемую структуру и функционирование системы. Также оно содержит характеристики внешних воздействий и окружающей систему среды. Таким образом, описание определяет предполагаемый алгоритм работы системы, который может быть формально рассмотрен как некоторая функция внешних воздействий.

В большинстве систем влияние всех факторов по отдельности не помогает оценить синергетического (совместного) эффекта от

поведения системы. Такие системы, в которых при вычленении компонент могут быть потеряны принципиальные свойства, а при добавлении компонент возникают качественно новые свойства, называют *сложными*. Модель сложной системы, основанная на принципах лишь покомпонентного анализа, будет неадекватна изучаемой системе. Возможным выходом из положения является построение модели на основе *синтеза* компонент. Синтетические модели являются практически единственной альтернативой в социологии, долгосрочных прогнозах погоды, в макроэкономике, медицине. В последнее время синтетические информационные модели широко используются и при изучении технических и инженерных систем. В ряде приложений информационные и математические компоненты могут составлять единую модель.

Сложные системы характеризуются выполняемыми процессами (функциями), структурой и поведением во времени. Для адекватного моделирования этих аспектов в автоматизированных информационных системах различают *функциональные*, *информационные* и *поведенческие* модели, пересекающиеся друг с другом. *Функциональная* модель системы описывает совокупность выполняемых системой функций, характеризует морфологию системы (ее построение) – состав функциональных подсистем, их взаимосвязи. *Информационная* модель отражает отношения между элементами системы в виде структур данных (состав и взаимосвязи). *Поведенческая* (*событийная*) модель описывает информационные процессы (динамику функционирования), в ней фигурируют такие категории, как состояние системы, событие, переход из одного состояния в другое, условия перехода, последовательность событий [32].

Можно выделить несколько типов информационных моделей, отличающихся по характеру запросов к ним, в частности:

1) моделирование отклика системы на внешнее воздействие;

2) классификация внутренних состояний системы;

3) прогноз динамики изменения системы;

4) оценка полноты описания системы и сравнительная информационная значимость параметров системы;

5) оптимизация параметров системы по отношению к заданной функции ценности;

6) адаптивное управление системой.

Основным принципом информационного моделирования явля-

ется принцип «*черного ящика*». В противоположность аналитическому подходу, при котором моделируется внутренняя *структура* системы, в синтетическом методе «черного ящика» моделируется внешнее *функционирование* системы. С точки зрения пользователя модели структура системы спрятана в черном ящике, который имитирует поведенческие особенности системы.

Кибернетический принцип «черного ящика» был предложен в рамках теории идентификации систем, в которой для построения модели системы предлагается широкий параметрический класс базисных функций (уравнений), а сама модель *синтезируется* путем выбора параметров из условия наилучшего при заданной функции ценности соответствия решений уравнений поведению системы. При этом структура системы никак не отражается в структуре уравнений модели [32, 33].

Можно выделить три основных области применения моделей – обучение, научные исследования, управление.

При обучении с помощью моделей достигается высокая наглядность отображения различных объектов и облегчается передача знаний о них. Это в основном модели, позволяющие описать и объяснить поведение системы. В научных исследованиях модели служат средством получения, фиксирования, упорядочивания новой информации, обеспечивая развитие теории и практики. В управлении модели используются для обоснования решений. Такие модели должны обеспечить как описание, так и объяснение (предсказание) поведения систем.

10.3. Виды и классификации моделирования систем

В основе моделирования лежит *теория подобия*. При моделировании абсолютное подобие не имеет места, и исследователи стремятся к тому, чтобы модель достаточно хорошо отображала интересующую их сторону функционирования объекта. При этом должен достигаться разумный компромисс между точностью воспроизведения и сложностью необходимых средств. Общими функциями моделирования являются описание, объяснение и прогнозирование поведения реальной системы. Типовыми целями моделирования могут быть поиск оптимальных или близких к оптимальным решений, оценка эффективности решений, определение свойств системы (чувствительности к изменению значений характеристик и

др.), установление взаимосвязей между характеристиками системы, перенос информации во времени [32].

Одна из возможных классификаций видов моделирования систем приведена на рис. 28.

Рис. 28. Классификация видов моделирования систем

В качестве одного из первых признаков классификации видов моделирования можно выбрать степень полноты модели и разделить модели на *полные*, *неполные* и *приближенные*. В основе полного моделирования лежит полное подобие, которое проявляется как во времени, так и в пространстве. В основе приближенного лежит подобие, при котором некоторые стороны функционирования реального объекта не моделируются совсем [30, 32].

В зависимости от характера изучаемых процессов в системе все виды моделирования могут быть разделены на детерминирован-

ные и стохастические, статические и динамические, дискретные, непрерывные и дискретно-непрерывные [30].

Детерминированное моделирование используется для моделирования детерминированных процессов, в которых предполагается отсутствие всяких случайных воздействий. *Стохастическое* (*статистическое* или *вероятностное*) *моделирование* воспроизводит вероятностные процессы и события. В этом случае анализируется ряд реализаций (набор однородных реализаций) случайного процесса и оцениваются его средние характеристики.

Статическое моделирование служит для описания поведения объекта в какой-либо момент времени, а *динамическое моделирование* отражает поведение объекта во времени (динамика может быть еще и *пространственной*, а для экономических систем иногда говорят о *финансовой* динамике).

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

В зависимости от формы представления объекта можно выделить *мысленное* (*абстрактное*) и *реальное* (*физическое*) моделирование. *Мысленное моделирование* связано с моделированием объектов, которые практически не допускают возможности проведения реального, физического эксперимента. Мысленное моделирование может быть реализовано в виде *наглядного*, *символического* и *математического*.

При *наглядном моделировании* создаются различные наглядные модели, воспроизводящие явления и процессы, протекающие в объекте, на базе представлений человека о реальных объектах. Здесь в основу может быть заложена некоторая гипотеза о закономерностях протекания процесса в реальном объекте, которая отражает уровень знаний исследователя об этом объекте и базируется на причинно-следственных связях между входом и выходом изучаемого объекта. Могут быть использованы аналогии различных уровней, обычно базирующиеся на причинно-следственных связях между явлениями и процессами в объекте. Существенное место при мысленном наглядном моделировании занимает и *макетирование*.

Если ввести условное обозначение отдельных понятий, т. е. знаки, а также определенные операции между этими знаками, то можно реализовать *знаковое моделирование* и с помощью знаков отображать набор понятий – составлять отдельные цепочки из слов и предложений. В принципе, любую программу, в которой на каком-либо языке программирования реализован какой-то алгоритм процесса функционирования системы, можно рассматривать как знаковую модель этого процесса.

В основе *языкового моделирования* лежит некоторый тезаурус – фиксированный набор понятий, где каждому слову может соответствовать лишь единственное понятие.

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

Под *математическим моделированием*, о котором мы уже говорили выше, понимается процесс установления соответствия данному реальному объекту некоторого математического объекта и его исследование, позволяющее получать характеристики рассматриваемого реального объекта. *Математическое моделирование* для исследования характеристик процесса функционирования систем можно разделить на *аналитическое, имитационное* и *комбинированное*.

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

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

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

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

При желании использовать аналитический метод идут на существенное упрощение первоначальной модели, чтобы иметь воз-

можность изучить хотя бы общие свойства системы. Такое исследование на упрощенной модели аналитическим методом помогает получить ориентировочные результаты для определения более точных оценок другими методами.

По сравнению с аналитическим методом численный метод позволяет исследовать более широкий класс систем, но при этом полученные решения носят частный характер.

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

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

Как правило, результаты имитации процесса функционирования системы возникают как компьютерные реализации случайных величин и функций и для нахождения надежных оценок характеристик процесса требуют многократного воспроизведения с последующей статистической обработкой результатов.

Не вдаваясь в историю математической статистики, отметим, что в середине 50-х годов XX столетия был разработан численный метод статистических испытаний (Монте – Карло) [27, 28], предназначавшийся для моделирования случайных величин и функций, вероятностные характеристики которых совпадали с решениями аналитических задач большой размерности (кратные интегралы, экстремумы функций многих переменных и др.). В последние годы

этот метод стали применять и для машинной имитации характеристик процессов функционирования систем, подверженных случайным воздействиям, и в этой сфере его стали называть *методом статистического моделирования*.

К математическому моделированию относят также новые виды моделирования, обозначенные на рис. 28 под заголовком «Другие виды». Особое место здесь занимает *информационное* (*кибернетическое*) *моделирование*, где стремятся отобразить лишь некоторую функцию и рассматривают реальный объект как «черный ящик», имеющий ряд входов и выходов, и моделируют некоторые связи между выходами и входами. Для построения имитационной модели в этом случае необходимо выделить исследуемую функцию реального объекта, формализовать в виде некоторых операторов связи между входами и выходами и воспроизвести на имитационной модели данную функцию, причем на базе совершенно иных математических соотношений [30].

Структурное моделирование базируется на некоторых специфических особенностях структур определенного вида, которые используются в качестве средства исследования систем или служат для разработки на их основе специфических подходов к моделированию с применением других методов формализованного представления систем (теоретико-множественных, лингвистических, кибернетических и т. п.). Развитием структурного моделирования является *объектно-ориентированное* моделирование.

Структурное моделирование включает:

1) методы сетевого моделирования;

2) сочетание методов структуризации с лингвистическими;

3) структурный подход в направлении формализации построения и исследования структур разного типа (иерархических, матричных, произвольных графов) на основе теоретико-множественных представлений и понятия номинальной шкалы теории измерений.

В структурном моделировании за последние два десятка лет сформировалась новая технология CASE. Аббревиатура CASE имеет двоякое толкование, соответствующее двум направлениям использования CASE-систем. Первое из них – Computer-Aided Software Engineering – переводится как автоматизированное проектирование программного обеспечения. Соответствующие CASE-системы часто называют инструментальными средами быстрой раз-

работки программного обеспечения (RAD – Rapid Application Development). Второе – Computer-Aided System Engineering – подчеркивает направленность на поддержку концептуального моделирования сложных систем, преимущественно слабоструктурированных. Такие CASE-системы часто называют системами BPR (Business Process Reengineering). В целом CASE-технология представляет собой совокупность методологий анализа, проектирования, разработки и сопровождения сложных автоматизированных систем, поддерживаемую комплексом взаимосвязанных средств автоматизации. CASE – это инструментарий для системных аналитиков, разработчиков и программистов, позволяющий автоматизировать процесс проектирования и разработки сложных систем, в том числе и программного обеспечения.

Что же касается так называемого *реального моделирования*, то оно использует возможность исследования некоторых характеристик либо на реальном объекте целиком, либо на его части.

Основные требования, предъявляемые к модели процесса функционирования системы:

1) *полнота модели* должна обеспечить пользователю возможность получения необходимого набора оценок характеристик системы с требуемой точностью и достоверностью;

2) *гибкость модели* должна давать возможность воспроизведения различных ситуаций при варьировании структуры, алгоритмов и параметров функционирования системы;

3) *длительность разработки* и реализации модели большой системы должна быть по возможности минимальной при учете ограничений на имеющиеся ресурсы;

4) *структура модели* должна быть блочной, т. е. допускать возможность замены, добавления и исключения некоторых частей без переделки всей модели;

5) информационное обеспечение должно предоставлять возможность эффективной работы модели с базой данных систем определенного класса;

6) программные и технические средства должны обеспечивать эффективную (по быстродействию и памяти) машинную реализацию модели и удобное общение с ней пользователя;

7) при наличии ограниченных вычислительных ресурсов проведение целенаправленных (планируемых) машинных эксперимен-

тов с моделью системы должно быть реализовано с использованием аналитико-имитационного подхода.

При получении новой информации об объекте его модель пересматривается и уточняется, т. е. процесс моделирования, включая разработку и программную реализацию модели, является итерационным. Этот процесс продолжается до тех пор, пока не будет получена модель, которую можно считать адекватной в рамках решения поставленной задачи.

Остановимся на некоторых особенностях разработки современных автоматизированных систем обработки информации и управления.

Все этапы существования таких систем (проектирование, внедрение, эксплуатация и эволюция) невозможны без использования различных видов моделирования. На всех этапах необходимо учитывать особенности систем: сложность структуры и стохастичность связей между элементами, неоднозначность алгоритмов поведения при различных условиях, большое количество параметров и переменных, неполноту и отсутствие детерминированности исходной информации, разнообразие и вероятностный характер воздействий внешней среды и т. д.

Независимо от разбиения конкретной сложной системы на подсистемы при проектировании каждой из них необходимо выполнить внешнее проектирование (макропроектирование) и внутреннее проектирование (микропроектирование).

На стадии макропроектирования должна быть разработана обобщенная модель процесса функционирования сложной системы, позволяющая разработчику получить ответы на вопросы об эффективности различных стратегий управления объектом при его взаимодействии с внешней средой. Стадию внешнего проектирования можно разбить на два этапа: *анализ* и *синтез*. При анализе изучают объект управления, строят модель воздействий внешней среды, определяют критерии оценки эффективности, имеющиеся ресурсы, необходимые ограничения. Конечная цель стадии анализа – построение модели объекта управления для оценки его характеристик. При синтезе на этапе внешнего проектирования решаются задачи выбора стратегии управления на основе модели проектируемого объекта, т. е. сложной системы.

На стадии микропроектирования разрабатывают модели с це-

лью создания эффективных подсистем. Причем используемые методы и средства моделирования зависят от того, какие конкретно обеспечивающие подсистемы разрабатываются: информационные, математические, технические, программные и т. д.

В последние годы основные достижения в различных областях науки и техники связаны с совершенствованием средств вычислительной техники.

Исторически первым сложился *аналитический подход* к исследованию систем, когда компьютер использовался в качестве вычислителя по предложенным аналитическим зависимостям. Однако трудности исследования больших систем с помощью только аналитических методов вели к существенному упрощению моделей, что могло привести к получению недостоверных результатов. Поэтому в настоящее время, наряду с построением аналитических моделей, большое внимание уделяется задачам оценки характеристик больших систем на основе *имитационных моделей,* реализованных на современных компьютерах с высоким быстродействием и большим объемом оперативной памяти. При этом перспективность такого подхода возрастает с ростом быстродействия и оперативной памяти компьютеров, с развитием математического обеспечения, совершенствованием банков данных и периферийных устройств для организации диалоговых систем моделирования. Это, в свою очередь, способствует появлению новых «чисто компьютерных» методов решения задач исследования больших систем на основе организации имитационных экспериментов с их моделями.

Однако применение современной ЭВМ не гарантирует возможность исследования системы любой сложности. Нельзя игнорировать то, что в основу любой модели положено трудоемкое по затратам времени и материальных ресурсов предварительное изучение явлений, имеющих место в системе-оригинале. И от детальности изучения реальных явлений, правильности формализации и алгоритмизации зависит в конечном итоге успех моделирования конкретной системы или процесса.

При создании больших систем их компоненты разрабатываются различными коллективами, которые используют средства моделирования при анализе и синтезе отдельных подсистем. При этом разработчикам необходимы оперативный доступ к программно-техническим средствам моделирования, а также оперативный обмен

результатами моделирования отдельных взаимодействующих подсистем. Появляется необходимость в создании диалоговых систем моделирования, для которых характерны возможность одновременной работы многих пользователей, занятых разработкой одной или нескольких систем, доступ пользователей к программно-техническим ресурсам системы моделирования, включая базы данных и знаний, пакеты прикладных программ моделирования, обеспечение диалогового режима работы с различными вычислительными машинами и устройствами, включая цифровые и аналоговые вычислительные машины, установки натурного и физического моделирования, элементы реальных систем и т. п., управление процессами в системе моделирования и оказание различных услуг пользователям, включая обучение работе с диалоговой системой моделирования при обеспечении дружественного интерфейса.

10.4. Моделирование как метод научного познания

Моделирование является основным методом исследований во всех областях знаний и научно обоснованным методом оценок характеристик сложных систем, используемым для принятия решений в различных сферах человеческой деятельности.

Определяя гносеологическую роль теории моделирования, т. е. ее значение в процессе познания, необходимо выделить то общее, что присуще моделям различных по своей природе систем реального мира. Это общее заключается в наличии некоторой структуры (статической или динамической, материальной или мысленной), которая подобна структуре исследуемого объекта.

Формально моделирование можно определить как *метод опосредованного познания*, при котором изучаемый объект-оригинал находится в некотором соответствии с другим объектом-моделью, причем модель способна в том или ином отношении замещать оригинал на некоторых стадиях процесса исследования. Стадии познания, на которых происходит такая замена, а также формы соответствия модели и оригинала могут быть различными:

1) моделирование как познавательный процесс, перерабатывающий информацию, поступающую из внешней среды, о происходящих в ней явлениях, в результате чего в сознании появляются образы, соответствующие объектам;

2) моделирование, заключающееся в построении системы-модели, связанной определенными соотношениями подобия с си-

стемой-оригиналом, причем в этом случае отображение одной системы в другую является средством выявления зависимостей между двумя системами, отраженными в соотношениях подобия, а не результатом непосредственного изучения поступающей информации.

С точки зрения философии моделирование – эффективное средство познания природы. Процесс моделирования предполагает наличие объекта исследования; исследователя, перед которым поставлена конкретная задача; модели, создаваемой для получения информации об объекте и необходимой для решения поставленной задачи. Причем по отношению к модели исследователь является, по сути дела, экспериментатором, только в данном случае эксперимент проводится не с реальным объектом, а с его моделью. Такой эксперимент для инженера является инструментом непосредственного решения организационно-технических задач.

Любой эксперимент может иметь существенное значение в конкретной области науки только при специальной его обработке и обобщении. Единичный эксперимент никогда не может быть решающим для подтверждения гипотезы, проверки теории. Поэтому исследователи должны быть знакомы с элементами современной методологии теории познания и, в частности, не должны забывать, что именно экспериментальное исследование, опыт, практика являются критерием истины [28].

В настоящее время нет области человеческой деятельности, где в той или иной степени не использовались бы методы моделирования. Особенно это относится к сфере управления различными системами (в том числе и экономическими), где основными являются процессы принятия решений на базе получаемой информации.

Одна из проблем современной науки и техники – разработка и внедрение в практику проектирования новейших методов исследования характеристик сложных информационно-управляющих и информационно-вычислительных систем различных уровней: автоматизированных систем научных исследований и комплексных испытаний, систем автоматизации проектирования, автоматизированных систем управления технологическими процессами (АСУ ТП), а также интегрированных автоматизированных систем управления (АСУ), вычислительных систем, комплексов и сетей, информационных систем, цифровых сетей интегрального обслуживания и т. д.

11. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Как было отмечено ранее, в середине 50-х XX столетия, практически одновременно с созданием первых ЭВМ, был разработан *метод статистических испытаний* (*Монте – Карло*) [27, 28], позволявший на основе моделирования случайных величин и функций искать с достаточной точностью приближенные решения так называемых многомерных задач (кратные интегралы, экстремумы функций многих переменных и др.), о численном решении которых другими методами в докомпьютерное время не приходилось и мечтать. Позднее этот метод в приложении для машинной имитации характеристик процессов функционирования систем, подверженных случайным воздействиям, стали называть *методом статистического моделирования*.

Естественно, что его основные задачи сводятся к следующему:

1) генерация случайных (псевдослучайных) чисел, равномерно распределенных на интервале (0, 1);

2) получение выборок с заданным распределением вероятностей (дискретным или непрерывным);

3) статистический анализ (большой самостоятельный раздел прикладной математики) вероятностных характеристик полученных последовательностей и их «улучшение» с целью использования при решении задач имитационного моделирования.

11.1. Дискретные и непрерывные случайные величины

Как известно, случайная величина X может принимать как дискретные значения, так и произвольные значения из некоторого диапазона с определенной вероятностью. Соответственно, говорят о *дискретной* или *непрерывной* случайной величине.

Разумеется, непрерывность случайной величины является идеализацией реальности из-за скромности человеческих требований к абсолютному знанию (едва ли вы пожелаете получать стипендию даже с точностью до долей копейки или знать свой рост с точностью до микрона) и вообще ограниченности человеческих и технических возможностей и представлений о великом и малом. Тем не менее, замена дискретной переменной на непрерывную, благодаря хорошо развитой технологии дифференцирования и интегрирования, позволяет вести статистический анализ с существенно меньшими затратами энергии [34].

