

Б.Б.РОЗИН
В.И.КОТЮКОВ
М.А.ЯГОЛЬНИЦЕР

ЭКОНОМИКО-СТАТИСТИЧЕСКИЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ЭКОНОМИКИ И ОРГАНИЗАЦИИ ПРОМЫШЛЕННОГО ПРОИЗВОДСТВА

Б. Б. РОЗИН, В. И. КОТЮКОВ,
М. А. ЯГОЛЬНИЦЕР

ЭКОНОМИКО-СТАТИСТИЧЕСКИЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

Ответственный редактор
канд. физ.-мат. наук *В. Г. Соколов*

ИЗДАТЕЛЬСТВО «НАУКА»
СИБИРСКОЕ ОТДЕЛЕНИЕ
Новосибирск·1984

Розин Б. Б., Котюков В. И., Яголь-ницер М. А. Экономико-статистические модели с переменной структурой.— Новосибирск: Наука, 1984.

В монографии на примерах типичных экономических ситуаций рассмотрены проблемы учета структурных изменений при статистическом моделировании. Анализируются причины изменчивости структуры экономических процессов; изучается взаимосвязь изменчивости структур и неоднородностей совокупностей; излагается методика построения моделей для выборок различного типа методами распознавания образов, теории сплайнов и многомерного статистического анализа; рассматриваются вопросы оценки качества моделей, их корректировки и адаптации.

Книга рассчитана на специалистов в области применения статистических методов в экономических исследованиях.

Р $\frac{0604020105-811}{042(02)-84}$ 77—84—I

© Издательство «Наука», 1984 г.

ВВЕДЕНИЕ

Динамичность процессов развития народного хозяйства, отраслевых систем и отдельных промышленных объектов на современном этапе остро ставит проблему их статистического описания и прогнозирования. Особенно актуальна она в планировании, когда речь идет о прогнозировании значений показателей на будущее. Традиционные приемы и методы статистического моделирования не вполне отвечают требованиям адекватного описания изменяющихся во времени и пространстве связей между показателями и факторами производства. Поэтому ощущается острая необходимость в методах, приспособленных к описанию и прогнозированию процессов с изменяющейся структурой.

Проблема изучения изменчивости показателей статистической связи во времени ставилась в советской статистике еще в 20-х годах, в частности С. П. Бобровым [1927], и позднее в трудах Б. С. Ястремского [1964], Н. С. Четверикова [1963] и др. В этих работах рассматривалось изменение коэффициента корреляции во времени. Например, Б. С. Ястремский рассмотрел случай изменения ковариации при неизменных значениях дисперсий переменных. В исследованиях Н. С. Четверикова и С. П. Боброва учитывалось еще и изменение дисперсии. Кроме того, С. П. Бобров применил метод скользящего коэффициента корреляции.

Все эти работы касались только изменения коэффициента парной корреляции и не затрагивали вопросов учета структуры модели в многомерном случае, в них также не рассматривалась возможность перелома структуры влияния и наличия участков стабильности. В последнее время сложились благоприятные условия для решения этой проблемы. К ним можно отнести возникновение новых и развитие традицион-

ных методов статистического исследования (многомерная классификация, методы распознавания образов, теория сплайнов и др.) и появление быстродействующих ЭВМ.

В основе построения экономико-статистической модели (ЭСМ) любого типа лежит определенная концепция моделируемой экономической системы (объекта), основанная на учете ее характерных свойств.

В настоящее время общепринята точка зрения на экономические объекты как на сложные большие системы, свойствами которых являются управляемость, неопределенность поведения, неполнота информации о будущих состояниях. Итак, первая характерная особенность (свойство), подлежащая учету при моделировании экономических систем,— *управляемость*. Под этим понимается целенаправленное изменение регулируемых входных показателей системы. Характерным свойством экономических систем является то, что, по крайней мере, часть их входных показателей поддается изменению в области допустимых значений. Конкретные значения регулируемых входных показателей представляют собой характеристики принимаемых управляющих решений.

Следующее свойство сложных экономических систем, подлежащее учету при моделировании их поведения,— *стохастический характер поведения* объектов. Под стохастичностью понимается невозможность однозначного определения будущих значений выходных показателей на основе информации об их предыстории и значениях входных показателей. Можно назвать несколько причин, обусловливающих стохастичность экономических процессов. Во-первых, неполны наши знания об изучаемом процессе; во-вторых, поведение определенной части входных показателей случайно, плохо поддается измерению и прогнозированию (природно-климатические и субъективные факторы и т. д.); в-третьих, любая математическая модель носит приближенный характер; в-четвертых, исходные статистические данные имеют погрешности.

Третье принципиальное свойство экономических систем — *инерционность*, т. е. зависимость состояния экономической системы в каждый данный момент времени от ее предыдущих состояний. Инерционность обусловлена невозможностью в небольшие отрезки времени коренным образом изменить технологическую структуру экономических процессов. Это объясняется несколькими причинами. Одна из них связана с лагами капитальных вложений; вторая — с запаздыванием информации о необходимости принятия новых управляющих решений; третья — с медленной реакцией управления на изменение ситуации. Степень инерционности экономических процессов зависит от уровня иерархии исследуемой системы: чем он выше, тем, как правило, больше инерционность.

В зависимости от того, какие свойства экономических систем учитываются, ЭСМ можно подразделить на два крупных класса: экономико-статистические модели со стабильной и ЭСМ с переменной структурой [1].

ЭСМ со стабильной структурой ориентированы в основном на учет свойства стохастичности, для них характерно описание будущих состояний экономического объекта при помощи экстраполяции закономерностей его поведения в прошлом. При построении ЭСМ такого типа принимается, что поведение экономической системы на всем анализируемом отрезке времени описывается одним и тем же случайным процессом, что равносильно предположению о постоянстве свойств экономической системы во времени. В основе корректного применения ЭСМ со стабильной структурой лежит одна из характерных особенностей экономических систем — их инерционность, невозможность существенного изменения состояния экономической системы за малые промежутки времени. Нужно определенное время для срабатывания управляющих воздействий. В пределах времени срабатывания управляющих воздействий применение ЭСМ со стабильной структурой вполне оправданно.

Модели с переменной структурой ориентированы на получение информации о возможных состояниях экономического объекта для различных управляющих решений, они учитывают возможные качественные изменения в поведении изучаемого объекта в результате воздействия новых управляющих решений. Эти модели одновременно отражают два основных свойства экономических систем — стохастичность и управляемость. С содержательной точки зрения применение ЭСМ с переменной структурой соответствует принятию гипотезы об изменяющейся во времени эффективности использования производственных ресурсов. Именно этому последнему классу моделей (способам их построения и областям приложения) и посвящена настоящая монография.

Применительно к задачам анализа структуры связей большое значение имеет разработка априорной гипотезы о ха-

[1] Под структурой ЭСМ понимается набор входных и выходных переменных, форма их взаимосвязи, параметры модели.

рактере переходов от одной структуры к другой. Если предварительный содержательный анализ позволяет принять гипотезу об относительной плавности переходов от одного состояния объекта к другому, то можно ориентироваться на непрерывные методы описания. Такая гипотеза может быть принята для объектов достаточно высокого уровня (макроэкономические показатели народного хозяйства, отрасли) или для временных выборок по отдельному объекту.

Другая ситуация возникает при пространственных (пространственно-временных) выборках, когда, как правило, выделяются отдельные типы объектов с различным характером зависимостей и резким, скачкообразным переходом от одного типа зависимости к другому. В этом случае целесообразно ориентироваться на комбинированные (дискретно-непрерывные) способы моделирования, базирующиеся на сочетании методов многомерной классификации и регрессионного анализа. Значение таких подходов особенно возрастает при исследовании многомерных процессов, а также тогда, когда отдельные переменные модели принимают дискретные значения или измеряются качественными признаками.

Особенность предлагаемой монографии — широкое использование методов теории распознавания образов (многомерной классификации) на всех этапах экономико-статистического исследования: при анализе структуры выборочной совокупности и в самом процессе построения моделей с переменной структурой. Нужно отметить, что применение методов корреляционно-регрессионного анализа в экономических исследованиях стало уже традиционным, тогда как приложение методов теории распознавания образов (многомерной классификации) к экономическим задачам представляет собой принципиально новый подход, позволяющий решить ряд острых проблем экономико-статистического моделирования, повысить адекватность моделей реальным процессам и расширить сферу их применения.

Основной результат анализа структуры исследуемой совокупности промышленных объектов методами распознавания образов есть разбиение ее на однородные классы (определение числа, характеристик и состава классов, описание тенденций изменения основных элементов классификации во времени). Это позволяет обоснованно подойти к выбору рационального типа модели, учесть при моделировании типологическую структуру совокупности в базисный период и изменение принадлежности объектов к определенным классам во времени под влиянием технического прогресса, при-

родных условий производства и других факторов производственно-хозяйственной деятельности.

Все это предопределило основную ориентацию нашей работы на использование методов кусочно-простой (в частности, кусочно-линейной) аппроксимации зависимостей с изменяющейся структурой.

Сущность построения кусочно-простых моделей заключается в разбиении всего множества возможных выборочных реализаций на ряд подмножеств, в пределах которых простые модели будут достаточно адекватны.

Кусочно-простые модели обладают рядом положительных качеств, выгодно отличающих их от других нелинейных моделей:

во-первых, они просты при интерпретации и использовании;

во-вторых, требуют сравнительно небольшого объема выборочных данных для «надежной» оценки параметров;

в-третьих, разработаны эффективные численные методы их оптимизации на основе анализа данных, позволяющие одновременно осуществлять и выбор информативной подсистемы исходных показателей, и учет разнотипности последних.

Эти качества подтверждаются большим практическим опытом применения дискретно-непрерывных моделей, неизменно обеспечивавших высокую точность описания и прогнозирования на их основе процессов с изменяющейся структурой.

Разработка методов кусочно-простого анализа экономических процессов с изменяющейся структурой ведется не на «пустом» месте — фундаментом являются теоретические результаты, полученные в «классическом» многомерном статистическом анализе. Однако разработанные методы и алгоритмы разбросаны по различным журнальным статьям, и попыток хотя бы как-то их систематизировать не предпринималось, во всяком случае, в отечественной экономико-статистической литературе соответствующих публикаций нет.

В данной работе отражены далеко не все методы кусочно-простого многофакторного анализа. Была сделана попытка изложить лишь наиболее универсальные алгоритмы, нашедшие широкое применение в практике, для которых создано программное обеспечение в виде комплексов и пакетов прикладных программ.

Большое внимание уделено методу «фиксированных элементов разбиений» (ФИЭР), обладающему многими положительными свойствами. При изложении методов ориентация

делается на два основных раздела многофакторной статистики, непосредственно связанных с построением моделей прогнозирования,— регрессионный анализ и дискриминантный анализ (распознавание образов). Нашли отражение и другие методы, в том числе аппроксимация зависимости сплайн-функциями.

Возможны два подхода к изложению в книге материала по экономико-статистическому моделированию. Первый заключается во всестороннем анализе определенного круга математико-статистических методов с последующим рассмотрением возможных областей их приложения. В этом случае акцент делается на методы, а области применения выступают как иллюстрация возможностей их применения.

Используется и другой подход, когда в основу положена содержательная экономическая проблема, а методы выступают как инструментарий ее решения. В настоящей монографии применялся преимущественно второй подход.

Работа носит экспериментально-методический характер, она основана на обобщении и систематизации опыта многочисленных научно-исследовательских работ, связанных общностью экономического содержания и используемого математико-статистического аппарата. Экономико-статистические методы выступают как инструментарий решения определенной экономической проблемы — учета структурных изменений.

Все методические обобщения в работе получены на основе конкретных исследований и иллюстрируются на их примере. Анализ результатов конкретных экономико-статистических исследований позволяет:

сделать некоторые методологические выводы, касающиеся экономико-статистического изучения структуры вообще;

разработать прикладную методику решения типичных экономических задач;

построить конкретные модели, на основе которых может быть выполнен анализ, приняты те или иные решения.

Книга состоит из пяти глав. В первой дается содержательная постановка проблемы выявления и учета структурных изменений экономических процессов при моделировании.

Вторая глава содержит обзор основных методов построения кусочно-простых моделей. Излагаются приемы типологической группировки как для случая последовательных разбиений, так и для одновременного учета множества факторов (кластер-анализ, таксономия). Большое внимание уделяется методу ФИЭР.

Третья глава посвящена изложению методики построения моделей с переменной структурой во временных выборках.

В четвертой главе изложена методика построения двух основных типов моделей для пространственно-временных выборок — классификационных и дискретно-непрерывных. Рассматриваются статический и динамический случаи, детерминистские и вероятностные разбиения.

Обсуждению проблем и методов оценивания качества моделей прогнозирования посвящена пятая глава работы. В ней также излагаются вопросы адаптации моделей.

Большое участие в обсуждении рассмотренных проблем и подходов к их решению приняла кандидат экономических наук, старший научный сотрудник Института экономики и организации промышленного производства Мэри Львовна Лукацкая, которой принадлежит большая заслуга в формировании позиций авторов по основным изложенным вопросам. Ее светлой памяти авторы посвящают эту книгу.

Работа над отдельными главами и параграфами распределилась следующим образом: Розин Б. Б., Ягольницер М. А.— введение, глава 1, глава 2 (§ 1, 2), глава 3, глава 4, глава 5 (§ 1, 3); Котюков В. И.— глава 2 (§ 1, 2, 3), глава 5 (§ 2).

Глава 1

СТРУКТУРНЫЕ ИЗМЕНЕНИЯ В ЭКОНОМИЧЕСКИХ СИСТЕМАХ И ИХ УЧЕТ ПРИ МОДЕЛИРОВАНИИ

§ 1. СТРУКТУРА ЭКОНОМИЧЕСКОГО ПРОЦЕССА

Статистическая модель может выступать в качестве надежного инструмента прогноза и управления только в том случае, если она достаточно адекватно отражает реальные причинно-следственные связи. Под причинно-следственными связями в наиболее простом случае понимают три взаимосвязанные категории: причину, следствие и оператор, переводящий первое во второе. Первые две категории достаточно подробно изучены. Для целей настоящей работы наибольший интерес представляет уточнение понятия оператора преобразования причины в следствие (и связанного с ним понятия структуры экономического процесса), а также способов его статистического описания.

«Оператором называется вещественное устройство или наделенный структурой процесс, который обеспечивает выполнение причинной связи» [Хейс, 1981, с. 18].

Следует заметить, что оператор преобразования объективно существует до реализации причинно-следственной связи. Важно подчеркнуть, что одна и та же причинно-следственная связь может реализоваться разными операторами, т. е. существуют классы преобразователей.

Можно представить два способа формирования операторов. Первый, «конструктивный», когда задан процесс организации и имеются его элементы. Например, задана технология сборки изделий и имеются детали. Тогда свойства оператора определяются в процессе конструирования изделия.

Второй подход «статистический», с которым чаще сталкиваются в экономических исследованиях, имеет дело с заранее существующими операторами, характеристики которых определяются по результатам наблюдения за их функционированием. Другими словами, имеется совокупность наблюдений со значениями причин и следствий, по которым требуется восстановить оператор. В качестве статистического аналога преобразователя может, например, выступать уравнение регрессии, параметры которого являются характеристиками оператора. По ним можно оценить существование связей, их силу и направление.

Исходя из рассмотренного определения оператора, его можно считать близким аналогом понятия *структуры экономических процессов*, под которой понимается то относительно устойчивое, неизменное, что сохраняется в экономической системе при вариации причин (входов). А под структурой статистической модели процесса следует понимать набор переменных и коэффициентов при них, которые преобразуют входные переменные модели в выходные [Пуарье, 1981].

Изложенное выше представляет собой общий, абстрактный подход к способу статистического описания причинно-следственных связей и необходим для введения некоторых общих понятий. В дальнейшем изложении представляется целесообразным перейти от общей терминологии к экономической с привязкой к типичному экономическому объекту. Наиболее часто в работе в качестве такого объекта будет использоваться промышленное предприятие (объединение).

Выбор этого объекта связан с ролью предприятия в общей системе промышленного производства. Предприятие не только производитель материальных ценностей и экономически самостоятельная единица, но и источник информации, используемой на всех уровнях планирования и управления народным хозяйством. Вышестоящие звенья (уровни) системы промышленного производства (тресты, комбинаты, главки, министерства) непосредственно не связаны с процессом материального производства, а осуществляют переработку информации о работе промышленных объектов и управляют ими на основе результатов этой переработки.

Представим рассмотренную выше общую схему причинно-следственных связей в терминах функционирования промышленного объекта.

В современном производстве можно выделить шесть структурных элементов. Это, во-первых, рабочий — главная производительная сила общества; во-вторых, материально-вещественные элементы производства (орудия труда, предмет труда и продукт труда); в-третьих, технология и организация производства, характеризующие взаимодействие всех элементов в производственном процессе.

Все структурные элементы промышленного производства взаимосвязаны, они образуют не произвольные, а определен-

ные типичные сочетания. Главенствующую роль в формировании этих комбинаций обычно играют орудия труда: конкретной технике производства обычно отвечают соответствующие предмет и продукт труда, технология и организация производства. В добывающих отраслях промышленности, где значительное влияние на остальные элементы процесса производства оказывают характер и качество предмета труда, в роли своеобразного структурообразующего элемента выступают горно-геологические условия. Вместе с тем каждый из остальных элементов, а также уровень технологии и организации производства оказывают значительное влияние на развитие техники и друг на друга и в определенных условиях могут выступать как база формирования типа предприятия.

Связи между структурными элементами производства не носят жесткого детерминированного характера. С развитием промышленного производства помимо увеличения количества разновидностей каждого элемента (специализация и дифференциация техники, расширение числа видов используемого сырья, топлива, многообразие технологических способов производства и т. д.) возрастает число возможных и реализуемых комбинаций этих элементов. Для большинства отраслей современного промышленного производства характерна неоднозначность технологии. Технологические решения стали чрезвычайно многообразными, в том числе и для определенных орудий труда. Как и выбор технологии, выбор варианта организации производства перестает быть однозначным. При данной технике и технологии становятся возможными многочисленные варианты организации производства. Сказанное относится и к другим структурным элементам производственного процесса.

Связи, направления влияния между элементами производства не однозначны. Взамен классической схемы формирования типа предприятия: орудия труда → технология → → организация производства — появляются новые схемы, где вершиной графа может выступать практически любой структурный элемент производства, а последовательность связей может быть различной. Например, требования к качеству продукта труда — необходимость получения особо чистых металлов — обусловили разработку новых технологических процессов и новой техники в металлургии (электровакуумная и электронно-лучевая плавка и т. п.).

В технико-экономических исследованиях, какими являются статистические исследования на уровне предприятия, работу промышленного объекта удобно рассматривать как

процесс преобразования предмета труда в продукт труда. Тогда предприятие можно представить в виде некоего преобразователя, задача которого сводится к переработке различных видов ресурсов в готовую продукцию.

Схема «предприятие — преобразователь»— аналогия технического преобразователя, и как всякая аналогия условна. В техническом преобразователе резко различаются преобразующее устройство и преобразуемый ресурс. На предприятии эти элементы близки. Основные фонды, рабочая сила носят двоякий характер: с одной стороны, выступают как ресурсы, с другой — являются элементами преобразователя.

Процесс функционирования промышленного объекта может быть представлен следующей схемой: на вход объекта поступают ресурсы: сырье, электроэнергия, затраты труда. При этом часть ресурсов — сырье, электроэнергия, топливо и т. д.— поступает из внешней среды, другая часть представляет собой использование элементов самого преобразователя — производственных фондов и рабочей силы.

На предприятии осуществляется процесс производства, который представляет собой воздействие труда при помощи орудий труда на предмет труда заданным технологическим способом при определенном уровне организации производства.

На выходе промышленного объекта мы вновь имеем материальный поток — готовую продукцию. Описанный процесс асимметричен, направлен от ресурсов к продукции и, как правило, необратим.

Процессу производства продукции сопутствует процесс преобразования состояния самого предприятия (изменения объема и состава его основных фондов, запасов сырья и полуфабрикатов и т. п.). Таким образом, при углубленном экономическом исследовании схема потоков на предприятии за определенный период должна включать помимо двух основных элементов (расхода ресурсов и выпуска продукции) также и третий элемент — изменение состояния предприятия. То есть предприятие можно рассматривать как двойной преобразователь: предмета труда в продукт труда и базисного состояния предприятия на начало исследуемого периода в конечное (схема 1.1).

Учет процесса изменения состояния предприятия имеет большое значение лишь в отдельные периоды его функционирования (при освоении, реконструкции и т. п.). Во время стационарного функционирования изменение состояния предприятия обычно оказывается незначительным и поэтому часто



Схема 1.1. Функционирование предприятия.

можно воспользоваться только основной частью схемы преобразователя, описывающей собственно процесс производства продукции.

Вполне понятно, что в приведенной схеме элементы весьма укрупнены. В реальных исследованиях каждый из элементов схемы описывается набором переменных, характеризующих его уровень, качество и структуру.

В ходе экономико-статистического исследования мы всегда имеем дело с информацией о материальных потоках в виде технико-экономических показателей и характеристик. Они могут быть поставлены в соответствие определенным частям материального потока: информация о ресурсах, информация о производственном потреблении ресурсов, информация о результатах производства. С этой точки зрения производство может рассматриваться как преобразователь входной информации в выходную.

Таким образом, все показатели, характеризующие деятельность предприятия, могут быть сведены в три группы: входные, условия производственного потребления и выходные. Входные показатели — это величина и структура затрат материальных ресурсов; условия производственного потребления ресурсов — технологические способы производства, природные условия и др.; выходные — результаты производства (объем готовой продукции, ее структура, качество и др.).

Между выходными показателями, с одной стороны, и всеми остальными, с другой, существует некая причинная связь,

выражающая влияние входных показателей и условий производственного потребления на выходные.

При исследовании непрерывный процесс функционирования промышленного объекта обычно заменяется набором дискретных актов производства. Для предприятия длительность одного такого дискретного акта принимается обычно равной году, но может быть и другой. Для объектов более низкого уровня величина квантов обычно меньше. Вход и выход по схеме преобразователя в этом случае представляют собой накопленные за рассматриваемый период величины, характеристики преобразователя — средние величины за этот же период.

Квантование процесса производства приводит к замене действительной линии развития технологического уровня производства на предприятии ступенчатой, дискретно-непрерывной.

Продолжая аналогию с преобразователем, можно выделить два вида оценок хода процесса — его интенсивность и эффективность.

Интенсивность процесса производства измеряется величиной выходного потока в единицу времени. В экономической интерпретации интенсивность процесса производства для предприятия — это объем выпуска продукции за год (или другой период), для агрегата — производительность в час, смену и т. д. Интенсивность — всегда именованная величина.

Эффективность процесса преобразования, его так называемый коэффициент полезного действия, представляет собой отношение выходного потока к входному, соотношение результатов и затрат производства. В экономической практике часто используются и обратные величины — отношение затрат к результату. Показатели эффективности могут быть и относительными, и именованными величинами.

Сопоставление результата с отдельными видами затрат дает множество частных показателей эффективности в прямой форме: производительность труда, коэффициенты использования оборудования и т. д.

Принципиальный недостаток всех частных показателей эффективности, используемых в традиционном технико-экономическом анализе,— неполнота. В них чаще всего недостаточно учитываются затраты, при этом общий результат производства относят к затратам одного вида ресурсов (например, показатель производительности труда — отношение всего выпуска продукции к затратам труда, фондоотдача — отношение всего выпуска к стоимости фондов и т. д.). В дейст-

вительности при определении эффективности производственного процесса невозможно изолированное измерение эффективности затрат отдельных ресурсов. Результат производства есть функция совместного воздействия всех ресурсов.

Подход к изучению экономической эффективности использования ресурсов должен базироваться на следующих основных принципах:

выпуск продукции является результатом совокупного действия всех затрат ресурсов;

отдельные свойства как ресурсов, так и результатов производства (а следовательно, и экономические показатели, характеризующие эти свойства) взаимосвязаны.

Характерным для этого подхода является то, что эффективность производства оценивается в рамках одной из модификаций общей модели «затраты — выпуск» путем сопоставления общего результата со всеми затратами.

Резюмируя, дадим краткую сводную характеристику предприятия как объекта экономико-статистического исследования.

Изучению и моделированию подлежит формирование показателей интенсивности и эффективности производственного процесса на отдельном промышленном объекте. При этом промышленное предприятие (объект) рассматривается как автономно действующий преобразователь ресурсов в продукцию, внешние условия функционирования которого создаются системой более высокого уровня. Исследованию подлежит совокупность искусственно сформированных дискретных актов производства, каждый из которых характеризуется начальным и конечным состоянием объекта, затратами ресурсов, условиями их производственного потребления, результатами производства и показателями экономической эффективности производственного процесса.

Пренебрегая изменением состояния предприятия в сравнительно короткие периоды времени и несколько упрощая, можно объединить ранее перечисленные характеристики схемы процесса производства в две группы:

производственные факторы (все виды ресурсов и характеристики самого процесса производства);

результативные экономические показатели (выпуск продукции, ее качество и структура, а также показатели экономической эффективности производственного процесса).

Переходя к геометрической интерпретации, можно соответственно двум группам характеристик работы объекта ввести два признаковых пространства — факторов производства и экономических показателей. Так как исследованию подлежит временна́я последовательность дискретных актов производства, в качестве одной из координат в обоих признаковых пространствах выступает календарное время.

Таким образом, каждое предприятие в определенный период времени можно представить двумя точками в соответствующих признаковых пространствах или одной точкой в комбинированном пространстве «факторы — показатели».

Поскольку для измерения каждого структурного элемента производственного процесса в практике исследований обычно используется много признаков, то и признаковое пространство многомерно.

Таким образом, под структурой процесса в экономических исследованиях промышленного производства понимаются характеристики преобразователя ресурсов в продукцию (предприятия). К числу основных таких характеристик относятся интенсивность процесса преобразования, эффективность преобразования, показатели, характеризующие влияние масштаба производства на его результаты, степень взаимозаменяемости ресурсов.

Статистическая модель, описывающая этот преобразователь, как правило, характеризует зависимость между входами (факторами производства) и выходами (результатами производства) исследуемой экономической системы, а также связь между ее входами. Описываемый моделью способ преобразования ресурсов в продукцию может быть назван абстрактной технологией производства. Следовательно, под изменением структуры экономического процесса в промышленности следует понимать изменение абстрактной технологии, ее основных характеристик.

§ 2. ВАРИАЦИЯ ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ И ТИПЫ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

Объектом экономико-статистического исследования является совокупность единичных актов производства. В образованной любым способом статистической совокупности актов производства имеет место вариация экономических показателей, они изменяют свое значение от одной реализации производственного процесса к другой.

Различные значения результирующих экономических показателей деятельности предприятий представляют собой статистические переменные, т. е. обладают вариацией, возникновение которой следует «объяснить» или найти источ-

ник возникновения вариации. Эта цель — выявление источников вариации и определение меры их участия в вариации исследуемого показателя — является содержанием задачи статистического моделирования экономических показателей. Отсюда следует, что изложению содержания этой задачи должен предшествовать анализ изменения экономических показателей и его причин.

Различают вариации экономических показателей определенного объекта во времени;

совокупности объектов в фиксированный момент (период) времени;

совокупности объектов в пространстве и во времени.

Первый вид вариации можно представить как изменение положения определенного объекта (или соответствующей ему точки) в пространстве экономических показателей в последовательные периоды времени. При моделировании одного экономического показателя пространство превращается в плоскость, траектория движения точки представляет собой динамический ряд.

Второй вид вариации ассоциируется с расположениями точек, соответствующих различным объектам в фиксированный период времени, т. е. с перемещением (движением) в пространстве перпендикулярно оси времени. В случае моделирования одного экономического показателя — это ряд распределения.

Третий, общий вид вариации является комбинацией первых двух частных видов и может трактоваться как дискретный случайный процесс. При такой трактовке динамику показателей отдельного предприятия можно считать одной реализацией этого случайного процесса, а вариацию показателей совокупности объектов в определенный период времени — пространственным сечением процесса.

Формирование общей пространственно-временной вариации показателей можно представлять двумя способами: 1) как общность динамических рядов показателей объектов, входящих в совокупность, и 2) как движение рядов распределения показателей объектов совокупности во времени. Выбор способа описания зависит от конкретной постановки задачи исследования и характера исходной информации. При моделировании экономических показателей предприятий, когда динамические ряды обычно короткие, а число объектов в совокупности значительно превышает число временных периодов, предпочтительнее второй способ описания.

Имеются существенные различия в уровне, степени вариации экономических показателей по видам: наибольшая

колеблемость их наблюдается при первом способе описания, вариация показателей отдельного объекта во времени значительно меньше.

Кроме этого имеются значительные различия в уровне вариации экономических показателей между отраслями промышленности. Наиболее высок уровень «пространственной» вариации в добывающей промышленности, меньше — в обрабатывающей. В пределах выделенных групп отраслей дифференциация уровня колеблемости также заметна.

С точки зрения статистического моделирования любой вид вариации экономического показателя можно подразделить на две части — *объяснимую* (*факторную*) и *остаточную*.

Объяснимая вариация вызывается изменением известных и поддающихся учету в модели факторов производства. Остаточная вариация — следствие воздействия не поддающихся учету, измерению и частично неизвестных факторов. Кроме влияния неучтенных факторов производства остаточная вариация может быть вызвана неадекватностью формы модели реальному процессу, погрешностями в измерении величины экономических показателей. В ряде случаев величина экономического показателя и по своему содержанию включает в себя элемент случайности или неопределенности.

Объяснимую вариацию экономических показателей можно в свою очередь разделить на «*динамическую*» и «*структурную*» части. Первая связана с изменением входа системы при сравнительно стабильной абстрактной технологии (преобразователе). Здесь экономико-статистическая модель (ЭСМ) имитирует воздействие на промышленный объект изменяющихся внешних условий его функционирования.

Вторая часть — «структурная»— связана с существенным изменением самой абстрактной технологии (переход к другому типу преобразователя). Между этими двумя причинами изменчивости экономических показателей имеются принципиальные различия. С. М. Вишнев ассоциирует первую причину с тактическим управлением, а вторую — со стратегическим. Он пишет: «Экономические (плановые и управляющие) решения можно условно разделить на тактические и стратегические. Тактические решения не меняют структуры системы и влекут за собой быстро затухающий эффект, стратегические же решения видоизменяют саму структуру, а последствия их сказываются в течении ряда лет и даже десятилетий» [Вишнев, 1977, с. 12].

Применительно к выделенным способам формирования совокупности актов производства (временной и пространственной) рассмотренную выше классификацию частей объ-

яснимой вариации можно трактовать следующим образом.

В совокупности последовательных периодов функционирования отдельного предприятия большее значение имеет динамическая вариация, так как основную часть своей истории развития предприятие остается «похожим» само на себя. Однако в развитии предприятия имеется ряд периодов, когда происходит резкое изменение характеристик абстрактной технологии и, следовательно, происходят изменения в характере статистической зависимости факторов и показателей производства. Это обычно имеет место при переходе предприятия с одного периода развития на другой (освоение → → стабильное функционирование → «старение»), реконструкции, изменения природных условий, используемых видов сырья и топлива и т. д. Характерной чертой временной выборки является непрерывность, сравнительная гладкость, относительная плавность переходов от одной абстрактной технологии к другой.