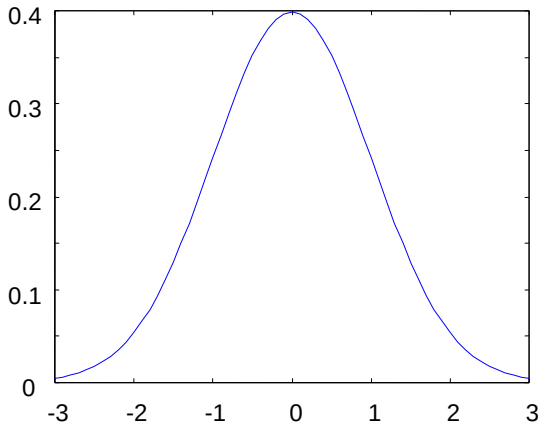
Если X принимает значения x_i (i = 1 .. n) (например, $x_1 \le x_2 \le ... \le x_n$) и вероятность выбора конкретного значения $P(X = x_i) = P_i$, то очевидно:

$$\sum_{i=1}^{n} P_i = 1;\ P_i \ge 0\ (i = 1\ ..\ n);\ P(x_l \le X \le x_k) = \sum_{i=l}^{k} P_k\ . \qquad (1)$$

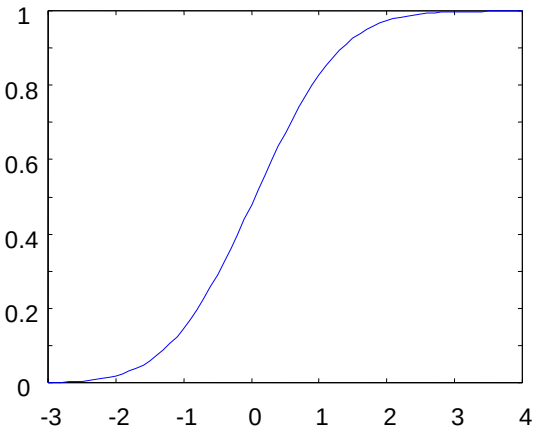
Если X принимает значения x в диапазоне [α, β] с вероятностями $P(x)$, то

$$\int_{\alpha}^{\beta} P(x)dx = 1;\ P(x) \ge 0,\ x \in [\alpha, \beta];\ P(X \le x) = \int_{\alpha}^{x} P(x)dx\ . \qquad (2)$$

Набор значений P_i в (1) и функцию $P(x)$ в (2) называют *плотностью распределения* вероятностей (рис. 29), а набор $F_k = P(X \le x_k) = \sum_{i=1}^{k} P_i$ и $F(X) = P(X \le x) = \int_{\alpha}^{x} P(x)dx$ – *функцией распределения* (рис. 30).



распределения вероятностей



вероятностей

11.2. Оценки параметров распределения и их свойства

Значение параметра (оценка), вычисленное по случайной выборке ограниченного объема, является случайной величиной, т. е. оно может меняться от выборки к выборке. Следовательно, в результате статистической обработки определяется не истинное значение какого-либо параметра, а лишь его приближенное значение – *статистическая оценка*. Получить статистическую оценку параметра теоретического распределения означает найти функцию от

имеющихся результатов наблюдения (случайных величин), которая и даст приближенное значение искомого параметра. Различают *точечные* и *интервальные* оценки, характеризующиеся одним числом или диапазоном возможного значения, соответственно.

Точечные оценки применяются при большом объеме выборки. Их качество характеризуется такими свойствами, как *состоятельность*, *несмещенность*, *эффективность* и *достаточность* [35].

Состоятельность характеризует сходимость по вероятности оценки к истинному значению параметра при неограниченном увеличении объема выборки.

Несмещенность характеризует отсутствие систематических отклонений оценки от теоретического значения параметра при любом конечном, в том числе и малом, объеме выборки.

Эффективность характеризует разброс случайных значений оценки около истинного значения параметра.

Достаточность характеризует полноту использования информации, содержащейся в выборке. Другими словами, оценка достаточна, если все другие независимые оценки, полученные на основе данной выборки, не дают дополнительной информации об оцениваемом параметре.

Базовыми *параметрами* всех теоретико-вероятностных распределений являются *моменты* [34]. Так величины $\int_{\alpha}^{\beta} x^k P(x)dx$ и $\sum_{i=1}^{n} x_i^k P_i$ называются моментами *k*-го порядка.

Момент первого порядка, определяемый в форме

$$Mx = \mu = \int_{\alpha}^{\beta} xP(x)dx \quad \text{или} \quad Mx = \mu = \sum_{i=1}^{n} x_i P_i \,, \qquad (3)$$

называют *математическим ожиданием* случайной величины (чаще средним значением).

Что касается моментов более высокого порядка, то предпочтительнее по соображениям вычислительного характера использовать *центральные моменты* (не относительно нуля, а относительно среднего):

$$M_k x = \int_{\alpha}^{\beta} (x-\mu)^k P(x)dx \quad \text{или} \quad M_k x = \sum_{i=1}^{n} (x_i - \mu)^k P_i \,, \qquad (4)$$

являющиеся основополагающими при поиске некоторых полезных характеристик распределений случайных величин.

Так центральный момент второго порядка определяет *дисперсию* распределения – меру вариабельности (степень разброса или размытости) значений случайной величины относительно среднего значения

$$Dx = \sigma_x^2 = \int_{\alpha}^{\beta} (x-\mu)^2 P(s)ds \quad \text{или} \quad Dx = \sigma_x^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P_i \,. \quad (5)$$

Кстати, может оказаться полезной и другая (тождественная) форма

$$\sigma_x^2 = \int_{\alpha}^{\beta} x^2 P(x)dx - \mu^2 \quad \text{или} \quad \sigma_x^2 = \sum_{i=1}^{n} x_i{}^2 P_i - \mu^2 \,. \qquad (6)$$

Часто вместо дисперсии используют квадратный корень из нее σ_x, называемый *среднеквадратическим* или *стандартным отклонением*.

Знание значений μ и σ_x позволяет создавать более удобные в анализе *стандартизованные* (центрированные и нормированные) значения

$$Z = \frac{x - \mu}{\sigma_x}$$

с нулевым математическим ожиданием и разбросом, определяемым числами из диапазона от 0 до нескольких единиц (по чисто вычислительным соображениям работа со стандартизованными величинами минимизирует погрешность вычислений, уменьшает шансы на потерю значности при умножении малых или переполнение при умножении больших значений).

Иногда используется оценка, называемая *стандартной ошибкой среднего*, характеризующая стандартное отклонение выборочного среднего:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \,. \qquad (7)$$

Момент третьего порядка определяет *асимметрию* распределения A_x (левую или правую) (Пирсон, 1895 г.) – меру несиммет-

ричности распределения (рис. 31) относительно ожидаемого значения (идеал – симметрия, т. е. $A_x = 0$).

$$A_x = \int_{\alpha}^{\beta} \left(\frac{s-\mu}{\sigma_x} \right)^3 P(s)ds \quad \text{или} \quad A_x = \sum_{i=1}^{n} \left(\frac{x_i-\mu}{\sigma_x} \right)^3 P_i \,. \tag{8}$$

Момент четвертого порядка определяет так называемый *эксцесс Ex* (*куртозис*) (Пирсон, 1905 г.) распределения (рис. 31) – меру сглаженности (остроты пика плотности распределения), которая определяется обычно в виде

$$E_x = \int_{\alpha}^{\beta} \left(\frac{s-\mu}{\sigma_x} \right)^4 P(s)ds \quad \text{или} \quad E_x = \sum_{i=1}^{n} \left(\frac{x_i-\mu}{\sigma_x} \right)^4 P_i \,. \tag{9}$$

Так для известного нормального распределения эксцесс равен 3, $Ex > 3$ свидетельствует об островершинности. Заметим, что часто (по соображениям отличия от нормальности) *Ex* берут уменьшенным на 3.

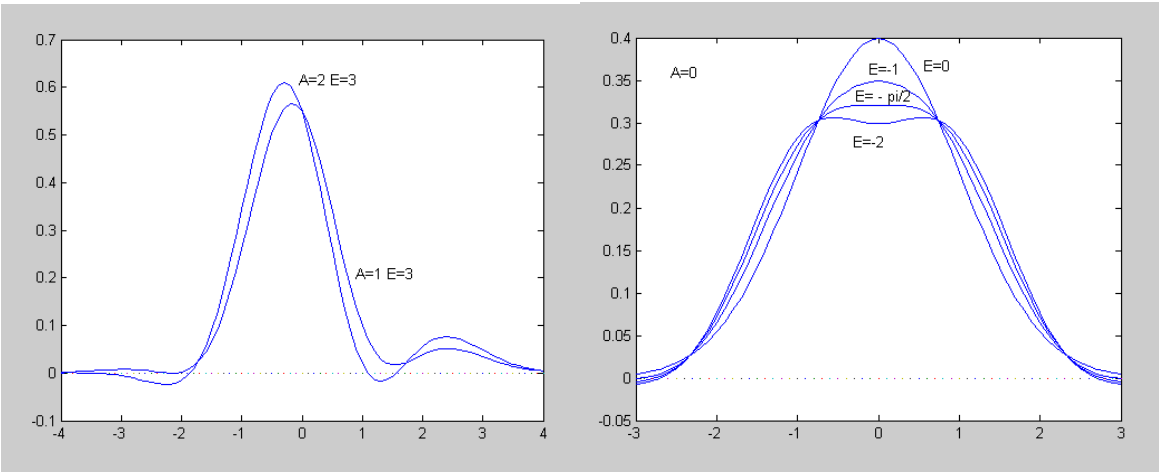


Рис. 31. Плотность распределения Лапласа – Шарлье при различных асимметрии и эксцессе

Иногда при ненулевых μ используют *коэффициент вариации*

$$V_x = \frac{\mu}{\sigma_x} \tag{10}$$

как показатель, характеризующий соотношение математического ожидания и показателя вариабельности случайной величины.

Естественно, что поиск указанных характеристик генеральной совокупности значений случайной величины при отсутствии априорной информации приходится строить на основе некоторых выборок, состоящих из *N* (объем выборки) элементов. По принципу не-

достаточного основания вероятность появления очередного значения берется равной $1/N$ и соответственно процесс вычисления оценок для выборочного распределения сводится к поиску:

$$Mx = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \; ; \; \sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \; ;$$

$$A_x = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3 \; ; \; E_x = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4 \; .$$

В случае малых выборок используется понятие *числа степеней свободы*. Так, если исходная выборка объема N обладает N степенями свободы, то при ее рассмотрении с учетом фиксированной оценки Mx число степеней свободы уменьшается на 1. С этой характеристикой связано понятие *несмещенных оценок* (оценок, математическое ожидание отклонений которых от оцениваемого параметра равно нулю или, другими словами, математическое ожидание которых равно истинному значению оцениваемой характеристики).

Соответственно несмещенные оценки имеют вид

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \; , \; \sigma_x^2 = \sqrt{\frac{N-1}{2}}\,\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)}\,\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \; ;$$

$$A_x = \frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3 \; ; \qquad (11)$$

$$E_x = \frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4 \; .$$

Здесь $\Gamma(k)$ – так называемая гамма-функция; $\Gamma(k+1) = k\,\Gamma(k)$; при целом $k > 0$ тождественна понятию факториала $\Gamma(k+1) = k!$; $\Gamma(0{,}5) = \sqrt{\pi}$.

Все характеристики случайных величин делят на три группы: характеристики *положения*, *рассеяния* (*разброса*) и *формы*.

Форма характеризуется асимметрией и эксцессом.

Наряду с дисперсией, среднеквадратическим отклонением и коэффициентом вариации к показателям разброса дискретных распределений относят и так называемый *размах* $R_x = x_{\max} - x_{\min}$.

К числу характеристик положения, наряду с математическим ожиданием, относятся медиана и мода.

Медиана (*Me*) делит распределение на две равновероятные половины и определяется как

$$\int_{\alpha}^{Me} p(s)ds = \int_{Me}^{\beta} p(s)ds. \qquad (12)$$

Для дискретных распределений достаточно упорядочить выборку и взять медиану равной *срединному значению* $x_{[N/2]+1}$ при нечетном N или полусумме $x_{[N/2]} + x_{[N/2]+1}$ – при четном; например, для совокупности значений {1, 2, 3, 4, 5} медианой будет 3, для {1, 2, 2, 7, 9, 13} медиана равна (2 + 7) /3 = 4,5.

Мода (*Mo*) соответствует такому значению случайной величины, при котором функция плотности распределения вероятностей достигает максимума. Очевидно, что найти моду однозначно возможно лишь в случае *унимодальных распределений* (плотность распределения имеет единственную точку максимума). Если максимума два, распределение называют *бимодальным*; в общем случае – при нескольких максимумах – *полимодальным*.

При проверке различных статистических гипотез используется такая характеристика распределения, как *квантиль*.

Так для случайной величины с функцией распределения $F(x)$ квантилью порядка α $(0 < \alpha < 1)$ называется максимальное (по модулю) число K_α такое, что $F(K_\alpha) \le \alpha$.

В случае непрерывной функции поиск квантили сводится к решению уравнения $F(x) = \alpha$.

Небесполезным в статистическом анализе является знание *закона больших чисел*. Интуитивно ясно, что чем больше объем статистической выборки, тем надежнее наши суждения об изучаемом явлении. Однако неограниченное его увеличение не всегда реально по разным причинам.

Исторически первая формулировка закона больших чисел принадлежит Якобу Бернулли (опубл. 1713 г.). Если N – число независимых испытаний и M – число «успехов», то вероятность P одиночного успеха можно оценить из условия $P(\left|\frac{M}{N} - p\right| \le \varepsilon) > 1 - \eta$ при любых ε, $\eta > 0$ и достаточно больших N. Отсюда можно установить, что N достаточно выбирать из условия

$$N > \frac{1+\varepsilon}{\varepsilon^2} \lg\frac{1}{\eta} + \frac{1}{\varepsilon}. \qquad (13)$$

Более простая оценка получена П. Л. Чебышевым (1846 г.):

$$N > \frac{1}{\varepsilon^2 \eta}. \qquad (14)$$

Например, при выборе $\varepsilon \sim 0{,}01$ и $\eta \sim 0{,}05$ эти неравенства дают оценки 13240 и 200000.

Более жесткие оценки получены С. Н. Бернштейном (1911 г.) и А. Н. Колмогоровым (1929 г.):

$$P\left(\frac{|\text{отклонение}|}{\sigma_x\sqrt{N}} > t\right) \le 2\exp\left(\frac{t^2}{2\left(1+\frac{\alpha}{3}\right)}\right), \alpha = L\,t\,\sigma\sqrt{N},$$

где L – максимальная по модулю оценка отклонений от среднего.

В процессе генерации случайных последовательностей (выборок) X и Y со значениями средних μ_x и μ_y не следует забывать о возможности наличия связей между ними. Одной из характеристик такой связи является *коэффициент корреляции* ($-1 \le r_{xy} \le +1$):

$$r_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} P(x,y)\,dx\,dy$$

или (15)

$$r_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{(x_i-\mu_x)(y_j-\mu_y)}{\sigma_x\sigma_y} p_{ij}.$$

Близость его к нулю не гарантирует отсутствия связи нелинейного характера, но близость $|r_{xy}|$ к 1 говорит о наличии явно неслучайной связи и ставит вопрос о допустимости использования таких выборок в одном имитационном эксперименте.

11.3. Датчик псевдослучайных чисел, равномерно распределенных в (0, 1)

Упомянутый датчик может быть аппаратным, основываясь на физических процессах (радиоактивный распад и т. п.), или программным. Разумеется, числа, создаваемые по какому-то алгоритму, но удовлетворяющие некоторым статистическим тестам, не могут

называться случайными и потому предпочтительнее называть их псевдослучайными.

Исторически первым решением проблемы генерации псевдослучайных чисел, равномерно распределенных в (0, 1), был *метод середины квадратов*, предложенный в 1946 г. Джоном фон Нейманом. Идея метода предельно проста. Берем некоторое n-значное число x_i. Возводим его в квадрат, получая 2 n-значное и выделяем средние n знаков, которые принимаем за следующее число x_{i+1}. Умножая получаемые таким образом числа на 10^{-n}, получаем числа из интервала (0, 1).

К сожалению, при некоторых начальных числах получаемая последовательность периодически повторяется («зацикливается»). Более того, как утверждали некоторые современники, получаемое распределение не удовлетворяло тестам на равномерность.

В последующие годы, начиная как с первых отечественных серийных ЭВМ «Стрела», «Урал» и М-20, так и зарубежных, все такие датчики имитировали итерационный процесс $x_{i+1} = \Phi(x_i)$, где в роли начального значения выбирается текущее время.

В частности, до сих пор остается популярным метод Лемера $x_{i+1} = (A\, x_i + C) \bmod M$, где A и C – некоторые константы, M выбирается с учетом длины разрядной сетки компьютера, например $M = 2^n - 1$.

В [36] утверждается эффективность *модифицированного метода Неймана*, смысл которого состоит в следующем. Если взять иррациональное число, например $\Theta = (\sqrt{5} - 1)/2$, то $x_i = \{i\,\Theta\}$ – дробная часть его произведения на целые числа i будет давать последовательность чисел, имеющую достаточно хорошую апериодичность.

Непрерывная случайная величина x называется равномерно распределенной на отрезке $[a, b]$, если ее плотность определяется функцией

$$p(x) = \begin{cases} 1/(b-a), & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}.$$

Соответственно функция распределения

$$F(x) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a \le x \le b \\ 1, & x > b \end{cases}.$$

Без особого труда можно оценить математическое ожидание

$$M(x) = \int_{-\infty}^{\infty} x\, p(x)dx = \int_{a}^{b} x\, p(x)dx = \int_{a}^{b} x\,(b-a)^{-1}dx = (a+b)/2,$$

совпадающее с медианой; дисперсию

$$D(x) = \int_{a}^{b} (x - M(\xi))^2 f(x)dx = (b-a)^2/12$$

и стандартное отклонение

$$\sigma_x = \sqrt{D(x)} = (b-a)/(2\sqrt{3}).$$

Асимметрия и эксцесс соответственно равны 0 и –6/5.

В настоящее время практически все системы программирования и пакеты прикладных программ обладают такими датчиками, в названии которых фигурирует *rand* или *random*. Поскольку не всякий выбор исходных констант дает хороший результат, все подобные датчики подвергаются проверке качества создаваемых псевдослучайных последовательностей. Основным *статистическим тестом* является проверка гипотезы согласия эмпирического распределения вероятностей с предполагаемым равномерным распределением.

Задавшись разбиением промежутка (0, 1) на s интервалов и выборкой фиксированного объема N, подсчитываем количества попаданий n_i ($i = 1 .. s$) в интервалы (можно для наглядности построить гистограмму) и итоговую оценку меры расхождения эмпирического и гипотетического равномерного в (0, 1) распределений:

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_i - N\ p_i)^2}{N\ p_i},$$

где $p_i = 1 / s$ (гипотетические вероятности). Величина χ^2 при неограниченном увеличении s имеет распределение хи-квадрат с $k = s - 1$ степенями свободы[20]. Таблицы данного распределения приведены во многих справочниках.

Если найденная оценка χ^2 меньше табличной для данного k и уровня значимости (вероятности ошибки) α, нет оснований отверг-

[20] Число степеней свободы равно числу s минус число линейных связей, наложенных на выборку. Одна связь существует в силу условия $\sum p_i = 1$. Если f параметров распределения заранее неизвестны и определяются по выборке, число степеней свободы $k = s - f - 1$.

нуть гипотезу равномерности. Критерий рекомендуется применять при $N > 200$ или хотя бы при $N > 40$.

Не бесполезна и проверка датчика на соответствие n-мерным равномерным распределениям – создание многомерного аналога гистограммы для квадрата, сектора круга, сферы, гиперкуба и т. п.

Существует целая система тестов (Кендалл), служащих для проверки закона распределения последовательности и ее случайности, состоящая из четырех тестов:

1) проверка частот (*frequency test*);
2) проверка пар (*serial test*);
3) проверка интервалов (*gap test*);
4) проверка комбинаций (*poker test*).

Проверка частот сводится к описанному выше подсчету количества случайных величин, попавших в интервалы разбиения. Предлагают и оригинальный тест проверки частот для отдельных разрядов псевдослучайного кода. *Проверка пар* состоит в проверке нулей и единиц в двоичном представлении чисел из последовательности.

Обязательной для псевдослучайных чисел является трудоемкая процедура *проверки апериодичности*. Отрезком апериодичности (длиной серии) называют наибольшее число членов случайной последовательности от начала генерации, среди которых нет повторения. В идеале датчик должен исчерпывать все множество чисел, представимых в «ячейке памяти» компьютера.

Проверка комбинаций сводится к проверке комбинаций двоичных разрядов в числах последовательности. Отрезки создаваемой последовательности (группы двух, трех и более чисел) должны быть статистически независимыми.

11.4. Моделирование случайных величин с известным законом распределения

Очевидно, что для преобразования случайных чисел x_i, равномерно распределенных в (0, 1), в аналогичные числа z_i из произвольного (a, b) достаточно выполнить преобразование:

$$z_i = a + x_i\,(b - a).$$

В случае преобразования к числам с другим распределением задача существенно усложняется.

Если x – случайная величина, равномерно распределенная на (0, 1), то искомая непрерывная случайная величина z получается с помощью преобразования $z = F^{-1}(x)$, где $F^{-1}(x)$ – функция, *обратная* к функции распределения вероятностей генерируемой случайной величины. Другими словами, задача сводится к поиску (см. рис. 32) величины z, являющейся решением уравнения

$$F(z) = x. \qquad (16)$$

Если учесть, что $F(z) = \int_{-\infty}^{z} P(z)dz$, $\frac{dF(z)}{dz} = P(z)$, то иногда (16) решается без труда. Так для показательного распределения $p(z) = \lambda e^{-\lambda z}, z > 0;$ $F(z) = 1 - e^{-\lambda z}$, и решение (16) дает $z = -\ln(1 - x) / \lambda$.

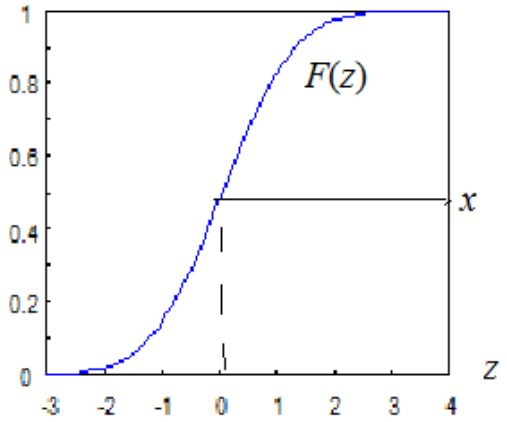


Рис. 32. К решению уравнения $F(z) = x$

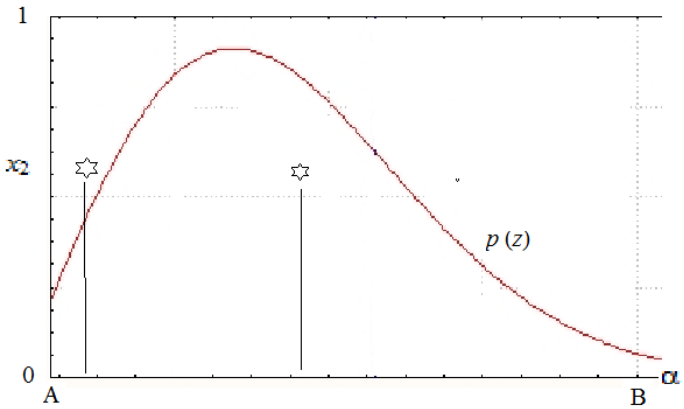


Рис. 33. Метод Неймана

Остается популярным способ получения последовательности случайных чисел с плотностью распределения $p(z)$ методом Неймана, в современной литературе фигурирующим под множеством «премудрых» названий.

В основе метода лежит выбор интервала (A, B), за пределами которого $p(z)$ практически близко к нулю. В дальнейшем реализуется процесс, где последовательно генерируются пары случайных чисел (x_1, x_2), равномерно распределенных в (0, 1), отыскивается $\alpha = A + (B - A)\,x_1$ и проверяется выполнение условия $x_2 \le p(\alpha)$ – точка (α, x_2) под кривой $p(z)$? Если условие выполняется, то α принимается за искомое число и в противном случае – отвергается (рис. 33).

Объем поисковых работ зависит от выбора A и B и может быть значительным. С другой стороны, решение уравнения (16) требует численного интегрирования $p(z)$, если нет приемлемой аппроксимации $F(z)$, и численного решения, например, методом дихотомии.

11.5. Моделирование эмпирических распределений

Всякому имитационному моделированию системы или какой-то ее характеристики предшествует обследование и получение некоторой *представительной* эмпирической выборки, которую подвергают статистической обработке. В процессе обработки стараются выяснить возможный диапазон вариации значений, основные оценки (хотя бы математического ожидания и дисперсии), плотность и функцию их эмпирического распределения.

Чтобы в дальнейшем не хранить в памяти компьютера таблично заданные функции, пытаются подобрать подходящее по соображениям точности аппроксимации аналитическое представление из списка известных распределений.

Завершающим этапом этой работы будет подбор способа будущей генерации случайных величин с выбранным законом распределения.

Алгоритмы моделирования эмпирических дискретных и непрерывных распределений различаются лишь техническими деталями.

11.5.1. Моделирование дискретных распределений

Пусть имеется выборка из генеральной совокупности дискретных случайных величин (чисел, объектов, состояний и т. д.). Если такая величина может принимать всего n различных значений, каждый вариант обозначаем номером от 1 до n. Далее составляется эмпирический ряд распределения: подсчитываются частоты появления каждого варианта в выборке и делением на объем выборки N оценивается вероятность каждого из состояний объекта.

Если k вариантов x_i ($i = 1, 2, \ldots, k$) не вошли в эмпирическую выборку (обычно редко встречающиеся, но гипотетически возможные), можно произвести корректировку. Поскольку $\sum_{i=1}^{k} p_i < 1/N$, берем $0 < p_i < \dfrac{1}{(N+1)\,k\,t}$, где $t \geq 1$ – показатель доверия к качеству и

полноте выборки, зависящий от объема выборки и вероятности ошибки (из таблиц распределения Стьюдента).

Вводя ненулевые частоты указанных вариантов, естественно пропорционально уменьшить частоты других вариантов, чтобы соблюдалось условие $\sum_{i=1}^{n} p_i = 1$.

Что касается будущего процесса моделирования, то он сводится к применению аналога метода обратной функции (рис. 34).

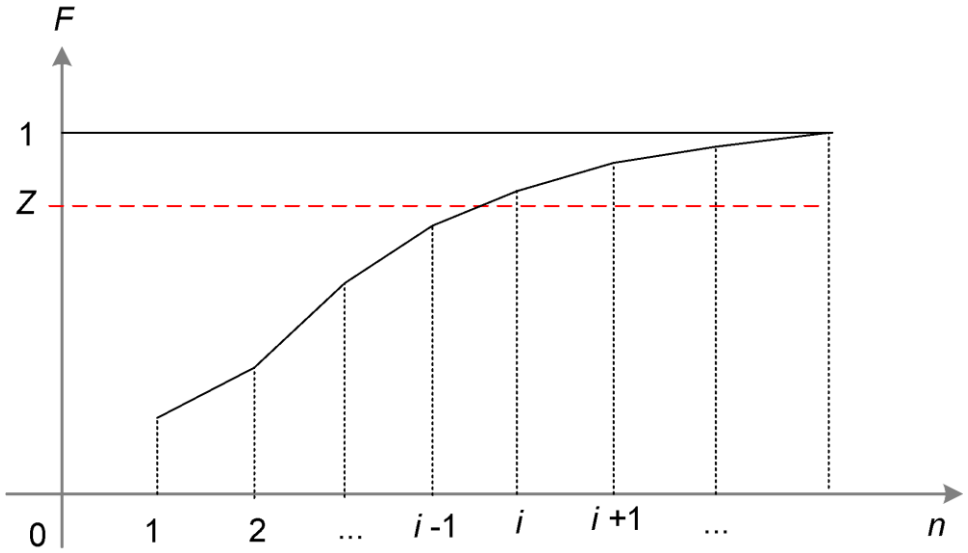


Рис. 34. Моделирование эмпирического дискретного распределения

Для генерации варианта дискретной величины датчиком случайных чисел, равномерно распределенных на (0, 1), создается значение z, по значениям функции эмпирического распределения определяется интервал, где выполняется соотношение $F(i-1) < z < F(i)$, и делается выбор между соответствующими вариантами.

11.5.2. Моделирование непрерывных распределений

В данном случае исходную выборку $\{x_i, i = 1, 2, \ldots, N\}$ упорядочивают по возрастанию, получая вариационный ряд.

Интервал $[x_{\min} - \varepsilon, x_{\max} + \varepsilon]$ разбивают на k каких-либо подынтервалов длиной Δ_j (может быть, первоначально равных). Значение k выбирается по каким-либо разумным соображениям, например, по формуле Стерджесса $k = 1 + 3{,}3219 \lg(N)$ (здесь $3{,}3219 = \log_2 10$), подсчитывают число элементов выборки, попавших в каждый из подынтервалов $\{N_j, \ j = 1, 2, \ldots, k\}$, и соответствующие частоты

$\{w_j = N_j / N,\ j = 1, 2, \ldots, k\}$, служащие основой для построения эмпирической плотности $p^*(x)$ и эмпирической функции (*кумуляты*) $F^*(x)$ распределения:

$$p_j^* = w_j / \Delta, \quad F_j^* = \sum_{i=1}^{j} w_j^* \ , \quad j = 1, 2, \ldots, k.$$

Во избежание «провалов», порождающих иллюзию отсутствия унимодальности, в функции плотности устанавливается «порог» – минимальное число попаданий в каждый из подынтервалов (обычно он берется равным 3 или определяется по субъективным соображениям). Подынтервал с числом попаданий, меньшим «порога», объединяется с соседним подынтервалом. Таким образом уточняется число подынтервалов $k > 2$, границы подынтервалов $[G_j, j = 1, k + 1)]$, число попаданий в каждый из них N_j и поинтервальные оценки плотности эмпирического распределения.

Далее эмпирическая функция распределения (рис. 35) аппроксимируется кусочно-линейной функцией, соединяющей середины верхних оснований прямоугольников полученной гистограммы.

Для последующего моделирования случайной величины с таким распределением генерируется равномерно распределенная величина z из (0, 1), выясняется ее соответствие интервалу $F_{i-1} < z < F_i$ и находится значение

$$x = x_1 + (z - F_{i-1}) \frac{x_2 - x_1}{F_i - F_{i-1}},$$

где x_1 и x_2 – середины соответствующих интервалов.

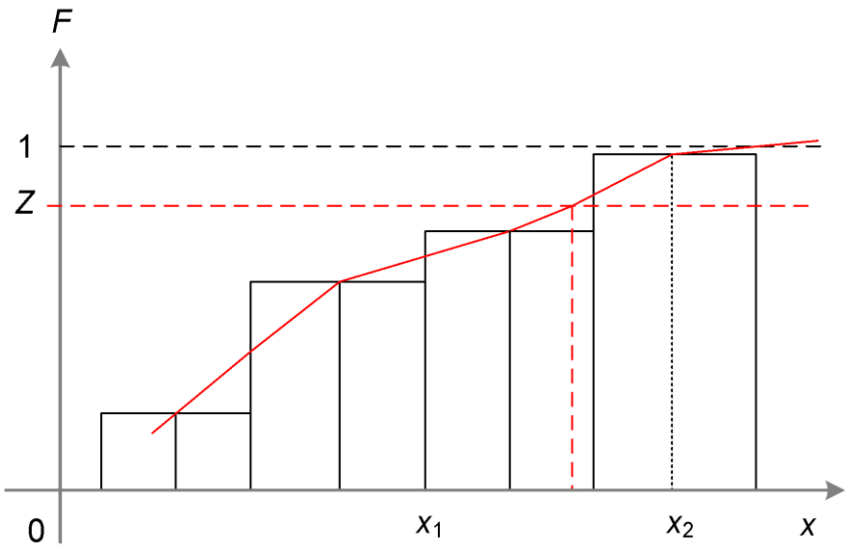


Рис. 35. Эмпирическая функция непрерывного распределения

Технология проверки соответствия найденного эмпирического распределения какому-либо гипотетическому из известных распределений (нормальному, Релея, Вейбулла и др.) состоит в следующем.

Левая граница первого подынтервала и правая граница последнего сдвигаются до нижней и верхней границ области определения случайной величины с таким распределением (бесконечность достаточно заменять отклонением от среднего не более чем на 3 – 4 стандартных отклонения). Затем находят оценки плотности гипотетического распределения вероятностей для каждого подынтервала

$$p_i^{\text{гип}} = \int_{G_{i-1}}^{G_i} P_{\text{гип}}(t)dt \qquad (17)$$

и в завершение вычислительной процедуры рассчитывают величину

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i}\,, \qquad (18)$$

оценку χ^2-критерия при числе степеней свободы $f = k - 1$ и выясняют уровень значимости для гипотезы близости эмпирического и гипотетического распределений (как велика вероятность ошибки, если мы гипотезу отвергнем). Другими словами, отыскивается оценка и выясняется, не превосходит ли она, например, 5%-ю точку χ^2-распределения с $k - 1$ степенями свободы.

Например, при 50-кратном бросании кости ($n = 50$) выпали 12 шестерок, 9 пятерок, 9 четверок, 6 троек, 9 двоек и 5 единиц (значения n_i). Проверяя гипотезу о том, что кость «правильная», т. е. сопоставляя имеющееся распределение с равномерным, где оценки вероятностей для каждого варианта равны 1 / 6, получаем оценку $\chi^2 = 3{,}76$, много меньшую табличной при $f = 5$. Вывод: отвергать гипотезу равномерности было бы ошибкой – нет оснований объявить кость фальшивой.

Разумеется, от выбора числа интервалов и правила группировки по порогу *при выборках небольшого объема* существенно зависит форма эмпирического распределения и последующие выводы, но ничего принципиально лучшего предложить нереально.

11.6. Распределения дискретных случайных величин

Большинство дискретных распределений основаны на так называемых независимых испытаниях Бернулли (биномиальные, Паскаля, геометрические) с двумя возможными исходами: «успех» ($x = 1$) и «неудача» ($x = 0$) и определяются параметром p – вероятностью «успеха». Они описывают результаты последовательности испытаний, прерываемой по тому или иному признаку, например биномиальное связано с числом «успехов» в заданной серии испытаний, а паскалево – с числом испытаний до обнаружения определенного количества «успехов».

Остановимся на дискретных распределениях, наиболее популярных в реальных приложениях.

11.6.1. Дискретное равномерное распределение

Соответствующая случайная величина аналогична непрерывной, задается двумя параметрами: – a, обычно целочисленного и сохраняющего смысл нижней границы области ее значений, и b – количества элементов n в области значений (параметра масштаба).

Плотность и функция распределения соответственно равны:

$$p(x) = 1 / b,\ F(x \mid a, b) = (x - a + 1) / b.$$

Математическое ожидание и дисперсия:

$$M(X) = a + (b - 1) / 2;\ D(X) = (b^2 - 1) / 12.$$

11.6.2. Биномиальное и отрицательное биномиальное распределения

Биномиальное распределение связано с именем Якоба Бернулли (1654 – 1705), одного из основоположников теории вероятностей.

Пусть в результате отдельного испытания событие A может осуществляться с вероятностью p (например, при бросании игральной кости вероятность выпадения шестерки равна 1 / 6). Тогда число x появления такого события в n «независимых испытаниях» будет случайной величиной, подчиненной биномиальному закону распределения. Можно интерпретировать эту величину и в «схеме извлечения с возвратом», где присутствует множество из N элементов, содержащее k элементов с признаком A. Естественно, при случайном извлечении одного элемента вероятность выбора A равна $p = k / N$. Число x извлечений элементов с признаком A в серии из n

попыток (не забывайте возвращать выбранный элемент в исходное множество) подчинено биномиальному закону с плотностью (рис. 36)

$$p(x \mid n, p) = C_n^x p^x (1-p)^{n-x},\ x = \{0,\ 1,\ \ldots,\ n\}$$

и функцией распределения (рис. 36)

$$F(x \mid n, r) = \sum_{i=0}^{x} C_n^i p^i (1-p)^{n-i}.$$

Здесь $Mx = n\,p$, $Dx = n\,p\,q$, мода совпадает с Mx, $Ax = {1-2p}\big/{\sqrt{np(1-p)}}$, $E_x = \frac{1-6p(1-p)}{np(1-p)} + 3$.

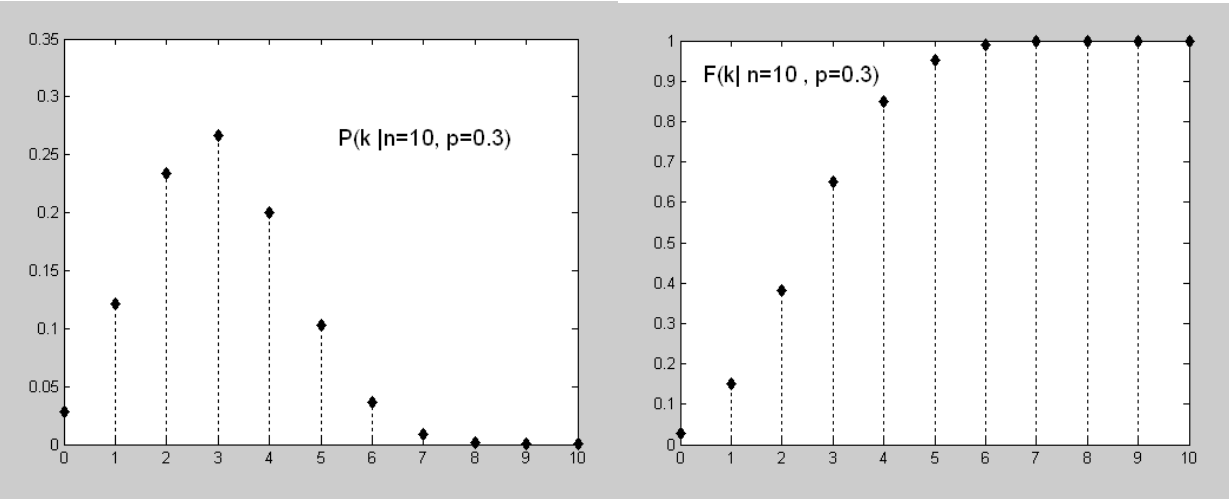


Рис. 36. Плотность и функция биномиального распределения

При большом n и малом p ($p < 0{,}1$) биномиальное распределение аппроксимируется распределением Пуассона (см. ниже):

$$C_n^k p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda},\ \lambda = n\,p.$$

Отрицательное биномиальное распределение связано с последовательностью испытаний Бернулли при вероятности успеха p и определяет число неудач до x-го успеха.

Плотность и функция распределения соответственно выглядят таким образом:

$$p(x \mid n, p) = C_{n+x-1}^x p^x (1-p)^n, x = \{0, 1, 2, \ldots\},$$

$$F(x \mid n, p) = (1-p)^n \sum_{x=0}^{\infty} C_{n+x-1}^x p^x.$$

Математическое ожидание $Mx = \frac{np}{q}$ и дисперсия $Dx = \frac{np}{q^2}$, $q = 1 - p$.

Оба вида биномиального распределения используются при разработке математических методов контроля качества промышленной продукции и играют существенную роль в моделях анализа и прогноза результатов испытаний. Так использование отрицательного биномиального распределения позволяет оценить объем испытаний, необходимый для достижения результата с вероятностью, не меньшей заданного уровня качества системы.

11.6.3. Распределение Паскаля

Распределение Паскаля отличается от биномиального тем, что биномиальная случайная величина определяет вероятность x успехов в n испытаниях, а случайная величина Паскаля – вероятность x неудач вплоть до m-го успеха (включая и этот успех).

Например, считывание некоторого фиксированного объема m двоичной информации с любого носителя время от времени дает сбои. Если известна вероятность p сбоя любого бита, то вероятность того, что потребуется $x = m + x$ попыток для считывания всех m бит, будет подчиняться распределению Паскаля.

Функция плотности распределения (рис. 37) выглядит как

$$p(x = n) = C^{n}_{m+n-1}\, p^{m}\, (1-p)^{n},$$

где m – число успехов; n – число неудач; $1 \le n < \infty$.