Совокупность пространственных реализаций (актов производства на различных предприятиях в фиксированный момент времени) отличается наличием различных типов абстрактных технологий и дискретностью (разрывностью) переходов от одного способа преобразования к другому.

Построение экономико-статистической модели любого типа базируется на гипотезе об основных свойствах моделируемой экономической системы, к числу которых можно отнести: управляемость, стохастичность и инерционность. Управляемость — возможность частичного регулирования входных параметров системы; стохастичность — неоднозначность связи выходных показателей с входными; инерционность — запаздывание реакции выхода системы на изменение ее входных характеристик и относительная устойчивость характеристик преобразователя.

В зависимости от того, на учет какого (или каких) свойства экономических систем делается акцент, ЭСМ можно подразделить на два крупных класса — со стабильной и переменной структурой.

ЭСМ со стабильной структурой ориентированы в основном на учет свойств стохастичности и инерционности, для них характерно описание будущих состояний экономического объекта при помощи экстраполяции закономерностей его поведения в прошлом. При построении ЭСМ такого типа принимается, что поведение экономической системы на всем анализируемом отрезке времени описывается одним и тем же случайным процессом. Это равносильно предположению о постоянстве свойств экономической системы во времени.

В основе корректного применения ЭСМ со стабильной структурой лежит одна из характерных особенностей экономических систем — их инерционность, невозможность существенного изменения состояния экономической системы за малые промежутки времени. В пределах времени срабатывания управляющих воздействий применение ЭСМ со стабильной структурой вполне оправданно.

Модели с переменной структурой ориентированы на получение информации о возможных состояниях экономического объекта для различных стратегических управляющих решений, они учитывают возможные качественные изменения в поведении изучаемого объекта в результате воздействия новых управляющих решений. Эти модели одновременно отражают два основных свойства экономических систем — управляемость и стохастичность. С содержательной точки зрения использование ЭСМ с переменной структурой соответствует принятию гипотезы об изменяющейся во времени эффективности использования производственных ресурсов. Учет свойства управляемости экономических систем при конструировании модели означает, что для описания поведения экономического процесса на разных отрезках времени или в разных областях признакового пространства должны использоваться модели с различной структурой.

В основе построения ЭСМ с переменной структурой лежит гипотеза о том, что поведение экономической системы на временном (пространственном) интервале ее развития описывается разными случайными процессами. Они реализуют более сложную вероятностную конструкцию по сравнению с моделями со стабильной структурой.

Выбор того или иного типа модели для описания конкретного экономического процесса зависит от степени инерционности и управляемости последнего. Естественно, что модели с переменной структурой являются более общим классом экономических моделей, включающим в себя и класс моделей со стабильной структурой. Однако, несмотря на большую универсальность моделей с переменной структурой, они не исключают использования первого типа моделей. Дело в том, что модели со стабильной структурой, реализующие гипотезу о стационарном или о приводимом к стационарному случайном процессе, отличаются сравнительной простотой конструкции. Если гипотеза о стационарности случайного процесса подтверждается, то применение моделей со стабильной структурой достаточно эффективно.

Построение моделей с переменной структурой — это один из основных путей повышения адекватности моделей реаль-

ным процессам. Однако при этом необходимо учесть резкое нарастание сложности конструкции модели в связи с введением требования изменчивости структуры. Поэтому в каждом конкретном случае выбранная форма модели является воплощением необходимого компромисса между ее содержательными, конструктивными и потребительскими свойствами.

§ 3. ОСНОВНЫЕ ПРИЧИНЫ ИЗМЕНЧИВОСТИ АБСТРАКТНОЙ ТЕХНОЛОГИИ. СТРУКТУРНАЯ НЕОДНОРОДНОСТЬ ОТРАСЛЕВЫХ СИСТЕМ

Главными причинами изменения характера статистических зависимостей во времени является реализация достижений научно-технического прогресса, этапность развития предприятий (освоение, стабильное функционирование, «старение»), изменение природных условий в добывающих отраслях промышленности, масштабов производства и уровня использования ресурсов, качества организации и управления, внешних условий функционирования экономического объекта.

Вообще говоря, причины вариации показателей как во времени, так и в пространстве могут быть одинаковыми. Но на практике факторы этих двух видов вариации относительно специфичны: отдельные из причин играют основную роль только для каждого из видов вариации. Например, природно-географические условия — в основном для пространственных сечений; изменения календарного времени, вернее, факторов, стоящих за ними,— только для временных или пространственно-временных совокупностей.

Диапазоны, характер вариации факторов и значимость изменения отдельных факторов обычно различны для временных реализаций и пространственных сечений. Например, в пространственном аспекте резко различаются объемы производства, уровни концентрации; во времени, для одного объекта, эти факторы обычно слабо варьируют. Зато больше сказываются тренды, изменение уровня организации и управления. Далее, во времени для отдельного предприятия размах изменения факторов значительно уже, чем для пространственной выборки, но он может быть разным для различных факторов.

Остановимся более подробно на характеристике первых трех источников вариации экономических показателей, влияние которых наиболее динамично.

В последние десятилетия главным фактором экономического развития является *научно-технический прогресс*. Возможность роста экономики за счет экстенсивных факторов резко сужается. Определенный темп научно-технического прогресса становится необходимым условием развития экономики. Под техническим прогрессом понимается процесс совершенствования средств, предметов и продуктов труда, технологии, организации и управления производством, ведущий к повышению эффективности последнего. В качестве интегральной характеристики реализации достижений технического прогресса на конкретном объекте в определенный период времени выступает показатель технического уровня производства, другими словами, технический уровень производства — это общий результат непосредственных проявлений технического прогресса в некоторый фиксированный момент времени для определенного объекта.

Реализация достижений научно-технического прогресса проявляется в изменении абстрактной технологии (способа преобразования), ее основных характеристик.

Характер влияния технического прогресса на абстрактную технологию (эффективность преобразования ресурсов в продукцию), а следовательно, и структуру модели зависит от типа и формы технического прогресса, масштаба изучаемого объекта, его положения на траектории развития и др.

Технический прогресс протекает во времени — как процесс прогрессивных изменений элементов производства, и в пространстве — как процесс распространения более совершенных элементов в некоторой производственной системе (например, отрасли), по терминологии Э. Янча [1974] соответственно «вертикальный» и «горизонтальный» технический прогресс [1]. На уровне предприятий преимущественно проявляется «вертикальная» форма, хотя для отрасли в целом это может быть проявлением «горизонтального» технического прогресса.

В рамках статистического моделирования мы имеем дело с преимущественно «горизонтальным» техническим прогрессом. Для совокупности действующих предприятий одного способа производства «горизонтальный» технический прогресс во многом может характеризоваться экономическими показателями предприятий, вступивших в строй в разные годы. Действительно, если экономические показатели предприятий с разным годом ввода в эксплуатацию привести к

[1] «Горизонтальный» технический прогресс также протекает во времени.

одинаковым значениям производственных факторов и «возраста» предприятия, то их различие будет характеризовать временную тенденцию «горизонтального» технического прогресса.

Проследить влияние распространения технических и технологических новшеств на характер вариации экономических показателей в совокупности удобно путем сопоставления рядов распределения за несколько фиксированных периодов (лет). Такой анализ позволяет оценить степень однородности и динамичности технико-организационного уровня объектов совокупности. В самом общем виде процесс влияния «горизонтального» технического прогресса на характер «пространственной» вариации экономических показателей описан в табл. 1.1.

Технический прогресс можно также подразделить на капиталоемкий (овеществленный), требующий существенных инвестиций и не требующий больших капитальных вложений (неовеществленный). Последний проявляется в совершенствовании методов организации и управления производством.

Рассмотрим влияние этих двух форм технического прогресса на абстрактную технологию отдельного промышленного предприятия и их совокупности (подотрасль, отрасль).

Таблица 1.1

Влияние «горизонтального» технического прогресса на характер распределения экономических показателей

Период	Характер технико-организационного уровня объектов совокупности	Вид распределения экономических показателей
1	Сравнительно однородная совокупность	Нормальное или обобщенно-нормальное
2	Появление отдельных объектов с резко отличающимся технико-организационным уровнем	Основная масса объектов — нормальное и обобщенно-нормальное, появление значительной асимметрии или выпадающих точек
3	Распространение передового технико-организационного уровня на другие предприятия	Появление многовершинных распределений или разложение общего распределения на несколько частных
4	Освоение прогрессивного уровня производства всеми предприятиями	Нормальное и обобщенно-нормальное распределение, но смещенное относительно периода 1

Для отдельного предприятия реальные возможности изменения абстрактной технологии зависят от положения рассматриваемого объекта на траектории развития. В период стабильного функционирования, когда производственные фонды освоены, возможность изменения масштабов и соотношений в затратах ресурсов весьма ограниченна. Здесь речь может идти в основном о реализации неовеществленного технического прогресса. Структура модели на таких участках траектории развития либо стабильна, либо меняется достаточно плавно.

Другая картина наблюдается при исследовании долгосрочной перспективы развития предприятия, где возможности маневра значительно шире и могут реализоваться обе формы технического прогресса — овеществленный и неовеществленный. Например, в период проектирования возможности маневра очень широки, меньше, но довольно существенны возможности маневра в период строительства предприятия, его реконструкции, коренного изменения технологии и т. д. Таким образом долгосрочное развитие предприятия включает несколько абстрактных технологий и областей перехода от одной к другой. Такое развитие может быть корректно описано с помощью моделей с переменной структурой, характеризующих каждую разновидность абстрактной технологии (технический уровень), а также переходные периоды между ними. Длительность переходного периода во многом определяется характером реализуемого технического прогресса. Естественно, что для случая овеществленного технического прогресса этот период более длителен и ярче выражен, чем для неовеществленного.

Таким образом, с точки зрения технического прогресса различие между моделями со стабильной структурой (краткосрочными) и переменной структурой (долгосрочными) для отдельного объекта заключается в том, что первые предполагают жесткие ограничения на наличие установленного оборудования и возможность его замены. Вторые предполагают лишь ограничения на накопленный фонд реализуемых технических альтернатив развития. Различия между типами моделей и лежащими в их основе гипотезами в некоторой степени относительны. Они зависят от скорости физического и морального износа оборудования, технологии и других элементов технического прогресса. Поэтому в различных отраслях, а также в одной отрасли в различные календарные периоды темп изменения технологии будет различным.

Для совокупности объектов процесс изменения абстрактной технологии разновременен: отдельные объекты находят-

ся на различных участках траектории развития, имеют неодинаковый технический уровень. Поэтому границы между технологическими периодами для системы в целом смазаны. Степень размытости границ зависит от числа объектов в системе. Чем их больше, тем более растянуты переходные периоды и тем плавнее ведет себя траектория развития. Однако если систему рассматривать на уровне объектов, то получим совершенно иную картину. В каждый фиксированный момент времени отдельные объекты имеют определенный технический уровень и переходы от объекта к объекту (от типа к типу) дискретны.

Вторая важная причина изменения структуры связей показателей и факторов производства во времени — это *предыстория развития и функционирования* предприятия.

Укрупненно история развития предприятия включает следующие стадии: проектирование, строительство, эксплуатация. Последняя, в свою очередь, может быть подразделена на периоды освоения, стационарного, стабильного функционирования и «старения». Каждой крупной стадии соответствуют свои характерные задачи исследования, степени свободы изменения масштабов и соотношений факторов производства, а следовательно, и специфические формы моделей. Каждый период эксплуатации предприятий характеризуется различным набором переменных, влияющих на моделируемый показатель, неодинаковой ролью отдельных переменных и разной скоростью изменения показателей во времени. Для каждой стадии развития промышленного объекта характерна своя динамика экономических показателей и специфические функции преобразования затрат ресурсов в готовую продукцию. Как правило, первая и третья стадии характеризуются быстрыми темпами изменения экономических показателей во времени и сравнительно низкой эффективностью производства, вторая — относительно медленными темпами изменения экономических показателей и наиболее высоким уровнем эффективности. Поэтому и структура моделей, соответствующих этим стадиям должна быть различной.

Анализ влияния «возраста» предприятия на вариацию экономических показателей производства покажем на примере леспромхозов, входящих в состав производственного объединения Красноярсклеспром (см. [Кулешов и др., 1972]). Поведение показателя «объем заготавливаемой древесины» изучалось путем анализа его динамики для различных по продолжительности функционирования групп леспромхозов. В качестве классификационного признака, формирующего такие группы, принималась дата основания леспромхоза.

Проведенный анализ динамики объема производства свидетельствует о том, что развертывание лесозаготовительного предприятия во времени носит отчетливо выраженный поэтапный характер и включает три производственных этапа функционирования.

Первый этап характеризуется быстрым наращиванием производственных мощностей леспромхоза. Здесь же осуществляется в некотором роде адаптация к местным условиям производства, оформление коллектива и т. д. Происходит становление леспромхоза как самостоятельной хозяйственной единицы.

Второй этап начинается после того, как леспромхоз достигнет проектных мощностей. На этом этапе увеличение объемов заготовки древесины (сравнительно незначительное) происходит в основном за счет роста производительности труда и выявления резервов производства.

Третий этап отражает постепенное уменьшение заготовок древесины. Снижается обеспеченность техническими средствами, растет износ производственных фондов, сокращается численность промышленно-производственного персонала.

Для каждого из этапов имеет место своя специфическая зависимость объема производства от производственных факторов. Такого же рода закономерность выявлена аналогичным исследованием по угледобывающей [Розин, Сергеева, 1970] и газовой промышленности [Пляскина, 1980].

Итак, характер связей и значение экономических показателей определяются не только достигнутым техническим уровнем производства, но и положением данного предприятия на траектории развития. Это обстоятельство представляется достаточно важным и должно учитываться при конструировании моделей.

Следующая важная причина нестабильности структуры взаимосвязей переменных во времени — *это изменение природных факторов.* Особенно отчетливо это обстоятельство проявляется в добывающих отраслях промышленности. Понятие «природные факторы» весьма широко — оно охватывает все элементы окружающей природы, влияющих на эффективность производственной деятельности промышленных объектов. Их можно подразделить на две основные группы: непосредственно участвующие в материальном производстве (естественные ресурсы) и не участвующие в производстве, но оказывающие влияние на его результаты (природные ус-

ловия). Как правило, наиболее динамичными являются качество естественных ресурсов, а также та часть природных условий, которая характеризует горно-геологические условия добычи естественных ресурсов. Как правило, с течением времени эта часть факторов претерпевает определенные изменения в основном в сторону ухудшения параметров (увеличивается глубина залегания, уменьшаются запасы, снижается пластовое давление, увеличивается обводненность и т. д.).

Помимо трех перечисленных укрупненных групп факторов есть еще целая группа менее четко выраженных причин изменения структуры связей в динамике. Все они часто объединяются термином *«фактор времени»*. Фактор времени — это своеобразный общий фактор, аккумулирующий (опосредующий) в себе влияние целого комплекса реальных причин развития процесса (частью измеримых, частью неизмеримых). В этом плане можно провести аналогию между фактором времени и географическим положением предприятия, за которым (особенно в добывающих отраслях) также стоит целый комплекс реальных факторов производства. Нагрузка на фактор времени в моделях во многом зависит от их структуры: чем больше факторов развития удается ввести в модель в явном виде, тем меньше нагрузка на фактор времени. Однако принципиально влияние этого косвенного, аккумулирующего измерителя не может быть полностью устранено из-за действия целого ряда неизмеримых факторов (например, повышение опыта и уровня знаний работников и т. п.). Время в моделях экономических показателей работы промышленных объектов может выступать в двух формах:

календарное время с общей начальной точкой отсчета для всех объектов совокупности;

«возраст» предприятия, время его функционирования.

В последнем случае начальные точки отсчета различны для объектов разного периода ввода в эксплуатацию, т. е. объектов разных «поколений».

На первый взгляд может показаться, что эти две формы измерения времени дублируют друг друга. Между ними действительно существует тесная взаимосвязь, однако их смысловые нагрузки неодинаковы. Календарное время, введенное в модель, попросту отражает все то, что связано с тенденцией развития моделируемого объекта и не находит явного отражения в независимых переменных модели. «Возрастной» фактор дифференцирует предприятия с одинаковым значением календарного времени на разные «поколения», тем самым учитывая в неявной форме многие качественные моменты, отличающие предприятия с разным периодом ввода в эксплуатацию.

Все указанные причины изменения структуры связей действуют не изолированно одна от другой, а, как правило, совместно. При этом существуют взаимосвязи между отдельными группами причин (например, те или иные природные условия определяют тип используемой технологии и техники). В конечном итоге получается довольно сложная картина изменения структуры влияний переменных, что определяет актуальность и важность разработки способов выявления и учета этого факта.

Таким образом, предметом экономико-статистического исследования является не просто совокупность явлений, а общие совокупности, состоящие из нескольких частных подсовокупностей. Общее качество частных подсовокупностей (типов), объединяемых общей совокупностью, в конечном счете одно (например, для предприятий одной отрасли). Но только в рамках отдельной частной подсовокупности (типа) имеет место стабильный характер зависимости показателя от факторов производства.

Следовательно, без выделения и изучения типов и без типологической классификации закономерности, существующие в общей совокупности, не могут быть изучены достаточно глубоко и точно. Процесс расслоения совокупности (изменения ее структуры) рассмотрим более детально на примере отраслевой системы, понимаемой как совокупность отдельных предприятий.

Совокупность объектов отрасли и даже подотрасли, как правило, неоднородна. Пространственные совокупности промышленных объектов одной отрасли (подотрасли) обладают определенной структурой, т. е. общая совокупность промышленных объектов распадается на ряд подсовокупностей, имеющих свой, отличный характер зависимости экономических показателей от входных переменных.

Объекты пространственной совокупности естественным образом объединяются в несколько типов (классов). Тип промышленного объекта характеризуется определенным диапазоном значений уровня структурных элементов производственного процесса и определенной их комбинацией. В пределах типа имеется значительная свобода вариации значений и комбинаций элементов, но только такая, при которой не изменяется характер их воздействия на величину экономических показателей.

В терминах многомерной классификации этот факт описывается следующим образом. Если признаковое пространство представить в виде n-мерного параллелепипеда, то точки, соответствующие объектам отрасли, будут занимать не весь объем параллелепипеда, а расположатся отдельными сгущениями, между которыми имеются разрывы. В общем случае предприятия (или другие промышленные объекты) располагаются в признаковом пространстве неравномерно, области сгущений чередуются с пустыми областями. Каждое такое четко выраженное сгущение можно ассоциировать с типом объектов. В его пределах объекты приближенно подобны друг другу, чего нельзя сказать об объектах разных типов.

Деление общей совокупности исследуемых предприятий на отдельные типы (подсовокупности) объективно обусловлено рядом причин.

Одна из таких причин — неравномерность развития отдельных единиц совокупности. Разные предприятия имеют разный «возраст», принадлежат к разным «поколениям». Каждому периоду создания предприятий, как правило, соответствует особая модификация исследуемого процесса, свой уровень развития техники и технологии производства. Объекты, построенные в одном или близких календарных периодах, имеют приблизительно равные уровни техники и технологии производства и находятся на примерно одинаковых точках траектории развития, и, наоборот, объекты разных «поколений» обычно существенно различаются между собой по своим характеристикам и по влиянию факторов производства на экономические показатели.

Далее, объекты одного «поколения» могут, в свою очередь, распадаться на группы. Здесь в качестве типообразующих фактов выступают мощность объекта (как правило, крупные и мелкие предприятия оснащены разной техникой и имеют различный уровень организации производства и, следовательно, разный тип зависимости), назначение и характер продукции. В добывающих отраслях один из важнейших типообразующих признаков — это природные условия.

Построение моделей с переменной структурой связано с исследованием изменения типологии совокупности во времени. Типологию можно представить как набор пространственных сечений (состояние совокупности объектов в фиксированные последовательные моменты времени t_0, t_1, ..., t_k, ...). Каждое пространственное сечение в неявном виде включает элементы динамики: объекты, входящие в него, находятся на разных точках траектории своего развития, имеют различную предысторию, различный технический уровень. Процесс формирования пространственных сечений и изменения их типологической структуры во времени представлен на рис. 1.1.

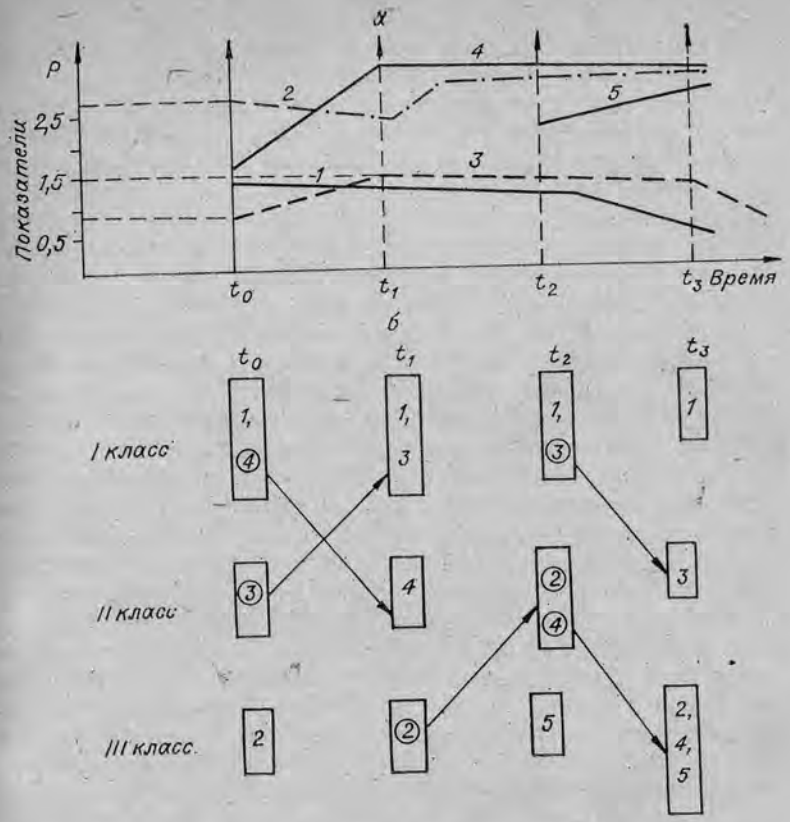


Рис. 1.1. Динамика типологической структуры совокупности.
а — траектория развития отдельных объектов; *б* — состав классов объектов по сечениям; 1—5 — номера объектов (предприятий).

На рис. 1.1 показаны условные траектории развития пяти предприятий. Уровень развития предприятия характеризуется некоторым показателем P (например, объемом производства). Предприятия различаются по уровню производственной мощности и времени ввода в эксплуатацию. Наиболее «старым» является предприятие № 1, которое в исследуемом периоде завершает свое функционирование, а самым «молодым» — предприятие № 5, которое в конце исследуе-

мого периода вводится в эксплуатацию. Построено четыре пространственных сечения в моменты времени t_0, t_1, t_2, t_3. В каждом сечении выделено три класса по показателю P. Классы ранжированы по величине этого показателя. Кружками отмечены объекты, которые переходят из класса в класс в исследуемые периоды времени.

Типологическую структуру каждого пространственного сечения можно характеризовать тремя элементами: составом классов; количественными характеристиками объектов, входящих в выделенный класс; числом классов. В случае, приведенном на рис. 1.1, число классов для упрощения принято постоянным. В общем случае все три элемента типологической структуры меняются во времени.

Первые два элементарных процесса (изменение состава классов и количественных характеристик объектов) носят систематический плавный характер, но скорость их протекания переменная. В отдельные периоды эти процессы протекают более интенсивно, в другие — менее. Третий элементарный процесс (изменение числа классов) имеет дискретный периодический характер.

Изменение базисной типологической структуры совокупности происходит за счет двух источников: развития действующих на период t_0 предприятий и ввода в эксплуатацию новых предприятий. Относительное значение этих двух источников изменения структуры совокупности различно для разных отраслей промышленности. Например, в металлургии пуск нового предприятия происходит сравнительно редко, более распространен путь увеличения мощности и реконструкции действующих предприятий. В отличие от этого в лесозаготовительной, угольной и нефтяной промышленности доля вводимых предприятий сравнительно велика.

Указанные два источника изменения типологической структуры совокупности оказывают различное влияние на выделенные элементы типологии. Процесс развития действовавших в базисный период предприятий не изменяет объема совокупности. Влияние этого процесса на типологическую структуру совокупности зависит от того, на каком участке своей траектории развития находится предприятие. Развитие отдельного предприятия на большей части его траектории происходит сравнительно медленно, плавно и, следовательно, не меняется его принадлежность к определенному классу (например, предприятие № 1 на рис. 1.1). Однако в истории развития предприятия есть ряд периодов, когда происходит резкое изменение его состояния. Это прежде всего реконструкция, после которой предприятие

может перейти в другой из существовавших классов и даже образовать новый класс (например, предприятие № 2, реконструкция в период $t_1 - t_2$). Относительно высокой скоростью изменения характеристик отличается также период освоения производственной мощности (например, предприятие № 4 на рис. 1.1). В ряде случаев изменения принадлежности действовавшего предприятия к классу могут быть вызваны радикальными изменениями технологии, характера сырья и выпускаемой готовой продукции.

В связи с тем, что предприятия, образующие базисную совокупность, как правило, находятся на разных этапах своей траектории и радикальные изменения их производственных способов происходят в разные периоды времени, траектории развития отдельных объектов оказываются не параллельными, а пересекающимися. Это приводит к изменению распределения объектов между классами: численность одних классов уменьшается, других возрастает. При этом набор классов, как правило, остается стабильным в течение сравнительно большого промежутка времени, так как радикальные изменения технологического способа в истории предприятия сравнительно редки. Исключение составляет период интенсивного внедрения в отрасли новых технологических способов производства, в который последовательно вступает большинство предприятий. Развитие объектов сказывается и на характеристиках классов — с течением времени они изменяются.

Ввод в эксплуатацию новых предприятий изменяет объем совокупности и элементы ее структуры (предприятие № 5, см. рис. 1.1). За счет новых предприятий увеличивается удельный вес соответствующих классов и могут появляться новые классы; последнее, как уже указывалось, происходит сравнительно редко.

Изменение базисной структуры совокупности вызывается не просто развитием входящих в нее объектов, а неравномерностью этого развития. Если — представим себе такой гипотетический случай — начальная точка (год пуска в эксплуатацию) была бы у всех объектов одна и та же, развивались они с одинаковой (близкой) скоростью и по одинаковой (близкой) траектории, то структура такой совокупности оставалась бы стабильной во времени. Изменение характера взаимосвязей показателей и факторов производства происходило бы у всех объектов в один и тот же момент и было бы обусловлено, во-первых, переходом предприятий с одного участка траектории развития на другой (например, с периода освоения на стационарный режим); во-вторых,

внедрением технико-технологических и организационных новшеств; в-третьих, действием «эволюционных» факторов (опосредованных фактором времени).

Из сказанного следует, что при исследовании динамики типологической структуры совокупности общая задача выделения однородных пространственно-временных блоков имеет два аспекта: выделение предприятий одного поколения и одного характера развития в рассматриваемом интервале времени; выделение для этих однородных (в смысле истории развития) предприятий участков траектории со стабильным характером взаимосвязи показателей и факторов производства.

§ 4. ИЗМЕНЧИВОСТЬ СТРУКТУР И НЕОДНОРОДНОСТЬ СТАТИСТИЧЕСКИХ СОВОКУПНОСТЕЙ

Проблему структурных изменений преобразователей (операторов) в статистическом плане можно рассматривать как задачу выявления неоднородностей статистических совокупностей с точки зрения моделируемых зависимостей.

Существует множество подходов к анализу и оценке степени однородности исследуемой совокупности. Однако несмотря на различие взглядов на эту проблему, большинство исследователей согласны, что экономические совокупности, как правило, обладают структурой и это обстоятельство должно учитываться в различных областях статистического анализа.

Необходимым условием решения задачи моделирования является конкретизация понятия неоднородности. В литературе предложено немало определений этого понятия. Однако устоявшегося термина нет: большинство определений, в том числе использованное и в данной работе можно трактовать лишь как «рабочие», «условные».

По нашему мнению, лучше говорить не об однородности вообще, а о степени, уровне этой однородности. В экономико-статистическом исследовании мы оперируем понятием относительной однородности и речь может идти лишь о достижении разумной степени однородности совокупности, такой, при которой обеспечивалась бы достаточная для данного исследования точность статистических выводов. В качестве характеристики степени однородности может выступать «близость», «похожесть» объектов совокупности по признакам их описания. Весь комплекс признаков, описывающих единицы наблюдения, можно разделить на два вида: общие

(близкие по значению), которые определяют сходство, однородность изучаемых единиц наблюдения, и варьирующие, влияние которых должно быть оценено в результате статистического исследования. Заметим при этом, что степень однородности характеризуется не только и не столько долей общих признаков во всем их комплексе, сколько их ролью в формировании характерных свойств объектов.

В зависимости от назначения и стадии исследования следует различать два понятия однородности: широкое — применяемое для выделения исходной совокупности из всей массы экономических явлений и используемое априори на стадии постановки задачи исследования; узкое — предназначенное для выделения однородных групп (типов) объектов уже после сбора информации, на стадии собственно экономико-статистического моделирования, т. е. апостериори.

Для выделения совокупностей, однородных в широком смысле, как правило, достаточно небольшого числа признаков, обычно носящих качественный характер, поскольку между ними есть причинно-следственные связи. Поэтому весьма полезным инструментом выбора признаков на первой, постановочной стадии исследования является причинный анализ структуры их связей [Бородкин, Лукацкая, 1968]. На основе такого анализа можно уже на стадии постановки задачи формировать относительно однородные совокупности объектов по ограниченному числу признаков, находящихся на входе графа связи. Например, в структурных схемах связи факторов, обусловливающих вариацию показателей работы угольных шахт, входными являются горно-геологические условия, в большой мере определяющие значения других варьирующих элементов (способ вскрытия, метод подготовки шахтного поля, система разработки угольных месторождений, способы механизации и т. д.). В этом случае достаточная начальная однородность совокупности обеспечивается группировкой шахт по горно-геологическим условиям.

Понятие однородности в узком смысле вводится для выделения таких типов объектов, внутри которых взаимосвязь показателей и факторов производства стабильна и может быть описана простыми аппроксимирующими функциями (например, линейными или функциями-константами), другими словами — выделения в признаковом пространстве областей с относительно постоянной структурой влияния факторов на экономические показатели производства. Оценка уровня этого показателя однородности производится апостериори, после сбора информации на основе анализа мно

гомерного распределения объектов исходной совокупности и выделения его «вершин». Последние трактуются как однородные в узком смысле группы объектов.

Введенное понятие однородности в широком и узком смысле в известной мере близко к делению Г. С. Кильдишевым и Ю. И. Аболенцевым [1978] однородности на качественную и количественную. Под первой понимается однородность, определяемая по признаку однокачественности, а под второй — однородность группы статистических наблюдений, выявляемая на основе количественных признаков. Оба эти понятия диалектически взаимосвязаны, причем количественная однородность имеет смысл только при наличии однокачественности явления. В то же время накопление количественной неоднородности может привести к изменению качества — формированию нового типа явления.