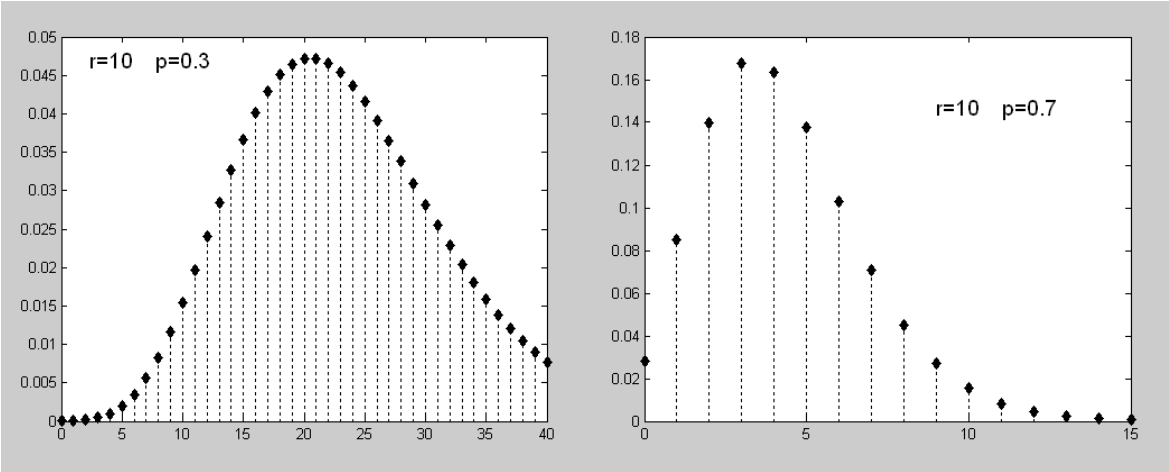


Рис. 37. Плотность распределения Паскаля при различных значениях p

Математическое ожидание $Mx = m / p$,
дисперсия $Dx = m\,(1 - p) / p^2$.

Функция распределения при $m = 0, 1, 2, \ldots$ может быть представлена формулой

$$F(m) = \frac{1}{\beta(n, m+1)} \int_0^p x^{n-1} (1-x)^m \, dx,$$

где $\beta(n, m + 1)$ – так называемая бета-функция.

11.6.4. Геометрическое распределение

Геометрическое распределение, впервые предложенное У. Феллером (1906 – 1970), отличается от биномиального тем, что здесь оценивается вероятность n неудачных попыток до первого успеха (включая первый успех). Если вероятность успеха равна p, то $p(x = n) = p\,(1 - p)^{n-1} = p\,q^{n-1}$; $F(x) = 1 - (1 - p)^n$. Очевидно, что $n \in [1, +\infty)$. Математическое ожидание и дисперсия $Mx = 1 / p$; $Dx = (1 - p) / p^2$.

11.6.5. Гипергеометрическое распределение

Смысл гипергеометрического распределения можно уловить из примера статистического приемочного контроля качества промышленной продукции, когда в партии из N изделий имеется M доброкачественных и $(N - M)$ дефектных. Случайным образом (без возврата) из всей партии выбирается контрольная группа из n изделий. Число m доброкачественных изделий в контрольной партии (случайная величина) подчинено гипергеометрическому распределению с вероятностями

$$p(X = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n},\ m = 1, 2, \ldots, \min\{n, M\},\ m \le N,\ n \le N.$$

Математическое ожидание и дисперсия случайной величины X, имеющей гипергеометрическое распределение с параметрами n, N, M, соответственно, равны:

$$Mx = \frac{n\,M}{N},\ Dx = \frac{nM}{N-1}\left(1 - \frac{M}{N}\right)\left(1 - \frac{n}{N}\right).$$

11.6.6. Распределение Пуассона

Это распределение пользуется исключительной популярностью в задачах теории массового обслуживания. Случайную величину, подчиненную распределению Пуассона, обычно интерпретируют как вероятность возникновения *k* событий (заявок) за интервал времени *L*. Примерами таких событий могут служить вызовы на АТС или неотложной медицинской помощи, прибытие самолетов в воздушное пространство аэропорта, отказы в работе устройств, дорожные происшествия и др. Обычно предполагается, что для пуассоновского потока заявок характерна *ординарность* (невозможно одновременное появление двух и более событий), *стационарность* (независимость наступления определенного числа событий за определенное время от начала его отсчета), *отсутствие последействия* (вероятность поступления определенного числа событий за отрезок времени не зависит от числа ранее поступивших требований).

Для распределения Пуассона (рис. 38)

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, (x = 0, 1, 2, ...), \; F(x) = \sum_{k=0}^{x} \frac{\lambda^k}{k!} e^{-\lambda},$$

где $\lambda > 0$ – параметр (интенсивность, ожидаемое число событий за единицу времени).

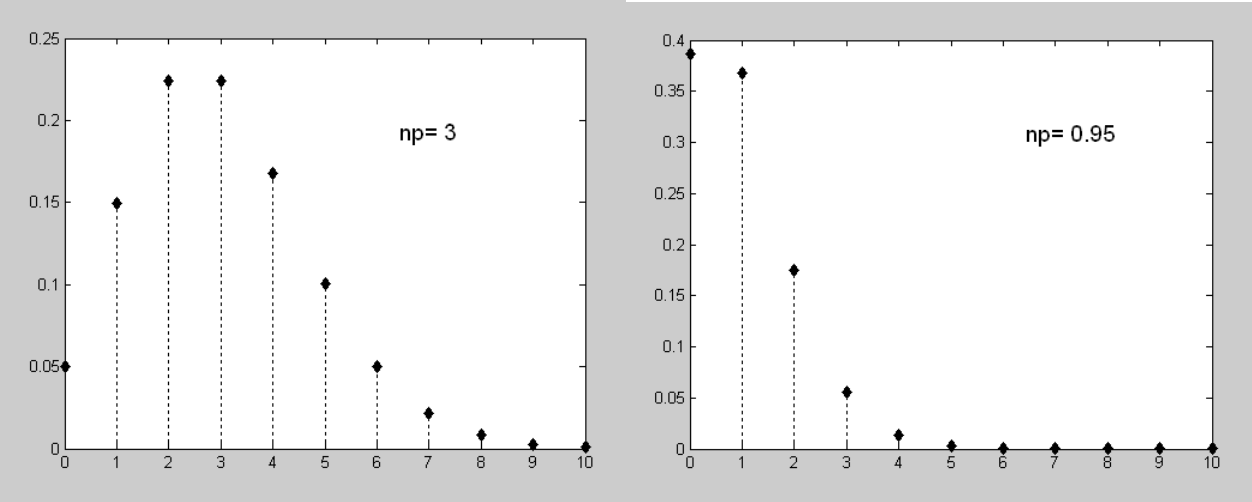


Рис. 38. Плотность распределения Пуассона при различных $\lambda = n\,p$

Математическое ожидание и дисперсия одинаковы и равны $Mx = Dx = \lambda$.

Примечательно, что при $\lambda > 9$ распределение Пуассона можно аппроксимировать нормальным распределением со средним и дисперсией, равными λ.

11.6.7. Распределение Маркова – Пойа

Это распределение впервые появилось в работе А. А. Маркова, опубликованной в 1917 г. В 1923 г. в работе Ф. Эггенбергера и Д. Пойа вводится такое же распределение.

Суть распределения сводится к модели серии экспериментов: из урны, содержащей N шаров ($N\,p$ черных и $N\,(1-p)$ белых), осуществляется выбор с возвратом, как и в схеме Бернулли, но кроме выбранного шара в урну возвращается c новых шаров того же цвета (процесс таких выборов продолжается n раз). Возникает *эффект последействия* (извлечение черного шара увеличивает вероятность извлечь такой же шар при следующей выборке). Подобные явления возникают при распространении заразных заболеваний, в профилактике преступности и т. п.

Соответственно, число x выборов черных шаров в серии n испытаний ($0 \le x \le n$) определяется формулой (рис. 39):

$$P(k \mid a,b,c,n) = \frac{C_n^k \cdot \prod_{j=0}^{k-1}(a+jc) \cdot \prod_{j=0}^{n-k-1}(b+jc)}{\prod_{j=0}^{n-1}(a+b+jc)},$$

где $a = N\,p$ – число шаров черного и $b = N\,(1-p)$ – белого цвета; p – начальная вероятность выбора шара черного цвета.

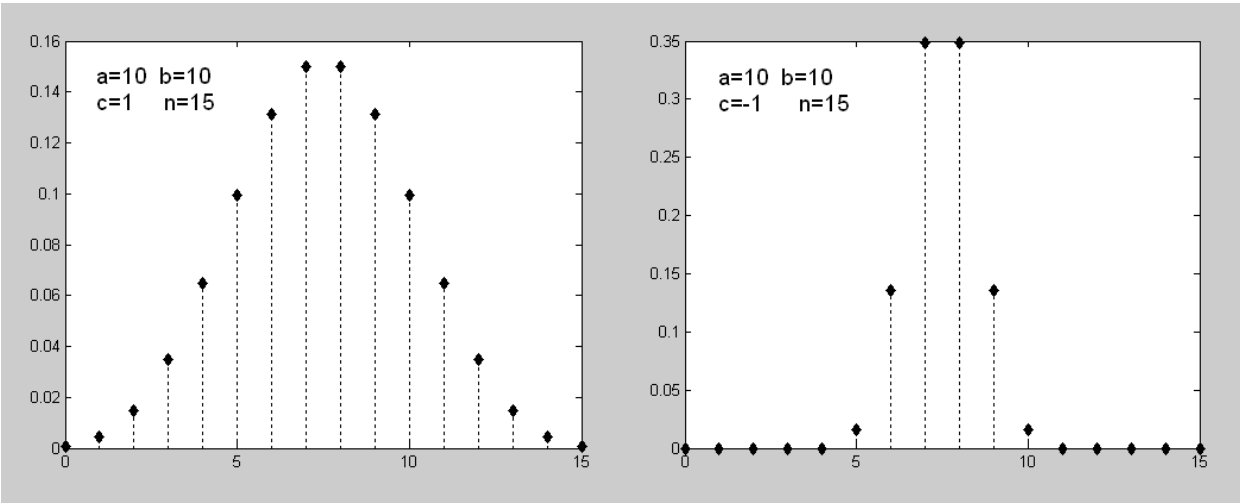


Рис. 39. Плотность распределения Маркова – Пойа при различных c

Математическое ожидание и дисперсия соответственно равны

$$Mx = np,\ Dx = np(1-p)\frac{1+nc/N}{1+c/N}.$$

При $c = 0$ распределение Маркова – Пойа превращается в биномиальное и при $c = -1$ – в гипергеометрическое.

11.7. Распределения непрерывных случайных величин

Многообразие известных непрерывных распределений велико, и большинство процессов, происходящих в природе, технике и обществе может быть описано с помощью какого-либо из них. Существует немало справочников по статистическим распределениям, из которых можно выделить работу [37]. Почти исчерпывающий список дискретных и непрерывных распределений читатель может обнаружить на сайте matlab.exponenta.ru в документации по системе MatLab [38].

Разумеется, здесь мы ограничимся упоминанием немногих из них, связанных с имитационным моделированием случайных процессов и не затронем распределений, используемых при рассмотрении статистических гипотез (Стьюдента, Фишера, ХИ-квадрат, Колмогорова и др.). Наряду с представлением функции плотности распределения $p(x)$ и функции распределения вероятностей $F(x)$, мы приведем основные статистические характеристики (по крайней мере, оценки математического ожидания Mx и дисперсии Dx) и методику прямого моделирования случайных величин, если легко отыскивается обратная для $F(x)$ функция.

11.7.1. Непрерывное равномерное распределение

Свойства этого распределения мы уже рассматривали при знакомстве с датчиком псевдослучайных чисел, равномерно распределенных в (0, 1), и для отрезка $[a, b]$ указывали, что

$$p(x) = \begin{cases} 0 & \text{при } x < a \\ \dfrac{1}{b-a} & \text{при } a \le x \le b \\ 0 & \text{при } x > b \end{cases}, \quad F(x) = \begin{cases} 0 & \text{при } x < a \\ \dfrac{x-a}{b-a} & \text{при } a \le x \le b \\ 1 & \text{при } x > b \end{cases};$$

$$Mx = \frac{a+b}{2};\ Dx = \frac{(b-a)^2}{12};\ \sigma_x = \frac{b-a}{2\sqrt{3}};$$

$$Ax = 0;\ E(x) = \frac{\mu_4}{\sigma_x^4} - 3 = -1{,}2;\ \mu_4 = \frac{(b-a)^4}{80}.$$

С таким распределением имеют дело при решении многомерных задач методом Монте – Карло, в ряде задач теории массового

обслуживания, при статистическом моделировании наблюдений, подчиненных заданному распределению.

11.7.2. Нормальное распределение

Это распределение, называемое и распределением Гаусса – Лапласа, наиболее распространено в статистической практике исследования явлений природы и общества. Так при обработке данных обнаруживается нормальный закон распределения случайных погрешностей измерений, возникающих из-за совместного воздействия множества случайных факторов. Особая роль нормального распределения определена и *предельными* теоремами теории вероятностей.

Функция плотности распределения вероятностей (рис. 40)

$$p(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma_x^2}}, -\infty < x < \infty,$$

математическое ожидание, мода и медиана равны μ, дисперсия равна σ_x^2, асимметрия и эксцесс равны 0. Что касается функции распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}z^2} dz$$

(*функция Лапласа*, *интеграл вероятностей*), то она имеет отличные полиномиальные аппроксимации.

Обобщением нормального распределения является *распределение Лапласа – Шарлье*

$$p(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{\frac{-R^2}{2}} \left[1 - \frac{A_x}{6}(3R - R^3) + \frac{E_x}{24}(R^4 - 6R^2 + 3)\right], R = \frac{x-\mu}{\sigma_x},$$

где A_x и E_x – коэффициенты асимметрии и эксцесса ($|A_x| < 3$, $-\pi / 2 < E_x < 4$). Медиана и мода зависят от указанных параметров и требуют численного поиска.

Практическое использование нормального распределения во многом связано с «*правилом трех сигм*», утверждающим более чем с 99,5%-й достоверностью, что отклонения случайной величины от μ не превышают $3\sigma_x$, т. е. лежат в диапазоне $[\mu - 3\sigma_x, \mu + 3\sigma_x]$.

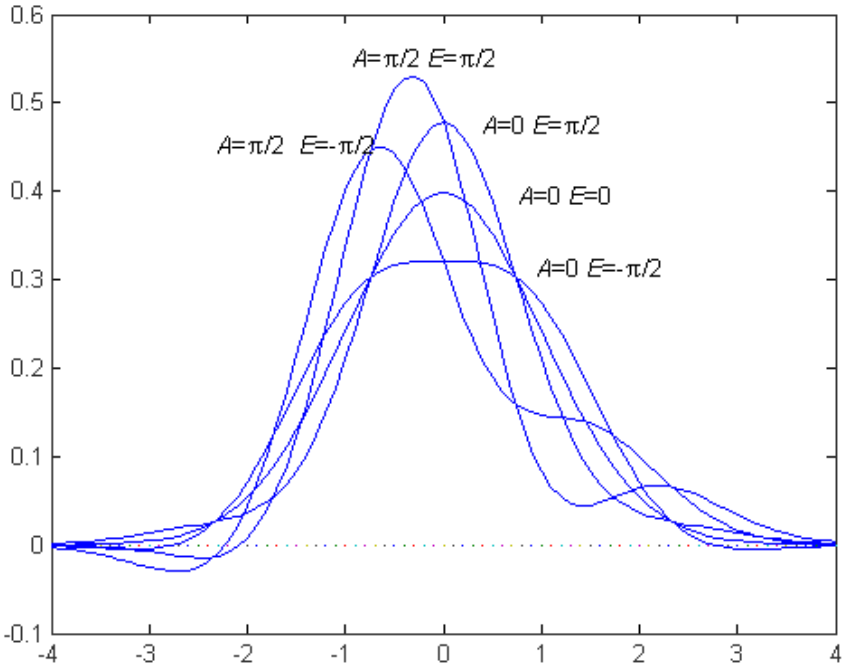


Рис. 40. Распределение Лапласа – Шарлье при малых асимметрии и эксцессе

Существуют разнообразные нетривиальные методы генерации нормального распределения на основе равномерного. Так метод Бокса – Мюллера предлагает получить две независимые нормально распределенные случайные величины N_1 и N_2 из равномерно распределенных на [0, 1] величин R_1 и R_2 с помощью соотношений:

$$N_1 \sim \sqrt{-2\ln(R_1)}\sin(2\pi R_2);\ N_2 \sim \sqrt{-2\ln(R_1)}\cos(2\pi R_2).$$

11.7.3. Экспоненциальное распределение

Экспоненциальное (*показательное*) распределение играет важную роль в теории массового обслуживания и теории надежности. Так, время t между двумя соседними событиями в простейшем потоке событий имеет это распределение с параметром, называемым интенсивностью потока.

Функции плотности и распределения вероятностей выглядят следующим образом:

$$p(x) = \lambda e^{-\lambda x}, x > 0;\quad F(x) = 1 - e^{-\lambda x}.$$

Математическое ожидание $Mx = 1 / \lambda$ и дисперсия $Dx = 1 / \lambda^2$.

Кстати, это одно из немногих распределений, моделируемых предельно просто методом обратной функции: $x = -\ln(z)/\lambda$, где z – равномерно распределенная на отрезке [0, 1] случайная величина.

11.7.4. Распределение Релея

Распределение Релея известно широким применением в теории связи.

Плотность и функция распределения вероятностей:

$$p(x)=\frac{x}{\beta^2}e^{-x^2/2\beta^2},\ x\in[0,\infty);\ F(x)=1-e^{-x^2/2\beta^2}.$$

Математическое ожидание и дисперсия

$$Mx=\beta\sqrt{\frac{\pi}{2}},\ Dx=\beta^2\frac{4-\pi}{2}.$$

Асимметрия и эксцесс

$$Ax=\frac{2(\pi-3)\sqrt{\pi}}{(4-\pi)^{3/2}},\ Ex=-\frac{6\pi^2-24\pi+16}{(\pi-4)^2}.$$

Аналогом распределения Рэлея в трехмерном пространстве является *распределение Максвелла*, для которого

$$p(x)=\sqrt{\frac{2}{\pi}}\frac{x^2}{\beta^3}e^{-\frac{1}{2}(\frac{x}{\beta})^2},\ x>0;$$

$$F(x)=-\sqrt{\frac{2}{\pi}}\frac{x}{\beta}e^{-\frac{1}{2}\frac{x^2}{\beta^2}}-1+2\Phi(x/\beta),\ \Phi(z)\text{ – функция Лапласа;}$$

$$Mx=\beta\sqrt{\frac{8}{\pi}},\ Dx=\frac{3\pi-8}{\pi}\beta^2,\text{ т. е. }\beta=\sigma_x\sqrt{\frac{\pi}{3\pi-8}}.$$

11.7.5. Распределение Вейбулла

Впервые данное распределение было предложено Е. Вейбуллом (1887 – 1979) для аппроксимации экспериментальных данных о прочности стали на разрыв при усталостных испытаниях. Распределение Вейбулла широко используется для описания закономерностей отказов шариковых подшипников, вакуумных приборов или элементов электроники. В теории надежности оно используется при оценке времени безотказной работы элементов машин. Известны примеры использования распределения при учете эффекта полураспада радиоактивных элементов, ядовитых веществ, остаточных знаний и др.

Для двухпараметрического распределения Вейбулла

$$p(x)=\frac{A}{B}x^{A-1}e^{-\frac{x^A}{B}},\ F(x)=1-e^{-\frac{x^A}{B}},\ x>0\ \ (A\ge 1;\ B>0).$$

Достаточно просто найти медиану $Me=(B\cdot\ln(2))^{1/A}$ и моду $Mo=[B\cdot(A-1)/A]^{1/A}$. Несколько сложнее выглядит математическое ожидание

$$Mx=\frac{A}{B}\int_0^\infty x^A e^{-\frac{x^A}{B}}dx=\left\langle x=z^{1/A};\ z=Bt\right\rangle=B^{\frac{1}{A}}\int_0^\infty t^{\frac{1}{A}}e^{-t}dt=B^{\frac{1}{A}}\Gamma(1+\tfrac{1}{A}),$$

где $\Gamma(\lambda)=\int_0^\infty t^{\lambda-1}e^{-t}dt$ – гамма-функция[21]. Учитывая

$$\sigma_x^2+(Mx)^2=\frac{A}{B}\int_0^\infty x^{A+1}e^{-\frac{x^A}{B}}dx=B^{\frac{2}{A}}\Gamma(1+\tfrac{2}{A}),$$

видим, что A – корень уравнения

$$L(A)=V_x^2+1-\Gamma(1+\tfrac{2}{A})/\Gamma^2(1+\tfrac{1}{A})=0;\ B=\left[\frac{Mx}{\Gamma(1+\frac{1}{A})}\right]^A.$$

(V_x – коэффициент вариации).