Следует заметить, что во всей проблеме однородности наиболее сложным и дискуссионным является именно вопрос о способах и критериях выделения таких однородных типов объектов в пределах однокачественной исходной совокупности. Этот вопрос в последние годы достаточно широко обсуждался в литературе, в частности на страницах журнала «Вестник статистики». Предложен ряд возможных критериев оценки однородности. Мы согласны с мнением Б. И. Пасхавера [1972], что полностью формализованный критерий однородности возможен только в одном случае — если теоретически, на основе качественного анализа явления, обоснован закон распределения. Тогда количественной оценкой однородности является степень близости распределения признака к теоретическому распределению. Но такие ситуации в социально-экономических исследованиях весьма редки. Как правило, априорный теоретический закон распределения неизвестен.

В. И. Сиськов [1971, 1975] предлагает использовать в качестве критерия однородности коэффициент вариации (V). Однако предложенная им и рядом других авторов методика базируется на двух достаточно сильных посылках: а) априори известен теоретический закон распределения единиц совокупности и б) исследуются преимущественно одномерные объекты. Обе эти посылки, как правило, не соблюдаются в прикладных исследованиях.

Проблему многомерности В. И. Сиськов предлагает решать двумя путями: использовать либо комбинационную группировку, в которой интервалы определены исходя из величины критерия V, либо методы последовательного уточнения состава групп по отдельным признакам. В обоих случаях набор учитываемых признаков при обычных объемах экономических совокупностей будет сравнительно невелик (порядка 3—5). Между тем развитие экономических объектов — процесс достаточно сложный и его особенности, как правило, не могут быть описаны несколькими показателями, необходим многомерный анализ по значительному числу признаков.

В связи с отсутствием априорных сведений о теоретическом законе распределения признаков в совокупности все формализованные способы выделения однородных групп будут в той или иной степени приближенными. Способы будут отличаться по степени учета конкретных условий функционирования экономических объектов.

С нашей точки зрения, выход для многомерного случая заключается: а) в использовании непараметрических методов оценивания плотности вероятности и б) в применении методов распознавания образов (многомерной классификации) для формирования однородных групп.

В наших исследованиях при построении плотности вероятности по выборке использовались непараметрические оценки Парзена (см. [Parsen, 1962]) с экспоненциальными ядрами. Выбор «нормальных» ядер обусловлен ориентацией на дальнейшее использование методов кусочно-линейной аппроксимации при построении экономико-статистической модели. Иначе говоря, мы считаем подсовокупность однородной в узком смысле слова, если многомерное распределение моделируемого показателя и существенно влияющих на него факторов близко к нормальному или к суперпозиции нормальных распределений.

Следовательно, при построении статистических моделей с переменной структурой помимо обычных этапов экономико-статистического исследования возникают дополнительные задачи. Они заключаются в следующем. Прежде всего, необходимы какие-либо содержательные предпосылки для выдвижения гипотезы о наличии структурных изменений исследуемой совокупности во времени или пространстве. Если есть основания для выдвижения такой гипотезы, то следующий этап — формальная проверка гипотезы. В зависимости от цели исследования, характера исходной информации, степени формализуемости, уровня алгоритмической и программной обеспеченности выбирается определенный метод исследования. При подтверждении гипотезы о наличии структурных изменений необходимо выявить области изменения структур и характер этого изменения. Что ка-

сается областей (точек) структурных изменений, то здесь возможны следующие ситуации:

точки (области) априорно заданы из содержательных представлений о моделируемом процессе;

расположение точек (областей) структурных изменений неизвестно и их необходимо найти на основе анализа выборочных значений.

В том и другом случае для выбора метода конструирования ЭСМ необходимо исследовать характер перехода от одной структуры влияния к другой. Теоретически можно рассмотреть два крайних случая: а) наблюдаемый структурный сдвиг гладкий, плавный и переходный режим может быть описан непрерывной функцией; б) структурный сдвиг носит качественный характер, является скачкообразным и переход описывается разрывной функцией. В прикладных исследованиях, как правило, имеют место некоторые смешанные переходные режимы с преобладанием того или иного характера перехода.

После решения перечисленного круга вопросов создаются предпосылки для выбора адекватного подхода к решению основной задачи — построению модели с переменной структурой.

Решение перечисленных задач (выдвижение гипотезы о наличии структурных изменений, определение точек и характера структурных сдвигов) по существу представляет собой комплексный анализ типологической структуры исходной совокупности.

Формальные способы решения задач типологического анализа будут рассмотрены нами в гл. 2. В настоящем параграфе остановимся лишь на приемах выдвижения предварительной гипотезы о неоднородности. Прежде всего необходимо обосновать содержательные посылки, которые позволяют предположить наличие структурных изменений в границах изучаемой статистической совокупности. К их числу, например, при изучении отдельного объекта можно отнести проведение реконструкции предприятия, существенное изменение технологического процесса, переход на переработку другого вида сырья, освоение новых видов продукции и т. п.

В случае пространственной совокупности объектов условия, предполагающие структурные изменения,— это наличие групп предприятий с разным периодом ввода в эксплуатацию, различия в природных условиях функционирования (например, шахты с пологими и крутыми углами падения пластов), в технологии преобразования сырья в готовую

продукцию (например, выплавка стали в мартеновском производстве на твердом и жидком чугуне), в масштабах производства и т. д.

Есть также ряд простых формальных приемов предварительного анализа исходной информации, позволяющих предположить структурную неоднородность совокупности. Во-первых, это исследование качественных признаков в описании объектов, существенно влияющих на моделируемый показатель. Влияние этих признаков проявляется двояко: в непосредственном воздействии качественного признака на уровень показателя и в опосредованном — через воздействие его на влияние других признаков. При этом может оказаться, что наиболее важные количественные зависимости в совокупности имеют разрывы и скачки, вызываемые изменением значений качественных признаков. Во-вторых, это выявление количественных признаков с особым характером распределения: а) признаков, значения которых располагаются только на определенных интервалах числовой оси, между которыми имеются разрывы; б) признаков с многомерными распределениями. Такие признаки по своему характеру очень похожи на качественные и поэтому могут также «ломать» исследуемые зависимости.

Еще один приближенный прием предварительного анализа степени однородности совокупности — построение полей корреляции и эмпирических линий регрессии каждого из включенных в исследование факторов с моделируемым показателем. Нелинейность парных эмпирических линий регрессий, а тем более наличие разрывов и скачков заставляют предположить неоднородность и для многомерного случая.

В некоторых случаях для выдвижения гипотезы о неоднородности многомерной совокупности можно прибегнуть к расчету различного рода коэффициентов криволинейности [Розанов, Френкель, 1967]. Получение значимого коэффициента криволинейности дает основания предположить структурную неоднородность совокупности, зависимость параметров модели от положения объекта в пространстве признаков.

Если предварительный анализ совокупности позволяет выдвинуть гипотезу о ее структурной неоднородности, то следующий этап — формальная проверка этой гипотезы.

Методы проверки гипотезы о наличии неоднородности совокупности (структурного сдвига) будут рассмотрены ниже в соответствующих главах работы.

Глава 2

ОСНОВНЫЕ ЗАДАЧИ И МЕТОДЫ КОНСТРУИРОВАНИЯ МОДЕЛЕЙ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

§ 1. ФОРМАЛЬНАЯ ПОСТАНОВКА ЗАДАЧИ. ТИПЫ МОДЕЛЕЙ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

Введем ряд понятий. Пусть описание любого объекта задается на языке значений «входных» признаков $X = \{X_1, \ldots, X_n\}$. Тогда каждое конкретное описание $x = (x_1, \ldots, x_n)$ есть точка в n-мерном признаковом пространстве R_x. «Выходной» результирующий показатель будем обозначать Y, а его конкретное значение — y. Моделью показателя Y является некоторый функционал $Y^* = g(X, \Theta)$, позволяющий на основе значения x получать прогнозируемое значение $Y - y^*$. Здесь Θ — вектор параметров.

Модель с переменной структурой есть преобразование $Y^* = g(X; \Theta(X))$, где вектор параметров Θ сам является функцией от X.

Такая модель должна строиться на основе анализа выборочной информации V, являющейся совокупностью m различных реализаций (y, x):

$$V = \{(y_i, x_i); \quad i = \overline{1, m}\}.$$

Каждая i-я реализация есть совокупность значений не только «входных» признаков $x_i = (x_{i1}, \ldots, x_{in})$, но и соответствующего значения результирующего показателя y_i. Выборка V есть конкретное проявление некоторого изучаемого и неизвестного механизма взаимодействия показателей (Y, X), которому можно условно сопоставить некоторую функцию распределения вероятностей $F(x, y)$ (или плотность распределения $f(x, y)$) в признаковом пространстве R_{yx}.

Если одним из параметров, например X_n, является время T, то модель будем называть *динамической*, в противном случае — *статической*. В данной главе будем рассматривать лишь статические модели и методы их построения. Динамическим моделям посвящены следующие две главы.

Исходя из характера изменения структуры в признаковом пространстве (т. е. в зависимости от вида функции $\Theta(X)$), можно выделить два больших класса моделей с переменной структурой: непрерывные кусочно-простые модели и дискретно-непрерывные кусочно-простые.

Непрерывный случай сводится по существу к построению нелинейной функции относительно параметров модели. Сложности построения, использования в задачах планирования и содержательной интерпретации таких моделей широко известны (см., например, [Бард, 1979]). Применение непрерывных моделей с переменной структурой позволяет обойти эти трудности путем использования кусочно-простых функций (как правило, линейных или приводимых к линейным).

Непрерывные модели описываются непрерывными функциями, когда при небольших изменениях аргумента значение функции изменяется также весьма мало, и график такой функции является «сплошной» непрерывной кривой (поверхностью).

Дадим определение непрерывной функции для случая одной переменной. Функция $y = f(x)$ называется непрерывной при значении $x = x_1$, если: 1) число x_1 принадлежит к области ее задания и 2) предел $\lim\limits_{x \to x_1} f(x)$ существует и равен $f(x_1)$.

Непрерывные модели с переменной структурой получили наибольшее распространение в управлении технологическими процессами в реальном масштабе времени. Характерным примером такого типа моделей являются адаптационные модели, в которых совмещены процессы построения модели и управления на ее основе. Пересчет параметров моделей производится по мере поступления новых наблюдений. В экономических исследованиях объектов более высокого уровня (предприятия, объединения и т. д.), связанных с периодическим анализом экономических показателей, область применения непрерывных моделей с переменной структурой значительно у́же. Однако и здесь они находят применение как элемент дискретно-непрерывных моделей. В частности, довольно распространен метод представления коэффициентов регрессии в виде функций времени в динамизированных статистических моделях, т. е.

$$y^*(t) = g(x(t), \alpha(t)).$$

Для линейного случая динамизированная регрессионная модель принимает следующий вид:

$$y^*(t) = \alpha_0(t) + \sum_{j=1}^{n} \alpha_j(t)\, x_j(t),$$

где $y^*(t)$ — прогнозируемое значение моделируемого показателя в момент времени t;
$x_j(t)$ — значение j-го фактора в момент времени t;
$\alpha_j(t)$ — значение коэффициента регрессии при j-м факторе в момент времени t;
$\alpha_0(t)$ — значение свободного члена уравнения в момент времени t.

В рамках вычислительной математики в последнее время был развит аппарат построения кусочно-простых функций, непрерывных в узлах «сшивания»,— методы сглаживающих сплайн-функций [Алберг и др., 1972; Василенко, 1978; Завьялов и др., 1980]. Простейшим видом сплайнов является линейный сплайн.

Введем следующие определения [Пуарье, 1981].

Сеткой называется произвольное множество точек оси абсцисс, $\Delta = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{k-1}\}$, где $\bar{x}_1 < \bar{x}_2 < \ldots < \bar{x}_{k-1}$. Точки $\bar{x}_j$ $(j = \overline{1, k-1})$ называются *внутренними узлами* или просто *узлами*.

Будем говорить, что y^* является *линейным сплайном* $S_\Delta(x)$ над Δ тогда и только тогда, когда y^* есть непрерывная кусочно-линейная функция от x, график которой состоит из k прямолинейных отрезков, расположенных над k интервалами $(-\infty, \bar{x}_1]$, $[\bar{x}_1, \bar{x}_2]$, . . ., $[\bar{x}_{k-1}, \infty)$. Входящие в Δ узлы называются также *точками перелома*, *точками стыковки*.

Определим k новых переменных:

$$\omega_1 = x;$$

$$\omega_j = (x - \bar{x}_{j-1})_+ = \max(x - \bar{x}_{j-1},\ 0) = \begin{cases} x - \bar{x}_{j-1}, & x > \bar{x}_{j-1}; \\ 0; & x \leqslant \bar{x}_{j-1}, \end{cases}$$

где $j = 2, 3, \ldots, k$. Тогда линейный сплайн записывается следующим образом:

$$S_\Delta(x) = \beta_0 + \beta_1\omega_1 + \beta_2\omega_2 + \ldots + \beta_k\omega_k, \qquad (*)$$

где β_1 — коэффициент наклона сплайна над первым интервалом, а каждый следующий коэффициент β_j $(j = \overline{2, k})$ показывает изменение углового коэффициента при переходе

от $(j-1)$-го интервала к j-му. На рис. 2.1 приведен пример кусочно-линейного сплайна с тремя узлами $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$.

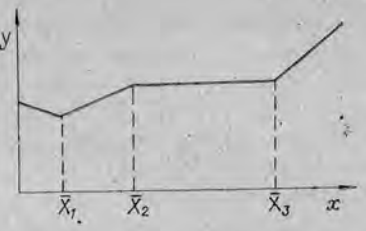


Рис. 2.1. Кусочно-линейный сплайн.

Использование линейного сплайна предполагает существование некоторой линейной зависимости, которая претерпевает изменения в узловых точках. Если же исходная зависимость не является линейной, то используются кусочно-простые сплайны (преимущественно кубические). При использовании нелинейных сплайнов в качестве аппроксимирующих функций можно представить функциональную зависимость в гораздо более «гладком» виде, чем это достигается с помощью линейного сплайна.

Использование сплайн-функций при конструировании моделей с переменной структурой представляется полезным и целесообразным в следующих ситуациях: а) исследуется одномерный случай (изменение одного показателя, обусловленное изменением другого показателя); б) рассматриваемая зависимость не имеет скачков и разрывов.

Следует, однако, отметить, что аппарат сплайн-функций значительно усложняется с увеличением размерности моделируемого показателя. Кроме того, при использовании сплайн-функций резко повышается значимость проверки гипотезы о наличии структурных изменений, поскольку построение сплайна осуществляется при фиксированных узлах сшивания. Определение же значимости структурных сдвигов для любых ситуаций в настоящее время еще затруднительно.

Вследствие этого методы построения сглаживающих сплайнов не нашли пока широкого применения в задачах многомерной статистики, но могут эффективно использоваться в задачах одномерного и двумерного анализа.

Перечисленные ограничения на использование аппарата сплайн-функций для моделирования многофакторных зависимостей присущи также и другим формам непрерывных моделей с переменной структурой.

Указанные недостатки непрерывных моделей, а также факт наличия разрывов (скачков) зависимостей в пространственных, а иногда и во временных выборках заставляют считать основным типом модели с переменной структурой ку-

сочно-простую модель, в которой отсутствует требование плавной стыковки зависимостей в точках перелома.

Такую модель можно назвать *дискретно-непрерывной* [Розин, 1973]. Ее отличие от непрерывных моделей заключается в том, что параметры модели определяются по информации для каждой характерной области, а не по всей совокупности данных.

Определим дискретно-непрерывную модель следующим образом.

При некоторых значениях аргумента непрерывность может нарушаться и график прерываться — функция имеет разрыв; те значения аргумента, при которых происходит разрыв, называются точками разрыва.

Если $g(x, \Theta)$ непрерывна во всех точках некоторого интервала, за исключением конечного числа отдельных его точек, в которых $g(x, \Theta)$ имеет конечные разрывы, то такая функция называется дискретно-непрерывной, ее график состоит из нескольких отрезков кривых линий. Здесь Θ — вектор параметров.

Приведем часто встречающиеся типы разрывов функций [Бронштейн, Семендяев, 1964]:

1) бесконечный разрыв, обращение функции в бесконечность (точки B, C, E, рис. 2.2);

2) конечный разрыв: при переходе x через значение x_1 функция перескакивает от одного конечного значения к

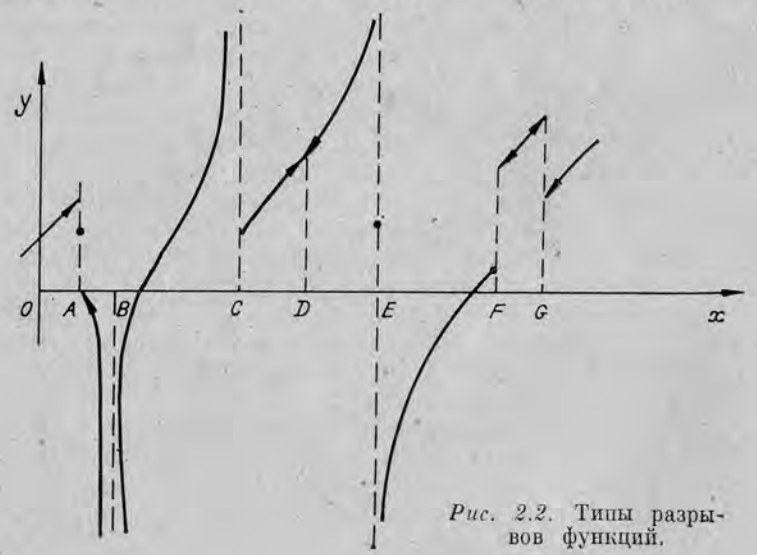


Рис. 2.2. Типы разрывов функций.

другому (точки A, F, G, рис. 2.2). Само значение $g(x, \Theta)$ при $x = x_1$ может быть не задано (точка G), может совпадать со значением $g(x_1 - 0)$ или $g(x_1 + 0)$ (точка G) и может быть отлично как от $g((x_1 - 0), \Theta)$ так и от $g((x_1 + 0), \Theta)$ (точка A);

3) устранимый разрыв, $\lim\limits_{x \to x_1} g(x, \Theta)$ существует: $g((x_1 - 0), \Theta) = g((x_1 + 0), \Theta)$, но при $x \neq x_1$ функция или не задана, или имеет значение $g(x_1, \Theta) = \lim g(x, \Theta)$ (точка D). Этот случай разрыва называется устранимым, так как, придавая $g(x_1, \Theta)$ значение $\lim g(x, \Theta)$, мы делаем функцию непрерывной.

В экономических приложениях наиболее часто встречаются случаи 2 и 3.

Дадим формулировку задачи построения кусочно-линейной модели (КЛМ) рассматриваемого (дискретно-непрерывного) типа.

Допустим, что все множество элементов пространства R_x разбито на k непересекающихся подмножеств (областей) $\{O_1, \ldots, O_k\}$ с помощью набора некоторых параметрических многомерных характеристических функций (гиперповерхностей) $\{\Psi(x; \beta_1), \ldots, \Psi(x; \beta_k)\}$. Обычно $k \ll m$, где m — объем выборки V. Пусть также имеется совокупность частных линейных моделей прогноза $\{g(x; \alpha_1), \ldots, g(x; \alpha_k)\}$. Тогда прогноз значения y по значению $x = (x_1, \ldots, x_n)$ по КЛМ имеет вид

$$y^* = g(x, \Theta) = \sum_{\nu=1}^{k} \Psi(x; \beta_\nu) \cdot g(x; \alpha_\nu),$$

где

$$\Psi(x; \beta_\nu) = \begin{cases} 1, & \text{если } x \in O_\nu; \\ 0, & \text{если } x \overline{\in} O_\nu; \end{cases}$$

$$\bigcup_{\nu=1}^{k} O_\nu = R_x; \; O_\nu \cap O_f \neq 0 \text{ при } \nu \neq f;$$

$$g(x; \alpha_\nu) = \sum_{j=1}^{n} \alpha_{\nu j} x_j + \alpha_{\nu 0}; \; \nu = \overline{1, k}. \qquad (2.1)$$

Иногда вместо (2.1) будем писать

$$g(x; \alpha_\nu) = \sum_{j=1}^{n_\nu} \alpha_{\nu j} x_j + \alpha_{\nu 0},$$

где n_ν — число информативных признаков $x_j \in x$, включенных в модель $g(x; \alpha_\nu)$ $(n_\nu \leqslant n)$.

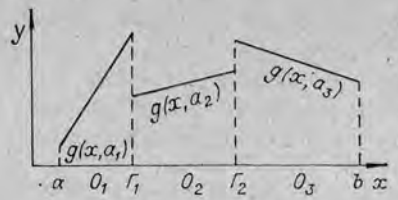


Рис. 2.3. Кусочно-линейная регрессия.

$R_x[a, b]$; $O_1 = [a, \Gamma_1]$; $O_2 = [\Gamma_1, \Gamma_2]$; $O_3 = [\Gamma_2, b]$; $\{\Gamma_1, \Gamma_2\}$ — границы областей.

Задача заключается в таком выборе значения вектора параметров $\Theta = \{k, \beta_\nu, \alpha_\nu\}$ на основе анализа обучающей выборки $V = \{(y_i, x_i);\ i = \overline{1, m}\}$, при котором достигает минимума некоторая оценка остаточной дисперсии прогноза Y

$$\bar{\delta} \to \min_{\Theta}.$$

На рис. 2.3 изображен пример одномерной КЛМ при $k = 3$. Однако часто функции $\{\Psi(x; \beta_\nu)\}$ определяются алгоритмом не непосредственно в виде параметрических разделяющих гиперповерхностей, а опосредованно, через промежуточную аппроксимацию истинной плотности распределения $f(x)$ конечной смесью [Исаенко, Урбах, 1976; Шлезингер, 1968].

$$f^*(x) = \sum_{\nu=1}^{k} \pi_\nu f(x; \beta_\nu);\ \sum_{\nu=1}^{k} \pi_\nu = 1.$$

Если $p(\nu/x) = \dfrac{\pi_\nu \cdot f(x; \beta_\nu)}{f^*(x)}$, то

$$\Psi(x; \beta_\nu) = \begin{cases} 1 \text{ при } p(\nu/x) = \max\limits_{q=\overline{1,k}} \{p(q/x)\}; \\ 0 \text{ — в противном случае.} \end{cases}$$

Обычно в качестве $\{f(x, \beta_\nu)\}$ используют нормальные плотности $\{N(\mu_\nu \Sigma_\nu)\}$.

В некоторых работах [Беккер и др., 1974; Исаенко, Урбах, 1976] учет вероятностной природы разбиения выборочной совокупности, заключающейся в неопределенности отнесения точек вблизи границ соседних классов, предлагается вести с помощью непосредственного определения значений оценок вероятностей отнесения объектов к классам. Заметим, что этот подход сложнее в смысле создания эффективных численных методов оптимизации Θ. Этот подход, по существу, представляет собой промежуточную форму (переходную форму) между детерминированными дискретно-непрерывными моделями в случае непересекающихся классов и непрерывными моделями с переменной структурой,

в которых классы не выделяются, а имеются переключения режимов.

Сравнительно просто аналитически задача построения КЛМ решается в одномерном случае $X = X_1$ либо многомерном, но тогда границы между областями $\{O_\nu\}$ устанавливаются по какому-то одному признаку $X_j \in X$. Эффективное решение одномерной задачи было дано Р. Беллманом на основе метода динамического программирования [Bellman, 1961]. Нас же будет интересовать лишь многомерный случай.

Остановимся на виде минимизируемого критерия $\hat{\delta}$. Истинное значение ошибки для модели $g(X; \Theta)$ определяется по формуле

$$\delta = \int_{R_{yx}} (y - g(x; \Theta))^2 \, dF(y, x).$$

Однако вычисление δ невозможно из-за незнания закона распределения вероятностей $F(y, x)$. Поэтому исследователи пользуются различными оценками величины δ.

Допустим, построенные функции $\{\Psi(x, \beta_\nu)\}$ осуществляют разбиение обучающей выборки V в пространстве R_{yx} на k подмножеств $\{V_\nu\}$ объемов $\{m_\nu\}$ согласно принципу: $x_i \in O_\nu \to (y_i, x_i) \in V_\nu$. Пусть также на основе каждого V_ν методом наименьших квадратов построена своя модель $g(X; \alpha_\nu)$.

В качестве оценки остаточной дисперсии КЛМ часто используется величина [Розин, 1973]

$$\delta_1 = \sum_{\nu=1}^{k} \frac{m_\nu}{m} \left[\frac{1}{m_\nu - n_\nu} \delta_{V_\nu} \right], \tag{2.2}$$

где

$$\delta_{V_\nu} = \sum_{i=1}^{m_\nu} (y_i - g(x_i; \alpha_\nu))^2. \tag{2.3}$$

В (2.3) суммирование идет по реализациям $(y_i, x_i) \in V_\nu$.

Однако величина δ_1 является заниженной оценкой δ, особенно если осуществляется одновременно оптимизация значений векторов β_ν и α_ν.

В работе В. Н. Вапника [1979] на основе теории равномерной сходимости частот к вероятностям получена следующая оценка:

$$\delta_2 = \sum_{\nu=1}^{k} \frac{\delta_{V_\nu}/m}{1 - \sqrt{\dfrac{n_\nu k_2 \left(\ln \dfrac{m}{n_\nu k_2} + 1 \right) + \ln C_n^{n_\nu} - \ln \eta}{m}}}, \tag{2.4}$$

где k_2 — априорно задаваемое, максимально допустимое, с точки зрения исследователя, а стало быть и анализируемое алгоритмом значение $k (k \leqslant k_2)$.

При некоторых предположениях с доверительной вероятностью $(1 - \eta)$ имеет место неравенство $\delta \leqslant \delta_2$.

Заметим, что величина δ_2 является обычно завышенной оценкой, и качество ее использования во многом зависит от значения η, априорное задание которого весьма затруднительно для пользователя.

Можно использовать и следующую эвристическую оценку:

$$\delta_3 = \sum_{v=1}^{k} \frac{m_v}{m} \left[\frac{1}{m_v - h_v - 1 - h_v(\rho(\sigma))} \delta_{V_v} \right], \qquad (2.5)$$

где $h_v(\rho(\sigma))$ — некоторая положительная функция от степени статистического разнообразия $\rho(\sigma)$ класса поверхностей σ, из которого были выбраны $\{\Psi(X; \beta_v), g(X, \alpha_v)\}$. Априорное задание $h_v(\cdot)$ субъективно и зависит от конкретного метода построения КЛМ.

Часто рекомендуется [Ивахненко и др., 1976] и такая просто вычисляемая оценка алгоритмического типа. Заранее некоторым образом исходная выборка V разбивается на обучающую подвыборку V^0 объемом m^0 и экзаменационную — V^e объемом m^e. Сначала выбирается наилучшая структура КЛМ путем построения различных $g(X; \Theta)$ из заданного класса на основе анализа V^0 с оцениванием их качества на V^e в виде

$$\delta_4 = \frac{1}{m^e} \sum_{(y_i, x_i) \in V^e} (y_i - g(x_i; \Theta))^2. \qquad (2.6)$$

После того как по минимуму δ_4 выбрана наилучшая структура $g(X; \Theta)$, ее параметры Θ могут быть уточнены на основе всей V.

В качестве $\hat{\delta}$ могут быть взяты и другие оценки аналитического или алгоритмического типа. Более подробно этот вопрос освещен в гл. 5. Для определенности же изложения методов синтеза регрессионных КЛМ будем полагать, что минимизируется одна из приведенных оценок $\hat{\delta}$. Структура рассматриваемых ниже алгоритмов инвариантна относительно широкого класса оценок.

Из анализа $\{\delta_1 - \delta_4\}$ следует, что минимизация $\hat{\delta}$ позволяет оптимизировать не только значения $\{\beta_v, \alpha_v\}$, но и величину k. Многие исследователи, чтобы четче определить оптимизационную проблему, решают задачу синтеза КЛМ при фиксированном k путем минимизации либо δ_1, либо

$$\delta_V = \sum_{v=1}^{k} \delta_{V_v}, \qquad (2.7)$$

либо еще какого-нибудь другого промежуточного критерия. При этом предполагается, что после нахождения своих оптимальных решений при различных k в заданном диапазоне $k_1 \leqslant k \leqslant k_2$ они (решения) будут оценены по некоторому иному, более надежному критерию $\hat{\delta}$ с целью окончательного выбора единственного наилучшего решения. Такая схема построения общего алгоритма также допустима и мы ей часто будем следовать.

Вне зависимости от того, какой тип кусочно-линейной модели строится и как отыскиваются оценки ее параметров, одним из основных и часто используемых элементов процесса построения модели выступает анализ типологической структуры совокупности и выделение однородных классов наблюдений. Поэтому в следующем параграфе мы остановимся на некоторых понятиях и наиболее употребительных классах алгоритмов классификации.

В качестве иллюстрации приведем наглядный пример, когда необходимость построения дискретно-непрерывной модели очевидна и вытекает из самого существа решаемой задачи.

При анализе расходования плановых фондов заработной платы необходимо учитывать уровень выполнения производственной программы. Особенно важно это при сдельно-премиальных и сдельно-прогрессивных системах оплаты труда. Обычно учет уровня выполнения производственной программы при анализе трудовых показателей осуществляется путем корректировки планового фонда заработной платы. Правильность этой корректировки зависит от точности коэффициентов пропорциональности (k_p), характеризующих соотношение темпов роста фонда заработной платы и уровня выполнения плана.

Величина k_p может быть получена на основе статистического анализа связи между уровнем выполнения плана и размером заработной платы. При определении формы уравнения связи необходимо учесть наличие разрыва функции в точке, соответствующей 100% выполнения производственного плана. Разрыв обусловливается построением систем премирования. Действующей системой премирования пре-

дусматривается выплата определенной премии за 100% выполнения плана и за каждый процент перевыполнения.

Наличие разрыва и разный темп изменения фондов заработной платы слева и справа от точки 100% выполнения плана определили необходимость построения по каждому цеху металлургического предприятия двух уравнений регрессии.

Обозначим:

y_1 и y_2 — размер заработной платы соответственно при невыполнении и перевыполнении плана;

x — процент выполнения плана;

a_1 и a_2 — постоянные составляющие заработной платы;

b_1 и b_2 — пропорциональные затраты заработной платы на каждый процент выполнения плана.

Тогда зависимость заработной платы от уровня выполнения плана может быть выражена формулами: при невыполнении плана $y_1 = a_1 + b_1x$; при перевыполнении плана $y_2 = a_2 + b_2x$.

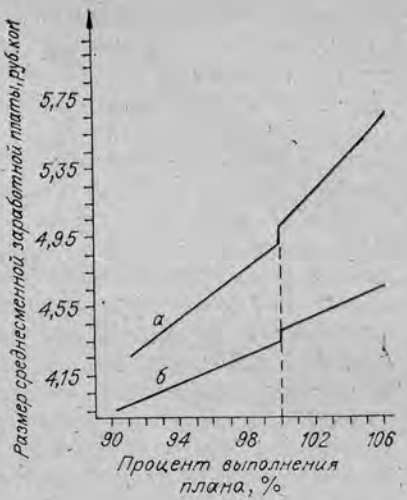
Величина коэффициента пропорциональности k_p вычисляется как отношение абсолютного темпа роста заработной платы b_1 и b_2 к уровню заработной платы при 100%-ном выполнении плана y_{100}, при этом необходимо брать y_{100} справа, что соответствует плановому фонду заработной платы.

Таблица 2.1

Зависимость заработной платы от уровня выполнения плана

Наименование цеха	Результаты статистического анализа			
	a, коп.	b, $\frac{\text{коп.}}{\%}$	y_{100}, справа, коп.	k_p, %
При невыполнении плана				
Мартеновский № 1	—158	6,5	500	1,30
Мартеновский № 2	452	0,14	512	0,03
Прокатный № 1	—	4,4	438	1,00
Прокатный № 2	378	0,2	457	0,04
Молотовый	392	0,03	442	0,07
Термокалибровочный	241	1,64	401	0,4
При перевыполнении плана				
Мартеновский № 1	—360	8,5	500	1,72
Мартеновский № 2	—248	7,6	512	1,49
Прокатный № 1	—192	6,3	438	1,48
Прокатный № 2	—343	8,0	457	1,75
Молотовый	12	4,3	442	0,96
Термокалибровочный	141	2,6	401	0,65

Рис. 2.4. Зависимость цехового фонда заработной платы от уровня выполнения месячного плана.
а — мартеновский цех № 1; *б* — прокатный цех № 1.



В табл. 2.1 приведены результаты статистического анализа по ряду цехов металлургического завода.

По двум цехам (мартеновский цех № 1, прокатный цех № 1) эти зависимости представлены графически (рис. 2.4). Данные табл. 2.1 и рис. 2.4 показывают, что по всем цехам имеются различия в характере зависимости при невыполнении и перевыполнении плана.

§ 2. МЕТОДЫ АНАЛИЗА ТИПОЛОГИЧЕСКОЙ СТРУКТУРЫ СОВОКУПНОСТИ

Цели и способы типологической группировки

Любая статистическая совокупность объединяет комплекс отличных друг от друга объектов или явлений, что позволяет делить ее на частные подсовокупности. Для выделения частных подсовокупностей используются методы группировки. В статистике различают следующие виды группировок по их целевому назначению: типологические, структурные, аналитические. При этом типологическая группировка занимает особое положение как основной прием выделения типов — групп однокачественных объектов или явлений. Типологическая группировка позволяет установить абсолютную и относительную важность рассматриваемых типов и позволяет объединять единичные данные по каждому типу для получения обобщенных статистических характеристик.

Типология может строиться для разных целей и по разным критериям. В наших исследованиях необходима такая типология, которая обеспечивает выделение групп наблюдений (объектов) с одинаковым (близким) характером зависимости экономических показателей от факторов производства, т. е. выделение из всей совокупности подобластей со стабильной структурой влияния.

Задача выделения типов из общей совокупности решается сравнительно просто только в тех случаях, если различия их (типов) очевидны и устойчивы и могут быть описаны одним или несколькими признаками. Однако на практике это бывает редко. Принадлежность группируемых явлений к общей совокупности приводит обычно к появлению у них некоторых общих особенностей, маскирующих различия между типами. Еще более усложняют группировку недостаточно четкое обособление отдельных типов друг от друга в реальной действительности, существование явлений промежуточного характера, множественность признаков описания объекта и ряд других обстоятельств. Поэтому задача проведения качественной типологической группировки совокупности весьма сложна и в ней, как, может быть, ни в какой другой задаче экономико-статистического исследования, должен соблюдаться принцип примата содержательного профессионального анализа над формальными построениями.

По способу формирования типологических групп следует различать две основные системы классификации: а) формирование *монотетических групп*, т. е. таких групп, все объекты которых имеют одинаковые значения классификационных признаков, и б) формирование *политетических групп*, т. е. таких, объекты которых могут иметь различные значения классификационных признаков.

Первый способ исторически более ранний. Типичный пример его реализации — комбинационная группировка, при которой формирование групп производится путем последовательного разбиения совокупности по одному признаку, полученных частей — по другому и т. д. Здесь строго соблюдается принцип иерархии групп. Второй способ классификации, когда группы формируются на основе близости объектов одновременно по большому числу признаков (многомерная классификация), получил широкое применение в последние десятилетия вместе с разработкой методов распознавания образов и появлением ЭВМ.

Попытаемся сравнить, сопоставить эти два метода типологической группировки.

Как и комбинационная группировка, методы формирования политетических групп могут выступать как инструмент решения одной из основных проблем экономико-статистического исследования — *классификации* явлений, объектов, процессов.

При использовании методов комбинационной группировки (или формировании монотетических групп) классификация осуществляется путем последовательного логического деления совокупности по отдельным признакам, сначала по одному, затем каждую из полученных групп по другому признаку и т. д. Сформированные в результате этого процесса логических делений группы характерны тем, что все элементы групп обладают одинаковыми значениями комплекса признаков группировки. Другими словами, *достаточным и необходимым* условием принадлежности единицы совокупности к данной группе является обладание соответствующими значениями комплекса группировочных признаков. В пределах набора признаков группировки элементы групп неразличимы.

В ходе развития научных исследований обнаружилось, что принципы чистой логики, лежащие в основе метода комбинационной группировки, нелегко применять к эмпирическому материалу. Опыт множества эмпирических исследований свидетельствует о существовании естественных типов явлений, каждый из которых объединяет индивидуальные явления, обладающие большим числом общих признаков и о том, что, как правило, для естественного типа невозможно найти жесткого определения, выраженного через небольшой набор совпадающих признаков. В некоторых случаях те или иные объекты можно без всяких сомнений отнести к определенному типу, несмотря на то, что у них отсутствуют или не совпадают несколько признаков из числа использованных при формировании группы. Все это обусловило необходимость разработки новых принципов группировки, отличных от классических.

Сущность этих новых принципов, лежащих в основе многомерной классификации, сводится к следующему. Классификация объектов производится не последовательно по отдельным признакам, а одновременно по большому числу признаков. Этот фиксированный набор признаков образует так называемое «признаковое пространство». Каждому признаку придается смысл координаты. Если исследователь оперирует n признаками, то любой объект рассматривается как точка в n-мерном признаковом пространстве и задача классификации сводится к выделению сгущений точек (объектов)

в этом пространстве. Выделение этих сгущений в разных алгоритмах многомерной классификации производится по-разному, но общим для всех них является то, что группы (таксоны, классы) формируются на основании «близости» объектов по большому числу признаков. При этом ни один из признаков, входящих в набор, не является необходимым или достаточным условием принадлежности объекта к данной группе, т. е. классификация происходит одновременно по всему комплексу признаков, описывающих объект.

Представляется, что подходы к формированию групп, используемые в многомерной классификации, лучше, чем комбинационные группировки, согласуются с ранее сформулированным положением о существовании естественных типов объектов, близких по комплексу признаков. Действительно, при использовании комбинационной группировки объект, отклоняющийся от нормы, характерной для группы, по одному единственному признаку набора, будет автоматически исключен из группы. Более того, если этот признак используется на первом шаге группировки, то объект может легко попасть в группу, очень далекую от той, с которой он в действительности имеет наибольшее сходство. Если использовать понятие признаковых пространств, то группы, получаемые при комбинационной группировке, представляют собой секторы такого пространства свойств. При этом границы между секторами-группами обычно параллельны осям признакового пространства. Осуществляя классификацию методом комбинационных группировок, исследователь зачастую искусственно «рубит» признаковые пространства, разрушая реально существующие в нем обособленно однородные классы жестко заданными интервалами признаков.

Этот основной недостаток делает комбинационные группировки не всегда эффективными для выделения типов объектов по комплексу признаков, так как с добавлением каждого нового признака опасность разрушения объективно существующих однородных групп возрастает. Следовательно, основное преимущество методов многомерной классификации заключается в том, что они позволяют с той или иной степенью приближения «нащупать» и выделить реально существующие в признаковом пространстве скопления точек — объектов, что связано с одновременной группировкой по большому числу признаков и использованием в качестве границ сложных поверхностей. Однако включение чересчур большого количества признаков в правило классификации также нежелательно.

Необходимо отметить, что если реализация методов комбинационной группировки (или ее модификаций) иногда возможна и без применения вычислительной техники, то методы многомерной классификации, как правило, требуют использования ЭВМ, в связи с чем их часто называют методами автоматической классификации. Между обычными методами группировки (комбинационная и ее модификации) и автоматической классификацией нет жесткой границы. Цели их совпадают, а основное различие состоит в том, что при использовании метода группировок исследователь имеет возможность априори сконструировать типы объектов на основе теоретического знания конструирующих признаков; при использовании методов автоматической классификации он лишь указывает направление поиска (заданием набора признаков, имеющих отношение к цели классификации), но отказывается на данном этапе исследования от самостоятельного формирования классов.

Выбор одной из указанных систем классификации во многом определяется характером признаков, составляющих описание объекта. В том случае, когда в описании объекта преобладают качественные признаки, их число не очень велико и априори известно, что они неравнозначны с точки зрения цели классификации (существует определенная иерархия признаков), целесообразнее ориентироваться на первый, последовательный способ классификации, на формирование монотетических групп. Правда, ограничение на его применение весьма жесткое — достаточный объем исходной совокупности (общее число объектов, по крайней мере, должно быть 60—70). Если же описание объекта включает большое количество примерно равнозначных признаков, особенно если эти признаки количественные и вопрос иерархии признаков и групп не столь важен, то следует ориентироваться на многомерную классификацию (формирование политетических групп). В отдельных случаях может оказаться полезной комбинация обоих указанных подходов. Например, предварительное разбиение совокупности на небольшое число монотетических подсовокупностей по содержательно наиболее важным признакам с последующим разбиением этих подсовокупностей на типы методами многомерной классификации.

Способы формирования монотетических групп

Наиболее широко для решения таких задач используется метод комбинационной группировки. Однако его практическое применение в исследованиях в тех случаях, когда

ставится задача получения достаточно однородных групп и число признаков — кандидатов для группировки превышает 3—4, наталкивается на серьезные трудности. Это происходит потому, что получающиеся группы малы по объему и не пригодны для статистических исследований.

Возникает проблема нахождения рационального способа построения иерархической системы монотетических группировок в условиях ограниченности совокупности и наличия довольно значительного числа признаков — кандидатов, предположительно вызывающих неоднородность совокупности. Таким способом может явиться модификация метода комбинационной группировки — *многошаговый метод последовательных разбиений совокупности* [Котюков, 1970; Розин, 1970].

Метод основан на анализе коэффициентов классификационности качественных признаков. Коэффициент классификационности $\gamma(x)$ характеризует способность признака различать отдельные элементы исследуемой выборочной совокупности. Его значения колеблются от нуля до единицы. Если все объекты совокупности имеют одно значение признака (одну его градацию), то признак не различает элементы совокупности и его коэффициент классификационности равен нулю. И наоборот, если коэффициент классификационности равен единице, то все объекты совокупности имеют разные значения признака.

Предположим, что совокупность m объектов распадается по градациям признака X на K групп $V_1, V_2, \ldots, V_k$. Введем величину $d_{ij}(x)$, которая равна нулю, если i-й и j-й объекты совокупности попадают в разные классы, и единице, если они попадают в один класс. В этом случае коэффициент классификационности признака принимает следующий вид:

$$\gamma(x) = \frac{\left[m^2 - \sum_{\nu=1}^{k} \sum_{i_\nu j_\nu} d_{i_\nu j_\nu}(x)\right]}{m(m-1)}, \qquad (2.8)$$

где i_ν, j_ν означают, что объекты принадлежат к классу V_ν. Если каждый класс состоит из идентичных объектов, что характерно для номинального признака, то формула для расчета коэффициента классификационности упрощается:

$$\gamma(x) = \frac{m^2 - \sum_{\nu=1}^{k} m_\nu^2}{m(m-1)},$$

где m_ν — число объектов в классе V_ν.

Коэффициент классификационности по своему содержанию и структуре близок к коэффициенту вариации качественных признаков $Q(x)$ [Мюллер, Шусслер, 1968], расчет значений которого основан на сопоставлении числа различных пар событий

$$Q(x) = \frac{\left[m^2 - \sum_{i=1}^{l} m_i^2\right] l}{(l-1)\, m^2}, \qquad (2.9)$$

где l — число градаций признака X;

m_i — число объектов, принимающих i-ю градацию признака ($i = 1, 2, \ldots, l$);

m — число объектов совокупности, $m = \sum_{i=1}^{l} m_i$.

Формальное различие между формулами расчета коэффициента $\gamma(x)$ и $Q(x)$ заключается в множителе

$$\frac{Q(x)}{\gamma(x)} = \frac{l(m-1)}{(l-1)\, m}.$$

Для статистических совокупностей с достаточным числом членов (30 и более) величина $\frac{m-1}{m}$ близка к единице и таким образом основное различие заключается в члене $\frac{l}{l-1}$. По мере возрастания числа градаций признака величина $\frac{l}{l-1}$ стремится к единице и величины коэффициентов $\gamma(x)$ и $Q(x)$ сближаются. Полностью они равны друг другу и равны единице в случае $m = l$. С содержательной точки зрения различие этих двух коэффициентов сводится к следующему: коэффициент вариации качественных признаков $Q(x)$ реагирует только на характер распределения объектов по градациям признака (классам). Он принимает максимальное значение (1) при равенстве частот классов, т. е. когда $m_i = m/l$, число классов на его значение влияния не оказывает.

Величина коэффициента классификационности $\gamma(x)$ зависит как от распределения объектов по градациям признака, так и от числа классов. Его значение тем больше, чем больше классов порождает разбиение совокупности по признаку x, и чем равномернее распределение объектов по полученным классам. Максимальное значение (1) $\gamma(x)$ принимает только в случае, когда $m = l$, т. е. когда каждый объект обладает своей отдельной градацией признака. Нулевое значение оба коэффициента принимают при одной градации

признака, т. е. тогда, когда вся совокупность образует по данному признаку один класс.

Нам представляется, что на стадии предварительного анализа вариации переменных предпочтение следует отдать коэффициенту классификационности $\gamma(x)$, так как он более полно характеризует структуру распределения совокупности по качественному признаку (либо количественному признаку, близкому по характеру к качественному). На стадии формирования монотетических групп одинаково эффективно могут применяться оба рассмотренных вида коэффициентов.

Получение малого значения коэффициента классификационности $\gamma(x)$ (близкого к нулю) свидетельствует о том, что данный признак разбивает совокупность на малое число классов и что объекты распределены по классам крайне неравномерно, т. е. большинство объектов обладает одной градацией изучаемого признака и лишь небольшое число имеет отличную градацию. В этом случае может быть принято несколько решений. Если позволяют содержательные соображения и связь признака с моделируемым показателем слаба, то признаки с малым значением $\gamma(x)$ вообще могут быть исключены из рассмотрения. Другой выход (тогда, когда нельзя просто отбросить этот признак) — исключить из совокупности те несколько объектов, которые обладают этим отличным, оригинальным значением признака и тем самым также исключить этот признак из набора. И наконец, если оба указанных выхода недопустимы, то признак, обладающий малым значением коэффициента классификационности, является кандидатом к исключению на последующих стадиях анализа. Что касается признаков с большим значением коэффициента классификационности, то им следует уделить особое внимание при анализе структуры выборочной совокупности и оценке степени ее однородности.

Используя коэффициент классификационности признака (или коэффициент вариации), можно предложить следующую процедуру формирования монотетических групп. Эта процедура наиболее наглядно представляется в случае дихотомических признаков.

По каждому признаку вычислить $\gamma_i(x)$ (или $Q_i(x)$) и разделить совокупность по тому признаку, который имеет максимальное значение вычисленной меры. Если таких признаков несколько, то выбор среди них производится по содержательному смыслу.

В результате получим два класса элементов выборочной совокупности. При этом признаки, сильно связанные с признаком, по которому произведено деление, на следующем

шаге исключаются из рассмотрения. После первого шага можно рассмотреть полученные группы как самостоятельные совокупности и повторить в них указанную процедуру, т. е. найти $\gamma_i^1(x)$ и $\gamma_i^0(x)$ (индексы 1 и 0 указывают класс) и разделить совокупность дальше по признаку с максимальным коэффициентом классификационности из набора оставшихся признаков.

Такое деление производится до тех пор, пока либо не будет достигнута желаемая степень однородности объектов, либо число элементов в группах не станет меньше заданного.

Рассмотрим пример использования многошагового метода последовательных разбиений.

При исследовании мартеновского производства [Экономико-статистическое..., 1977] было выделено девять существенных признаков — кандидатов для разбиения совокупности из 66 единиц на однородные классы. В их число вошли семь качественных (тип печей, вид футеровки, тип процесса, способ выплавки, вид применяемого топлива, тип завалочных средств, организация разливки) и два количественных с многовершинными распределениями (годовой объем выплавки стали и доля углеродистой стали в сортаменте).

Использование всех признаков — кандидатов для комбинационной группировки совокупности практически неприемлемо, так как приводит к формированию большого числа малочисленных групп.

Для решения задачи был использован метод последовательных разбиений. Формальным критерием выбора признаков выступал коэффициент качественной вариации $Q(x)$. Величины этого коэффициента для первого шага процесса последовательных разбиений приведены в табл. 2.2.

Таким образом, по величине коэффициента качественной вариации выделяются два кандидата на роль группировочного признака для первого шага процесса последовательных разбиений — тип процесса и организация разливки. Содержательный анализ заставляет отдать предпочтение первому из них.

Выбор признаков для разбиения на последующих шагах процесса производится таким же образом, за исключением того, что рассматриваются несколько подсовокупностей и для каждой из них подбирается свой группировочный признак.

Уже после первого шага процесса достигается существенное увеличение однородности частей. Группировка по типу процесса позволила для группы цехов на твердом чугуне (1_1) получить однородность по пяти признакам из девяти,

Таблица 2.2

Характеристики классификационных признаков

Номер признака	Наименование признака	Частоты градаций				Коэффициенты качественной вариации, $Q(x)$
		m_1	m_2	m_3	m_4	
1	Тип печей	65	1	—	—	0,06
2	Вид футеровки	65	1	—	—	0,06
3	Тип процесса	31	35	—	—	0,99
4	Способ выплавки	49	17	—	—	0,77
5	Вид применяемого топлива	19	35	5	7	0,83
6	Тип завалочных средств	14	52	—	—	0,67
7	Организация разливки	31	35	—	—	0,99
8	Годовой объем выплавки стали	52	14	—	—	0,66
9	Характеристика сортамента цеха	53	13	—	—	0,63

по группе 1_2 повышение однородности было меньшим. Здесь, помимо группировочного, существенно уменьшился коэффициент вариации по двум признакам — способу разливки и сортаменту.

После второго шага отдельные группы имели следующее число совпадающих признаков: $2_1 - 6$; $2_2 - 6$; $2_3 - 3$; $2_4 - 5$, т. е. для трех из них процесс разбиения можно было бы остановить. Численности групп на этом шаге примерно равны (16, 19, 17 и 14 цехов) и допускают применение статистических приемов анализа.

Следующий шаг (третий) главным образом необходим для группы 2_3 (цехов мощностью до 1,1 млн. т в год, работающих на жидком чугуне), по которой имеется значительное число несовпадающих признаков, в частности различие по такому важному признаку, как вид топлива (мазут и природный газ). Однако результатом этого шага является появление мелких групп, включающих по три-четыре цеха (3_2, 3_5, 3_7, 3_8, 3_{10}), что служит сигналом к остановке процесса разбиения.

Помимо рассмотренного метода последовательных разбиений, ориентированного в основном на качественные признаки, к подобным процедурам, но для количественных признаков относится метод, предложенный Г. С. Кильдишевым и Ю. И. Аболенцевым [1978] (метод $U(\rho^2)$-группировки).

Этот метод позволяет определять оптимальные границы интервалов группировки; дает статистические оценки однородности (неоднородности) исходной совокупности и получаемых на каждой итерации групп. Метод имеет, таким

образом, обоснованный критерий остановки процесса разбиений.

Алгоритм основан на идее проверки гипотезы о различии средних значений двух многомерных выборок (двух произвольных разбиений одной совокупности). Он построен на базе известной T^2-статистики Г. Хотеллинга (см. [Андерсон, 1963]) при условиях независимости признаков группировки, нормальности распределения выборки, ранжированности исходной совокупности и использования последовательного дихотомического деления выборки.

Критерий $U(\rho^2)$-группировки может быть представлен следующим образом:

$$U(\rho^2) = \frac{m-1}{m(m-l)\,l} \sum_{j=1}^{n} \frac{\left[(m-l)\sum_{i=1}^{l} x_{ij} - l\sum_{i=l+1}^{m} x_{ij}\right]^2}{\sum_{i=1}^{m} x_{ij}^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_{ij}\right)^2},$$

где m — объем выборочной совокупности;

l — номер объекта в ранжированной совокупности, по которому выборка делится на две произвольные части ($l = 1, 2, \ldots, m-1$);

n — число признаков;

x_{ij} — значение j-го признака на i-м объекте ($i = \overline{1, m}$; $j = \overline{1, n}$).

Алгоритм проверки однородности совокупности и разбиения ее на однородные группы включает следующие операции:

1. Совокупность ранжируется по значению любого из n исходных признаков (для нашего случая предпочтительнее по значению выходного показателя).

2. Для каждого возможного разбиения ранжированной совокупности на две подсовокупности (общее число разбиений для m объектов $m-1$) вычисляются значения $U(\rho^2)$.

3. Для проверки гипотезы об однородности вычисленные значения $U(\rho^2)$ сравниваются с табличным значением χ^2-распределения с m степенями свободы при выбранном уровне значимости.

Гипотеза об однородности принимается, если для всех $m-1$ разбиений справедливо неравенство $U(\rho^2) \leqslant \chi^2_{\alpha,n}$.

Если хотя бы для одного из $m-1$ последовательного разбиения это неравенство не выполняется, т. е. $U(\rho^2) > \chi^2_{\alpha,n}$, то совокупность не является однородной и переходят к формированию однородных групп (операция 4).

4. Отыскивается максимальное значение $U(\rho^2)$ и ранжированная совокупность разбивается на два класса по строке, соответствующей этому значению.

5. Каждая из полученных групп подвергается операциям 1—4.

Процесс последовательных делений продолжается до достижения полной однородности получаемых последовательными дихотомическими делениями групп.

Рассмотренный метод может использоваться для проверки однородности исходной совокупности; формирования однородных групп объектов; для объединения малочисленных групп с $U(\rho^2) \leqslant \chi^2_{\alpha,n}$ (объединение производится по минимальному значению критерия).

Предложенный метод может быть также применен для организации процесса последовательных разбиений по одному признаку. Для этого на первом шаге $U(\rho^2)$ вычисляется для каждого признака в отдельности. Разбиение производится по признаку и в точке с максимальным значением $U(\rho^2)$. Процесс таких последовательных делений продолжается до тех пор, пока не будет достигнут приемлемый уровень однородности.

С точки зрения построения моделей с переменной структурой большим достоинством метода является его способность к обнаружению скачкообразных изменений в моделируемой совокупности и установлению границ групп в районе резких изменений стабильного характера поведения признаков.

Несколько ограничивают область применения метода $U(\rho^2)$-группировки априорные предположения о независимости признаков и нормальности их распределения. Особенности метода (независимость исходных признаков и довольно быстрое уменьшение численности групп в результате последовательных дихотомических делений) заставляют при его использовании в процессе построения моделей с переменной структурой ориентироваться только на часть признаков набора или комбинировать его с методами многомерной классификации.

Способы формирования политетических групп (многомерная классификация)

Часто требования к качеству классификации заставляют использовать при построении классов не один признак (упорядоченную последовательность признаков), а полную систему информативных признаков. Например, при изучении производственно-хозяйственной деятельности предприятий отрасли исследователь сталкивается с большим набором характеристик, описывающих различные условия их функционирования (экономические, социальные, природно-климатические, технологические и т. д.).

Можно выделить несколько обстоятельств, обусловливающих необходимость использования набора группировочных признаков:

изменение смысла и значения выделенного группировочного признака в зависимости от взаимосвязанных с ним признаков;

недостаточность отдельного группировочного признака, как бы важен он ни был, для полной характеристики группируемых явлений;

явления с нулевым значением определенного группировочного признака сами могут различаться по другим дополнительным признакам.

Понятно, что типологические группировки по небольшому числу признаков в такой ситуации могут оказаться несопоставимыми. Это, в свою очередь, приводит к трудностям при интерпретации и использовании результатов классификации. Выходом из создавшегося положения является разработка методов классификации, позволяющих группировать объекты одновременно по большому набору признаков.

Классификация — это распределение явлений, «похожих» объектов по группам на основе определенных критериев. Всякая классификация относительна, она направлена на достижение определенной цели. Трудность формализации таких задач связана с тем, что цель классификации может быть недостаточно четко определена. Из сказанного следует, что сама по себе никакая вычислительная машина не может произвести «объективную» классификацию, поскольку все предъявленные объекты равнозначны. Непременным участником процесса классификации является некто (назовем его учителем), который может выразить цель классификации в формальном виде и передать ее машине (процесс обучения). В соответствии с этой целью вычислительная машина осуществляет классификацию и выдает решение, угодное учителю. Таким образом, поставить задачу классификации — это значит в строгих математических терминах сформулировать цель классификации.

Классом (таксоном, кластером) будем называть некоторое подмножество объектов, выделенное на основании некоторого правила. Это правило называется *решающим*.

Всякая «естественная» классификация основана на интуп-

тивном понятии «близости» или «сходства». Удачным считается такое разбиение выборочной совокупности V, при котором похожие объекты собираются в один класс, а непохожие разносятся по разным классам.

Различия между схемами и приемами решения задачи классификации зависят от формализации этих понятий на языке математики.

Введение понятия «близости» между элементами некоторого множества соответствует заданию в нем некоторой топологии. Однако задания одной топологии еще недостаточно для решения задачи классификации, так как одну и ту же выборку можно разбить на классы многими способами в зависимости от *цели классификации*.

Итак, поставить задачу классификации в самом общем виде — это задать *целевую функцию* классификации. Различные практические способы такого задания определяют разнообразие конкретных постановок задачи классификации.

Функция, которая указывает степень согласованности данной классификации с введенной топологией, называется мерой компактности, критерием качества группировок, функцией качества классификации.

Таким образом, целевую функцию можно определить, если заданы топология в пространстве описаний, мера компактности классификации.

Мера сходства. Топологию пространства описаний можно задать при помощи некоторой функции ρ, ставящей в соответствие каждой паре точек x_1, x_2 в n-мерном пространстве R_x вещественное число $\rho(x_1, x_2)$. В задачах классификации эту функцию называют мерой сходства или мерой близости между объектами. К мере сходства предъявляются следующие требования:

она должна быть положительной величиной, т. е. $\rho(x_1, x_2) \geqslant 0$;

должна обладать свойствами симметрии: $\rho(x_1, x_2) = \rho(x_2, x_1)$;

мера сходства объекта с самим собой должна быть максимальной: $\rho(x_1, x_1) \geqslant \rho(x_1, x_2)\ \forall x_1, x_2 \in R_x$.

Выделяют три типа мер сходства [Айвазян и др., 1974; Дорофеюк, 1971; Жидков, 1970; Загоруйко, 1981; Фукунага, 1979]: коэффициенты подобия, коэффициенты связи (корреляции), расстояние в метрическом пространстве.

Первые две меры можно назвать мерами близости: чем больше их величина, тем «ближе» объекты друг к другу. Обратное положение с показателями расстояния: чем больше их величина, тем больше «различие» между объектами и

Таблица 2.3

Коэффициенты подобия

Название	Формула	Комментарий
1	2	3
Общепринятого названия нет	$S_{ij} = \frac{p_{ij}}{n}$ $(0 \leqslant S_{ij} < 1)$	Применяется при предположении о «равноважности» нулевых и единичных признаков
Коэффициент Рао	$S_{ij} = \frac{n_{ij}^{(1,1)}}{n}$ $(0 \leqslant S_{ij} \leqslant 1)$	
Коэффициент Хаммана	$S_{ij} = \frac{p_{ij} - q_{ij}}{n}$	При $p_{ij} = q_{ij} \to S_{ij} = 0$
Коэффициент Роджерса и Танимото	$S_{ij} = \frac{n_{ij}^{(1,1)}}{n_i^{(1)} + n_j^{(1)} - n_{ij}^{(1,1)}}$ $(0 \leqslant S_{ij} \leqslant 1)$	
Коэффициент Джекарда	$S_{ij} = \frac{n_{ij}^{(1,1)}}{n_{ij}^{(1,1)} + q_{ij}}$ $0 \leqslant S_{ij} \leqslant 1$	
Коэффициент Дейка	$S_{ij} = \frac{2n_{ij}^{(1,1)}}{2n_{ij} + q_{ij}}$ $0 \leqslant S_{ij} \leqslant 1$	Придает вдвое больший вес совпадающим единичным признакам
Общепринятого названия нет	$S_{ij} = \frac{n_{ij}^{(1,1)}}{n_{ij}^{(1,1)} + 2q_{ij}}$ $0 \leqslant S_{ij} \leqslant 1$	Придает больший вес несовпадающим признакам

Окончание табл. 2.3

1	2	3
Коэффициент композиционного сходства	$S_{ij} = \dfrac{p_{ij}}{2m - p_{ij}} = \dfrac{p_{ij}}{m + q_{ij}}$ $0 \leqslant S_{ij} \leqslant 1$	Модификация коэффициента Роджерса — Танимото. Отличается включением совпадения нулевых признаков

Примечание. $n_{ij}^{(1,1)}$ — число совпадающих единичных признаков у обоих объектов; $n_{ij}^{(0,0)}$ — число совпадающих нулевых признаков у обоих объектов; $n_{ij}^{(1,0)}$ — число совпадающих единичных признаков у i-го и нулевых признаков у j-го объектов; $n_{ij}^{(0,1)}$ — число совпадающих нулевых признаков у i-го и единичных признаков у j-го объектов; $p_{ij} = n_{ij}^{(1,1)} + n_{ij}^{(0,0)}$; $q_{ij} = n_{ij}^{(1,0)} + n_{ij}^{(0,1)}$; n — общее число признаков, по которым осуществляется сравнение.

поэтому их содержательная трактовка — показатели различия объектов. Меры сходства могут определяться как между векторами — строками (объектами) матрицы исходных данных, так и между векторами — столбцами (признаками). Хотя по сути эти операции принципиально одинаковы, есть предпочтение в использовании определенных видов мер сходства для первого и второго случая (об этом подробнее будет сказано ниже).

Коэффициенты подобия S_{ij} используются для измерения степени близости между объектами, каждая из n компонент которых принимает значение 0 или 1, т. е. признак x_j — булевый ($j = 1, 2, \ldots, n$). Наиболее употребительные коэффициенты подобия приведены в табл. 2.3.

Выбор того или иного коэффициента зависит от содержательных соображений (например, что важнее — совпадение или несовпадение данного признака) и от вида представления исходных данных (дихотомические или многовариантные признаки) и т. п. Преимущественная область применения коэффициентов подобия — многомерная классификация объектов, описываемых набором качественных признаков.

Часто в качестве мер сходства используются *коэффициенты корреляции*. В случае количественных признаков эта мера имеет следующий вид:

$$r_{qj} = \frac{\sum_{i=1}^{m} (x_{iq} - \bar{x}_q)(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^{m} (x_{iq} - \bar{x}_q)^2 \sum_{i=1}^{m} (x_{ij} - \bar{x}_j)^2}},$$

где $\bar{x}_q, \bar{x}_j$ — средние значения признаков x_q и x_j соответственно.