Из предельных оценок следует ограничение $V_x<1$. В результате табулирования $L(A)$ при A в диапазоне [1, 10] нами получена последовательность значений V_x^2, приемлемых для данного распределения: 0,27 0,13 0,08 0,05 0,04 0,03 0,02 0,018 0,014.

Частными случаями распределения Вейбулла являются экспоненциальное распределение при $A=1$ и распределение Релея при $A=2$.

11.7.6. Распределение Эрланга

Распределение, названное по имени А. Эрланга (1878 – 1920), было одним из первых, использованных в задачах теории массового обслуживания, и применялось к промежуткам между случайными событиями, распределению времени обслуживания в телефонии, к количеству посещений веб-сайта (утверждается существование

[21] При целочисленном $x>0$ $\Gamma(x)$ совпадает с факториалом $\Gamma(n+1)=n!$, $\Gamma(1)=\Gamma(2)=1$, в общем случае $\Gamma(x+1)=x\cdot\Gamma(x)$, $\Gamma(0{,}5)=\sqrt{\pi}$, $\Gamma(x\to 0)\to\infty$ ($\Gamma(10^{-10})=10^{10}$, $\Gamma(30)=8{,}84(10\ 30)$. Существует отличная полиномиальная аппроксимация гамма-функции.

факта, что следующий по посещаемости сайт имеет посещений вдвое меньше, чем предыдущий).

Его плотность и функция распределения имеют вид

$$p(x) = \frac{(x/b)^{c-1}}{b\, e^{\frac{x}{b}}\ (c-1)!};\ F(x) = 1 - e^{-\frac{x}{b}} \sum_{k=0}^{c-1} (x/b)^k / k!.$$

Математическое ожидание и дисперсия: $Mx = b\ c$; $Dx = b^2\ c$.

При $c = 1$ распределение Эрланга совпадает с экспоненциальным.

11.7.7. Гамма-распределение

Гамма-распределение является одним из наиболее распространенных в статистической теории радиотехники при разработке систем обработки сигналов, излучений сложных радиоисточников и др.

Функция плотности распределения вероятностей имеет вид

$$p(x) = \frac{1}{b^a\ \Gamma(a)}\ x^{a-1}\ e^{-x/b};\ x \in (0,\ 1);\ a,\ b > 0$$

где $\Gamma(a)$ – гамма-функция. Функция распределения, к сожалению, не выражается в элементарных функциях:

$$F(x) = \frac{1}{b^a\ \Gamma(a)} \int_0^x t^{a-1}\ e^{-t/b} dt;\ x \in (0,\ 1).$$

Математическое ожидание, дисперсия, асимметрия, эксцесс и мода распределения равны:

$$Mx = a\ b;\ Dx = \sigma_x^2 = b\sqrt{a};\ Ax = \frac{2}{\sqrt{a}};\ Ex = 6 / a;\ Mo = b\ (a - 1).$$

Можно показать, что

$$a = \left[\frac{Mx}{\sigma_x}\right]^2 = \frac{1}{V_x^2};\quad b = \frac{\sigma_x^2}{Mx} = \sigma_x V_x.$$

Для построения соответствующих случайных величин можно брать $x \sim (-a)\ \ln(\prod_{i=1}^{b} z_i)$, где z_i – равномерно распределенные на (0, 1) случайные величины.

При $a = 1$, $b = 1$ гамма-распределение превращается в экспоненциальное.

11.7.8. Бета-распределение

Функция плотности бета-распределения:

$$p(x) = \frac{1}{\beta(a,b)}\, x^{a-1}\,(1-x)^{b-1},\ x \in [0,1],$$

где $\beta(a,b) = \int_0^1 t^{a-1}\,(1-t)^{b-1}\,dt$ – так называемая полная бета-функция.

Функция бета-распределения также не выражается в элементарных функциях:

$$F(x) = \int_0^x \frac{1}{\beta(a,b)}\, t^{a-1}\,(1-t)^{b-1}\,dt.$$

Математическое ожидание и дисперсия вычисляются по формулам $Mx = a / (a + b)$; $Dx = a\,b / ((a + b)^2(a + b + 1))$.

При $a = b = 1$ бета-распределение превращается в равномерное, при $a = b = 0{,}5$ – в распределение арксинуса (см. ниже).

11.7.9. Распределение арксинуса

Распределение арксинуса задается следующими функциями плотности и распределения вероятностей:

$$p(x) = \frac{1}{\pi}\left(\sqrt{x\,(1-x)}\right)^{-1} \text{ при } x \in (0,\,1);$$

$$F(x) = \frac{2}{\pi}\arcsin\sqrt{x} \text{ при } x \in (0,\,1).$$

Распределение арксинуса используется в теории случайных блужданий (броуновского движения).

Обобщенное распределение арксинуса представляет собой частный случай бета-распределения при $a = b$. Функция плотности обобщенного распределения арксинуса:

$$f(x \mid \alpha) = \begin{cases} \dfrac{\sin \pi\alpha}{\pi}\, x^{-\alpha}(1-x)^{\alpha-1}, & 0 < x < 1. \end{cases}$$

Математическое ожидание и дисперсия:

$$Mx = 1 - \alpha;\ Dx = \frac{1}{2}(1-\alpha)\alpha.$$

11.7.10. Распределение Коши

Распределение Коши характеризуется двумя параметрами и является частным случаем распределения Стьюдента, популярного при проверке гипотез:

$$p(x)=\frac{\lambda}{\pi[\lambda^2+(x-a)^2]};F(x)=\frac{1}{\pi}\operatorname{arctg}(\tfrac{x-a}{\lambda})+0{,}5,$$

где a – параметр сдвига, мода и медиана; λ – параметр масштаба.

Для $F(x)$ сравнительно просто находится обратная функция:

$$F^{-1}(x)=x_0+\lambda\cdot\operatorname{tg}(\pi(x-0{,}5)),$$

что способствует легкой генерации соответствующих псевдослучайных величин, но не существует ни один из моментов этого распределения, даже математическое ожидание.

11.7.11. Распределение Лапласа

Это распределение, иногда называемое *двусторонним показательным*, имеет функцию плотности распределения вероятностей:

$$p(x)=\frac{1}{2b}e^{-\left|\frac{x-a}{b}\right|},-\infty<x<\infty,$$

где $A = Mx$, $B=\sigma_x/\sqrt{2}$, эксцесс равен 3 и асимметрия отсутствует.

Функция распределения имеет вид

$$F(x)=\begin{cases}\frac{1}{2}e^{\frac{x-a}{b}}, x<a\\ 1-\frac{1}{2}e^{\frac{a-x}{b}}, x>a\end{cases}.$$

11.7.12. Логарифмически нормальное распределение

Для иллюстрации использования этого распределения рассмотрим лазерный луч привода считывания дисков, преодолевающий некоторое расстояние перед считыванием [35]. Разобьем это расстояние на заданное число участков. На каждом таком участке интенсивность луча лазера ослабевает (например, из-за наличия пыли в воздухе), таким образом можно оценить коэффициент потерь на каждом участке. Итоговый коэффициент потерь на всем расстоянии будет распределен по логнормальному закону.

Случайная величина x называется распределенной логарифмически нормально, если ее логарифм $Z = \ln(x)$ распределен по нормальному закону

$$p(x) = \frac{1}{x\,B\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln(x)-A}{B}\right)^2};\ F(x) = \Phi\left(\frac{\ln(x)-A}{B}\right),\ x > 0.$$

Установив равенства

$Mx = e^{A+\frac{B^2}{2}}$, $\sigma_x^2 = e^{2A+B^2}(e^{B^2}-1)$, $Mo = \exp(A - B^2)$, $Me = e^A$, имеем

$$A = \ln\frac{Mx^2}{\sqrt{Mx^2+\sigma_x^2}},\ B = \sqrt{\ln\frac{Mx^2+\sigma_x^2}{Mx^2}}.$$

11.7.13. Степенное распределение

Степенное распределение является самым простым. Здесь плотность и функция распределения выглядят следующим образом:

$$p(x) = c\ x^{c-1};\ F(x) = x^c.$$

Математическое ожидание и дисперсия:

$$Mx = \frac{c}{c+1}; Dx = \frac{c}{(c+1)^2(c+2)}.$$

Генерация случайных чисел этого распределения производится посредством простой обратной функции $x = z^{1/c}$, где z – равномерно распределенная на (0, 1) случайная величина.

11.7.14. Логистическое распределение

Логистическое распределение достаточно популярно в экономических исследованиях. Распределение определяется двумя параметрами: $a = Mx$, $b = \sigma_x$ – стандартное отклонение. При использовании обозначения $k = b\sqrt{3}/\pi$ плотность и функция распределения вероятностей представимы в виде:

$$p(x) = \frac{e^{\frac{x-a}{k}}}{k\left(1+e^{\frac{x-a}{k}}\right)^2};\ F(x) = \frac{1}{\left(1+e^{\frac{-(x-a)}{k}}\right)}.$$

Для генерации распределения методом обратной функции используется моделирующая формула $x = a + k \ln\left(\frac{z}{1-z}\right)$, где z – равномерно распределенная на (0, 1) случайная величина.

11.7.15. Распределение Парето

Данное распределение получило широкое распространение в различных задачах экономической статистики с появлением работ Парето о распределении доходов (основа *критерия оптимальности по Парето*).

Функции плотности и распределения вероятностей выглядят следующим образом:

$$p(x) = cx_0^c x^{-(1+c)};\ F(x) = 1 - \left[\frac{x_0}{x}\right]^c,\ x \geq 0.$$

Математическое ожидание и дисперсия конечны при $c > 1$ и c > 2 и соответственно равны:

$$Mx = \frac{c\,x_0}{c-1};\ Dx = \frac{c\,x_0^2}{(c-1)^2(c-2)}.$$

Используется и однопараметрический вариант при $x_0 = 1$

$$p(x) = \frac{c}{x^{c+1}},\ F(x) = c\int_0^x \frac{dx}{x^{c+1}} = 1 - \frac{1}{x^c},$$

где $c = \frac{Mx}{Mx-1} > 1$.

Моделирование распределения Парето здесь представимо соотношением $x = (1/z)^{1/c}$, где z – случайная величина, равномерно распределенная на (0, 1).

12. АЗБУКА ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

12.1. Задачи теории массового обслуживания

Буквально с момента рождения каждому человеку приходится сталкиваться с очередями. Ваши родители сидят в очереди в ЗАГСе, чтобы документально зафиксировать радостный факт вашего рождения, и это ожидание может завершиться через 5 минут или 0,5 часа. Потом они становятся в очередь на ваше попадание в детский сад и эта очередь иногда растягивается на годы. Вы набираете телефонный номер вашей подруги и слышите продолжительные гудки или вежливое приглашение оставить сообщение. Не дозвонившись, вы решаете для экономии времени воспользоваться собственным лимузином и попадаете в традиционную «пробку». Ваш самолет запросил посадку в Рио-де-Жанейро и, получив отказ, совершил посадку в Буэнос-Айресе. В старости отправляетесь на прием к офтальмологу в поликлинику по месту жительства и узнаете, что на ближайший месяц «очередей» нет.

Очереди возникают практически во всех системах массового обслуживания (СМО) и *теория массового обслуживания* (*теория очередей*) занимается оценкой функционирования системы при заданных параметрах и поиском параметров, оптимальных по некоторым критериям. Эта теория представляет особый раздел теории случайных процессов и использует, в основном, аппарат теории вероятностей. Первые публикации в этой области относятся к 20-м годам XX века и принадлежат датчанину А. Эрлангу, занимавшемуся исследованиями функционирования телефонных станций – типичных СМО, где случайны моменты вызова, факт занятости абонента или всех каналов, продолжительность разговора. В дальнейшем теория очередей нашла развитие в работах К. Пальма, Ф. Поллачека, А. Я. Хинчина, Б. В. Гнеденко, А. Кофмана, Р. Крюона, Т. Саати и других советских и зарубежных математиков.

В качестве основных элементов СМО (рис. 41) следует выделить входной поток заявок, очередь на обслуживание, механизм обслуживания и выходящий поток. В роли заявок (требований, вызовов) могут выступать покупатели в магазине, телефонные вызовы, поезда при подходе к железнодорожному узлу, вагоны под разгрузкой, автомашины на станции техобслуживания, самолеты в ожидании разрешения на взлет, штабель бревен при погрузке на автотранспорт. Роль обслуживающих приборов (каналов, линий) играют

продавцы или кассиры в магазине, таможенники, пожарные машины, взлетно-посадочные полосы, экзаменаторы, ремонтные бригады.

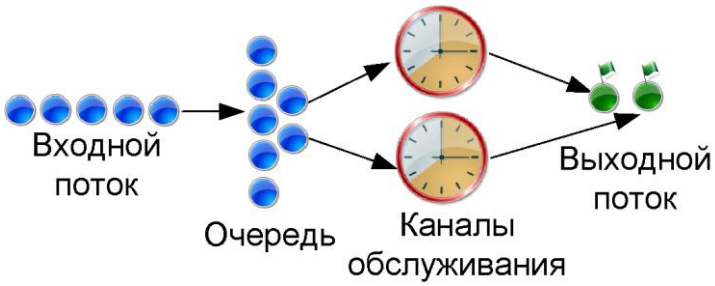


Рис. 41. Схема системы массового обслуживания

В зависимости от характеристик этих элементов (рис. 41) СМО классифицируются следующим образом.

1. *По характеру поступления заявок*. Если интенсивность входного потока (количество заявок в единицу времени) постоянна или является заданной функцией от времени, поток называют *регулярным*. Если параметры потока независимы от конкретного момента времени, поток называют *стационарным*.

2. *По количеству одновременно поступающих заявок*. Поток с нулевой вероятностью одновременного появления двух и более заявок называется *ординарным*.

3. *По связи между заявками.* Если вероятность появления очередной заявки не зависит от количества предшествующих заявок, имеем дело с потоком *без последействия*.

4. *По однородности заявок* выделяют *однородные* и *неоднородные потоки*.

5. *По ограниченности потока заявок* различают *замкнутые* и *разомкнутые* системы (система с ограниченной клиентурой называется замкнутой). Так универсальный магазин является разомкнутой системой, тогда как оптовый магазин с постоянными клиентами – замкнутая система.

6. *По поведению в очереди* системы делятся на системы *с отказами* (заявка покидает систему, если нет мест в очереди), *с ограниченным или неограниченным по времени ожиданием*.

7. *По дисциплине выбора на обслуживание*. Здесь можно выделить системы с обслуживанием в порядке поступления, в случайном порядке, в порядке, обратном поступлению (последний пришел – первым обслужен) или с учетом приоритетов.

8. *По числу каналов обслуживания* системы разделяют на одно- и многоканальные.

9. *По времени обслуживания* выделяют системы *с детерминированным* и *случайным* временем.

10. *По количеству этапов обслуживания* различают *однофазные* и *многофазные* системы.

12.2. Математический аппарат анализа простейших систем массового обслуживания

Рассмотрим *стационарный поток однородных заявок* без последействия. Пусть $P_k(\tau)$ вероятность появления k заявок в интервале времени τ. Эта вероятность зависит только от τ и не зависит от начала отсчета времени, от поступления заявок в предыдущих временных интервалах. Пусть к тому же поток является ординарным, т. е. $P_k(dt)$ при $k > 1$ бесконечно мала в сравнении с малым интервалом dt. Если обозначить через λ число заявок в единицу времени (интенсивность потока), то можно показать, что такой *простейший* поток подчинен *распределению Пуассона*

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

Для пуассоновского потока можно обнаружить, что промежутки времени T между поступлениями заявок распределены по экспоненциальному (показательному) закону

$$P(T < t) = 1 - e^{-\lambda t} \qquad (2)$$

(вероятность того, что промежуток времени T не превышает значения t) .

Естественно, что входной поток может описываться не только пуассоновским, но и другими распределениями (Эрланга, гиперэкспоненциальным и т. п.).

Аналогичная ситуация имеет место и для выходного потока. Чаще всего используется показательный закон распределения времени обслуживания заявки:

$$P(t) = 1 - e^{-\mu t}, \qquad (3)$$

где $\mu = 1 / t_{обс}$ – интенсивность обслуживания (среднее число обслуживаний в единицу времени); $t_{обс}$ – среднее время обслуживания.

Обозначим через S множество состояний системы. Пусть $P(l, t + \tau / i, t)$ – вероятность того, что система, находившаяся в момент t в состоянии i, в момент $t + \tau$ окажется в состоянии l. Для мар-

ковских систем (привлекательных отсутствием последействия) можно записать *уравнения Чепмена – Колмогорова*:

$$P(l, t + \tau / i, t) = \sum_{j \in S} P(j, t + \tau^* / i, t)\, P(l, t + \tau / j, t + \tau^*). \qquad (4)$$

Если под состояниями понимать число заявок, то эти уравнения можно записать в виде:

$$P_k(t + \tau) = \sum_{i + j = k} P_i(t)\, P_j(\tau). \qquad (5)$$

Рассмотрим *случай разомкнутой системы* с простейшим входным потоком интенсивности λ и одним каналом обслуживания с интенсивностью μ.

Возьмем интервал времени $[t, t + dt]$. В силу разомкнутости системы множество ее состояний $S = \{S_0, S_1, S_2, ..., S_k, S_{k+1}, ...\}$, где S_k – состояние, когда в системе находится k заявок.

Попробуем оценить вероятности перехода между состояниями с учетом того, что вероятность появления заявки в этом интервале времени равна $\lambda\, dt$ и вероятность завершения обслуживания предшествующей заявки равна $\mu\, dt$.

Очевидно, что вероятность перехода $S_0 \to S_1$ равна $\lambda\, dt$ и вероятность перехода $S_1 \to S_0$ равна $1 - \lambda\, dt$. Если в системе присутствовали $k > 0$ заявок (состояние S_k), то для перехода в состояние S_{k-1} необходимо, чтобы заявка была обслужена и не поступило новой заявки; отсюда вероятность перехода $S_k \to S_{k-1}$ равна $\mu\, dt\,(1 - \lambda\, dt) \cong \mu\, dt$. Для перехода из состояния S_k в состояние S_{k+1} необходимо, чтобы поступила новая заявка, но ни одна из ранее поступивших не обслужена: вероятность перехода $S_k \to S_{k+1}$ равна $\lambda\, dt \cdot (1 - \mu\, dt) \cong \lambda\, dt$. Вероятность остаться в том же состоянии составит $1 - (\lambda + \mu)\, dt$.

Тогда из (5) имеем

$$P_0(t + dt) = (1 - \lambda\, dt)\, P_0(t) + \mu\, dt\, P_1(t),$$

$$P_k(t + dt) = \lambda\, dt\, P_{k-1}(t) + (1 - \lambda\, dt - \mu\, dt)\, P_k(t) + \mu\, dt\, P_{k+1}(t),\ k > 0.$$

Используя предельный переход при $dt \to 0$, получаем систему обыкновенных дифференциальных уравнений для описания состояний СМО:

$$\begin{cases} \dfrac{d}{dt}P_0(t) = -\lambda\ P_0(t) + \mu\ P_1(t)\ , \\ \dfrac{d}{dt}P_k(t)\ = \lambda\ P_{k-1}(t) - (\lambda + \mu\ )\,P_k(t) + \mu\, P_{k+1}(t)\,, k > 0. \end{cases} \quad (6)$$

Решение (6) при заданных начальных условиях для непрофессионала в численном анализе может оказаться затруднительным (воспользоваться преобразованием Лапласа или прибегнуть к численному решению задачи Коши для системы обыкновенных дифференциальных уравнений большого порядка).

Если же ограничиться рассмотрением *установившегося* режима, признаком которого является существование предела

$$\lim_{t\to\infty} P_k(t) = P_k,\ k = 0, 1, 2, ..., \quad (7)$$

система (6) приведется к бесконечной системе линейных алгебраических уравнений с трехдиагональной матрицей коэффициентов:

$$\lambda\, P_0 = \mu\, P_1,\ (\lambda + \mu)\, P_k = \lambda\, P_{k-1} + \mu\, P_{k+1},\ k = 1, 2, 3, ... \quad (8)$$

Обозначив $\rho = \lambda / \mu$, имеем $P_1 = \rho\, P_0$, $P_2 = \rho^2\, P_0$, ..., $P_k = \rho^k\, P_0$, ..., откуда с учетом $P_0 + P_1 + P_2 + ... + P_k + ... = 1$ получаем при $\rho < 1$ $P_0\,(1 + \rho + \rho^2 + \rho^3 + ... + \rho^k + ...) = P_0 / (1 - \rho) = 1$. Тогда

$$P_0 = 1 - \rho,\ P_k = \rho^k\, P_0 \text{ для } k = 1, 2, ...,\ \rho = \lambda / \mu < 1. \quad (9)$$

Обратите внимание на требование $\rho < 1$. Если это требование нарушено, ни о каком установившемся режиме не может быть речи: очередь растет неограниченно (средняя продолжительность обслуживания больше среднего интервала времени между заявками).

Теперь обратимся к аналогичной *замкнутой системе* с числом заявок, не превышающим n. Здесь система уравнений (6) приведется к конечной системе

$$\begin{cases} \dfrac{d}{dt}P_0(t)\ = -\lambda\ P_0(t)\ +\ \mu\ P_1(t)\ , \\ \dfrac{d}{dt}P_k(t)\ = \lambda\ P_{k-1}(t) - (\lambda + \mu\ )\,P_k(t) + \mu\, P_{k+1}(t)\,, k = 1,...,n-1, \\ \dfrac{d}{dt}P_n(t)\ = -\mu\ P_n(t)\ +\ \lambda\ P_{n-1}(t), \end{cases} \quad (10)$$

которая для установившегося режима дает конечную систему линейных алгебраических уравнений

$$\begin{gathered} \lambda\, P_0 = \mu\, P_1; \\ (\lambda + \mu)\, P_k = \lambda\, P_{k-1} + \mu\, P_{k+1},\ k = 1, 2, ..., n - 1; \\ \lambda\, P_{n-1} = \mu\, P_n\,. \end{gathered} \quad (11)$$

Решение системы

$$P_1 = \rho\, P_0,\ P_2 = \rho^2\, P_0 / 2,\ P_3 = \rho^3\, P_0 / (2\cdot 3),\ P_4 = \rho^4\, P_0 / (2\cdot 3\cdot 4)$$

дает

$$P_0 = \left[\sum_{k=0}^{n} \frac{\rho^k}{k!} \right]^{-1},\ P_k = P_0\, \rho^k / k! \text{ для } k = 1, 2, \ldots, n.$$

Полученные решения можно обобщить на случай многоканальных систем с ограниченным ожиданием. Так, при N однотипных каналах обслуживания (интенсивность обслуживания равна $N\cdot\mu$), m мест в очереди и числе n возможных заявок большем $N + m$ (в противном случае нет проблем) возникает система

$$\begin{cases} \frac{d}{dt} P_0(t) = -\lambda\, P_0(t) + \mu\, P_1(t); \\ \frac{d}{dt} P_k(t) = \lambda\, P_{k-1}(t) - (\lambda + k\mu)\, P_k(t) + (k+1)\mu\, P_{k+1}(t),\ k = 1, \ldots, N-1; \\ \frac{d}{dt} P_k(t) = \lambda\, P_{k-1}(t) - (\lambda + N\mu)\, P_k(t) + N\mu\, P_{k+1}(t),\ k = N, \ldots, N+m-1; \\ \frac{d}{dt} P_{N+m}(t) = -N\mu\, P_{N+m}(t) + \lambda\, P_{N+m-1}(t), \end{cases}$$

из которой для установившегося режима

$$\begin{gathered} \lambda\, P_0 = \mu\, P_1; \\ (\lambda + k\,\mu)\, P_k = \lambda\, P_{k-1} + (k+1)\,\mu\, P_{k+1},\ k = 1, 2, \ldots, N-1; \qquad (12) \\ (\lambda + N\,\mu)\, P_k = \lambda\, P_{k-1} + N\,\mu\, P_{k+1},\ k = N, N+1, \ldots, N+m-1; \\ \lambda\, P_{N+m-1} = N\,\mu\, P_{N+m}. \end{gathered}$$

Решение этой системы дает

$$P_k = P_0\, \rho^k / k! \text{ для } k = 1, 2, \ldots, N-1; \qquad (13)$$

$$P_k = P_0\, \rho^k / (N^{k-N}\, N!) \text{ для } k = N, N+1, \ldots, N+m; \qquad (14)$$

$$P_0 = \left[\sum_{k=0}^{N} \frac{\rho^k}{k!} + \frac{\rho^{N+1}}{N\, N!} \frac{(\rho / N)^m - 1}{(\rho / N) - 1} \right]^{-1}. \qquad (15)$$

Умение найти значения P_k дает возможность отыскать и ряд основных характеристик СМО.

12.3. Основные характеристики систем массового обслуживания

Значение P_0 определяет вероятность того, что все каналы обслуживания свободны (находятся в состоянии простоя).

Значение P_k определяет вероятность того, что в системе (в очереди и на обслуживании) находятся k заявок. Если k не превышает числа каналов N, то все заявки находятся на обслуживании и очередь отсутствует; в противном случае все каналы заняты и $k - N$ заявок находится в очереди.

Вероятность отказа в обслуживании $P_{\text{отк}} = P_{N+m}$ (ситуация занятости всех N каналов и всех m мест в очереди).

Среднее число занятых каналов $N_{\text{зан}}$ определяется математическим ожиданием дискретной случайной величины [39]:

$$N_{\text{зан}} = \sum_{k=1}^{N} k\,P_k + \sum_{k=N+1}^{N+m} N\,P_k = \rho\left[1 - \frac{\rho^{N+m}}{N!\cdot N^m}P_0\right] \qquad (16)$$

(мы опускаем здесь достаточно простые преобразования).

Среднее число свободных каналов

$$N_{\text{своб}} = N - N_{\text{зан}}. \qquad (17)$$

Коэффициент простоя каналов

$$K_{\text{прост}} = N_{\text{своб}} / N. \qquad (18)$$

Коэффициент занятости каналов

$$K_{\text{занят}} = N_{\text{зан}} / N. \qquad (19)$$

Относительная пропускная способность (доля обслуженных заявок в общем числе поступавших в систему)

$$q = 1 - P_{\text{отк}}. \qquad (20)$$

Абсолютная пропускная способность (среднее число заявок, обслуживаемых в единицу времени) определяется величиной

$$A = \lambda\, q. \qquad (21)$$

Средняя длина очереди

$$L_{\text{очер}} = \sum_{k=N+1}^{N+m} (k-N)P_k = \frac{\rho^{N+1}}{N!\cdot N}\,\frac{1-(\rho/N)^m(m+1-m\rho/N)}{(1-\rho/N)^2}\,P_0. \qquad (22)$$

Среднее число заявок, находящихся в системе:

$$L = N_{\text{зан}} + L_{\text{очер}}. \qquad (23)$$

Среднее время пребывания заявки в очереди

$$T_{\text{очер}} = L_{\text{очер}} / \lambda. \qquad (24)$$

Общее время пребывания заявки в очереди будет складываться из $T_{\text{очер}}$ и среднего времени обслуживания:

$$T_{\text{сист}} = T_{\text{очер}} + q / \mu. \qquad (25)$$

Полученные характеристики дают возможность анализа замкнутых и разомкнутых систем с отказами ($m = 0$), с очередью или с ожиданием ($m \to \infty$) при простейшем входном потоке и однотипных параллельных каналах обслуживания с показательным законом длительности обслуживания (в частности, с фиксированной длительностью).

12.4. Примеры систем с ограниченной очередью

Пример 1. На аэродром самолеты прибывают с интенсивностью 27 самолетов в час, время приземления составляет 2 минуты, допустимо нахождение над аэродромом не более $m = 10$ самолетов. Нужно определить число N посадочных полос, гарантирующее вероятность отказа, не превышающую 0,05, и среднее время ожидания, не превышающее 5 минут.

Здесь $\lambda = 27$, $\mu = 30$, $\rho = \lambda / \mu = 0{,}9$.

Вероятность простоя службы посадки согласно (15):

$$P_0 = \left[\sum_{k=0}^{N} \frac{0{,}9^{\,k}}{k!} + \frac{0{,}9^{\,N+1}}{N \cdot N!} \frac{(0{,}9/N)^{10} - 1}{(0{,}9/N) - 1} \right]^{-1}.$$

Вероятность отказа в посадке равна

$$P_{\text{отк}} = P_0 \, 0{,}9^{\,N+10} / (N^{10} \, N\,!) \,.$$

Среднее время ожидания в воздухе согласно (24) и (22)

$$T_{\text{очер}} = L_{\text{очер}} / \lambda,$$

где

$$L_{\text{очер}} = \sum_{k=N+1}^{N+m} (k - N) \, P_k = \frac{\rho^{N+1}}{N\,! \cdot N} \frac{1 - (\rho/N)^m (m + 1 - m\rho/N)}{(1 - \rho/N)^2} P_0 =$$

$$= \frac{0{,}9^{N+1}}{N!\,N} \frac{1 - (0{,}9/N)^{10} (11 - 10 \cdot 0{,}9/N)}{(1 - 0{,}9/N)^2} P_0 .$$

Выполняя арифметические действия при $N = 1$, обнаруживаем, что

$$P_0 \cong 0{,}14; \; P_{\text{отк}} \cong 0{,}04; \; L_{\text{очер}} \cong 0{,}045; \; T_{\text{очер}} \cong 0{,}9 \text{ мин}$$

и что одной посадочной полосы при указанных условиях вполне достаточно.

Пример 2. Имеются станки, которые могут выходить из строя с частотой в среднем 2 раза за смену. Продолжительность ремонта

одним оператором составляет около трех часов (оператор одновременно может ремонтировать лишь один станок и не переходит к другому, не отремонтировав предыдущий). Хотелось бы определить число операторов, при котором потери от простоя станков и оплаты лишнего числа операторов были бы минимальны.

Такую замкнутую систему можно представить системой с N каналами (операторами) и очередью с m местами ожидания (совпадает с числом станков). Если известны потери C_n от простоя станка в течение часа и оплата $C_р$ часа работы оператора, то при семичасовой смене задача сводится к нахождению значения N, которое минимизировало бы выражение

$$C_n\, T_{очер} + C_р \cdot 7 \cdot N,$$

где $T_{очер}$ определяется из (22) и (24) при $\lambda = 2 / 7$, $\mu = 1 / 3$, $\rho = \lambda / \mu = 6 / 7$.

Можно привести множество подобных задач для определения числа кассиров в универмаге, наилучшего с позиций минимума потерянных покупателей, числа бригад грузчиков на железнодорожной станции, минимизирующего штрафы за простой вагонов, числа полос движения на проектируемой автомагистрали и т. п.

12.5. Дисциплина ожидания и приоритеты

Выше рассматривался простейший поток однотипных заявок с дисциплиной выборки из очереди на обслуживание *в порядке поступления*.

Можно показать, что и в ситуации *случайного выбора на обслуживание* полученные выше оценки не претерпят изменения, но их дисперсия (разброс относительно ожидаемой величины) возрастет. Среднее время сидения в очереди не изменится, если кто-то пройдет без очереди, но для отдельных клиентов время ожидания увеличится. Так отношение дисперсий времени ожидания в неупорядоченной и упорядоченной очереди имеет порядок $(2 + \rho) / (2 - \rho)$, где $\rho = \lambda / \mu$ (мы обычно предпочитаем систему с жесткой дисциплиной обслуживания из-за предсказуемости ее поведения и всякое «возмущение» в ее работе отрицательно действует на нашу психику).

Существует много систем, в которых присутствует N входных потоков с различной интенсивностью λ_i ($i = 1, .. N$) и время обслуживания заявок которых распределено по показательному закону с

параметрами μ_i. Здесь при условии пуассоновости входных потоков можно считать, что и суммарный поток будет пуассоновским с интенсивностью $\Lambda = \sum\lambda_i$; функция распределения времени обслуживания заявок суммарного потока в одноканальной системе

$$S(t) = \sum_{i=1}^{N} \frac{\lambda_i}{\Lambda}(1 - e^{-\mu_i t});$$

среднее время ожидания по формуле Полачека – Хинчина:

$$W = \frac{\Lambda / 2 \int_0^{\infty} t^2 dS(t)}{1 - R_N}, \quad R_N = \Lambda \int_0^{\infty} t\, dS(t),$$

которая для данного случая дает $R_N = \sum[\lambda_i / \mu_i] = \sum\rho_i$ и в случае стационарности режима ($R_N < 1$)

$$W = \frac{1}{1 - R_N} \sum_{i=1}^{N} \frac{\rho_i}{\mu_i}.$$

Можно показать [39], что поскольку время ожидания заявки с приоритетом k складывается из времени завершения обработки требования, вошедшего в канал, времени обслуживания ранее поступивших требований приоритета от 1 до $k - 1$ (наивысший приоритет определяется $k = 1$) и ранее поступивших требований с приоритетом k, то его среднее значение равно

$$W_k = \frac{\sum_{i=1}^{N} \rho_i / \mu_i}{(1 - R_k)(1 - R_{k-1})}; \quad R_k = \sum_{i=1}^{k} \rho_i \ (k = 1 .. N).$$

На этой основе можно определить среднюю длину очереди заявок k-го приоритета $L_k = \lambda_k \cdot W_k$ и среднее число таких заявок в системе $L_k + \rho_k$. Показано, что введение приоритетов улучшает функционирование системы, если более высокое преимущество присваивается заявкам с меньшей длительностью обслуживания. Если учитывать стоимостные характеристики, то такое преимущество предоставляется заявкам с большим значением $C_k \cdot \mu_k$, где C_k – средняя стоимость ожидания [40].

Существуют системы с *абсолютными приоритетами*, где появление заявки более высокого уровня прерывает обслуживание те-

кущей заявки, которая возвращается в очередь и потом снова поступит на обслуживание с места прерывания (или с начала). Здесь среднее время ожидания заявки с приоритетом k

$$W_k = \frac{\sum_{i=1}^{N} \rho_i / \mu_i}{(1-R_k)(1-R_{k-1})} + \frac{R_{k-1}}{\mu_k (1-R_{k-1})}; \quad R_k = \sum_{i=1}^{k} \rho_i \ (k = 1 .. N).$$

Исключительно сложно установить разумные приоритеты в случае многофазных систем, где заявка проходит обслуживание в нескольких последовательных подсистемах [40, 41]. Здесь относительно простые выводы удается сделать лишь для случая двух подсистем и для получения выводов для более сложных систем приходится прибегать к статистическому моделированию.

12.6. Статистическое моделирование систем массового обслуживания

До сих пор мы рассматривали СМО, для которых удавалось описать результаты обследования в аналитическом виде. Однако многие реальные системы характеризуются отнюдь не пуассоновскими входными потоками или экспоненциальным распределением длительности обслуживания. Получение аналитических оценок для большинства других распределений, тем более для эмпирических, не столь тривиально. Осложняется получение решений и в случае нестационарности процесса (время функционирования системы не позволяет ей войти в стационарный режим). Практически нет простых решений для многофазных систем и систем с оригинальными приоритетами.

Поэтому поступают следующим образом. Время функционирования СМО разделяется на достаточно большое количество подынтервалов (единиц времени, в течение которых не может возникнуть более одной заявки или завершиться выполнение более одной заявки). Для каждого такого подынтервала последовательно моделируется факт появления новой заявки (да/нет), проверяется наличие свободного канала (закончено ли обслуживание какой-то заявки) и загрузка его заявкой из очереди, проверяется наличие мест в очереди с последующим выводом (принять в очередь/отказать в обслуживании) и т. д. При этом фиксируется число отказов, время

ожидания заявок в очереди и в системе вообще, число заявок в очереди в каждый момент и другие значения, которые позволяют найти вероятность отказа, распределение времени ожидания и его среднее время, вероятность простоя каналов и т. п. Для надежности выводов такой акт однократного моделирования повторяется достаточно много раз.

Этот подход (статистическое моделирование) позволяет учесть уже упоминавшиеся многофазность, приоритеты и другие нетривиальные факторы. Более того, появляется возможность учета и получения и стоимостных оценок.

13. МОДЕЛИ ЭКОНОМИЧЕСКИХ СИСТЕМ И ПРОЦЕССОВ

Экономическая теория долгое время оставалась научным рассказом о сложной системе реальной экономики и разные авторы рассказывали о ней по-разному [43]. Достаточно назвать общеизвестных А. Смита (1723 – 1790, «Исследование о природе и причинах богатства народов»); К. Маркса (1818 – 1883, «Капитал»); Д. Кейнса (1883 – 1946, «Общая теория занятости, процента и денег»). Каждый из них создал свою теорию, свою модель экономики, исходя из принципов, которые он считал верными. Обнаружив некоторые закономерности в реальной экономике, он возводил их в принципы и на их основе пытался вывести частные закономерности.

Аналогично и *математическая экономика* (математические модели экономических явлений) строится на основе некоторых постулатов (*системы аксиом*) и правил рассуждения, ведущих к доказательству тех или иных выводов (*теорем*). Так большинство математических моделей микроэкономики принимает за аксиомы то, что потребитель принимает решения, исходя лишь из своей системы предпочтений, а производитель ставит своей целью лишь максимум получаемой прибыли.

Известны различные подходы к классификации экономико-математических моделей. Так их весьма условное разделение на микро- и макроэкономические предполагает, что микроэкономика занимается исследованием взаимодействия отдельных потребителей и производителей, тогда как макроэкономика интересуется экономикой региона в целом (условия равновесия на рынке, общий объем продукции, общий уровень цен, развитие во времени и т. п.).

Разделяются модели и по назначению: балансовые модели соответствия наличия и использования ресурсов; трендовые модели, задающие тенденции основных показателей моделируемой системы; оптимизационные модели для выбора наилучшего из определенного числа вариантов производства или потребления; имитационные модели изучаемых процессов и др. Кроме того модели разделяют на статические и динамические, по типу исходной информации, по используемому математическому аппарату и т. п.

Для иллюстрации одной из значимых сфер приложений метода имитационного моделирования остановимся на характеристике некоторых известных моделей экономических систем и процессов.

13.1. Модели фирмы

Теория фирмы является одной из самых популярных в математической экономике.

Фирма определяется как организация, производящая затраты экономических факторов (товары, услуги и *первичные товары*, к которым относятся труд, земля и т. д.) для производства продукции и услуг, которые она продает потребителям или другим фирмам.

Серьезное математическое описание аксиоматики теории фирмы и ее следствий с обсуждением вопросов ценообразования и теории спроса заинтересованный читатель может найти в работе А. Ф. Терпугова [42]. В дальнейшем описании мы ограничимся лишь наиболее простыми ее элементами, хотя и здесь имитационное моделирование сопряжено с рядом проблем, связанных с получением достоверных эмпирических данных для моделирования, построением адекватной численной модели, убедительной интерпретацией результатов [44].

13.1.1. Вероятностные паутинообразные модели ценообразования

Под *вероятностными* (*стохастическими*) моделями понимают модели, параметры которых (например, спрос на каждом этапе) функционально зависят от их значения на предыдущем этапе времени и экзогенной (внешней) случайной величины, распределенной по известному закону с заданными параметрами. Все эти величины, как правило, распределены по нормальному закону с нулевым математическим ожиданием и известной дисперсией.

Паутинообразная модель ценообразования описывает динамический процесс – траекторию корректировки цен и объема производства при переходах от одного состояния к другому на пути к *состоянию равновесия* и используется для описания колебаний цен на рынках, где предложение реагирует на изменения цен с некоторым запозданием.

В основе модели лежит *закон спроса и предложения* (окончательно сформулирован А. Маршаллом в 1890 г.), устанавливающий зависимость объемов спроса и предложения товаров на рынке от их цен. Так для большинства товаров (исключение немногочисленные так называемые товары Гиффена, спрос на которые растет с ростом цены) при прочих равных условиях (размеры рынка, рост доходов

инфляция, сезон и др.) с понижением цены на товар величина спроса растет, но величина предложения понижается (рис. 42). В реальной экономике в процессе колебания цены устанавливается равновесие между предложением и спросом. Если цена на рынке выше равновесной, то возникает перепроизводство товаров и предложение превышает спрос, а при цене ниже равновесной товар становится дефицитным и спрос превышает предложение.

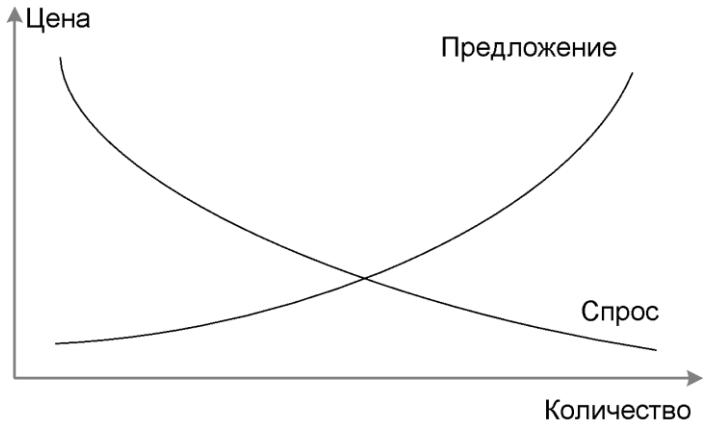


Рис. 42. Кривые спроса и предложения

Естественно, что поведение кривых гиперболического типа (рис. 42), определяющих зависимости спроса и предложения от цены, для разных товаров зависит от степени их эластичности. Например, большая эластичность спроса – изменение спроса в отношении к изменению цены (обе величины в процентах) – вызывает сильную реакцию покупателя («крутое падение кривой спроса») и для реальной экономики подлежит учету в первую очередь.