Однако встречаются объекты, обладающие такими признаками, которые не поддаются точной количественной оценке, но позволяют сравнивать объекты друг с другом. В результате сравнения всю совокупность можно «ранжировать», приписав каждому объекту порядковый номер, соответствующий итогам сравнения с другими объектами. Если объекты обладают не одним, а двумя признаками подобного типа, то каждому объекту приписывают два порядковых номера в соответствии с «ранжировками» по обоим признакам.

Выборочными мерами связи признаков служат коэффициенты ранговой корреляции. Один из наиболее распространенных коэффициентов ранговой корреляции ρ_{ij} (для случая строгого упорядочения) введен Спирменом:

$$\rho_{ij} = 1 - 6 \frac{\sum_{l=1}^{n} (S_{li} - S_{lj})^2}{n^3 - n},$$

где S_{li} — ранг, приписываемый i-м экспертом l-му признаку;

S_{lj} — ранг, приписываемый j-м экспертом l-му признаку;

n — число признаков.

Во многих случаях роль меры сходства играет *функция расстояния*. С введением функции расстояния мы приходим к понятию *метрического пространства*. Это множество элементов какой угодно природы, в котором для любой пары элементов x_1 и x_2 определено вещественное число $d(x_1, x_2)$, называемое расстоянием и обладающее следующими свойствами:

$d(x_1, x_2) = 0$ тогда и только тогда, когда x_1 и x_2 совпадают;

для любых трех точек x_1, x_2, x_3 $d(x_1, x_2) \leqslant d(x_1, x_3) + d(x_3, x_2)$; $d(x_1, x_2) = d(x_2, x_1)$.

Приведем примеры часто используемых расстояний.

Хеммингово расстояние — для булевых признаков

$$d_{ij} = \sum_{l=1}^{n} |x_{il} - x_{jl}|.$$

Пространство признаков в данном случае представляет n-мерный двоичный куб, расстояние между вершинами которого равно числу несовпадающих разрядов соответствующих n-разрядных двоичных векторов описаний объектов x_i и x_j.

Евклидово расстояние

$$d_{ij} = \sqrt{\sum_{l=1}^{n} (x_{il} - x_{jl})^2},$$

где x_{il} — значение l-го признака на i-м объекте. Однако евклидово расстояние обладает существенным недостатком: в нем не учитывается возможная неравномерность осей пространства. При ненормированных осях возможен такой случай, когда два объекта, сходные по всем признакам кроме одного, по которому они сильно различаются, будут далеки друг от друга в евклидовом пространстве. Поэтому часто вводят *взвешенное евклидово расстояние*, где подбором весов ω_l пытаются нормировать оси пространства

$$d_{ij} = \sqrt{\sum_{l=1}^{n} \omega_l (x_{il} - x_{jl})^2}; \quad 0 \leqslant \omega_l \leqslant 1. \qquad (2.10)$$

Обычно величина ω_l обратно пропорциональна среднеквадратическому отклонению значения признака x_l.

Обобщением евклидова расстояния является *расстояние Махаланобиса*

$$d_{ij} = (x_i - x_j)\Sigma^{-1}(x_i - x_j)^T;$$
$$x_i = (x_{i1}, \ldots, x_{in}),$$

где Σ — ковариационная матрица связи признаков (размерности $n \times n$);

T — знак транспонирования.

В распространенном для задач классификации методе потенциальных функций используется *обобщенное расстояние*

$$d_{ij} = K(x_i, x_j) + K(x_j, x_i) - 2K(x_i, x_j),$$

где $K(x_i, x_j)$ — потенциальная функция вида $K(x_i, x_j) = \sum_{l=1}^{n} \lambda_l^2 \varphi_l(x_i)\varphi_l(x_j)$, $\varphi_l(x)$ $(l = 1, \ldots, m)$ — некоторая система функций на множестве, к которому принадлежат объекты x_i и x_j. Это расстояние позволяет производить разделение линейно-несепарабельных множеств, в то время как, например, евклидово расстояние позволяет осуществить лишь разделение линейно-разделимых множеств.

В ряде задач применяются и другие метрики. Выбор метрики определяется в основном структурой признакового пространства и целью классификации. Полностью формализовать этот этап решения задачи многомерной классификации пока не представляется возможным.

Для определения целевой функции классификации кроме топологии пространства описаний, определяемой, например, мерой сходства, нужно задать еще и *меру компактности* группировки.

Классификацию критериев группировки можно проводить по степени «организованности» пространства, в котором действует критерий.

Общее топологическое пространство. Оно является самым низкоорганизованным, поскольку в нем не определены ни метрика, ни операции над элементами. Класс критериев группировки в таких пространствах также сравнительно беден. Примером общего топологического пространства является случай, когда для каждой точки перечисляется совокупность других точек пространства, которые близки к данной. На практике этот способ может быть использован, если выборочное пространство состоит из конечного числа точек. Такая «дискретная» топология совершенно естественно выражается в терминах теории графов. Каждой точке выборочного пространства ставится в соответствие вершина обыкновенного графа, причем если пара точек является близкой в введенной топологии, то соответствующая пара вершин соединяется ребром. Таким образом, задание графа полностью определяет топологию дискретного пространства.

Если топология выборочного пространства представляется графом, то классификация заключается в разрезании графа на систему непересекающихся подграфов. Критерий классификации (или, что то же, мера компактности) сводится тогда к оценке качества разрезания. Описанный выше способ задания топологии пространства является самым примитивным. Значительно более «богатым» и общим является подход, основанный на количественном определении меры сходства классифицируемых объектов.

Общее псевдометрическое пространство. Введение в пространство описаний меры «сходства» определяет в нем «почти» метрику или псевдометрику, которая при выполнении неравенства треугольника превращается в «настоящую» метрику. Здесь при построении критериев качества классификации обычно исходят из нескольких концепций: рассеяния [Ямпольский, Макаров, 1971], энтропии [Жидков, 1970],

«диаметра» классов и «рассеяния между классами» [Wilks, 1960] и др.

Линейное метрическое пространство. Очень многие подходы к классификации основаны на концепции «среднего элемента классов», «эталона» или «центра», вычисляемого как средний элемент.

Все (или почти все) меры компактности в линейном пространстве основаны на концепции выборочной дисперсии как внутри классов, так и между классами (дисперсия средних). Критерий качества классификации в этих случаях обычно строится как некоторая функция ряда параметров, характеризующих межклассовую и внутриклассовую структуру совокупности. Подробный обзор используемых критериев можно найти в работах [Айвазян и др., 1974; Дорофеюк, 1971; Жидков, 1970; Загоруйко, 1972; Фукунага, 1979].

Часто используются комбинированные критерии качества классификации, например основанные на сочетании концепций «компактности» классов и максимальной «удаленности» их друг от друга и др.

В работе Н. Г. Загоруйко [1972] предпринята попытка построения критерия качества классификации, основанного на экспертном анализе разбиения совокупности. Сконструированный в этой работе критерий классификации является функцией четырех параметров: «близости» точек внутри класса, удаленности классов друг от друга, «одинаковости структуры» точек внутри класса и равномерности распределения точек по классам.

Разные критерии качества (меры компактности), вообще говоря, могут дать различающиеся результаты классификации. Поэтому для получения более строгих данных относительно структуры классов желательно использовать различные критерии, а результаты разбиения сравнить.

Модели автоматической классификации (кластер-анализ, таксономия). В общем случае методы кластерного анализа (таксономии, автоматической классификации) решают задачу оптимального — с точки зрения некоторого критерия T — разбиения исходного множества реализаций $V = \{x_i;\ i = \overline{1,\ m}\}$ на k подмножеств (кластеров, таксонов, классов) $\{V_\nu = \{x_{(\nu)i};\ i = \overline{1,\ m\nu}\};\ \nu = \overline{1,\ k}\}$. Обычно величина k априори неизвестна, известен лишь диапазон ее возможных значений $1 < k_1 \leqslant k \leqslant k_2 < m$.

Задачи кластерного анализа можно классифицировать разными способами. В данной работе рассматриваются два классификационных признака — критерий разбиения и реализующие его алгоритмы. По первому признаку ниже выделяются три основных типа моделей: дисперсионные, дискриминантные, задачи разделения смесей. Реализующие их алгоритмы можно разделить на детерминированные и статистические, хотя некоторые из них могут иметь как одну, так и другую интерпретацию. Статистические в свою очередь делятся на параметрические, непараметрические и аппроксимационные.

Детерминированные алгоритмы применяются в предположении четкой разделимости классов в признаковом пространстве, и для каждого объекта существует строго однозначное (с вероятностью, равной единице) отнесение к определенному классу.

Для статистических алгоритмов предполагается, что каждый объект с некоторой вероятностью может принадлежать к любому из полученных классов. Решение в данной ситуации обычно выдается в виде набора вероятностей отнесения объекта к классам.

Как правило, при использовании определенных моделей многомерной классификации предпочтение отдается конкретным видам алгоритмов. Однако жесткой связи здесь нет.

В настоящее время предложено много различных критериев оптимальности T и алгоритмов решения задачи таксономии, написаны обзорные работы [Айвазян и др., 1974; Дорофеюк, 1971; Загоруйко, 1972; Фукунага, 1979]. Здесь будут рассмотрены лишь некоторые из наиболее популярных постановок и методов решения задач.

Большое разнообразие используемых исследователями критериев качества разбиения T объясняется либо отсутствием четкого (формализованного) представления о дальнейшем использовании результатов таксономии, либо сложностью использования такого формализованного представления на данном этапе исследований. Иными словами, в задачах таксономии так или иначе отсутствует «суперцель» решения задачи [Загоруйко, 1972].

Учитывая сказанное, очевидно, следует использовать те критерии T, интерпретация которых проста и естественна для специалиста, для оптимизации которых имеются эффективные численные методы и которые приводят к результатам таксономии, также просто интерпретируемым.

Обычно требуется, чтобы в одну и ту же группу попадали объекты, «похожие» (по значению признаков X) друг на друга (реализации) и мало похожие на объекты других групп. Наиболее просто эти требования выражаются на языке так называемых *«дисперсионных» критериев* качества таксономии.

Очень часто в качестве критерия T используется средний по группам разброс реализаций в группе относительно ее «центра тяжести» (оценка математического ожидания группы)[1]. Под «центром тяжести» ν-й группы будем понимать $\bar{x}_{(\nu)} = \bar{x}_{1(\nu)}, \ldots, \bar{x}_{n(\nu)}$, где $\bar{x}_{j(\nu)} = \frac{1}{m(\nu)} \sum_{x_i \in V_\nu} x_{ij}$, $j = \overline{1, n}$. Заметим, что при минимизации такого критерия происходит одновременно и максимизация среднего разброса «центров тяжести» групп относительно общего «центра тяжести» всей совокупности V. Оптимизации этого критерия и будет посвящена первая часть изложения.

Введем дополнительно еще одно понятие. Под *эталоном* группы $V_{(\nu)}$ будем понимать тот из выборочных объектов группы $x_{\ni(\nu)} \in V_{(\nu)}$, суммарное расстояние от которого до других объектов группы минимально, $\sum_{x_i \in V_\nu} d_{i\ni(\nu)} \to \min$. Здесь $d_{i\ni(\nu)}$ — расстояние от $x_i \in V_\nu$ до $x_{\ni(\nu)} \in V_\nu$. Нетрудно показать, что эталон является наиболее близким к «центру тяжести» объектом из $V_{(\nu)}$. Использование же понятия эталона вместо «центра тяжести» позволяет создать более эффективные алгоритмы оптимизации таксономии. Кроме того, область приложений такой схемы кластерного анализа шире.

Ограничимся пока рассмотрением случая фиксированного значения числа групп k. Дисперсионный критерий качества таксономии имеет вид

$$T_1 = \sum_{\nu=1}^{k} \sum_{x_i \in V_\nu} d_{i\ni(\nu)}. \qquad (2.11)$$

Минимизация T_1 позволяет автоматически выбрать и эталон в каждой из k групп.

Пусть задана матрица $[d_{ij}]$ $i = \overline{1, m}$; $j = \overline{1, (m+1)}$, где подматрица $[d_{ij}]$, $i, j = \overline{1, m}$, есть совокупность вычисленных взаимных расстояний между объектами $x_i \in V$, и $\{d_{i,(m+1)} = 0;\ i = \overline{1, m}\}$. В случае детерминированного разбиения введем целочисленные переменные (неизвестные)

$$z_{ij} = \{0, 1\};\ i = \overline{1, m};\ j = \overline{1, (m+1)}.$$

Если $z_{ij} = 1$, то x_i и x_j принадлежат одной группе, в противном случае $z_{ij} = 0$. Тогда минимизация T_1 эквивалентна

[1] Мерой компактности классификации в этом случае чаще всего выступает функция расстояния.

минимизации

$$T_0 = \sum_{i=1}^{m} \sum_{j=1}^{m+1} d_{ij} z_{ij} \qquad (2.12)$$

при следующих ограничениях на искомые неизвестные $[z_{ij}]$:

$$\sum_{i=1}^{m} z_{ij} = 1;\ \ j = \overline{1, m};$$
$$\sum_{i=1}^{m} z_{i(m+1)} = k; \qquad (2.13)$$

$$\sum_{j=1}^{m} z_{ij} - m z_{i(m+1)} \leqslant 0;\ \ i = \overline{1, m}. \qquad (2.14)$$

Таким образом, мы пришли к решению задачи целочисленного линейного программирования специального вида [Котюков, 1972]. Ограничения (2.13) гарантируют отнесение каждого объекта $x_i \in V$ к одной и только к одной из групп. Число ненулевых строк в найденной после решения матрице $[z_{ij}]$ равно числу групп, это гарантирует неравенство (2.14). Если i-я строка $[z_{ij}]$ получилась ненулевой, то x_i является эталоном соответствующей группы, а единичные значения z_{ij} показывают ее состав. На рис. 2.5 изображена возможная реализация разбиения совокупности на три группы.

Для случая больших m может быть предложен следующий простой итерационный алгоритм [Шлезингер, 1968]:

1. Задается начальное разбиение $\{V_\nu;\ \nu = \overline{1, k}\}$ всех объектов на k групп.

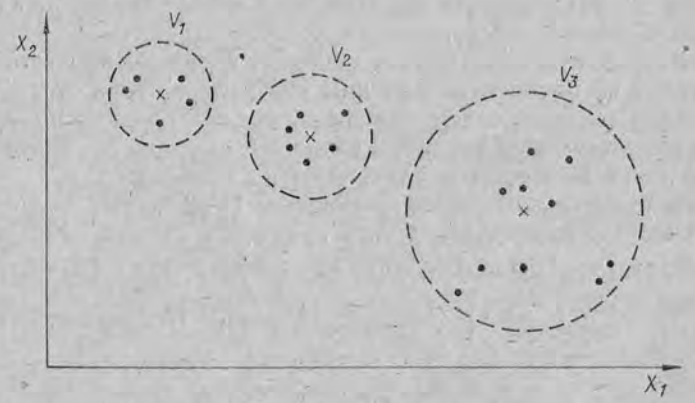


Рис. 2.5. Группировка объектов (крестиком отмечен эталонный объект группы).

2. В группах $\{V_\nu\}$ определяются свои эталонные объекты.

3. Распределяем заново все m исходных объектов по группам $\{V_\nu\}$ согласно принципу — x_i относится к V_ν, если расстояние от x_i до эталона V_ν минимально, т. е. по принципу ближайшего эталона.

4. Если объекты $\{x_i\}$ иначе распределились в группах, то пункты 2, 3 повторяются, в противном случае работа алгоритма заканчивается.

Нетрудно показать, что значение критерия T_1 на каждом шаге алгоритма лишь уменьшается. В силу ограничения $T_1 \geqslant 0$ мы имеем сходимость итерационного процесса за конечное число шагов. При этом достигается лишь локальный минимум — T_1, многое зависит от результатов 1-го шага алгоритма.

Изложенная оптимизационная процедура является модификацией алгоритма М. И. Шлезингера и алгоритма ISODATA [Ball, Hall, 1965]. Методология последнего явилась основой большинства современных итерационных методов кластерного анализа. К этому же классу методов можно отнести известный алгоритм ФОРЭЛЬ [Елкина, Загоруйко, 1966]. Получаемые таксоны являются попарно линейноразделимыми множествами.

Заметим, что рассмотренная схема задачи таксономии подходит и для решения задачи разбиения признаков $\{X_j\}$ на группы так, чтобы в одну и ту же группу попадали сильно коррелированные между собой признаки, а в разные группы — слабо коррелированные. Для этого можно положить $d_{ij} = 1 - |\rho_{ij}|$, где ρ_{ij} — коэффициент корреляции X_i и X_j, и воспользоваться изложенными алгоритмами оптимизации [Котюков, 1972].

Результаты кластерного анализа, согласно критерию T_1, зависят от исходного значения весового вектора ω в (2.10), которое задается специалистом до решения задачи на основе его представлений о степени важности каждого X_j. Приведенные ниже критерии и алгоритмы не зависят от ω.

В качестве критерия разбиения $\{V_\nu\}$ может быть взят *суммарный дискриминантный критерий Уилкса* для s ортогональных дискриминантных признаков $\left\{Z_l = \alpha_l X = \sum_{j=1}^{n} \alpha_{lj} X_j;\ l = \overline{1, s}\right\}$ $(1 \leqslant s \leqslant (k-1))$ [Андерсон, 1963],

$$T_2 = \sum_{l=1}^{s} W_l = \sum_{l=1}^{s} \frac{\alpha_l' \Sigma^b \alpha_l}{\alpha_l' \Sigma^\omega \alpha_l}, \qquad (2.15)$$

где Σ^b, Σ^ω — оценки матриц ковариаций признаков соответственно для межгруппового и внутригруппового рассеяния реализаций, определенные на основе разбиения $\{V_\nu\}$. Рассмотрим метод максимизации T_2 [Котюков, 1972]. Предварительно сделаем ряд замечаний. Суммирование в (2.15) определено в силу ортогональности $\{Z_l\}$. Критерий T_2 перепишем в виде

$$T_2 = \sum_{l=1}^{s} \frac{\sigma_l^{2(b)}}{\sigma_l^{2(\omega)}} = \sum_{l=1}^{s} \frac{\sigma_l^2 - \sum_{\nu=1}^{k} m_\nu \sigma_{l\nu}^{2(\omega)}}{\sum_{\nu=1}^{k} m_\nu \sigma_{l\nu}^{2(\omega)}}, \qquad (2.16)$$

где σ_l^2 — общий разброс (дисперсия) всех m реализаций на шкале признака $Z_l = \alpha_l' \cdot X$; $\sigma_{l_\nu}^{2(\omega)}$ — собственный разброс (дисперсия) m_ν реализаций группы V_ν по признаку Z_l. Величина σ_l^2 для Z_l не зависит от разбиения $\{V_\nu\}$ относительно оценки математического ожидания группы V_ν на шкале Z_l.

Итерационный алгоритм максимизации T_2 следующий.

1. Задается начальное разбиение $\{V_\nu\}$ всех m объектов.

2. Согласно численному методу (см. [Котюков, 1972]) определяется система ортогональных дискриминантных признаков $\{Z_l = \alpha_l' \cdot X;\ l = \overline{1, s}\}$. Вычисляется значение T_2. Исходные координаты $\{x_i\}$ всех m объектов пересчитываются в систему координат $\{Z_l\}$ — $\{Z_i = \{Z_{i_1}, \ldots, Z_{i_s}\};\ i = \overline{1, m}\}$.

3. Перераспределяем все m реализаций $\{Z_i\}$ по k группам согласно следующему принципу. Каждая реализация Z_i относится к той группе, присоединение к которой дает наибольшее увеличение T_2. В силу простоты выражения (2.16) такие расчеты выполнить нетрудно.

4. Выполнение пунктов 2, 3 повторяется до тех пор, пока распределение объектов по группам $\{V_\nu\}$ не стабилизируется.

При использовании рассмотренного алгоритма наблюдается монотонное возрастание значения T_2 и конечная сходимость к локальному максимуму критерия.

Данная процедура таксономии по своим результатам теоретически близка известным методам максимизации величины $(|\hat{\Sigma}^b|/|\hat{\Sigma}^b + \hat{\Sigma}^\omega|)$ [Фукунага, 1979], однако приведенный метод имеет бо́льшую численную устойчивость.

Следующая рассматриваемая здесь постановка задачи кластерного анализа является достаточно общей и связана с так называемой *задачей разделения смесей*. Каждой из k групп поставим в соответствие некоторую генеральную сово-

купность объектов, определяемую своей плотностью распределения $f(x; \Theta_\nu)$. Тогда исходная совокупность объектов $V = \{x_i;\ i = \overline{1, m}\}$ может рассматриваться как выборка из общей генеральной совокупности, определяемой следующей конечной смесью:

$$f(x; \Theta) = \sum_{\nu=1}^{k} \pi_\nu f(x; \Theta_\nu),$$

где π_ν — априорная вероятность ν-й группы и $\sum_{\nu=1}^{k} \pi_\nu = 1$ [Шлезингер, 1968]. В параметрическом случае задача заключается в определении по V оценок параметров $\{\pi_\nu, \Theta_\nu\}$. Для ее решения может быть использована следующая итерационная процедура метода максимума правдоподобия [Шлезингер, 1968]. Оценки $\{\widehat{\pi}_\nu, \widehat{\Theta}_\nu\}$ определим, исходя из максимума величины $L = \sum_{i=1}^{m} \ln\left(\sum_{\nu=1}^{k} \pi_\nu f(x_i; \Theta_\nu) \right)$. Считаем вид плотностей распределения $\{f(x;\ \theta_\nu)\}$ фиксированным.

1. Пусть на t-м шаге алгоритма получены оценки $\{\pi_\nu^t, \Theta_\nu^t\}$.

2. Определим апостериорную вероятность каждой группы для каждой реализации $x_i \in V$ на t-м шаге:

$$p^t(V_\nu/x_i) = \frac{\pi_\nu^t f\left(x_i; \Theta_\nu^t\right)}{\sum_{\nu=1}^{k} \pi_\nu f\left(x_i; \Theta_\nu^t\right)}; \quad \nu = \overline{1, k};\ i = \overline{1, m}.$$

3. Определим оценки для $\{\pi_\nu\}$ на $(t+1)$-м шаге:

$$\pi_\nu^{t+1} = \frac{1}{m} \sum_{i=1}^{m} p^t(V_\nu/x_i).$$

Обозначим $h_\nu^{t+1} = \sum_{i=1}^{m} p^t(V_\nu/x_i)$. Очевидно, что $\sum_{\nu=1}^{k} \pi_\nu^{t+1} = 1$.

4. Определим оценки параметров $\{\Theta_\nu\}$ на $(t+1)$-м шаге, максимизирующие величины $\left\{ \sum_{i=1}^{m} p^t(V_\nu/x_i) \ln f(x_i; \Theta_\nu) \right\}$. Например, для многомерного нормального закона они определяются так:

$$\mu_{(\nu)}^{t+1} = \frac{1}{h_\nu^{t+1}} \sum_{i=1}^{m} p^t(V_\nu/x_i)\, x_i;$$

$$\Sigma_\nu^{t+1} = \frac{1}{h_\nu^{t+1}} \sum_{i=1}^{m} p^t(V_\nu/x_i) \left(x_i - \mu_{(\nu)}^{t+1}\right)^t \left(x_i - \mu_{(\nu)}^{t+1}\right).$$

Далее выполнение пунктов 2, 3, 4 многократно повторяется до тех пор, пока не получим $\{\pi_\nu^{t+1}, \Theta_\nu^{t+1}\} \approx \{\pi_\nu^t, \Theta_\nu^t\}$.

Справедливо соотношение $L^{t+1} > L^t$ при $\{\Theta_\nu^{t+1}\} \neq \{\Theta_\nu^t\}$ [Шлезингер, 1968]. Это обеспечивает сходимость процедуры к локальному максимуму L при $t \to \infty$. Правда, возможны случаи отсутствия сходимости, когда функция L не ограничена сверху.

Н. Г. Загоруйко [1972] предлагает критерий и соответствующий алгоритм, позволяющий выделять группы, области которых могут иметь весьма сложную конфигурацию в R_x. В работе В. В. Александрова, Н. Д. Горского и А. О. Полякова [1978] изложена процедура кластерного анализа, использующая отображение на основе кривой Гильберта h-мерного пространства R_x в одномерное.

Использование параметрических методов в практике часто наталкивается на затруднения, связанные с тем, что заранее бывает невозможно установить вид функции распределения. В связи с этим в последние годы усиленно разрабатываются модели кластерного анализа, основанные на непараметрических оценках плотности распределения.

Рассмотрим вкратце идею непараметрических алгоритмов. Во-первых, для непараметрических процедур классификации не делается никаких предположений о виде распределения. Для них достаточно существования некоторой вероятностной меры в пространстве описаний и принадлежности ее к достаточно широкому классу вероятностных мер. Во-вторых, в основу всех непараметрических процедур положены различные непараметрические факты. Например, при аппроксимации функции плотности гистограммой используется тот факт, что с ростом объема выборки частота наступления некоторого события стремится к его вероятности.

Один из непараметрических алгоритмов классификации основан на аппроксимации распределения вероятностей некоторыми n-мерными гистограммами [Себестиан, 1965]. Предлагаются два метода аппроксимации. Первый состоит в том, что задается аналитический вид аппроксимирующей функции и для каждого объекта из обучающей выборки вычисляется соответствующее значение функции, которое и принимается равным значению искомой функции плотности в данной точке. Второй метод заключается в том, что запоминается достаточно большое количество значений функции плотности в различных точках. При поступлении очередного объекта его координаты в пространстве описаний сравниваются с ближайшей узловой точкой и затем либо значение

функции в узловой точке принимается за значение плотности в точке x, либо производится интерполяция по соседним точкам. В указанной работе также рассматривался вопрос о построении оптимальной гистограммы; о выборе числа узловых точек и размеров интервалов гистограмм.

Несколько иные идеи, чем Себестиан, использует в своем алгоритме Честер [Chester, 1969], который строит оценки плотности вероятности по точкам, расположенным внутри гиперсферы определенного радиуса в пространстве описаний. Радиус гиперсферы выбирается так, чтобы минимизировать средние потери от применения решающего правила.

В экономико-статистических исследованиях получил распространение алгоритм многомерной классификации, который позволяет получать однородные группы объектов на основе непараметрической оценки плотности вероятности [Беккер и др., 1974].

Можно предполагать, что, если построить совместную плотность распределения, то ее вершины будут соответствовать «центрам» групп, а поверхности, проходящие в долинах этого рельефа,— границам групп; число существенных максимумов функции плотности вероятности определяет число групп.

Пусть задана выборка, т. е. начальный признак $V^0 = \{x_1^0, \ldots, x_m^0\}$. Тогда оценка для функции плотности вероятности, построенная по данной выборке, имеет вид (см. [Parsen, 1962]):

$$f_m(x) = c(h)\frac{1}{m}\sum_{i=1}^{m}\exp\left\{-\frac{1}{2h^2}(x_i - x_i^0)\Sigma^{-1}(x_i - x_i^0)'\right\},$$

где $c(h) = 1/((h\sqrt{2\pi})^{n+m}\sqrt{|\Sigma|})$;
$n + m$ — размерность выборочного пространства;
Σ — выборочная ковариационная матрица;
h — некоторое положительное число.

Исходным для алгоритма является первоначальный признак $V^0 = \{x_1^0, \ldots, x_m^0\}$. Алгоритм представляет собой итерационную процедуру, определяющую по признаку $V^t = \{x_1^t, \ldots, x_m^t\}$ новый признак $V^{t+1} = \{x_1^{t+1}, \ldots, x_m^{t+1}\}$, по которому объекты более выраженно группируются в классы, чем по признаку V^t. Причем если по признаку V^0 каждая точка входит с весом $1/m$, то по признаку V^t каждая точка входит с некоторым весом p_i^t, где $0 \leqslant p_i^t \leqslant 1$; $i = \overline{1, m}$.

Одна итерация алгоритма состоит из шести этапов.

Этап 1. Вычисление ковариационной матрицы связи одномерных признаков, составляющих признак V^t, по формуле

$$\Sigma^{(t)} = \sum_{j=1}^{m} p_j^t\left(x_j^{(t)} - \sum_{i=1}^{m} p_i^{(t)}x_i^{(t)}\right)\left(x_j^{(t)} - \sum_{i=1}^{m} p_i^{(t)}x_i^{(t)}\right).$$

Этап 2. Вычисление меры связи между объектами по формуле

$$\rho(x_j^{(t)}, x_l^{(t)}) = \exp\left\{-\frac{1}{2h^2}(x_j^{(t)} - x_l^{(t)})\Lambda\Sigma^{(t)-1}\Lambda'(x_j^{(t)} - x_l^{(t)})'\right\},$$

где Λ — диагональная матрица весов одномерных признаков. Чем существеннее объекты группируются в классы по некоторому признаку, тем больше его вес.

Этап 3. Определение величин $\alpha_{il}^{(t)}$, имеющих смысл вероятности принадлежности объекта x_i к некоторому классу V_l, по формуле

$$\alpha_{il}^{(t)} = \frac{\rho(x_i^{(t)}, x_l^{(t)})}{\sum_{j=1}^{m}\rho(x_i^{(t)}; x_j^{(t)})}.$$

Этап 4. Вычисление $p_l^{(t+1)}$, $l = \overline{1, m}$, по формуле

$$p_l^{(t+1)} = \frac{1}{m}\sum_{i=1}^{m}\alpha_{il}^{(t)}.$$

Этап 5. Определение признака V^{t+1} по формуле

$$x_l^{(t+1)} = \sum_{i=1}^{m}\alpha_{il}^{(t)}x_i^{(t)}\Big/\sum_{i=1}^{m}\alpha_{il}^{(t)};\ l = \overline{1, m}.$$

Этап 6. Проверка критерия, по которому прекращается итеративный процесс, т. е.

$$\max(x_i^{(t)} - x_i^{(t+1)})\Sigma^{(t)-1}(x_i^{(t)} - x_i^{(t+1)})' \leqslant \varepsilon,$$

где $\varepsilon > 0$ — заданное число.

Аппроксимационные методы не требуют знания распределения выборки, а исходят лишь из самых общих предположений о виде распределения — непрерывности, разложимости в ряд и т. п.

Для нахождения экстремума функционала целевой функции классификации можно воспользоваться и процедурой типа стохастической аппроксимации [Невельсон, Хасьминский, 1972].