Единым для любых описываемых ниже моделей можно назвать набор параметров – характеристик, по которым оценивается состояние экономической системы на *t*-м интервале времени:

P_t – цена (*price*) продукции;

S_t – предложение (*supply*) продукции;

D_t – спрос (*demand*) на продукцию.

В результате анализа составляется уравнение функционирования модели, описывающее на основе имеющихся величин (в т. ч. случайных) либо текущее состояние системы, либо следующее ее состояние на основании текущего (с запаздыванием).

В моделях функционирования нескольких фирм, действующих в одной отрасли, для каждой фирмы *i* необходим учет величин C_{it} и PP_{it} (*pure profit*) – ее затрат и полной прибыли на *t*-м интервале времени соответственно.

Теория фирмы допускает, что в отрасли действует одна или несколько фирм, ставящих своей целью получение максимальной

прибыли. Так в схеме экономики по Вальрасу предполагается наличие нескольких производителей, покупателей и типов товаров.

При рассмотрении паутинообразной модели ценообразования вводятся существенные упрощения. Так предполагается, что имеется один продукт и может изменяться только его цена, а все остальные факторы, от которых зависит спрос на данный товар (цены других товаров, налоги и дотации, применяемые технологии и пр.), неизменны.

Различают два подхода к описанию математической модели: непрерывная динамика цен (процесс описывается дифференциальными уравнениями) и дискретная (переменные в течение одного периода неизменны). При дискретном подходе, более простом в моделировании, каждому последовательному интервалу времени t соответствуют значения цены P_t, спроса D_t и предложения S_t и в зависимости от используемых гипотез предполагается либо запаздывание предложения $S(P_{t+1}) = D(P_t)$, либо запаздывание спроса $D(P_{t+1}) = S(P_t)$. Само моделирование сводится к построению кривых спроса, предложения и получению оценок P_t в заданном интервале времени.

13.1.1.1. Классическая вероятностная модель

Классическая вероятностная модель – одна из исторически первых динамических моделей рынка, отражающих поведение участников. Она служит хорошей иллюстрацией применения метода имитационного моделирования для анализа экономических процессов. Значение этой модели определяется еще и тем, что многие современные модели динамики цен, а также динамические модели макроэкономики приводят к «паутинообразному» процессу.

Гипотезы, которые лежат в основе этой модели:

1) производитель товара, принимая решение об объеме предложения (выпуска продукции), ориентируется на цену предыдущего периода;

2) рынок всегда находится в состоянии локального равновесия.

Другими словами:

1) объем предложения на рынке в каждый период времени определяется ценой предыдущего периода при помощи функции предложения $S_{t+1} = S(P_t)$;

2) в каждый период времени $t + 1$ устанавливается равновесная цена P_{t+1}, являющаяся решением уравнения $D(P_{t+1}) = S_{t+1}$;

3) потребитель предъявляет спрос, который при цене P_{t+1} в каждый момент времени равен предложению S_{t+1}, вследствие чего потребитель приобретает все, что ему предложено [45].

Уравнения функционирования модели:

$$S_t = C + D\,P_{t-1} + V_t; \quad (1)$$

$$S_t = D_t + W_t; \quad (2)$$

$$D_t = A - B\,P_t + U_t, \quad (3)$$

где V_t, W_t, U_t – экзогенные (внешние) случайные величины с нормальным распределением при нулевых средних; A, B, C, D – параметры, характеризующие отрасль. Уравнение (2) выступает как условие локального равновесия (спрос совпадает с предложением с точностью до случайной величины).

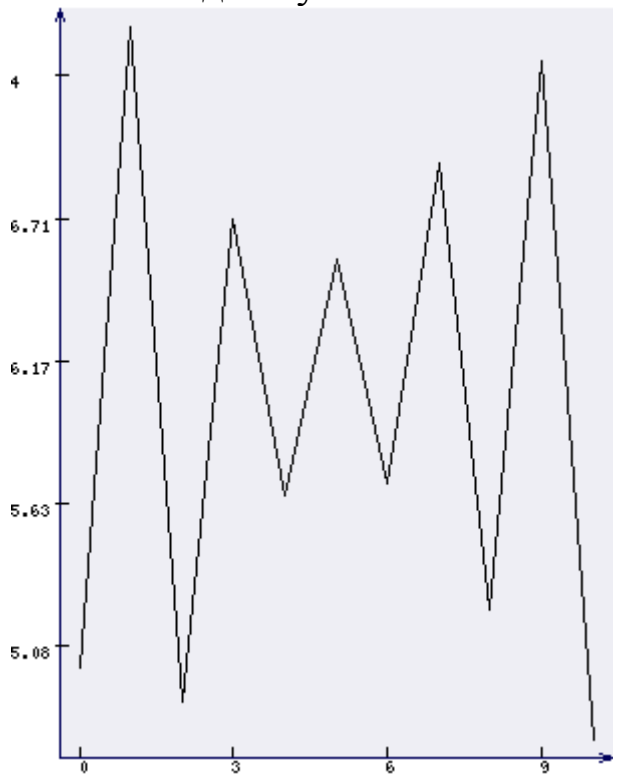


Рис. 43. Траектория цены в модели, B=D

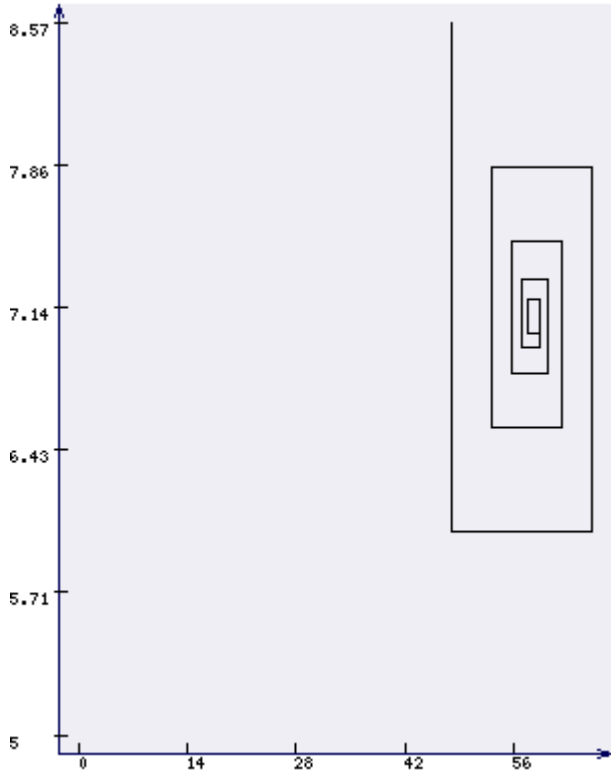


Рис. 44. Траектория динамики спроса и предложения

Алгоритм моделирования состоит из нескольких этапов. Задав некое начальное значение цены P_0, предварительно определив значения всех экзогенных переменных модели, вычисляем S_t из (1), D_t из (2). Далее выражаем P_t из (3):

$$P_t = \frac{A - C - D\,P_{t-1} - V_t + U_t + W_t}{B}. \quad (4)$$

В результате имитационного моделирования строятся различ-

ные графики траекторий изменения показателей во времени. Например, график зависимости цены от периода времени (рис. 43), на котором изображается изменение цены во времени, и цены от объемов продукции (рис. 44), на котором отображаются траектории кривых спроса и предложения во времени. Характер динамики цен (4) зависит в данной модели от отношения угловых коэффициентов прямых для линейных функций спроса и предложения (рис. 43). Поэтому теоретически равновесное положение паутинообразной модели может быть и неустойчивым. Если $D > B$, амплитуда колебаний неограниченно растет. Если $D = B$ (рис. 43), колебания имеют относительно постоянную амплитуду с точностью до случайной величины. Если $D < B$, то колебания затухают [44].

13.1.1.2. Модель с обучением

Эта модель (*модель с запаздыванием предложения*) строится на следующих гипотезах:

1) определяя объем предложения в каждый период, товаропроизводитель ориентируется на спрос в предыдущем периоде;

2) цена предлагаемого товара устанавливается производителем на уровне, определяемом согласно функции предложения;

3) объем потребления не может превосходить ни объема предложения, ни объема спроса.

Если говорить формально, то поставщик учитывает предыдущую тенденцию изменения цен и планирует свой выпуск на очередной отрезок времени t, ожидая, что цена P_t на этом отрезке будет равна $P_t = P_{t-1} - \rho\,(P_{t-1} - P_{t-2})$, где $\rho \in [0, 1]$ – константа, характеризующая значение, которое поставщики придают наблюдаемым колебаниям цен, т. е. константа учета тенденции изменения цен.

Параметрами вероятностной модели ценообразования с обучением являются:

1) A, B, C, D – параметры, характеризующие отрасль;

2) ρ – степень реакции производителя на колебания цен.

Уравнения функционирования модели (при $\rho = 0$ совпадают с уравнениями предыдущей модели):

$$S_t = C + D\,(P_{t-1} - \rho\,(P_{t-1} - P_{t-2})) + V_t;$$

$$S_t = D_t + W_t;$$

$$D_t = A - B\,P_t + U_t.$$

Процедура построения всех траекторий S_t, D_t, P_t по сравнению с предыдущей моделью меняется несущественно [44, 45].

13.1.1.3. Модель с запасами

В предыдущих двух моделях цены устанавливались на таком уровне, чтобы обеспечить локальное равновесие рынка, как бы за счет текущего производства, и никаких запасов продукции не создавалось (например, для скоропортящихся продуктов) [44]. В этой модели вводится новый механизм, изменяющий обеспечение потребителей – хранение товаров (запасы). При определении цены продукции на текущем отрезке времени она повышается относительно ее величины на предыдущем отрезке, если в течение последнего запасы уменьшились:

$$P_t = P_{t-1} - \lambda\,(Q_{t-1} - Q_{t-2}) = P_{t-1} - \lambda\,(S_{t-1} - D_{t-1}) \qquad (5)$$

где Q_t – запас к концу t-го отрезка времени; λ – степень реакции на изменение запасов, $0 \le \lambda \le 1$ (коэффициент учета запасов).

В данной модели спрос и предложение определяются без запаздывания. Уравнения функционирования модели с учетом случайных возмущений и преобразованного уравнения (5) имеют следующий вид

$$S_t = C + D\,P_t + V_t;$$
$$D_t = A - B\,P_t + U_t;$$
$$Q_t - Q_{t-1} = S_t - D_t;$$
$$P_t = \lambda(A + U_{t-1} - C - V_{t-1}) + \left[1 - \lambda(D - B)\right]P_{t-1}.$$

Процедура численного построения траекторий P_t, S_t, D_t и Q_t полностью определена этими уравнениями. Для вычисления значений переменных в произвольный момент времени достаточно задать начальные значения P_0 и Q_0 [44, 45].

13.1.2. Модель олигополии

В модели *олигополии* (*модели конкурентной отрасли*) фирмы через фиксированные отрезки времени планируют свои текущие объемы производства и реализации. Конкретные значения выпуска (управляемые значения) администрация фирмы выбирает, руководствуясь своими *неформальными* соображениями. В этом и состоит принципиальное отличие данной модели от паутинообразной, в которой значения выпусков целиком определяются формулами, основанными на различных предположениях относительно характера

функционирования фирм и отраслей.

Предположим, что в отрасли действуют k фирм с одинаковыми технологиями производства, производящих однородную продукцию [45].

Управляемыми значениями, как уже было сказано, являются величины X_{it} объемов продукции, выпускаемой каждой i-й фирмой на каждом отрезке времени t.

В модели олигополии используются следующие параметры:

A_i – номинальная мощность i-й фирмы;

B, C – технологические параметры;

D, E, F, G – константы.

Эндогенные параметры модели для каждого отрезка времени t:

1) S_t – полный выпуск отрасли;

2) D_t – потребление продукции отрасли;

3) P_t – цена товара;

4) C_{it} – производственные затраты i-й фирмы;

5) PP_{it} – полная прибыль i-й фирмы.

Экзогенными переменными являются случайные величины U_{it} и V_t, учитывающие влияние каких-либо факторов на фирмы:

Тождества модели:

общий выпуск продукции $S_t = \sum_{i=1}^{k} X_{it}$; 6)

условие локального равновесия рынка $S_t = D_t$; (7)

полная прибыль i-й фирмы $PP_{it} = P_t X_{it} - C_{it}$. (8)

Уравнения функционирования модели олигополии:

$$C_{it} = (X_{it} - A_i)^2 + B A_i^2 + C + U_{it}; \quad (9)$$

$$P_t = D - E D_t - F D_{t-1} - G D_{t-2} + V_t. \quad (10)$$

Поскольку условия модели запрещают фирмам вступать в кооперацию, при выборе управлений на текущий отрезок времени каждая фирма располагает только следующей информацией:

1) коэффициенты своей функции затрат;

2) вид функций затрат своих конкурентов при неизвестных значениях коэффициентов;

3) функция спроса на продукцию отрасли;

4) объемы выпусков на предыдущих двух отрезках времени;

5) цена на продукцию отрасли, которая установилась на предыдущем отрезке времени.

Расчет цены текущего периода, обеспечивающей равновесие на рынке, производится в соответствии с (10). После этого по формулам (8) – (9) вычисляется текущая прибыль каждой фирмы [45].

Возможны различные вариации процесса моделирования. Например, можно организовать эксперимент с созданием новых и банкротством существующих фирм в том случае, если прибыль фирмы оказалась отрицательной. Номинальная мощность созданных фирм выбирается из условия максимума ее потенциальной прибыли при заданной цене на текущий момент и нормальном режиме производства, когда выпуск $X_{it} = A_i$ [45].

Другой вариацией модели олигополии является формализация правил или введение ограничений на принятие решений, что существенно может расширить и приблизить модель к реальным процессам, например, за счет добавления процесса планирования производства каждой фирме на основе объема выпуска всей отрасли.

13.1.3. Модель дуополии

Дуополия – это частный случай олигополии. В дуополии рассматриваются две конкурирующие фирмы. Причем каждая из них при выборе объема выпуска учитывает не только прямое влияние на рынке, но и косвенное влияние конкурента.

В данной модели используются следующие допущения:

1) обе компании производят однородный товар;

2) цены с объемом выпуска связаны линейно;

3) каждая фирма должна выбрать такой объем выпуска, который максимизирует прибыль;

4) обе фирмы принимают решение одновременно.

Уравнения функционирования модели дуоплии:

$$P = a - b(y_1 + y_2);\ a > 0, b > 0;$$

$$C_1 = cy_1 + d; C_2 = cy_2 + d;\ c > 0; d > 0;$$

$$\Pi_1 = Py_1 - C_1 = \left[a - b(y_1 + y_2)\right] y_1 - cy_1 - d;$$

$$\Pi_2 = Py_2 - C_2 = \left[a - b(y_1 + y_2)\right] y_2 - cy_2 - d;$$

$$\frac{\partial \Pi_1}{\partial y_1} = \left[a - by_1 - by_2\right] - by_1 - b\frac{\partial y_2}{\partial y_1} y_1 - c = 0;$$

где P – цена;

y_1, y_2 – объем выпуска каждой фирмы;

C_1, C_2 – издержки каждой фирмы;

Π_1, Π_2 – прибыль каждой фирмы;

c – удельные издержки в отрасли;

d – фиксированные издержки;

a, b – величины, определяющие состояние рынка и изменение спроса;

$\frac{\partial y_2}{\partial y_1}$ – предположительная вариация (реакция второй фирмы на изменение объема выпуска первой фирмы).

Существует несколько моделей, описывающих поведение фирм, входящих в дуополию.

13.1.3.1. Модель Курно

Согласно этой модели каждый из дуополистов считает, что изменения в его собственном выпуске продукции не повлияют на конкурента (предполагается, что *объем выпуска конкурента постоянен*). Потому имеет место так называемое равновесие Курно:

$$\frac{\partial \Pi_1}{\partial y_1} = 0, \qquad \frac{\partial \Pi_2}{\partial y_2} = 0;$$

$$\frac{\partial y_2}{\partial y_1} = 0, \qquad \frac{\partial y_1}{\partial y_2} = 0.$$

Кривые реализации первой и второй фирмы соответственно

$$y_1 = \frac{a - 2by_1 - by_2}{2b}; y_2 = \frac{a - c - by_1}{2b}.$$

Оптимальный объем выпуска первой фирмы в зависимости от объема выпуска конкурента:

$$y_1 = \frac{a - c}{3b}; \qquad y_2 = y_1 = \frac{a - c}{3b}; \quad P = \frac{a + 2c}{3}.$$

13.1.3.2. Модель Стэкельберга

В данной модели первая фирма предполагает, что вторая фирма будет реагировать соответственно кривой реакции Курно:

$$y_2 = \frac{a - c - by_1}{2b}.$$

Соответственно для предположительных вариаций объемов и прибыли:

$$\frac{\partial y_2}{\partial y_1} = -0{,}5; \quad \frac{\partial \Pi_1}{\partial y_1} = \left[a - b(y_1 + y_2)\right] - by_1 + \frac{by_1}{2} - c = 0;$$

$$y_1 = \frac{a - c - by_2}{1{,}5b} = \frac{a - c}{2b}; \quad y_2 = \left[a - c - b\frac{a - c}{2b} \right] : 2b = \frac{a - c}{4b}.$$

13.1.3.3. Договорное решение

В данной модели фирмы договариваются с целью максимизации прибыли. Критерий оптимизации – максимизация $П = П_1 + П_2$, а с учетом

$$\frac{\partial П}{\partial y} = \left[a - b(y_1 + y_2) \right] - c = 0$$

объемы выпуска фирм равны

$$y = \frac{a - c}{2b}; y_1 = y_2 = \frac{a - c}{4b}.$$

Выполнив оценки $П = П_1 + П_2$, обнаруживаем наиболее проигрышный вариант работы фирмы без оглядки на конкурента, и наиболее выгодная с математической точки зрения политика договорных сделок – организация картелей. С экономической точки зрения оптимальны действия по модели Стэкельберга, так как организация картелей является недопустимой с точки зрения антимонопольного законодательства любой страны.

13.2. Модели развития отрасли

Модели развития отрасли описывают отрасль как единое целое. Модели некоторых отраслей, построенные на основе экономики США середины XX века, со ссылками на первоисточники подробно описаны в [44].

Модель текстильной промышленности. На основании официальной статистики по отраслям было подобрано девять *эндогенных переменных* (это такие переменные, значения которых задаются вне модели) и построены их траектории. Многие из них не вошли в модель, так как статистические оценки соответствующих коэффициентов оказались близки к нулю. Функционирование модели представлено обычной системой рекуррентных соотношений, параметры которых и их уровни значимости оценены по методу наименьших квадратов. Текущие значения эндогенные переменных вычислялись на основе их запаздывающих значений для использования в последующих итерациях.

Модель К. Коэна для *кожевенной и обувной промышленности*. Созданы две модели: долгосрочного и краткосрочного прогноза.

Основными экзогенными переменными являются индексы цен на потребительские товары, доходы населения и запасы товара и полуфабрикатов. Разбив все факторы на пять разных функциональных уровней (потребление, розничная торговля, производство, торговля полуфабрикатами и обработка полуфабрикатов), Коэн ввел два основных вида эндогенных переменных: цены и материальные потоки. Обе модели Коэна представляют собой системы взаимозависимых нелинейных разностных уравнений *с запаздываниями*. Условной единицей времени был выбран месяц.

Модель лесоразрабатывающей промышленности. Авторы имитационной модели пытались показать влияние на экономику отрасли ограничений на информацию о рынке, децентрализации планирования рыночной деятельности и регламентирующих ее правил.

13.3. Макроэконометрические модели

Макроэкономические модели рассматривают сложную экономическую систему: регион или страну в целом. Для создания такой модели прибегают к тем же методологическим приемам, что и при имитации фирмы или отрасли. Определяют структуру изучаемой системы, выбирают соответствующие экзогенные и эндогенные переменные, строят систему их взаимосвязей и определяют процедуры моделирования. Однако имитационные модели глобальных экономических систем сильно отличаются от микроэкономических в связи с проблемой вывода адекватных уравнений функционирования экономики в целом.

Во-первых, эндогенные переменные макродинамической системы (такие как национальный продукт, численность работающих) зависят от очень большого числа значимых факторов. Во-вторых, для использования укрупненных показателей необходимо разработать приемлемую теорию для агрегатирования переменных. В-третьих, между эндогенными переменными существуют очень сложные взаимодействия и обратные связи. В-четвертых, необходимо обладать глубокими знаниями макроэкономических процессов, что может привести исследователя «скорее к модели собственного невежества, а не реального мира» [44]. В-пятых, эконометрическая обработка данных и оценка параметров с большим числом уравнений порождают серьезные трудности. В-шестых, получить данные для построения математической модели такой системы намного сложнее, чем для микроэкономических систем.

В [44] можно найти краткое описание некоторых больших эконометрических моделей: брукгинская модель (230 уравнений, из них 118 – уравнения функционирования); модель ОВЕ Министерства торговли США (56 уравнений, 46 тождеств, 75 экзогенных переменных), уортонская модель М. Эванса и Л. Клейна (47 уравнений и 29 тождеств).

13.4. Методологические проблемы, анализ и интерпретация результатов имитационного моделирования

Достаточно часто создатели модели экономической системы, задаваясь проблемами проведения машинного имитационного эксперимента, не уделяют должного внимания вопросу оценки пригодности этой модели и результатов эксперимента. Это, в первую очередь, связано тем, что истинность результатов требует подтверждения с помощью доступных наблюдений (практика – критерий истины), что не всегда возможно, и потому большинство моделей, заявляемых в информационном пространстве, остаются чисто гипотетическими.

Существуют основные методологические точки зрения на проблему оценки пригодности экономических моделей:

1) *рационализм* – утверждение возможности существования положений, определяющих поведение системы априори, которые просто требуют поиска и формулировки;

2) *эмпиризм* – утверждение абсолютной ценности наблюдений и отрицание логических выводов, которые не подкреплены опытами;

3) «*позитивная экономическая наука*» – практический подход к модели, оценка не правильности посылок и выводов, а способности предсказывать поведение зависимых переменных.

Существует еще один подход – многоэтапная оценка пригодности. Ни одна из указанных выше методологических точек зрения в отдельности не достаточна для решения проблемы. Последовательное комплексное рассмотрение проблемы с разных точек зрения может приблизить к пониманию адекватности модели.

В первую очередь, возникает проблема нахождения критерия, указывающего, что совпадение траекторий, построенных с помощью имитационной модели, с наблюдаемыми или наблюдавшимися показателями нельзя было считать просто случайным. Р. Сайерт предложил использовать для этого показатели, связанные с числом

точек экстремума, их распределением во времени и градиентом изменения в этих точках, одновременностью экстремальных точек для различных переменных, амплитудой возмущений на одних и тех же временных отрезках, средними значениями и точными совпадениями значений переменных.

К этому списку можно добавить различия в параметрах вероятностных распределений переменных: в средних значениях, дисперсиях, асимметриях, эксцессах.

Из наиболее важных методов проверки точности и адекватности данных, полученных с помощью модели, можно выделить эконометрические методы: дисперсионный анализ, критерий хи-квадрат, факторный анализ, критерий Колмогорова – Смирнова, непараметрические критерии, регрессионный анализ, спектральный анализ, коэффициент несовпадения Тейла. Ссылки на оригиналы публикаций о применении этих методов и проблеме в целом можно найти в [44].

Что касается методологии имитационного моделирования, то заслуживают упоминания следующие проблемы [44].

Имитация и аналитическое решение. Часто бывает очень сложно или даже невозможно получить в явном виде приведенную форму системы взаимозависимых нелинейных стохастических разностных уравнений. По этой причине экономисты вынуждены прибегать к численным методам анализа пригодности и оценки динамических свойств таких моделей. Неоднозначен и ответ на вопрос о целесообразности использования имитационного подхода не только в эконометрических, но и любых других экономических моделях.

Ошибочные результаты имитации. Эконометрические модели, оцененные корректным способом, основанные на верных экономических предпосылках, могут, тем не менее, определять бессмысленные траектории. Положительные по своему смыслу переменные могут в процессе счета принимать отрицательные значения, приводя к результатам, совершенно противоречащим реальности и здравому смыслу.

Неадекватные методы оценки. Несмотря на то, что большая часть методов по оценке результатов моделирования известна, эконометрическая теория не исследует вопрос соответствия результатов динамического моделирования реальности. Встает вопрос о получении таких методов оценки, в которых «качество оценки опре-

деляется тем, *как хорошо модель имитирует*, а не тем, *насколько* хорошо она предсказывает на один шаг вперед».

Переменные коэффициенты. Вопрос о рассмотрении коэффициентов эконометрической модели как случайных величин приводит к существенному усложнению процесса моделирования. Результаты такого моделирования существенно отличаются от траектории детерминированной модели. Кроме того, возможно изменение структуры модели со временем, что также приводит к необходимости пересмотра предположения о постоянстве коэффициентов модели.

Целесообразность создания и применения модели экономической системы не может быть определена однозначно. Положительной может быть оценка такой модели, за счет которой достигнута экономия и приняты правильные управленческие решения, но объективная такого рода оценка зачастую бывает невозможной.

13.5. Основные возможности веб-портала «Виртуальная случайность»

Для информационной поддержки дисциплин, связанных с имитационным моделированием, был создан веб-портал [35] «Виртуальная случайность» (http://vtit.kuzstu.ru/stat/).

В ходе реализации процесса дистанционного образования информационная поддержка приобретает особую актуальность. Приоритетным направлением здесь становится обеспечение достаточным материалом при подготовке и осуществление консультативного взаимодействия преподавателей-специалистов с обучающимися.

Одним из способов решения этой актуальной задачи является создание специализированного образовательного портала, который с технологической точки зрения представляет собой системное многоуровневое объединение образовательных ресурсов и сервисов в интернете. С содержательной точки зрения портал представляет собой учебно-методический центр. Портал создается с целью разработки новых стандартов организации и информационного обеспечения образовательного процесса на всех уровнях образования.

Для простоты работы с порталом при его создании ставилась задача поместить всю требуемую функциональность в максимально ограниченный набор разделов и, соответственно, использовать минимальное количество специфичных терминов. Вследствие этого было выделено всего шесть крупных разделов портала [35].

1) *Новости и обсуждения* – раздел, в котором располагаются все текстовые сообщения пользователей, размещенные на портале. Каждое сообщение связано с каким-либо ресурсом. Одним из ресурсов являются сами новости портала.

2) *Раздел ресурсов* является информационной базой, предназначенной для хранения в иерархической структуре файлов, ссылок на сайты в сети, то есть ресурсов. Этот раздел динамически пополняется в соответствии с правами пользователя. Пополнение позволяют осуществлять так называемые функции управления содержимым (*content management*).

3) *Раздел статистического моделирования* позволяет моделировать различные выборки, распределенные по разным (*дискретным* и *непрерывным*) законам с заданными параметрами. Кроме того, этот раздел позволяет оценить основные статистические характеристики полученного ряда и сопоставить графики теоретических и эмпирических, построенных на основе модельных выборок, функций распределения и функций плотностей распределения вероятностей.

4) *Раздел моделирования эмпирических распределений* позволяет пользователю сгенерировать выборку на основе *эмпирических данных* (*дискретных* или *непрерывных*). В результате можно оценить полигоны частот, нанесенные одновременно на один график, и графики накопленных частот, что позволяет визуально оценить соответствие сгенерированного ряда исходному.

5) В *разделе моделирования экономических процессов* реализованы простейшие экономические модели. Этот раздел позволяет провести имитационный эксперимент, изменяя параметры моделей, оценить результаты моделирования, визуализированные в форме графиков и таблиц. Модуль моделирования предназначен для демонстрации возможностей самых простых моделей и не может претендовать на использование в качестве имитационного инструментария для моделирования реального экономического процесса, количество и природа влияющих факторов в котором значительно отличаются от реализованных.

6) *Раздел администрирования доступа*. Поскольку большинство функций портала зависят от прав работающего пользователя, на первый план встает управление правами. Таким образом, под термином администрирование в данном случае понимаются не кон-

кретные функции, доступные администратору, а контроль доступа к этим функциям. Такой подход позволяет объединить управление порталом в одном разделе с помощью системы контроля доступа с использованием иерархических структур.

Веб-портал «Виртуальная случайность» спроектирован модульно [35], что позволяет в определенных рамках расширять его функциональность.

Для реализации некоторых видов вероятностных распределений достаточно будет добавить их описания, воспользовавшись одним из реализованных методов генерации случайных величин. Для реализации других видов моделирования случайных величин необходимо создавать потомков существующего, функционально абстрактного класса.

Для реализации новых экономических моделей необходимо также создать новый класс или воспользоваться существующим для его расширения. Подготовленный инструментарий (шаблоны вывода, построения графиков) может существенно сократить время разработки.

ЦИТИРОВАННАЯ И РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Dantzig, G. B. Reminiscencesabout the Origins of Linear Programming // Mathematical Programming:The State of the Art. – Berlin: Springer Verlag, 1983. – pp. 79–86.

2. Данциг Д. Линейное программирование, его применения и обобщения. – Москва : Прогресс, 1966. – 600 с.

3. Вершик, А. М. О Л. В. Канторовиче и линейном программировании. – [Электронный ресурс]. – Режим доступа: http://www.mmonline.ru/message/1893/print/, свободный.

4. Канторович, Л. В. Математические методы организации и планирования производства. – Ленинград : Изд-во ЛГУ, 1939. – 67 с.

5. Гасс, С. Линейное программирование. – Москва : Физматгиз, 1961. – 304 с.

6. Гасс, С. Путешествие в Страну Линейного Программирования. – Москва : Мир, 1973. – 176 с.

7. Беллман, Р. Динамическое программирование. – Москва : Издательство иностранной литературы, 1960. – 400 с.

8. Беллман, Р. Прикладные задачи динамического программирования / Р. Беллман, С. Дрейфус. – Москва : Наука, 1965. – 460 с.

9. Юдин, Д. Б. Линейное программирование (теория, методы и приложения) / Д. Б. Юдин, Е. Г. Гольштейн. – Москва : Наука, 1969. – 424 с.

10. Гольштейн, Е. Г. Новые направления в линейном программировании / Е. Г. Гольштейн, Д. Б. Юдин. – Москва : Советское радио, 1966. – 524 с.

11. Тынкевич, М. А. Экономико-математические методы (исследование операций). – 3-е изд., испр. и доп. – Кемерово, 2010. – 216 с.

12. Форд, Л. Потоки в сетях / Л. Форд, Д. Фалкерсон. – Москва : Мир, 1966. – 277 с.

13. Прим, Р. К. Кратчайшие связывающие сети и некоторые обобщения // Кибернетический сборник. – Москва : Издательство иностранной литературы, 1961. – Вып. 2. – С. 95–107.

14. Ху, Т. Целочисленное программирование и потоки в сетях. – Москва : Мир, 1974. – 520 с.

15. Берж, К. Теория графов и ее применения. – Москва : Издательство иностранной литературы, 1962. – 320 с.

16. Кюнци, Г. П. Нелинейное программирование / Г. П. Кюнци, В. Крелле. – Москва : Советское радио, 1965. – 304 с.

17. Хедли, Д. Нелинейное и динамическое программирование. – Москва : Мир, 1967. – 509 с.

18. Вагнер, Г. Основы исследования операций : в 3 т. – Москва : Мир, 1973. – т. 1 – 336 с.; т. 2 – 488 с.; т. 3 – 503 с.

19. Таха, Х. Введение в исследование операций : в 2 кн. – Москва : Мир, 1985. – кн. 1 – 479 с.; кн. 2 – 496 с.

20. Моудер, Д. Исследование операций / Д. Моудер, С. Элмаграби. Т. 1. – Москва : Мир, 1981. – 716 с.

21. Ховард, Р. Динамическое программирование и марковские процессы. – Москва : Советское радио, 1964. – 190 с.

22. Калихман, И. Л. Динамическое программирование в примерах и задачах / И. Л. Калихман, М. А. Войтенко. – Москва : Высшая школа, 1979. – 125 с.

23. Нейман, Дж. Теория игр и экономическое поведение / Дж. Нейман, О. Моргенштерн. – Москва : Издательство иностранной литературы, 1970. – 708 с.

24. Карлин, С. Математические методы в теории игр, программировании и экономике. – Москва : Мир, 1964. – 838 с.

25. Льюс, Р. Игры и решения / Р. Льюс, Г. Райфа. – Москва : Издательство иностранной литературы, 1961. – 641 с.

26. Костевич, Л. С. Теория игр. Исследование операций / Л. С. Костевич, А. А. Лапко. – Минск : Вышэйшая школа, 1982. – 230 с.

27. Бусленко, Н. П. Метод статистических испытаний (Монте – Карло) / Н. П. Бусленко, Ю. А. Шрейдер. – Москва : Физматгиз, 1961. – 228 с.

28. Соболь, И. М. Метод Монте – Карло. – Москва : Наука, 1985. – 80 с.

29. Самарский, А. А. Математическое моделирование и вычислительный эксперимент // Вестник АН СССР. – 1979. – № 5. – С. 38–49. – [Электронный ресурс]. – Режим доступа: http://samarskii.ru/articles/1979/1979-002ocr.pdf, свободный.

30. Емельянов, А. А. Имитационное моделирование экономических процессов / А. А. Емельянов, Е. А. Власова, Р. В. Дума ; под ред. А. А. Емельянова. – Москва : Финансы и статистика, 2009. – 416 с.

31. Советов, Б. Я. Моделирование систем / Б. Я. Советов, С. А. Яковлев. – Москва : Высшая школа, 2009. – 344 с.

32. Анфилатов, В. С. Системный анализ в управлении / В. С. Анфилатов, А. А. Емельянов, А. А. Кукушкин ; под ред. А. А. Емельянова. – Москва : Финансы и статистика, 2009. – 368 с.

33. Тарасенко, Ф. П. Прикладной системный анализ. – Москва : Кнорус, 2010. – 224 с.

34. Тынкевич, М. А. Статистический анализ данных на компьютере / М. А. Тынкевич, А. Г. Пимонов, А. М. Вайнгауз. – Кемерово, 2013. – 124 с.

35. Пимонов, А. Г. Имитационное моделирование экономических систем / А. Г. Пимонов, С. А. Веревкин. – Кемерово, 2013. – 138 с.

36. Голенко, Д. И. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах. – Москва : Физматлит, 1965. – 228 с.

37. Хастингс, Н. Справочник по статистическим распределениям / Н. Хастингс, Дж. Пикок ; пер. с англ. А. К. Звонкина. – Москва : Статистика, 1980. – 95 с.

38. Список функций Statistics Toolbox. – [Электронный ресурс]. – Режим доступа: http://matlab.exponenta.ru/statist/book2/index.php, свободный.

39. Костевич, Л. С. Теория игр. Исследование операций / Л. С. Костевич, А. А. Лапко. – Минск : Вышэйшая школа, 1982. – 230 с.

40. Кофман, А. Массовое обслуживание. Теория и приложения / А. Кофман, Р. Крюон. – Москва : Мир, 1965. – 302 с.

41. Кокс, Д. Теория очередей / Д. Кокс, У. Смит. – Москва : Мир, 1966. – 218 с.

42. Терпугов, А. Ф. Экономико-математические модели : учеб. пособие. – Барнаул : Алтайский экономико-юридический институт, 1999. – 115 с.

43. Малыхин, В. И. Математическое моделирование экономики : учеб.-практич. пособие. – Москва : Издательство УРАО, 1998. – 160 с.

44. Нейлор, Т. Машинные имитационные эксперименты с моделями экономических систем. – Москва : Мир, 1975. – 504 с.

45. Лебедев, В. В. Математическое моделирование социально-экономических процессов. – Москва : Изограф, 1997. – 224 с.

46. Аллен, Р. Математическая экономия. – Москва : Издательство иностранной литературы, 1963. – 667 с.

47. Ланкастер, К. Математическая экономика. – Москва : Советское радио, 1972. – 464 с.

48. Саати, Т. Л. Элементы теории массового обслуживания и ее приложения. – Москва : Советское радио, 1965. – 505 с.

49. Афанасьев, М. Ю. Исследование операций в экономике: модели, задачи, решения : учеб. пособие / М. Ю. Афанасьев, Б. П. Суворов. – Москва : ИНФРА-М, 2003. – 444 с.

50. Афанасьев, М. Ю. Прикладные задачи исследования операций : учеб. пособие / М. Ю. Афанасьев, К. А. Багриновский, В. М. Матюшок. – Москва : ИНФРА-М, 2006. – 352 с.

51. Кремер, Н. Ш. Исследование операций в экономике. – Москва : Юрайт, 2013. – 438 с.

Тынкевич Моисей Аронович
Пимонов Александр Григорьевич
Веревкин Сергей Анатольевич

Исследование операций и имитационное моделирование

Учебное пособие

Редактор З. М. Савина

Подписано в печать 14.10.2015. Формат 60×84/16
Бумага офсетная. Гарнитура «Times New Roman». Уч.-изд. л. 16,0
Тираж 500 экз. Заказ № 30

КузГТУ, 650000, Кемерово, ул. Весенняя, 28

Издательский центр УИП КузГТУ, 650000, Кемерово, ул. Д. Бедного, 4А

Тынкевич Моисей Аронович – родился 28 февраля 1937 г. в г. Новосибирске, кандидат физико-математических наук, доцент, профессор кафедры прикладных информационных технологий. В 1959 году окончил механико-математический факультет Томского государственного университета в группе 24 первых за Уралом выпускников по новой специальности «Вычислительная математика». В 1959 – 1966 гг. работал на кафедре вычислительной математики Томского университета. С 1966 г. старший преподаватель кафедры экономики и организации производства Кузбасского политехнического института, с 1969 г. старший преподаватель новой в ВУЗе кафедры вычислительной техники и промэлектроники. Внес значительный вклад в становление и развитие кафедры прикладных информационных технологий. Подготовил более 80 научных работ и учебных пособий, несколько циклов методических разработок. Почетный работник высшего образования России (1997 г.). Награжден медалями «За особый вклад в развитие Кузбасса» (2001 г.), «За достойное воспитание детей» (2010 г.). Ведет занятия по курсам «Численные методы анализа», «Исследование операций и методы оптимизации», «Экономико-математические методы и модели».

Пимонов Александр Григорьевич – родился 23 ноября 1959 г. в селе Чапаево Хакасской автономной области, профессор, доктор технических наук, профессор кафедры прикладных информационных технологий. В 1981 г. с отличием окончил факультет прикладной математики и кибернетики Томского государственного университета. В Кузбасском государственном техническом университете работает с 1985 г. (старший инженер, старший преподаватель, доцент, профессор, заместитель заведующего кафедрой, заведующий кафедрой, исполняющий обязанности декана факультета информационных технологий и менеджмента). Подготовил более 160 научных работ и учебно-методических разработок. Почетный работник высшего профессионального образования Российской Федерации (2010 г.), лучший профессор КузГТУ (2014 г.), лучший руководитель научно-исследовательской работы студентов КузГТУ (2014 г.), научный руководитель магистерской программы по направлению подготовки «Прикладная информатика». Ведет занятия по дисциплинам «Теория систем и системный анализ», «Математические и инструментальные методы поддержки принятия решений», «Имитационное моделирование экономических систем», «Математическое и имитационное моделирование».

Веревкин Сергей Анатольевич – родился 18 сентября 1983 г. в городе Кемерово, старший преподаватель кафедры прикладных информационных технологий. В 2005 г. с отличием окончил инженерно-экономический факультет Кузбасского государственного технического университета. Тема дипломной работы – «Разработка минипортала «Виртуальная случайность» для информационного обеспечения курса «Имитационное моделирование экономических систем». Окончил аспирантуру по специальности «Математическое моделирование, численные методы и комплексы программ». В Кузбасском государственном техническом университете работает старшим преподавателем с 2006 г. Ведет занятия по дисциплинам «Высокоуровневые методы информатики и программирования», «Объектно-ориентированное программирование», «Информационная безопасность». Руководит проектированием и разработкой программного обеспечения в подразделении ФГУП ГНИВЦ ФНС России. В сферу профессиональных интересов входят аспекты проектирования и реализации систем массового обслуживания, проблемы обработки больших данных в высоконагруженных системах.

ISBN 978-5-906805-12-6