До сих пор мы не рассматривали важный вопрос о выборе наилучшего значения величины k (числа классов). Этот выбор можно сделать различными способами [Загоруйко, 1972; Исаенко, Урбах, 1976]. Если у нас есть наилучшие разбиения $\{V_\nu; \nu = \overline{1, k}\}$ при различных значениях k, то для их сравнения можно воспользоваться известным критерием дисперсионного анализа для проверки гипотезы об эквивалентности k нормальных совокупностей $\{N(\mu_\nu, \sum_\nu); \nu = \overline{1, k}\}$, где соответствующие оценки математических ожиданий и ковариационных матриц $\{\mu_\nu, \sum_\nu\}$ определены на основе результатов разбиения $\{V_\nu\}$ при данном значении k [Андерсон, 1963]. Таким критерием, как известно, может служить величина

$$W_k = \left(\prod_{\nu=1}^{k} |\sum_\nu|^{\frac{1}{2}(m_\nu - 1)}\right) \Big/ |\sum|^{\frac{1}{2}(m-k)},$$

распределение которой при гипотезе об эквивалентности всех $N(\mu_\nu, \sum_\nu)$ (нулевая гипотеза) изучено [Андерсон, 1963]. Здесь $\sum$, $\sum_\nu$ — оценки матриц ковариаций соответственно общей совокупности V и ν-й группы, а m_ν — объем подвыборки V_ν.

Обозначим через $W_k(\gamma)$ величину, определяющую критическую область для W_k при заданных уровне значимости γ и величинах $\{n, m, k\}$. Очевидно, чем меньше величина W_k, тем менее справедлива нулевая гипотеза. Для абсолютного сравнения результатов таксономии при различных значениях k предлагается использовать нормированную величину [Котюков, 1974]

$$W_k^* = \frac{W_k(\gamma) - W_k}{W_k(\gamma)}. \qquad (2.17)$$

В качестве окончательного решения задачи таксономии берется оптимальная группировка при том значении k, для которого величина W_k^* максимальна.

При вероятностном разбиении выборки иногда целесообразно использовать помимо строгих критериев и эвристические критерии качества разбиения. Один из них описан в работе [Экономико-статистическое..., 1977].

Критерий качества разбиения основан на анализе матрицы вероятностей принадлежности объектов к классам. На базе ее исследования делаются важные выводы относительно компактности классов, степени изолированности и т. п. Достаточно корректно определить приемлемые разбиения позволяют сравнительно простые характеристики: коэффициент диагональности D, равный отношению суммы диагональных элементов матрицы к сумме всех ее элементов, и средняя величина неопределенности H, рассчитанная как сумма неопределенностей всех строк данной матрицы. Исследование качества классификации по характеристикам D и H показало, что, во-первых, выбор приемлемой классификации по значению средней величины неопределенности H должен быть пороговым, поскольку H есть монотонно убывающая функция в зависимости от числа классов; во-вторых, значение величины коэффициента диагональности также зависит от числа классов, но в отличие от первой характеристики не является монотонной функцией.

§ 3. МЕТОДЫ ПОСТРОЕНИЯ КУСОЧНО-ЛИНЕЙНЫХ МОДЕЛЕЙ

Классификация методов

Кусочно-линейные модели (КЛМ) являются наиболее распространенным типом статических ЭСМ с переменной структурой. Напомним наиболее часто используемые формулировки КЛМ.

Допустим, что все множество элементов пространства R_x разбито на k непересекающихся подмножеств (областей) $\{O_1, \ldots, O_k\}$ с помощью набора некоторых параметрических многомерных характеристических функций (гиперповерхностей) $\{\Psi(X; \beta_1), \ldots, \Psi(X; \beta_k)\}$, а также определены частные линейные многофакторные регрессионные модели $\{g(X; \alpha_k), \ldots, g(X; \alpha_k)\}$. Тогда прогноз значения y по значению $x = (x_1, \ldots, x_n)$ по КЛМ имеет вид

$$y^* = g(x; \Theta) = \sum_{\nu=1}^{k} \Psi(x; \beta_\nu)\, g(x; \alpha_\nu), \qquad (2.18)$$

где

$$\Psi(x; \beta_\nu) = \begin{cases} 1, & \text{если } x \in O_\nu; \\ 0, & \text{если } x \overline{\in} O_\nu; \end{cases} \qquad (2.19)$$

$$g(x; \alpha_\nu) = \sum_{j=1}^{n} \alpha_{\nu j} x_j + \alpha_{\nu 0}. \qquad (2.20)$$

Построение КЛМ такого типа на основе анализа выборки $V = \{(y_i, x_i);\ i = \overline{1, m}\}$ заключается в выборе значения вектора параметров $\Theta = \{k, \beta_\nu, \alpha_\nu\}$, при котором достигает

минимума некоторая оценка остаточной дисперсии прогноза по КЛМ

$$\hat{\delta} \to \min_{\Theta}.$$

Иногда функции $\{\Psi(X;\ \beta_\nu)\}$ могут определяться не непосредственно в виде параметрических разделяющих гиперплоскостей, а опосредованно, через промежуточную аппроксимацию истинной плотности распределения $f(x)$ конечной смесью

$$f^*(x) = \sum_{\nu=1}^{k} \pi_\nu f(x;\ \beta_\nu), \tag{2.21}$$

где $\{f(x;\ \beta_\nu)\}$ — частные плотности, а $\{\pi_\nu\}$ — их априорные вероятности; $\sum_{\nu=1}^{k} \pi_\nu = 1$. Если апостериорная вероятность ν-й компоненты в точке x равна

$$p(\nu/x) = \frac{\pi_\nu(x;\ \beta_\nu)}{f^*(x)}, \tag{2.22}$$

то

$$\Psi(x;\ \beta_\nu) = \begin{cases} 1, & \text{при } p(\nu/x) = \max\limits_{q=\overline{1,k}} \{p(q/x)\}; \\ 0, & \text{в противном случае.} \end{cases} \tag{2.23}$$

Заметим, что прогноз по модели может быть и рандомизированным:

$$y^* = \sum_{\nu=1}^{k} p(\nu/x)\, g(x;\ \alpha_\nu). \tag{2.24}$$

Такая КЛМ в отличие от (2.18) непрерывна на границах областей, но обычно и менее точна.

Задача также заключается в таком выборе значений $\Theta = \{k,\ \pi_\nu,\ \beta_\nu;\ \alpha_\nu\}$ на основе анализа выборки V, при котором имеем

$$\hat{\delta} \to \min_{\Theta}.$$

Некоторые возможные виды оценок $\hat{\delta}$ остаточной дисперсии прогноза по КЛМ нами обсуждены в § 1 данной главы. Для определенности дальнейшего изложения будем полагать, что минимизируется одна из оценок (2.2), (2.4), (2.5), (2.6).

Конкретные методы синтеза КЛМ на основе анализа V различаются между собой лишь формулировкой класса разделяющих области $\{O_\nu\}$ границ и способом выбора из него наилучших $\{\Psi(X;\ \beta_\nu)\}$. Необходимо подчеркнуть, что при любой разумной постановке синтез КЛМ есть сложная многоэкстремальная оптимизационная задача.

Если для простоты рассмотрения значение k фиксировать, то схемы различных известных алгоритмов можно условно классифицировать на следующие три типа.

1. Сначала определяется в окончательном виде значение вектора $\{\beta_\nu\}$, т. е. осуществляется разбиение R_x на области $\{O_\nu\}$. При известном $\{\beta_\nu\}$ определяется наилучший вектор $\{\alpha_\nu\}$, т. е. методом наименьших квадратов для каждой области O_ν на основе соответствующих выборочных реализаций $V_\nu = \{(y_i,\ x_i) \in V;\ x_i \in O_\nu\}$ строится своя частная регрессионная модель $g(X,\ \alpha_\nu)$.

2. Реализуется итерационный процесс поочередного корректирования значений векторов $\{\beta_\nu\}$ и $\{\alpha_\nu\}$. Некоторое начальное значение $\{\beta_\nu\}$ задается.

3. Единый оптимизационный процесс поиска наилучших значений $\{\beta_\nu\}$ и $\{\alpha_\nu\}$ одновременно.

Исходя из общих представлений о возможностях решения оптимизационных задач, можно сделать следующие заключения относительно указанных схем алгоритмов:

методы первого типа сравнительно просты, но часто дают решение, с точки зрения минимизации погрешности $\hat{\delta}$ далеко не лучшее;

методы третьего типа, наоборот, наиболее точные, но и достаточно сложные;

методы второго типа занимают промежуточное положение.

В данном параграфе будут рассмотрены лишь алгоритмы, широко используемые при построении многофакторных КЛМ. Изложение методов проводится в порядке, соответствующем приведенной их классификации.

Методы предварительного кластерного анализа

Эти методы относятся к первому типу приведенной выше классификации и являются исторически одними из первых среди эффективных методов построения многофакторных КЛМ [Дорофеюк, и др., 1970; Розин, 1973; Розин и др., 1971].

Рассмотрим случай фиксированного k.

Построение КЛМ (2.18) осуществляется в два этапа. На первом этапе исходные реализации в пространстве R_x разбиваются на k подмножеств $\{V_\nu;\ \nu = \overline{1,\ k}\}$ с помощью того или иного метода кластерного анализа. Здесь $V_\nu = \{(y_i,$

$x_i) \in V_\nu; x_i \in O_\nu\}$. При этом целесообразно использовать такие методы таксономии, которые кроме чистого разбиения реализаций из V на группы разбивают одновременно и все R_x, осуществляя характеризацию каждого подмножества V_ν в R_x, т. е. ставя ему в соответствие определенную математическую модель, из которой непосредственно следует вид характеристической функции $\Psi(X; \beta_\nu)$.

Например, при минимизации критерия таксономии T_1 (2.11), характеризующего средний разброс реализаций групп относительно своих эталонов, одновременно осуществляется и выбор эталонов в группах, которые и являются их математической моделью. Тогда

$$\Psi(X; \beta_\nu) = \begin{cases} 1, \text{ если } d_{x\ni(\nu)} = \min\limits_{q=\overline{1,k}} \{d_{x\ni(q)}\}; \\ 0, \text{ в противном случае.} \end{cases}$$

Здесь $d_{x\ni(q)}$ — расстояние между точкой x и эталоном q-й группы. Вектор β_ν есть совокупность коэффициентов в формуле расстояния.

Если же метод кластерного анализа осуществляет аппроксимацию плотности $f(x)$ смесью k нормальных плотностей $\{N(\mu_\nu, \Sigma_\nu)\}$, то вид каждой $\Psi(X; \beta_\nu)$ определяется согласно (2.23), где $\beta_\nu = \{\pi_\nu, \mu_\nu, \Sigma_\nu\}$.

Наличие $\{\Psi(X; \beta_\nu)\}$ позволяет группировать не только выборочные реализации V, но и любые из R_x, т. е. иметь разбиение $\{O_\nu\}$.

На втором этапе методом наименьших квадратов для каждой области O_ν строится линейная модель $g(X; \alpha_\nu)$ на основе выборки V_ν.

Для построенной КЛМ определяется значение той или иной оценки критерия $\hat{\delta}$.

Так как обычно величина k априори неизвестна, а известен лишь диапазон ее допустимых, с точки зрения специалистов, значений $k_1 \leqslant k \leqslant k_2$, то подобные построения выполняются для всех k в этом диапазоне. Окончательно выбирается та КЛМ, для которой значение $\hat{\delta}$ минимально.

Сделаем ряд замечаний.

1. На первом этапе может использоваться метод таксономии, осуществляющий одновременно с группировкой объектов и поиск для описания каждой группы своего наилучшего признакового подмножества $X^* \subset X$ [Загоруйко, 1972]. На втором этапе также целесообразно использовать методы включения лишь своих информативных $\{X_l\}$ признаков в каждую модель $g(X; \alpha_\nu)$.

2. При аппроксимации $f(x)$ смесью $N(\mu_\nu, \Sigma_\nu)$ необходимо помнить следующее. С одной стороны, желательно определять оценки «полных» матриц ковариаций $\{\Sigma_\nu\}$, а не диагональных, как это часто делают в задачах чисто кластерного анализа — в противном случае на первом этапе будет утерян «эффект линейных корреляций» признаков X, необходимый для построения $\{g(X; \alpha_\nu)\}$. С другой стороны, при этом возрастает сложность, а в некоторых случаях и ухудшается численная устойчивость решения задачи таксономии. Все зависит от размерности n пространства R_x и свойств выборки V.

3. Методы данной группы дают хорошие результаты, обычно совпадающие с результатами других методов, при решении тех прикладных задач, для которых характерно наличие в R_x «компактных» и «изолированных» групп. Обычно такая группировка выборочных реализаций не является следствием действия случайного механизма и каждой группе соответствует свой статистический механизм взаимодействия признаков $\{X_j\}$. Если же компактность и изолированность групп отсутствует (в этом случае критерий типа W_k^* (2.17) не имеет ярко выраженного максимума) — а класс таких прикладных задач весьма широк,— то результат решения задачи таксономии является случайным по отношению к оптимизации $\hat{\delta}$ на втором этапе. В этих случаях желательно использовать другие методы построения КЛМ.

Недостатком этого класса методов является отсутствие непосредственной связи критерия качества решения задачи таксономии с общим критерием $\hat{\delta}$.

Итерационные методы

Под итерационными методами, относящимися к алгоритмам второго типа, здесь мы будем понимать методы поочередного корректирования векторов $\{\beta_\nu\}$ и $\{\alpha_\nu\}$ с помощью процедур кластерного и регрессионного анализов. Алгоритмы же итерационного типа, связанные с корректировкой значений $\{\beta_\nu, \alpha_\nu\}$, в процессе последовательного поступления данных $\{(y_i; x_i)\}$, нами вынесены в особый класс так называемых адаптивных процедур и будут рассмотрены позже.

Рассмотрим первый метод итерационного типа [Касавин, 1972]. На первом шаге задается некоторое начальное разбиение $\{V_\nu; \nu = \overline{1, k}\}$, например, с помощью метода кластерного анализа.

Итерационная процедура в общем виде выглядит так. Имея разбиение на t-м шаге $\{V_\nu\}$, определяем методом регрессионного анализа параметры каждой частной модели $g(X; \alpha_\nu^t)$ на основе соответствующих подвыборок $\{V_\nu^t\}$. Имея $g(X; \alpha_\nu^t)$, на $(t+1)$-м шаге перераспределяем все m реализаций (y_i, x_i) по группам $\{V_\nu^{t+1}; \nu = \overline{1, k}\}$ согласно следующему принципу:

$$(y_i, x_i) \in V_\nu^{t+1}, \text{ если } (y_i - g(X_i; \alpha_\nu^t))^2 = \min_{q=\overline{1,k}} \{(y_i - g(X_i; \alpha_q^t))^2\}; \quad i = \overline{1, m}.$$

Иными словами, точка относится к ν-й группе, если прогноз по $g(X; \alpha_\nu)$ значения y наилучший. На основе $\{V_\nu^{t+1}\}$ заново определяются $g(X; \alpha_\nu^{t+1})$, и так далее до тех пор, пока реализации не перестанут перераспределяться по группам. На каждом шаге имеем лишь уменьшение эмпирической ошибки δ_V (2.7):

$$\delta_V^{t+1} < \delta_V^t, \text{ если } \{V_\nu^{t+1}\} \neq \{V_\nu^t\}.$$

На третьем шаге выполняется окончательная корректировка разбиений $\{V_\nu\}$ и частных моделей. Дело в том, что совокупности подмножеств $\{V_\nu\}$ оказались не охарактеризованы математически, т. е. им пока нельзя сопоставить систему характеристических функций $\{\Psi(X; \beta_\nu)\}$. Более того, структура областей подмножеств $\{V_\nu\}$ в R_x может быть весьма сложной и «разрывной». Корректировка выполняется следующим образом. Сопоставим каждому подмножеству V_ν свою компоненту смеси $N(\mu_\nu, \Sigma_\nu)$, $\pi_\nu = m_\nu/m$ (m_ν — объем подвыборки V_ν) или свой эталон (центр тяжести), параметры которых вычисляются на основе V_ν соответствующим образом. Характеризация $\{V_\nu\}$ позволяет построить систему функций $\{\Psi(X; \beta_\nu)\}$, например, согласно (2.23). Наличие $\{\Psi(X; \beta_\nu)\}$ позволяет окончательно разбить исходные реализации $\{(y_i, x_i)\}$ на k групп $\{V_\nu\}$. На основе результатов корректировки разбиения выборки $\{V_\nu\}$ окончательно определяются параметры частных моделей $g(X; \alpha_\nu)$.

Окончательная корректировка может сильно изменить результаты итерационной процедуры. Поэтому трудно утверждать что-либо определенно относительно данного метода.

Рассмотрим второй алгоритм итерационного типа [Котюков, 1972], не обладающий недостатком предыдущего метода.

Также предполагаем, что на первом шаге задано некоторое разбиение $V = \{V_\nu; \nu = \overline{1, k}\}$. Каждое V_ν охарактеризуем с помощью компоненты смеси $N(\mu_\nu, \Sigma_\nu)$, $\pi_\nu = m_\nu/m$. Оценки $\{\mu_\nu, \Sigma_\nu\}$ вычисляются стандартным образом. Заметим, что приводимые ниже рассуждения можно провести и для случая, когда каждая группа охарактеризована, например, своим эталоном.

Пусть на t-м шаге итерационного процесса мы имеем $\{\pi_\nu^t, N(\mu_\nu^t, \Sigma_\nu^t)\}$. Соответствующие функции $\{\Psi(X; \beta_\nu^t)\}$, определенные согласно (2.23), порождают разбиение исходной выборки $\{V_\nu^t\}$. На основе $\{V_\nu^t\}$ построены $g(X; \alpha_\nu^t)$ и определена ошибка прогноза δ^t.

Тогда на $(t+1)$-м шаге осуществляется следующая процедура. Допустим, для (y_i, x_i) имеют место соотношения

$$(y_i, x_i) \in V_\nu^t; \; p^t(q/x_i) = \max_{\substack{l=\overline{1,k} \\ l \neq \nu}} \{p^t(l/x_i)\};$$

$$(y_i - g(x_i; \alpha_\nu^t))^2 > (y_i - g(x_i; \alpha_q^t))^2.$$

Тогда для $(y_i; x_i)$ выполняются следующие действия:

(y_i, x_i) удаляется из V_ν^t и присоединяется к V_q^t; рекуррентно пересчитываются параметры $\{\pi_\nu^t, \mu_\nu^t, \Sigma_\nu^t, \pi_q^t, \mu_q^t, \Sigma_q^t\}$ [Невельсон, Хасьминский, 1972];

при новых $\{N(\mu_\nu^t, \Sigma_\nu^t), \pi_\nu^t \; \nu = \overline{1, k}\}$, согласно (2.23), все m точек заново распределяются по k группам — $\{V_\nu^t; \nu = \overline{1, k}\}$;

для реализаций $\{V_\nu^t\}$ относительно $\{g(X; \alpha_\nu^t)\}$ подсчитывается оценка ошибки δ^t. Если новое значение δ^t меньше предыдущего, то результаты указанных действий принимаются за окончательные, в противном случае присоединение (y_i, x_i)-й точки никакой коррекции не вызывает.

Аналогичные действия на $(t+1)$-м шаге выполняются для всех $(y_i, x_i) \in V$, удовлетворяющих вышеприведенным условиям. После этого получившиеся $V_\nu^t, N(\mu_\nu^t, \Sigma_\nu^t), \pi_\nu^t$ принимаются в качестве $\{V_\nu^{t+1}, N(\mu_\nu^{t+1}, \Sigma_\nu^{t+1}) \pi_\nu^{t+1}\}$. На основе $\{V_\nu^{t+1}\}$ определяются $g(X, \alpha_\nu^{t+1})$ и находится δ^{t+1}. Если $\{V_\nu^{t+1} = V_\nu^t, \forall_\nu\}$, то на этом работа алгоритма заканчивается, в противном случае аналогично выполняется $(t+2)$-й шаг.

Нетрудно показать, что если $\{V_\nu^{t+1}\} \neq \{V_\nu^t\}$, то имеет место $\delta^{t+1} < \delta^t$. Таким образом, в силу ограничения $\delta \geqslant 0$

мы имеем сходимость к локальному минимуму δ. Качество решения зависит от $\{V_\nu^1\}$.

Метод последовательного укрупнения групп

Данный метод [Дорофеюк и др., 1976] также может быть отнесен к алгоритмам второго типа, хотя и имеет особую структуру.

На первом этапе одним из методов кластерного анализа осуществляется разбиение исходных объектов на некоторое достаточно большое количество групп $\{V_\nu;\ \nu = \overline{1,\ c}\}$, где $k_1 \leqslant k \leqslant k_2 \leqslant c < m$.

На втором шаге каждой паре групп $(V_\nu,\ Vq)$ поставим в соответствие некоторую меру близости в R_x:

$$d_{\nu q} = \frac{1}{m_\nu m_q} \sum_{x_i \in V_\nu} \sum_{x_j \in V_q} K(x_i x_j),$$

где $K(x_i,\ x_j)$ — некоторая мера близости точек x_i и x_j в R_x, например «потенциальная функция».

На третьем шаге каждой из c групп V_ν поставим в соответствие частную модель $g(X;\ \alpha_\nu)$ и определим ее ошибку δ_ν.

После этого начинается процесс последовательного укрупнения групп.

Находится пара самых «близких» групп $(V_\nu,\ V_q)$, для которых величина $d_{\nu q}$ максимальна. Проверяется гипотеза о статистической неразличимости моделей $g(X;\ \alpha_\nu)$ и $g(X;\ \alpha_q)$. Для этого вычисляется статистика

$$h = \frac{\left[\sum_{x_i \in V_q} (y_i - g(x_i;\ \alpha_\nu)) + \sum_{x_i \in V_\nu} (y_i - g(x_i;\ \alpha_q))^2 - \delta_{V_\nu} \delta_{V_q}\right] n}{(\delta_{V_\nu} + \delta_{V_q})/(m_\nu + m_q - 2n)},$$

которая сравнивается с h_0, определяемом при $\{n,\ (m_\nu + m_q - 2n)\}$ степенях свободы [Дорофеюк и др., 1976]. Если $h \leqslant h_0$, то гипотеза считается допустимой и происходит объединение V_ν и V_q в одну группу, для которой определяются $g(X;\ \alpha)$, δ, а также мера близости d со всеми остальными группами. Если же $h > h_0$, то объединения не происходит.

В любом случае процедура объединения групп далее продолжается аналогичным образом до тех пор, пока не найдется ни одной пары групп, объединение которых допустимо.

После окончания процедуры объединения на основе вычисленных для групп оценок $\{f(x;\ \Theta_\nu),\ \pi_\nu\}$ согласно (2.23)

корректируются подмножества $\{V_\nu\}$, по которым, в свою очередь, окончательно определяются $g(X;\ \alpha_\nu)$.

Было проведено сравнение данного метода с алгоритмом первого типа на реальных данных ($m = 174$, $n = 4$, $k = 4$). Результаты сравнения показали преимущества: а) рассмотренного метода перед методом предварительного кластерного класса (примерно в 1,5 раза по величине δ_V) и б) обоих КЛМ перед общей линейной моделью (приблизительно в 2,5—3 раза).

Необходимо отметить, что построение на третьем этапе моделей $g(X;\ \alpha_\nu)$ для каждой из большого количества начальных групп $\{V_\nu;\ \nu = \overline{1,\ c}\}$ выдвигает требования данного алгоритма к сравнительно большому объему исходных данных.

Вариационный метод

В работе В. И. Опойцева [1970] приведены необходимые и достаточные условия оптимальности КЛМ $g(X;\ \Theta) \equiv \equiv \{\Psi(X;\ \beta_\nu),\ g(X;\ \alpha_\nu)\}$ типа (2.18). Рассмотрим их. Пусть каждая пара областей $\{Q_\nu,\ Q_q\}$ разделяется гиперплоскостью

$$h(X;\ \beta_{V_q}) = \sum_{j=1}^{n} \beta_{\nu q j} X_j + \beta_{\nu q 0} \quad \text{так, что}$$

$$x \in O_\nu, \text{ если } h(x;\ \beta_{\nu q}) \leqslant 0;$$

$$x \in O_q, \text{ если } h(x;\ \beta_{\nu q}) > 0.$$

Заметим, что $h(X;\ \beta_{q\nu}) = -h(X;\ \beta_{\nu q})$. Тогда характеристическая функция области O_ν равна

$$\Psi(X;\beta_\nu) = \prod_{q=1;\ q \neq \nu}^{k} \operatorname{sign}(h(X;\beta_{\nu q})),$$

где

$$\operatorname{sign}(a) = \begin{cases} 1 \text{ при } a \leqslant 0; \\ 0 \text{ при } a > 0. \end{cases}$$

Здесь $\nu = \overline{1,\ (k-1)}$. Кроме того,

$$\Psi(X;\beta_k) = 1 - \sum_{\nu=1}^{k-1} \Psi(X;\beta_\nu).$$

Суммарная эмпирическая ошибка КЛМ $g(X;\ \Theta)$ равна

$$\delta_V = \sum_{i=1}^{m} (y_i - g(x_i;\ \Theta))^2.$$

Необходимые условия оптимальности $g(X;\ \Theta)$ получаются, если приравнять частные производные δ_V нулю:

$$\frac{\partial \delta_V}{\partial \alpha_\nu} = 0;\ \frac{\partial \delta_V}{\partial \beta_{\nu q}} = 0.$$

Согласно первому условию,

$$\Sigma_\nu \alpha_\nu = \delta_{(\nu)xy};\ \alpha_{\nu 0} = \mu_{(\nu)y} - \alpha'_\nu \mu_{(\nu)x};\ \nu = \overline{1, k},$$

где μ_ν, $\delta_{(\nu)}$, Σ_ν — оценки первых и вторых моментов функции распределения, определенные на основе реализаций $\{(y_i,\ x_i):\ x_i \in O_\nu\}$.

Из второго условия получаем

$$H(y, x; \alpha_\nu, \alpha_q) = (y - g(x;\ \alpha_\nu))^2 - (y - g(x;\ \alpha_q))^2 = 0;$$

$$h(x;\ \beta_{\nu q}) = 0;\ \nu,\ q = \overline{1, k}. \tag{2.25}$$

Достаточные условия оптимальности КЛМ в общем случае получаются из положительной определенности квадратичной формы, задаваемой матрицей вторых производных

$$\left[\frac{\partial \delta_V}{\partial \beta_{\nu q} \cdot \partial \beta_{ep}}\right].$$

Однако специальных численных методов совместного решения линейных и нелинейных уравнений не разработано [Опойцев, 1970]. В работе А. Д. Касавина [1972] предложен алгоритм решения (2.25) на основе процедуры стохастической аппроксимации, который будет рассмотрен далее.

Вероятностный подход

В тех случаях, когда можно сделать какие-либо предположения относительно распределений отклонений в каждой области $\{\varepsilon_{(\nu)} = y - g(x, \alpha_\nu);\ \nu = \overline{1, k}\}$, задача построения кусочно-линейной аппроксимации формулируется в статистической постановке. Такая постановка более обоснованна, так как рассматриваемые переменные имеют случайный характер. Вместе с тем решение этой задачи просто сводится к решению задачи классификации.

Обозначим вектор всех m случайных ошибок для ν-й области через $\varepsilon_{(\nu)}$, его математическое ожидание $M\varepsilon_{(\nu)} = 0$, ковариационную матрицу случайной величины $\varepsilon_{(\nu)}$ через Σ_ν. Будем считать этот вектор случайным, подчиняющимся распределению $f_\nu(\varepsilon)$.

Предположим далее, что априорные вероятности появления объектов в области O_ν равны π_ν. Тогда последовательность m поступивших возмущений $\varepsilon_1, \ldots, \varepsilon_m$ можно рассматривать как выборку из генеральной совокупности, описываемой суммарным распределением вероятностей $f(\varepsilon) = \sum_{\nu=1}^{k} \pi_\nu f_\nu(\mathrm{E})$.

Задача состоит теперь в отыскании по точкам $\{x_1, \ldots, x_m\}$, истинная принадлежность которых к рассматриваемым областям априорно неизвестна, статистических характеристик рассматриваемого распределения.

Таким образом, мы приходим к задаче классификации в ее статистической постановке [Розин, Ягольницер, 1981]. Будем обозначать в дальнейшем общее число подлежащих оценке параметров d, а совокупность самих параметров через $A = (\alpha_1, \ldots, \alpha_d)$. Распределение $f_\nu(\varepsilon)$ представим в виде

$$f_\nu(\mathrm{E}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma_\nu|}} \exp\left\{-\frac{1}{2}(y - \alpha'_\nu x)\Sigma_\nu^{-1}(y - \alpha'_\nu x)'\right\}.$$

Целевая функция классификации при статистической постановке следующая:

$$f(\mathrm{E}) = \frac{1}{(\sqrt{2\pi})^n} \sum_{\nu=1}^{k} \frac{\pi_\nu}{\sqrt{|\Sigma_\nu|}} \exp\left\{-\frac{1}{2}(y - \alpha'_\nu x)\Sigma_\nu^{-1}(y - \alpha'_\nu x)'\right\}. \tag{2.26}$$

В нашем случае совокупность неизвестных параметров состоит из вероятностей π_ν, элементов матриц Σ_ν и параметров регрессий α_ν, $\nu = \overline{1, k}$.

Для решения задачи нахождения оценок указанных параметров обычно предполагаются выполненными следующие условия:

известно число линейных составляющих кусочно-линейной регрессии;

рассматриваемая последовательность объектов является стационарной, т. е. параметры A стабильны.

Обозначим искомые оценки параметров A через $\hat{A}$. Для распределения вероятностей вида (2.26) задача оценки параметров имеет некоторые особенности. Так как все частные распределения $f_\nu(\mathrm{E})$ описываются одинаковыми законами, то суммарное распределение $f(\varepsilon)$ не зависит от порядка нумерации $f_\nu(\mathrm{E})$ и π_ν. Поэтому существует $k!$ оценок, отличающихся друг от друга только индексами при параметрах.

При разделении имеющейся совокупности объектов на классы, соответствующие областям O_ν, $\nu = \overline{1, k}$, все векторы $\widehat{A}$ равноценны. Таким образом, задача определения оценок усложняется неоднозначностью решения.

Так как в нашем случае число параметров A довольно велико, то известный метод моментов имеет ограниченные возможности и практически непригоден для оценивания параметров совместного распределения.

При применении метода наибольшего правдоподобия оценки обычно определяются как решение системы уравнений

$$\sum_{l=1}^{d} \frac{\partial \ln f(\mathrm{E}, A)}{\partial \widehat{\alpha}_l} = 0; \quad l = \overline{1, d}, \qquad (*)$$

где $\widehat{\alpha}_l$ — неизвестные параметры, d — количество параметров.

Реализация метода наибольшего правдоподобия в таком виде связана с определенными трудностями. Дело в том, что система (*) имеет решение, соответствующее седловым точкам поверхности $L = \{f(\mathrm{E}_1; A), \ldots, f(\mathrm{E}_m; A)\}$ в пространстве параметров. Эти решения не являются состоятельными оценками искомых параметров и требуются дополнительные вычисления для их опознавания и отбрасывания.

Для нахождения оценок $\widehat{A}$ можно использовать различные итерационные алгоритмы, сходящиеся к оценкам наибольшего правдоподобия. Более подробно с этими методами можно познакомиться в работе [Розин, Ягольницер, 1981].

Метод фиксированных элементов разбиений

Метод фиксированных элементов разбиений (ФИЭР) относится к классу поисково-оптимизационных методов и представляет собой универсальный подход к построению не только регрессионных, но и других кусочно-простых моделей.

Для метода ФИЭР [Котюков, 1974; Котюков, Буторин, 1979] характерно свое представление способа задания и оптимизации характеристических функций $\{\Psi(X; \beta_\nu)\}$.

Основной предпосылкой метода является задание на основе анализа V системы различных гиперповерхностей $\Gamma = \{\Gamma^1(X), \ldots, \Gamma^s(X)\}$ (списка фиксированных элементов разбиений) до синтеза КЛМ. Относительно каждой «границы» $\Gamma^l(X) \in \Gamma$ для любой реализации $x \in R_x$ выполняется

либо $\Gamma^l(x) \leqslant 0$, либо $\Gamma^l(x) > 0$. Эти две ситуации можно изобразить и как $\Gamma^l(x) = 0$, и $\Gamma^l(x) = 1$ соответственно.

Иными словами, каждая граница $\Gamma^l(X)$ разбивает R_x на два подмножества — $\{R_0^l, R_1^l\}$.

Синтез функций $\{\Psi(X; \beta_\nu)\}$ осуществляется на основе элементов из Γ. Вектор булевых признаков Γ может иметь различную интерпретацию. Рассмотрим две из них.

I. В простейшем и практически наиболее часто используемом варианте системе Γ соответствует список $\{Z^1, \ldots, Z^s\}$ некоторых фиксированных значений признаков $\{X_j\}$. Пусть Z^l — некоторое значение количественного X_j. Тогда для реализации $x' = (x_1, \ldots, x_n)$ имеем $\Gamma^l(x) = 0$ при $x_j \leqslant Z^l$ и $\Gamma^l(x) = 1$ при $x_j > Z^l$. Если X_j — качественный признак, то $\Gamma^l(x) = 0$ при $x_j = Z^l$ и $\Gamma^l(x) = 1$ при $x_j \leqslant Z^l$.

II. Наряду со значениями $\{Z^l\}$ на шкалах исходных признаков $\{X_j\}$ список Γ дополняется совокупностью фиксированных значений $\{Z^q\}$ на шкалах вторичных признаков $\left\{Z_p = \sum_j \alpha_{pj} X_j,\ p = \overline{1, r}\right\}$. Обычно $r \approx k \ll m$. Признаки $\{Z_p\}$ могут быть получены на основе предварительного анализа выборки V, например методом главных компонент.

При определении системы Γ желательно, чтобы элементы $\{\Gamma^l(X)\}$ давали по возможности различные разбиения V. Величина s ограничена сверху возможностью последующего решения на ЭВМ сложной оптимизационной задачи синтеза КЛМ. Обычно $s \approx 10n$. На рис. 2.6 изображена система Γ.

Сделаем ряд замечаний. При первой интерпретации система разделяющих поверхностей $\{\Psi(X; \beta_\nu)\}$, как будет видно из дальнейшего, реализует монотетическую классификацию, имеет простую и естественную интерпретацию. Иногда это очень важно. При второй интерпретации можно получить политетическую или комбинированную классификацию, и модель может быть более точной.

Априорная фиксация системы Γ, представляющей собой некоторую «ε-сеть» в R_x, позволяет

четко фиксировать схему глобальной оптимизации критерия качества;

использовать эффективные методы программирования для ЭВМ.

Рассмотрим две различные схемы допустимых разбиений $\{O_1, \ldots, O_k\}$ пространства R_x, имеющих древовидную струк-

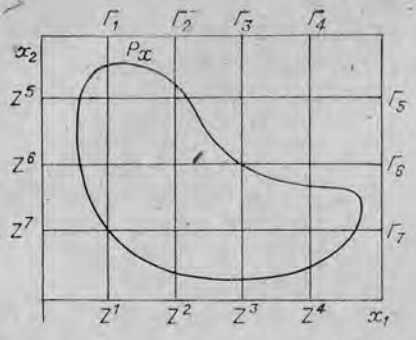


Рис. 2.6. Фиксированные элементы разбиений Г.

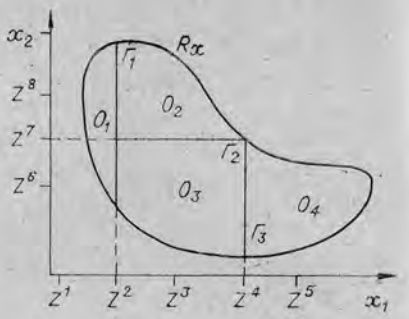


Рис 2.7. Пример основного допустимого разбиения.

туру. Заметим, что любое разбиение пространства булевых признаков Г порождает соответствующее разбиение R_x.

Нижний индекс у элементов из Г будем использовать для их нумерации в определенной последовательности.

Запись $\tilde{\Gamma}(X) \to R_t = (R_q \cup R_l)$ будет означать, что граница $\Gamma(X) \in \Gamma$ реализует разбиение некоторого множества R_t на указанные подмножества $\{R_q, R_l\}$, а выбранный символ $q \in \{0,1\}$ определяет эти подмножества: либо $Rq = (x : \Gamma(x) = 0\}$, либо $R_q = \{x : \Gamma(x) = 1\}$; $R_l = R_t/R_q$.

Пусть $R_1 = R_x$. Тогда выбранная на основе списка Г последовательность $(\Gamma_1(X), \ldots, \Gamma_{k-1}(X))$ определяет *«основное допустимое разбиение»* (ОДР), если

$$\tilde{\Gamma}_\nu(X) \to R_\nu = (O_\nu \cup R_{\nu+1}); \quad \nu = \overline{1,\ k-1}; \qquad (2.27)$$
$$O_k = R_k.$$

Иными словами, при разбиении типа ОДР каждая граница $\Gamma_\nu(X)$ «отделяет» от «оставшегося» множества R_ν очередную область O_ν. Построение ОДР соответствует используемому в одномерном и двумерном случаях принципу последовательного отделения «хороших» областей. На рис. 2.7 приведен пример ОДР при первой интерпретации Г, а на рис. 2.8 изображена его структура.

Приведем пример гипотетической КЛМ, которая может быть построена, согласно схеме ОДР, при $k = 4$.

Если $X_1 \leqslant 0{,}75$, то $Y^* = 3{,}42\ X_2 - 12{,}6\ X_7 - -0{,}0182\ X_8 + 22{,}3$.

Если $X_1 > 0{,}75$ и $X_4 > 124$, то $Y^* = 2{,}71\ X_1 + +258\ X_3 + 18{,}9$.

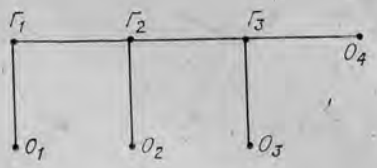


Рис. 2.8. Структура основного допустимого разбиения.

Если $X_1 > 0{,}75$ и $X_4 \leqslant 124$, и $X_5 \leqslant 3{,}5$, то $Y^* = =4{,}32\ X_5 - 16{,}8$.

Если $X_1 > 0{,}75$ и $X_4 \leqslant 124$, и $X_5 > 3{,}5$, то $Y^* = =5{,}17\ X_2 - 0{,}025\ X_6 + 20{,}7$.

Заметим, что при первой интерпретации вектора Г границы получившихся областей $\{O_\nu\}$ наиболее просто интерпретируемы и мы имеем дело с монотетической классификацией.

«Сложное допустимое разбиение» (СДР) определим следующим образом.

Запись $\left(\tilde{\Gamma}_1(X) \to R'_q = R_q \cup R_l\right)$ будет означать, что одно из получившихся при разбиении подмножеств (R_q) имеет индекс (q) разбиваемого множества $\{R'_q\}$, которое, в свою очередь, далее рассматриваться уже не будет. Обозначим также $A_1 = R_1 = R_x$ и $A_\nu = \{R_1, R_2, \ldots, R_\nu\}$.

Тогда выбранная на основе списка Г последовательность $(\tilde{\Gamma}_1(X), \ldots, \tilde{\Gamma}_{k-1}(X))$ порождает СДР, если $\tilde{\Gamma}_\nu(X) \to R'_q = = (R_q \cup R_{\nu+1})$, где $R'_q \in A_\nu$; $1 \leqslant q \leqslant \nu$, $\nu = \overline{1, k-1}$. При $\nu = k-1$ имеем $\{O_q = R_q;\ q = \overline{1, k}\}$. (2.28)

Иными словами, на ν-м шаге граница $\Gamma_\nu(X)$ реализует разбиение некоторого из имеющихся на данном этапе множеств $R'_q \in A_\nu$ на два подмножества.

СДР можно рассматривать и как конечную композицию разбиения типа ОДР.

На рис. 2.9 изображен пример СДР при второй интерпретации вектора Г, а на рис. 2.10 приведена его структура.

Сделаем важное замечание. Будем допускать эквивалентность частных моделей $g(X; \alpha_\nu)$ и $g(X; \alpha_{\nu+1})$ для «соседних» по индексу ν областей $\{O_\nu, O_{\nu+1}\}$ в схемах ОДР и СДР. Это может уменьшить оценку ошибки прогноза по КЛМ $\hat{\delta}$.

Обычно любое разбиение должно удовлетворять еще и ряду априорных ограничений, часто имеющих вид

$$m_\nu \geqslant m_0;\ V^*_\nu \subset O_\nu;\ \nu = \overline{1, k}, \qquad (2.29)$$

где m_ν — объем подвыборки V_ν;

m_0 — минимально допустимое значение m_ν;

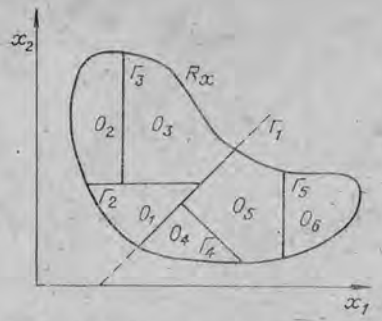


Рис. 2.9. Пример сложного допустимого разбиения.

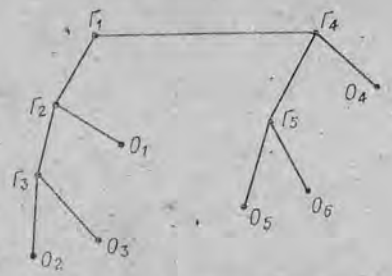


Рис. 2.10. Структура сложного допустимого разбиения.

V_v^* — совокупность некоторых исходных объектов, которые должны принадлежать одной и той же области.

ОДР является наиболее простым разбиением и для его глобальной оптимизации удается создать наиболее эффективные численные методы. Анализ (2.27) показывает, что классы поверхностей, определяющих границы разных областей, различны. Однако этот методологический недостаток ОДР не является существенным в силу наличия процедур его глобальной (по всем $\{O_v\}$) оптимизации. Кроме того, опыт решения большого количества разнообразных прикладных задач не позволил выявить существенных преимуществ схемы СДР перед ОДР. Целесообразность схемы СДР в большей степени обусловлена классом задач, для решения которых необходимо осуществлять структурную адаптацию КЛМ в процессе последовательного поступления новой информации $\{(y_i; x_i)\}$. О таких задачах речь будет идти в последней главе книги.

Несколько слов относительно оптимизируемого критерия δ — оценки остаточной дисперсии КЛМ. Излагаемый ниже метод оптимизации разбиения предполагает его аддитивную структуру:

$$\hat{\delta} = \sum_{v=1}^{k} \hat{\delta}_v, \qquad (2.30)$$

где $\hat{\delta}_v$ соответствует оценке качества $g(X; \alpha_v)$. Видно, что оценки (2.2), (2.4), (2.5), (2.6) удовлетворяют этому требованию, и при дальнейшем изложении под $\hat{\delta}$ можно подразумевать любую из них. При использовании (2.5) в качестве $h_v(\rho(\sigma))$ может быть взята, например, величина $(n/10 + s/n)$.

Для простоты изложения далее ограничимся случаем фиксированного значения k.

Оптимизация ОДР. Условия (2.27) сводят задачу синтеза КЛМ к выбору оптимальной последовательности элементов $(\Gamma_1(X), \ldots, \tilde{\Gamma}_{k-1}(X))_0$ на основе списка Г. В основу предлагаемого алгоритма [Котюков, 1974] положен метод «ветвей и границ». Процесс выбора имеет «ветвящуюся» структуру: на первом шаге рассматриваются все размещения из $2s$ исходных элементов $\{\tilde{\Gamma}^1(X), \ldots, \tilde{\Gamma}^{2s}(X)\}$ по одному $\{(\tilde{\Gamma}^i(X))\}$ — напомним, что каждый элемент $\Gamma^v(X) \in \Gamma$ порождает два элемента $\tilde{\Gamma}(X)$ в зависимости от выбора ориентации области относительно границы $(\Gamma^v(X))$; на втором шаге — по два $\{(\tilde{\Gamma}^i(X), \tilde{\Gamma}^j(X))\}$ и т. д. При этом мы для всех областей, построенных согласно (2.27), каждой ветви $(\tilde{\Gamma}^i(X), \tilde{\Gamma}^j(X)\ldots)$ строим свои модели $g(X; \alpha_v)$ и оцениваем их ошибки $\hat{\delta}_v$. Анализируемые ветви должны удовлетворять ограничениям (2.29).

Индекс p будем использовать, если необходимо выделить некоторую конкретную ветвь $(\tilde{\Gamma}_{1p}(X), \ldots, \tilde{\Gamma}_{lp}(X))$ длины l, соответствующие ей области $\{O_{1p}, \ldots, O_{lp}\}$, частные модели прогноза $g(X; \alpha_{1p}), \ldots, g(X; \alpha_{lp})$ и оценки их ошибок $\{\hat{\delta}_{1p}, \ldots, \hat{\delta}_{lp}\}$.

Пусть $1 \leqslant l < k - 1$. Оценка ошибки для объединения уже «выделенных» областей $\{O_{1p}, \ldots, O_{lp}\}$, согласно (2.30), равна

$$\hat{\delta}_p = \sum_{v=1}^{l} \delta_{vp}.$$

Однако при переходе от l-го шага к $(l+1)$-му не все ветви длины l служат основой для образования ветвей длины $(l+1)$. Для любой p-й ветви можно определить оценку заведомо наилучшего $(\hat{\delta}_{p\min})$ и наихудшего из приемлемых $(\hat{\delta}_{p\max})$ вариантов ее продолжения. Обозначим через O_p^* область, «оставшуюся» от p-й ветви и подлежащую дальнейшему разбиению,

$$O_p^* = R_x \Big\backslash \bigcup_{v=1}^{l} O_{vp},$$

а через $\hat{\delta}_p^*$ — оценку ошибки для модели $g(X; \alpha_p^*)$, построенной на реализациях $\{(y_i, x_i); x_i \in O_p^*\} \subset V$. Тогда имеем

$$\hat{\delta}_{p\max} = \hat{\delta}_p + \hat{\delta}_p^*;$$

$$\hat{\delta}_{p\min} = \hat{\delta}_p + 0 = \hat{\delta}_p. \qquad (2.31)$$

Заметим, что оценки (2.31) могут быть несколько улучшены [Котюков, Буторин, 1979]. Определим величину «рекорда» для всех построенных на данном шаге ветвей:

$$\rho = \min_p \{\hat{\delta}_{p\max}\}. \qquad (2.32)$$

Ветви, у которых $\hat{\delta}_{q\min} > \rho$, в дальнейшем не рассматриваются и не служат основой для образования ветвей большей длины. Полученная на $(k-1)$-м шаге ветвь, для которой величина $\hat{\delta}$ минимальна, является искомой.

Знание ограничений типа (2.29) также позволяет значительно сокращать переборы при поиске $(\tilde{\Gamma}_1(X), \ldots, \tilde{\Gamma}_{k-1}(X))_0$.

Заметим, что схема переборов в данном методе может быть организована по различным принципам: «в ширину» (как описано), «в глубину» (на каждом шаге анализируется очередная ветвь максимальной длины) и комбинированным способом.

Однако при больших n и m программная реализация на ЭВМ метода «ветвей и границ» может встретить затруднения. Ниже будет изложен приближенный метод «гарантированного включения» (ГАРВИ). Если считать, что выбор $\tilde{\Gamma}_l(X)$ зависит лишь от выбора $\tilde{\Gamma}_{l-1}(X)$, то вычислительная процедура упрощается и сводится к решению рекуррентных уравнений типа уравнений динамического программирования. Алгоритмическая структура метода ГАРВИ проста. Поясним ее сущность. После первого шага имеем все $2s$ возможных ветвей длины 1, каждая из которых в качестве своей единственной компоненты содержит некоторое $\tilde{\Gamma}^i(X)$. Пусть после l-го шага имеем $2s$ ветвей длины l. Тогда на $(l+1)$-м шаге выполняются следующие операции. Элемент $\tilde{\Gamma}^i(X)$ пытаемся присоединить ко всем из имеющихся, но не содержащих $\tilde{\Gamma}^i(X)$ ветвей. Из полученных ветвей длины $(l+1)$, содержащих $\tilde{\Gamma}^i(X)$ в качестве последней компоненты, запоминается лишь одна — наилучшая (с минимальной $\hat{\delta}_p$). Такие же операции выполняются и для всех других $\tilde{\Gamma}^i(X)$. В результате $(l+1)$-го шага будем иметь также $2s$ ветвей, причем каждая $\tilde{\Gamma}^i(X)$ будет включена по крайней мере в одну из них (принцип гарантированного включения), и т. д. Лучшая ветвь, полученная на $(k-1)$-м шаге, является искомой.

Алгоритм ГАРВИ можно дополнять принципами «отсечения» ветвей оптимального метода.

При построении ОДР может быть реализован и простой алгоритм локальной оптимизации, основанный на широко используемом принципе «последовательного включения». Сущность его заключается в следующем: на первом шаге окончательно определяется O_1, имеющая наименьшее значение $\hat{\delta}_1$; на втором шаге из подмножества (R_x/O_1) выделяется O_2, имеющая $\hat{\delta}_2 = \min$; и т. д. до $(k-1)$-го шага.

Заметим, что в процессе работы того или иного переборного алгоритма построение каждой частной модели $g(X; \alpha_{\nu p})$ целесообразно осуществлять широко используемым методом «последовательного включения» в $g(X; \alpha_\nu p)$ значимых $\{X_j\} \subset X$.

Оптимизация СДР. Для синтеза СДР (2.28) модифицируем ранее рассмотренный метод «ветвей и границ». На первом шаге исходные границы $\{\Gamma^i(X)\}$ порождают s различных разбиений, содержащих по две области, для каждой из которых строится $g(X; \alpha_\nu)$ и определяется $\hat{\delta}_\nu$. На l-м шаге каждое из полученных на $(l-1)$-м шаге разбиений $\{O_{1p}, \ldots, O_{lp}\}$ порождает $(l \times s)$ новых разбиений длины $(l+1)$ путем разбиения с помощью $\Gamma^i(X) \in \overline{\Gamma}$ $(i = \overline{1, s})$ каждой области $O_{\nu p}$ на две подобласти $(\nu = \overline{1, l})$. И так далее. Для каждого p-го разбиения $\{O_{1p}, \ldots, O_{lp}\}$ можно, как и в случае оптимизации ОДР, определить величины

$$\hat{\delta}_{p\max} = \sum_{\nu=1}^{l} \hat{\delta}_{\nu p};$$

$$\hat{\delta}_{p\min} = \begin{cases} \min \sum_{\nu=1}^{r} \hat{\delta}_{\nu p} & \text{при } r \geqslant 1; \\ 0 & \text{при } r \leqslant 0, \end{cases} \qquad (2.33)$$

где r — целая часть частного $(2l - k)/2$; $\min \sum_{\nu=1}^{r}$ — сумма наименьших r компонент из $\{\hat{\delta}_{1p}, \ldots, \hat{\delta}_{lp}\}$. Оценки (2.33) могут быть несколько улучшены [Котюков, Буторин, 1979]. Принцип отсечения на каждом шаге неперспективных ветвей на основе $\{\hat{\delta}_{p\max}, \hat{\delta}_{p\min}\}$ аналогичен принципу для ОДР. Введение ограничений (2.29) повышает эффективность алгоритма. Очевидно, что трудоемкость оптимизации СДР намного больше, чем ОДР,— число возможных вариантов больше и оценки (2.33) «слабее» оценок (2.31).

Для целей синтеза СДР при больших n и m может быть предложена в качестве приближенного метода модификация

рассмотренного ранее метода ГАРВИ [Котюков, Буторин, 1979].

Можно использовать и следующий простой метод локальной оптимизации: на первом шаге определяется $\Gamma_1(X) \in \Gamma$, наилучшим образом разбивающая R_x на две области $\{O^1, O^2\}$ $(\hat{\delta}^1 + \hat{\delta}^2 = \min)$; на каждом l-м шаге $(l \leqslant k-1)$ разбивается на $\{O^{v1}, O^{v2}\}$ та из имеющихся областей O^v, для которой найдется $\Gamma_l(X) \in \Gamma$, обеспечивающая

$$\left(\sum_{q=1, q\neq v}^{l} \hat{\delta}_q + \hat{\delta}^{v1} + \hat{\delta}^{v2}\right) = \min.$$

Оценки $\hat{\delta}$ типа (2.2), (2.4), (2.5), (2.6) и ограничения (2.29) позволяют осуществить тривиальное обобщение рассмотренных алгоритмов на случай заранее не фиксированного значения k.

Некоторые аспекты метода ФИЭР близки принципам построения так называемых *линейно-логических функций* [Лбов, 1981; Манохин, 1976], которые являются обобщением логических решающих правил для разнотипных признаков [Лбов и др., 1973]. Методология построения линейно-логических функций в значительной степени ориентирована именно на случай разнотипных признаков X [Лбов, 1981], а сами алгоритмы их синтеза являются разновидностью ранее приведенного метода локальной оптимизации СДР.

Глава 3

МЕТОДЫ ПОСТРОЕНИЯ МОДЕЛЕЙ С ПЕРЕМЕННОЙ СТРУКТУРОЙ ДЛЯ ВРЕМЕННЫХ ВЫБОРОК

Методические подходы к построению моделей с переменной структурой во многом определяются исходной информационной базой и объектом моделирования. Поэтому их рассмотрение целесообразно начать с некоторой классификации динамических моделей.

В основу выделения типов динамических моделей положено два классификационных признака: метод прогнозирования показателей и объект моделирования (как и что моделируется). По первому признаку следует различать прогнозирование путем экстраполяции

закономерностей изменения самого значения показателя;

модели прошлого периода и тенденций изменения независимых переменных;

закономерностей изменения во времени параметров модели и независимых переменных.

По объекту моделирования различаются динамические модели отдельного объекта и совокупности объектов.

Сочетание двух указанных классификационных признаков приводит к выделению следующих типов динамических моделей; одномерные и многомерные модели отдельного объекта и многомерные модели совокупности объектов. Следует подчеркнуть, что последний тип моделей также относится к уровню отдельного объекта. Отличие заключается лишь в форме представления исходной информации и способах построения моделей. Обозначим: m — число изучаемых объектов, y_{il}^t — значение l-го моделируемого показателя i-го объекта в t-м периоде $(i = \overline{1, m}; l = \overline{1, n}; t = \overline{1, \omega})$; x_{ir}^t — значение r-го фактора производства для i-го объекта в t-м периоде $(i = \overline{1, m}; r = \overline{1, p}; t = \overline{1, \omega})$. Тогда исходная информация для одномерной модели отдельного объекта

характеризуется $m = 1$, $n = 1$, $p = 0$, $t \gg 1$; для многомерной модели отдельного объекта — $m = 1$, $n \geqslant 1$, $p \geqslant 1$, $t \gg 1$; для моделей, построенных на основе данных совокупности объектов, — $m \gg 1$, $n \geqslant 1$, $p \geqslant 1$, $t > 1$. В данной главе рассматриваются лишь одномерная и многомерная модели отдельного объекта.

Одномерные динамические модели отдельных объектов могут либо иметь самостоятельное значение, как инструмент анализа динамики показателей и их прогнозирования, либо являться вспомогательным элементом при построении общей динамической модели совокупности объектов.

Исследование динамических рядов отдельного показателя относится к одному из традиционных статистических методов. Имеется довольно богатый формальный аппарат выявления тенденций: методы скользящих средних, наименьших квадратов, линейного и нелинейного программирования в случае наличия ограничений, прогнозирование случайной компоненты по авторегрессионной модели, выделение сезонных и циклических колебаний, методы экспоненциального сглаживания и гармонических весов, учитывающие нарастание веса более близких наблюдений, и др.

Но область прямого использования этого подхода в моделировании и прогнозировании экономических показателей весьма ограниченна. Это связано с сущностью метода прогнозирования по отдельным динамическим рядам — прогноз осуществляется по собственной предыстории изменения показателя без вскрытия причин развития и представляет собой прямую экстраполяцию. Она возможна только при том условии, что динамика показателя носит стабильный характер и для будущего можно предположить сохранение этой динамики. Такое условие соблюдается, как правило, для агрегированных, интегральных показателей больших систем в целом, например для спроса на массовую традиционную продукцию, но не для экономических показателей отдельного предприятия. Кроме того, область применения этого приема также ограничена размерами рядов — в экономических исследованиях обычно можно набрать сравнительно небольшое число годовых реализаций. Далее, однофакторные динамические модели не позволяют прогнозировать показатели новых предприятий, так как они жестко привязаны к конкретному объекту. Не играя существенной самостоятельной роли в прогнозировании экономических показателей, методы анализа динамических рядов в то же время — обязательный составной элемент всех других методов описания динамики.

Необходимо отметить, что методы анализа одномерных временных рядов довольно хорошо разработаны и широко освещены в литературе, тогда как по методам построения моделей на основе анализа многомерных рядов литературные материалы более скудны. Отличие второго подхода от первого заключается в переходе к многокомпонентной модели, описывающей развитые системы рядов. Динамический ряд показателя представляется как функция динамических рядов производственных факторов с учетом сдвигов во времени (лагов). Достоинствами этого типа модели являются учет причин изменения показателя и хорошая приложимость к конкретному объекту (временна́я информация по отдельному объекту более однородна, чем пространственная по группе объектов). Особо следует отметить, что при анализе многомерных временных рядов можно выявить величину запаздывания во времени действия отдельных производственных факторов. Последнее играет большую роль при любых способах построения динамических моделей и прогнозирования по ним. Поэтому такие модели — помимо их самостоятельного применения — могут оказаться полезными и как вспомогательные элементы построения динамических моделей совокупности объектов. Недостатки моделей обусловлены характером информации — при коротких рядах не представляется возможным включать в модель достаточное число переменных, а главное, малая вариация отдельных факторов и их соотношений во времени не позволяет отразить в модели достаточную область свободы, что делает динамические модели отдельного объекта непригодными для прогноза при существенных изменениях условий работы объекта (изменение объема и структуры фондов, технологии и т. п.); нельзя их также использовать для других объектов. Хотя необходимо заметить, что область свободы здесь все-таки шире, чем у жестко привязанных динамических рядов отдельного показателя.

Динамические модели многомерных временных рядов чаще всего комбинированные — они включают явную функцию времени для учета тренда, значения независимых переменных в разные периоды времени и авторегрессионную часть (значения моделируемого показателя в предшествующие периоды времени). В общем виде модель может быть записана следующим образом:

$$y^{*t} = g\left(t, y^{t-1}, \ldots, y^{t-\tau}, x_1^t, \ldots, x_1^{t-\tau_1}, x_2^t, \ldots \right.$$
$$\left. \ldots, x_2^{t-\tau_2}, \ldots, x_p^t, \ldots, x_p^{t-\tau_p} \right) + \varepsilon^t,$$

где τ_j — время запаздывания (лаг) показателя (фактора) x_j;

ε^t — случайное возмущение на момент t.

Выше мы рассмотрели наиболее простой случай построения динамической модели на основе многомерного временного ряда — в качестве моделируемого показателя выступал один признак ($n = 1$) — т. е. случай одномерного выхода. При этом предполагается, что результаты работы исследуемого объекта могут быть описаны одной или несколькими автономными (независимыми) моделями.

Понятно, что это предположение носит условный характер, поскольку в действительности любой экономический процесс характеризуется целой системой зависимостей между переменными, причем некоторые из них, отнесенные к числу независимых, на самом деле могут выступать и в роли функции. Решение проблемы заключается в построении системы структурных (совместных) уравнений [Маленво, 1976], дающих более полное и всестороннее представление об объекте.

Однако построение структурных моделей для каждого предприятия отраслевой системы — чрезвычайно громоздкая и трудоемкая процедура, к тому же не всегда обеспеченная достаточным объемом информации. Поэтому здесь обычно идут на следующее упрощение: строится не общая структурная модель, а ее специальный вид — рекурсивная модель.

§ 1. ОДНОМЕРНЫЕ ВРЕМЕННЫЕ РЯДЫ

В общем случае проблема построения моделей с переменной структурой включает решение следующих задач:

выявление точек перелома зависимости;

установление характера перехода (плавное или скачкообразное изменение);

построение модели с переменной структурой;

проверка гипотезы о наличии структурных изменений.

Решение указанных задач представляет значительные трудности, более того, отдельные задачи еще не имеют надежного формального решения. Однако в прикладных исследованиях довольно часто встречаются ситуации, когда не требуется формальное решение первых двух задач, поскольку априори с той или иной степенью точности могут быть сформулированы содержательные гипотезы о положении точки перелома и характере перехода. Например, при

изучении изменения во времени какого-либо экономического показателя отдельного предприятия (объем выпуска, себестоимость продукции, производительность труда и т. д.) заранее можно высказать предположение, что в большинстве случаев переходы будут плавные, а точки перелома зависимости будут располагаться на временной оси в моменты существенных изменений условий производства (изменение технологии, характера сырья, переход на выпуск нового вида продукции и т. д.). В ряде случаев, когда нет достаточной содержательной априорной информации, местоположение точки перелома можно ориентировочно определить по виду графика временного ряда.

Рассмотрение методов построения моделей с переменной структурой целесообразно начать с самой простой ситуации.

Случай известных точек перелома и непрерывного перехода

Наиболее простым и наглядным методом выявления тренда в данной ситуации является метод фиктивных переменных [Дрейпер, Смит, 1973]. Сущность метода сводится к следующему. Пусть, для простоты, имеются две линии регрессии, образующие перелом

$$y^*_{iv} = a_v + b_v x_{iv} + \varepsilon_{iv} \qquad (v = 1, 2;\ i = 1, 2, \ldots, m_v).$$

Можно ввести фиктивную переменную Θ_i, равную единице, если наблюдение соответствует второй линии регрессии, и нулю — в противном случае. Тогда обе модели можно привести к одной

$$y^*_i = a_1 + b_1 x_1 + (a_2 - a_1)\,\Theta_1 + (b_2 - b_1)\,\Theta_i x_i + \varepsilon_i = \\ = \alpha_0 + \alpha_1 z_{i1} + \alpha_2 z_{i2} + \alpha_3 z_{i3} + \varepsilon_i,$$

где $$(x_i, y^*_i) = \begin{cases} (x_{i1}, y^*_{i1}), & i = 1, 2, \ldots, n_1; \\ (x_{i2}, y^*_{i2}), & i = (n_1 + 1), \ldots, (n_1 + n_2) \end{cases}$$

и $$\Theta_i = \begin{cases} 0, & i = 1, 2, \ldots, n_1; \\ 1, & i = (n_1 + 1), \ldots, (n_1 + n_2). \end{cases}$$

В случае трех прямых вводим две фиктивные переменные

$$\Theta_1 = \begin{cases} 1, \text{ если наблюдение соответствует второй прямой,} \\ 0 \text{—в противном случае;} \end{cases}$$

$$\Theta_2 = \begin{cases} 1, \text{ если наблюдение соответствует третьей прямой,} \\ 0 \text{ — в противном случае.} \end{cases}$$

При этом получаем следующую объединенную модель:

$$y_i^* = a_1 + b_1 x_i + (a_2 - a_1)\,\Theta_{i1} + (a_3 - a_1)\,\Theta_{i2} + \\ + (b_2 - b_1)\,D_{i1}x_i + (b_3 - b_2)\,D_{i2}x_i + \varepsilon_i.$$

Дальнейшее обобщение проводится просто. Коэффициенты модели оцениваются на основе выборки с помощью метода наименьших квадратов.

Правда, рассмотренный метод применим в случае небольшого числа точек перелома, поскольку при значительном количестве переломов требуется введение большого числа фиктивных переменных. Способ введения фиктивных переменных нами апробирован на ряде конкретных примеров и оказался весьма удобным практическим приемом кусочно-линейной аппроксимации трендов.

Рассмотрим способ описания тенденции показателя комплексной выработки Ингузетского леспромхоза за 11 лет (рис. 3.1). Полученное способом фиктивных переменных уравнение регрессии при задании точки пересечения в седьмом году (две фиктивные переменные) имело следующий вид:

$$y^* = 608{,}75 + 37{,}3x_1 - 35{,}8x_2;\ R^2 = 0{,}81,$$

где y^* — значение комплексной выработки;

x_1, x_2 — соответственно значения t для периода до точки пересечения и после точки пересечения. x_1 принимает значения 1, 2, 3, 4, 5, 6 до точки пересечения и 0 после нее. $x_2 = 0$ до точки пересечения и 1, 2, 3, 4 — после.

Как видим, получено хорошее качество аппроксимации. Приведенный способ описания структурного изменения можно трактовать как частный случай построения линейного сплайна.

Перейдем к изложению методов описания зависимостей сплайн-функциями для одномерных временных рядов.

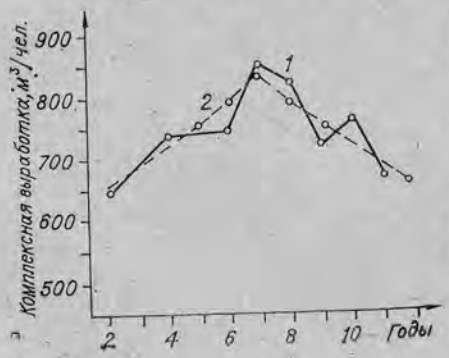


Рис. 3.1. Тенденция изменения показателя комплексной выработки по Ингузетскому леспромхозу.
1 — эмпирическая регрессия;
2 — теоретическая кривая.

Ранее (§ 1 гл. 2) были введены основные определения линейного сплайна. Отметим, что представление (*) линейного сплайна удобно использовать как составную часть множественной линейной регрессии относительно неизвестных параметров β_j ($j=0, \ldots, q$). Последнее связано с тем, что на параметры не накладывается никаких ограничений; коэффициенты β_j имеют t-распределение и можно определить статистическую значимость изменения коэффициента наклона при переходе от (j—1)-го к j-му интервалу. Значимое отличие β_j от нуля свидетельствует о том, что в узле x_{j-1} происходит «структурное изменение».

Построение линейного одномерного сплайна с k узлами равносильно построению множественной регрессии с k факторами, где роль отдельных факторов играют преобразованные переменные из представления (*).

В более общем случае сплайн-функция может выступать как часть статистической модели, описывающая действие фактора времени. Тогда экономико-статистическая модель может быть представлена в виде

$$y^{*t} = a_0 + \sum_{j=1}^{n} b_j x_j + \sum_{i=1}^{p} \beta_i (t - \bar{t}_i) + \varepsilon^t,$$

где a_0 — свободный член уравнения регрессии;

$\sum_{j=1}^{n} b_i x_j$ — факторная часть регрессионной модели;

$\sum_{i=1}^{p} \beta_i (t - \bar{t}_i)$ — кусочно-линейный сплайн с p узлами;

ε^t — случайный остаток.

Использование линейного сплайна предполагает существование некоторой линейной зависимости, претерпевающей в узлах структурные изменения. Если же аппроксимируемая зависимость существенно нелинейна, то для более «гладкого» и гибкого описания используются кубические многочлены. Если в узлах сетки кубический многочлен меняет структуру то мы получаем так называемый *кубический сплайн.*

Математическая формулировка кубического сплайна дана в работе Д. Пуарье [1981].

Сеткой называется множество точек оси абсцисс $\Delta = \{\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_k\}$ на отрезке $[\bar{x}_0, \bar{x}_k]$, а каждая из $(k+1) \geqslant 3$ точек x_j ($j = 0, 1, \ldots, k$) — *узлом*. Пусть $y = \{y_0, y_1, \ldots, y_k\}$ — набор соответствующих значений ординат. *Кубическим сплайном* над Δ, интерполирующим набор y, называется функция $S_\Delta(x)$ такая, что

1) $S_\Delta(x)$ и ее первая и вторая производные непрерывны на $[\bar{x}_0, \bar{x}_k]$;

2) $S_\Delta(x)$ на каждом интервале $[x_{j-1}, x_j]$ $(j = 1, 2, \ldots, k)$ совпадает с некоторым многочленом степени не больше трех;

$$3)\ S_\Delta(\bar{x}_j) = y_j\ (j = 0, 1, \ldots, k).$$

Условия непрерывности удовлетворяются путем наложения на сплайн *граничных условий* в концевых узлах $\bar{x}_0$ и $\bar{x}_k$. Эти условия могут быть введены различными способами [Пуарье, 1981].

Общий вид кусочно-кубического сплайна для отрезка $[x_{j-1}, x_j]$ можно представить следующим образом:

$$S_\Delta(x) = a_j + b_j(x - \bar{x}_{j-1}) + c_j(x - \bar{x}_j)^2 + d_j(x - x_j)^3,$$

где неизвестные параметры a_j, b_j, c_j и d_j находятся с учетом условий 1)—3).

Если решается задача аппроксимации с точными условиями (вектор ординат y известен), то не возникает проблем в определении параметров кусочно-кубического сплайна — задача решается с помощью хорошо известного метода решения системы линейных уравнений, матрица коэффициентов которой имеет трехдиагональную форму [Алберг и др., 1972]. Сложности возникают при построении статистической модели, включающей кубический сплайн. В этом случае вектор ординат $y = \{y_0, y_1, \ldots, y_k\}$ неизвестен, он определяется методом наименьших квадратов.

Кубические сплайны нашли широкое распространение в задачах аппроксимации экспериментальных данных потому, что они являются «самыми гладкими» функциями, проходящими через данные точки (свойство наилучшего приближения [Алберг и др., 1972]).

Рассмотрим некоторые примеры приложения теории сплайн-функций к задачам моделирования основных этапов развития промышленного предприятия. Все задачи решены в предположении, что точки изменения структуры заданы. Задачи решались как для отдельного, так и для группы объектов. В последнем случае для обеспечения сопоставимости необходимо было соизмерять динамики развития отдельных объектов — приводить кривые к единому масштабу.

Представляется перспективным использование сплайнов («чистых» и комбинированных) для описания процесса развития объекта, так как оно неравномерно (периоды быстрого изменения показателей чередуются с периодами сравнительно устойчивого функционирования объекта) и в то же время для отдельного объекта, как правило, очень редки скачки в динамике экономических показателей.

Каждому этапу развития промышленного предприятия (проектирование, строительство, эксплуатация) соответствует своя динамика экономических показателей. На проектной стадии оценке подлежат проектные характеристики будущего предприятия, в период строительства — показатели этого процесса (удельные капиталовложения, сроки строительства, распределение вложений по срокам строительства), на этапе эксплуатации для каждой стадии строятся специфические модели зависимости экономических показателей от производственных факторов.

Рассмотрим методические подходы к решению задач моделирования производственного процесса на различных этапах развития предприятия.

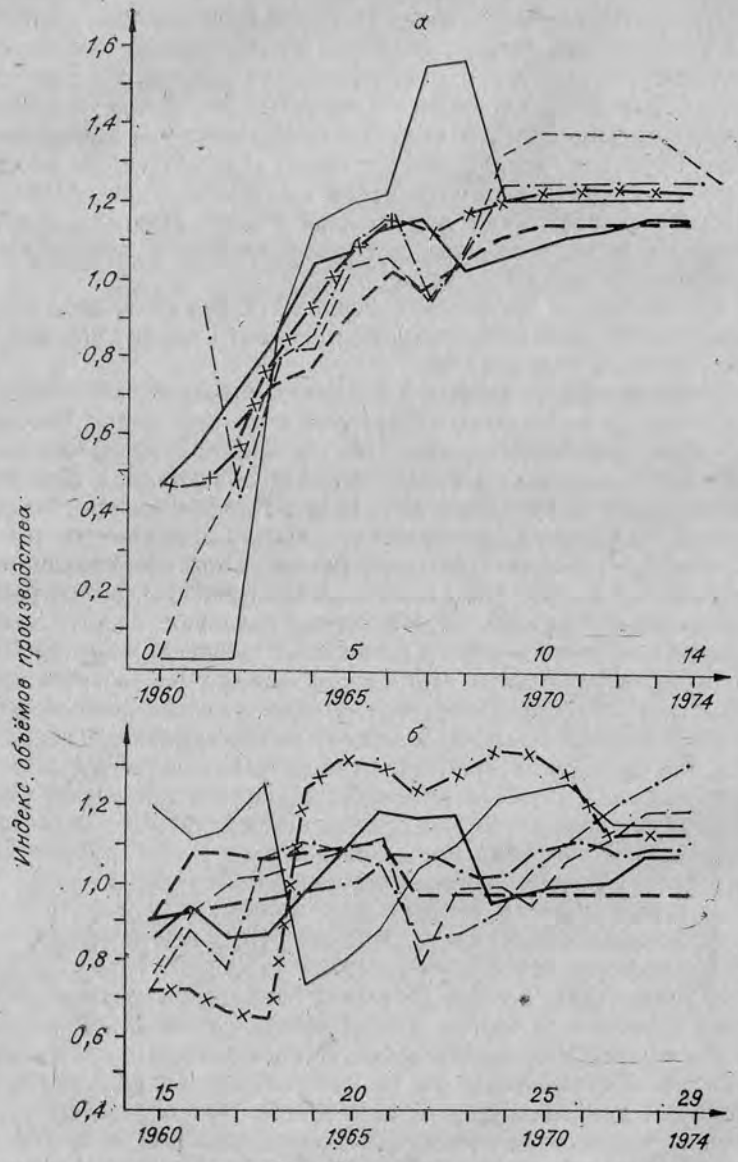
Моделирование динамики развития предприятия. Характер влияния «возраста» предприятия на экономические показатели производства покажем на примере леспромхозов, входящих в состав производственного объединения Красноярсклеспром [Кулешов и др., 1972]. Поскольку проследить полностью динамику развития отдельного предприятия из-за длительных сроков функционирования не представляется возможным, в данной задаче пришлось прибегнуть к искусственному конструированию истории развития леспромхоза. Для этого функционирующие в период обследования (1960—1975 гг.) леспромхозы были подразделены на «возрастные» группы. Поведение показателя «Объем заготавливаемой древесины» изучалось путем анализа его динамики для различных по продолжительности функционирования групп леспромхозов. В качестве классификационного признака, формирующего такие группы, принималась дата основания леспромхоза. В конечном счете было выделено три группы леспромхозов, формирование которых относится

к периоду после 1957 г. (А);

к послевоенному периоду (1946—1956 гг.) (Б);

к предвоенному периоду (1930—1940 гг.). (В).

Предприятия перечисленных групп имеют разный «возраст» и находятся на различных этапах развития. Для того чтобы провести сравнительный анализ зависимости показателя объема производства от «возраста» для разных предприятий, результирующий показатель был переведен в условные единицы (отношение объема производств в данном году к среднему по леспромхозу за период его функционирования). Проведенный графический анализ динамики объема производства (рис. 3.2) свидетельствует о том, что развертывание лесозаготовительного предприятия во времени носит отчетливо выраженный поэтапный характер. График



отражает динамику изучаемого показателя для трех групп леспромхозов.

Первая группа (А) — леспромхозы, пущенные в эксплуатацию после 1957 г. Можно считать, что динамика их показателей характеризует этап освоения производственной мощности и переход к стабильному функционированию. Их «возраст» 0—15 лет.

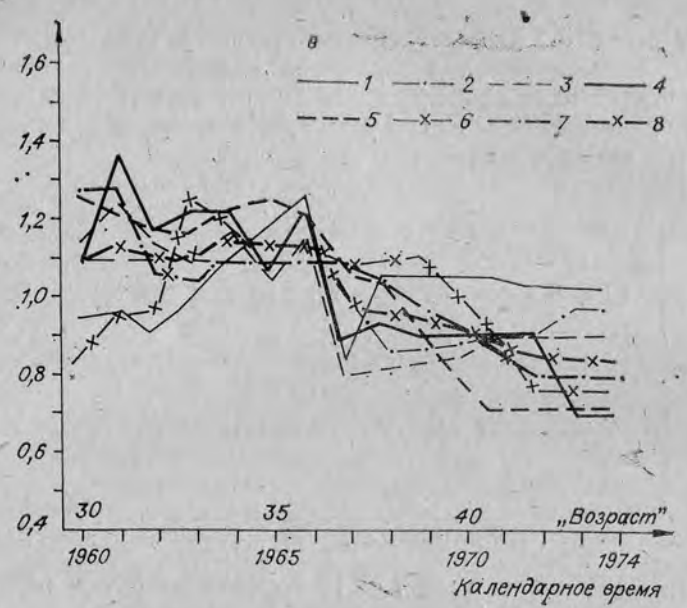


Рис. 3.2. Динамика объемов производства леспромхозов.

а — предприятия группы А: *1* — Большемуртинский; *2* — Пинчугский; *3* — Боготольский; *4* — Верхнетомский; *5* — Манзенский; *6* — среднее значение для группы А. *б* — предприятия группы Б: *1* — Нижнеказульский; *2* — Тугачинский; *3* — Тасеевский; *4* — Кежемский; *5* — Верх-Енисейский; *6* — Нижне-Ингашский; *7* — среднее значение для группы Б. *в* — предприятия группы В: *1* — Она-Чунский; *2* — Партизанский; *3* — Матурский; *4* — Унгутский; *5* — Уйбатский; *6* — Июсский; *7* — Советский; *8* — среднее значение для группы В.

Вторая группа (Б) — леспромхозы послевоенного периода. Их динамику, по существу, отражает особенность периода стабильного функционирования объектов. «Возраст» предприятий 15—30 лет.

И наконец, третья группа (В) — предприятия довоенного периода, сырьевая база которых истощается, а динамику их можно ассоциировать с периодом сворачивания производства. «Возраст» предприятий более 30 лет.

Для построения кривой, характеризующей изменение показателя «Объем заготавливаемой древесины» по всем этапам функционирования предприятия, был использован кусочно-линейный сплайн с заданными узлами в точках $t = \{6, 13, 36\}$. Предварительно данные по имеющимся леспромхозам были усреднены для каждого года рассматриваемого периода функционирования (45 лет).

На рис. 3.2 представлены эмпирические кривые для каждого леспромхоза и усредненная для всей совокупности. На оси абсцисс приведены две шкалы отсчета — календарное время (1960—1975 гг.) и «возраст» предприятия с момента его пуска в эксплуатацию (0—45).

Полученные кусочно-линейные сплайны имеют вид

$$y^* = \begin{cases} \underset{(10,55)}{0,19 + 0,14t}, & \text{при } 0 \leqslant t \leqslant 6; \\ \underset{(7,46)}{0,19 + 0,14t - 0,14(t-6)}; & \text{при } 6 < t \leqslant 13; \\ \underset{(0,88)}{0,19 + 0,14t - 0,14(t-6) - 0,007(t-13)}, & \text{при } 13 < t \leqslant 36; \\ 0,19 + 0,14t - 0,14(t-6) - 0,007(t-13) - 0,02(t-36), & \\ & \text{при } t > 36; \end{cases}$$

$$R^2 = 0,86; \ \sigma_{ост} = 7,07\ 10^{-2}. \qquad (3,64)$$

Поскольку для узла $t = 13$ перелом незначим (эмпирическое значение t-критерия меньше теоретического), эта точка была исключена из рассмотрения. Окончательная форма модели

$$y^* = \begin{cases} \underset{(13,14)}{0,17 + 0,15t}, & 0 < t \leqslant 6; \\ \underset{(12,47)}{0,17 + 0,15t - 0,15(t-6)}, & 6 < t \leqslant 36; \\ \underset{(4,29)}{0,17 + 0,15t - 0,15(t-6) - 0,03(t-36)}, & t > 36; \end{cases}$$

$$R^2 = 0,85; \ \sigma_{ост} = 7,05\ 10^{-2}$$

Таким образом, описание кривой развития усредненного предприятия четко выделяет три этапа. Первый — освоение производственной мощности в течение 6 лет. В этот период имеет место нарастание производственной мощности с темпом 15% в год. Второй период — стабильное функционирование предприятия в течение 30 лет. В этот период объем производства практически не меняется. Темпы изменения $\beta_1 + \beta_2 = 0,1487 - 0,1496 = -0,0009$, т. е. снижение составляет 0,09% в год. И наконец, третий период продолжительностью 9 лет — постепенное свертывание производственного процесса. Темпы изменения $\beta_1 + \beta_2 + \beta_3 = -0,00278$, что составляет примерно 0,3% в год.

[1] В круглых скобках приведены эмпирические значения t-критерия Стьюдента.

Таблица 3.1

Распределение ошибок аппроксимации

Вид функции	Абсолютная ошибка, %						
	До 5	5—10	10—15	15—20	Свыше 20	Итого	Максимальная
Парабола второго порядка	16	10	9	7	3	45	27,0
Кусочно-линейный сплайн	23	15	7	—	—	45	13,8

Для сравнения точности кусочной и непрерывной аппроксимации была построена квадратичная функция

$$y^* = 0,63 + 0,04t - 0,0008t^2; \ R^2 = 0,54; \ \sigma_{ост} = 0,12.$$

В табл. 3.1 приведены распределения ошибок аппроксимации для кусочно-линейного сплайна и квадратичной параболы.

Видно, что кусочно-линейный сплайн имеет лучшие показатели качества модели (R^2, $\sigma_{ост}$); резко снижает число больших ошибок аппроксимации; уменьшает максимальную ошибку примерно в два раза и, кроме того, обеспечивает

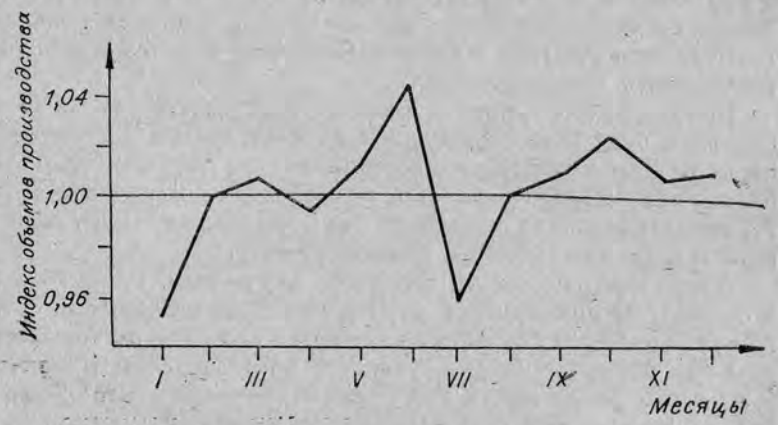


Рис. 3.3. Внутригодовая динамика объема производства.

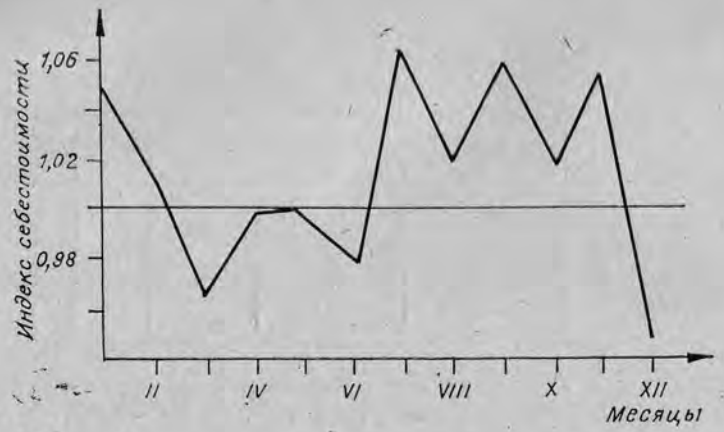


Рис. 3.4. Внутригодовая динамика себестоимости продукции.

более равномерное приближение к эмпирической кривой развития (квадратичная парабола дает систематическую ошибку на этапе освоения).

Исследование внутригодовой динамики экономических показателей. Как правило, внутри отдельных календарных периодов функционирования промышленных объектов наблюдается изменчивость экономических показателей. При этом темпы изменения показателей не остаются стабильными. Вследствие этого представляется целесообразным использовать для изучения динамики показателей модели с переменной структурой. Методику их использования покажем на примере исследования сезонной динамики показателей объема выпуска и себестоимости для одного из заводов тракторного машиностроения.

Исходная информация представляла собой помесячные данные о значении исследуемых показателей на пятилетний период. Для обеспечения сопоставимости годовых данных по каждому показателю были исчислены месячные индексы I_i, представляющие отношение фактического значения показателя в данном месяце к среднегодовому.

Характер изменения индексов изучаемых показателей в среднем за шестилетний период приведен на рис. 3.3, 3.4. Анализ графиков позволяет выдвинуть гипотезу о квартальном изменении темпов изучаемых показателей и принять в качестве узлов последние месяцы кварталов. Это равносильно построению кусочно-линейного сплайна.

Приведем результаты моделирования по одному году.

Модель показателя «Объем выпуска» (шт.)

$$y_1^* = \begin{cases} \underset{(1,82)}{3416{,}80 - 258{,}12t}, & 0 \leqslant t \leqslant 3; \\ 3416{,}80 - \underset{(1,82)}{258{,}12t} - \underset{(1,41)}{287{,}45}\,(t-3), & 3 < t \leqslant 6; \\ 3416{,}80 - 258{,}12t - 287{,}45\,(t-3) - \underset{(0,89)}{135{,}74}\,(t-6), & 6 \leqslant t \leqslant 9; \\ 3416{,}80 - 258{,}12t - 287{,}45\,(t-3) - 135{,}74\,(t-6) + \\ \quad + \underset{(1,91)}{295{,}70}\,(t-9), & 3 < t \leqslant 12; \end{cases}$$

$$R^2 = 0{,}75;\ \sigma_{ост} = 219{,}78.$$

Поскольку перелом в узле $t = 6$ оказался незначимым, то была построена модель без этого узла:

$$y_1^* = \begin{cases} \underset{(2,41)}{3348{,}8 + 309{,}19t}, & 0 \leqslant t \leqslant 3; \\ 3348{,}8 + \underset{(2,41)}{309{,}19t} - \underset{(2,69)}{408{,}02}\,(t-3), & 3 < t \leqslant 9; \\ 3348{,}8 + 309{,}19t - 408{,}02\,(t-3) + \underset{(1,83)}{197{,}19}\,(t-9), & 9 < t \leqslant 12; \end{cases}$$

$$R^2 = 0{,}71;\ \sigma_{ост} = 216{,}98$$

Модель показателя «Себестоимость продукции» (руб./шт.)

В связи с тем, что перелом в узле $t = 6$ также оказался незначимым, модель имеет вид

$$y_2^* = \begin{cases} \underset{(1,84)}{5405{,}28 - 241{,}46t}, & 0 < t \leqslant 3; \\ 5405{,}28 - \underset{(1,84)}{241{,}46t} + \underset{(2,76)}{383{,}68}\,(t-3), & 3 < t \leqslant 6; \\ 5405{,}28 - 241{,}46t + 383{,}68\,(t-3) - \underset{(3,34)}{329{,}39}\,(t-9), & 9 < t \leqslant 12; \end{cases}$$

$$R^2 = 0{,}79;\ \sigma_{ост} = 198{,}80.$$

Построение аналогичных моделей за все годы исследуемого периода позволило выявить величину темпов изменения показателей по кварталам года. Соответствующие данные по показателю «Объем производства» приведен в табл. 3.2, а по показателю «Себестоимость» — в табл. 3.3.

Во все годы исследуемого периода в первом квартале наблюдается рост месячного объема выпуска. Во втором квар-

Таблица 3.2

Квартальные темпы изменения объема выпуска тракторов

Год	Квартал			
	I	II	III	IV
1976	32,5	106,7	—49,7	141,8
1977	112,7	16,1	49,6	141,4
1978	72,3	23,6	70,9	—93,0
1979	258,2	—29,2	—164,9	130,8
1980	48,6	—82,6	—26,4	37,1

Таблица 3.3

Квартальные темпы изменения себестоимости тракторов

Год	Квартал			
	I	II	III	IV
1977	—123,0	97,0	—62,0	—62,9
1978	—176,7	43,4	3,3	—110,5
1979	—241,4	39,8	155,1	—258,0
1980	—195,2	47,7	137,3	—225,5

тале в первые три года имело место увеличение месячного производства тракторов, а в последние два — падение. В третьем квартале устойчивого изменения месячного производства не выявлено, а в четвертом, как правило, имеет место наращивание объемов производства.

Внутриквартальные изменения характеризуются следующими величинами: первый квартал — устойчивое снижение себестоимости; второй — повышение себестоимости; четвертый — устойчивое снижение себестоимости. Аналогичный подход может быть также использован при изучении внутримесячной изменчивости показателей.

§ 2. ПОИСК ТОЧЕК ПЕРЕЛОМА В ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДАХ

В случае одномерных рядов решение задачи, вынесенной в название параграфа, существенно облегчается тем, что имеется возможность предварительного визуального анализа графика временного ряда. Такой анализ позволяет выдвинуть априорные предположения о числе, характере и расположении точек перелома зависимости. Формальные методы служат лишь для проверки существенности выдвинутых гипотез.

Самый простой случай — аппроксимация зависимости кусочно-линейным сплайном.

Естественный путь в этой ситуации — это включение всех предположительных точек перелома в качестве узлов, построение кусочно-линейного сплайна и оценка значимости коэффициентов модели с помощью t-критерия Стьюдента. Если один (или несколько) коэффициент оказывается незначимым, то исключается узел, в котором соответствующий β-коэффициент имеет наименьшую значимость. Строится сплайн по оставшимся узлам, и процедура повторяется до тех пор, пока все β-коэффициенты не будут значимыми.

Приведем пример такого анализа. В ранее выполненном исследовании динамики развития лесозаготовительных предприятий (гл. 3, § 1) первоначально на основе визуального анализа графика были отобраны три точки, в которых мог бы произойти перелом зависимости ($t = 6$, $t = 13$, $t = 36$). В результате построения кусочно-линейного сплайна получены следующие характеристики значимости коэффициентов регрессии и уравнения в целом:

$$y^* = 0{,}19 + 0{,}14t - 0{,}14(t - 6) - 0{,}007(t - 13) - \\ -0{,}02(t - 36);$$

$$(3{,}11)(10{,}55)(7{,}46)\ \ (0{,}88)\ \ (3{,}64)$$

$$R^2 = 0{,}86;\ \sigma_{ост} = 7{,}07 \cdot 10^{-2}.$$

Как видно, коэффициент модели в узле ($t = 13$) является несущественным, и поэтому эту точку перелома следует исключить из дальнейшего анализа.

Приведем кусочно-линейный сплайн без указанного узла:

$$y^* = 0{,}17 + 0{,}15t - 0{,}15(t - 6) - 0{,}03(t - 36);$$

$$(2{,}99)\ \ (13{,}14)\ \ \ (12{,}47)\ \ \ \ (4{,}29)$$

$$R^2 = 0{,}85;\ \sigma_{ост} = 7{,}05 \cdot 10^{-2}.$$

Таким образом, без ухудшения точности аппроксимации получен сплайн, в котором все коэффициенты значимы.

Описанная процедура, являясь наглядной и легко реализуемой, в то же время не совсем строга. Это обусловлено тем, что она требует включения всех предположительных точек перелома, что при ограниченной выборке является затруднительным и, кроме того, процедура существенно зависит от порядка исключения незначимых коэффициентов. По су-

ществу, она представляет собой аналог метода пошаговой регрессии.

Разработка более тонких и строгих приемов поиска точек перелома зависимости является предметом активных поисков.

Здесь можно упомянуть исследования Пауэлла [Powell, 1968], де Боора и Райса [Boor, Rice, 1968], Хадсона [Hudson, 1966], касающиеся вычислительных проблем. Статистические же подходы к этой проблеме представлены в работах Спрента [Sprent, 1961], Робинсона [Robinson, 1964], Маленво [1976] и др. Наиболее полно и обобщенно эти результаты представлены в работах Федера [Feder, 1975a, b]. Однако, несмотря на интересные и глубокие теоретические результаты, их практическое использование пока затруднительно. Почти единственный прикладной результат, описанный Федером, заключается в проверке числа существенных сегментов.

Все вышесказанное относилось к случаю «гладкого переключения» и, следовательно, к возможности построения сплайн-функций.

Если априорно такую гипотезу принять не представляется возможным, то альтернативой является апроксимация зависимости разрывными кусочно-простыми функциями. В этой ситуации можно рассмотреть два наиболее распространенных метода проверки гипотезы о наличии перелома зависимости: дисперсионный двух или многих регрессий (сравнение существенности различия) и байесовский. Практически наиболее легко реализуемым является первый, дисперсионный, метод.

По своему содержанию процедура выделения периодов со стабильной структурой связей во многом совпадает с проблемой анализа устойчивости параметров моделей, построенных на базе временных рядов. Идейно она заключается в последовательном увеличении «длины» информационной базы (или ее перемещении во времени), на которой строится модель одной и той же структуры с анализом идентичности полученных на каждом шаге оценок параметров. Если на каком-то шаге наблюдаются статистически значимые различия в оценках параметров модели, то это является сигналом о возможном резком изменении структуры влияния.

Реализация этих методов основана на идеях дисперсионного анализа отклонений.

Для проверки устойчивости во времени значений параметров модели можно использовать методы проверки идентичности двух или нескольких регрессий. Проверяются следующие гипотезы относительно величины коэффициентов

регрессий (b) и свободного члена (a)[2]:

1. $a^{(1)} = a^{(2)}$, $b^{(1)} = b^{(2)}$ — регрессии идентичны по сдвигу и наклону;
2. $a^{(1)} \neq a^{(2)}$, $b^{(1)} = b^{(2)}$ — регрессии различаются только по сдвигу;
3. $a^{(1)} = a^{(2)}$, $b^{(1)} \neq b^{(2)}$ — регрессии различаются по наклону;
4. $a^{(1)} \neq a^{(2)}$, $b^{(1)} \neq b^{(2)}$ — регрессии различны по сдвигу и наклону.

Гипотеза об идентичности двух уравнений регрессии (или, что то же, об отсутствии структурного изменения) может проверяться методом, применяемым к общим линейным гипотезам [Chow, 1960].

При верной нулевой гипотезе (регрессии идентичны) статистика

$$F = \frac{S_1 - (S_2 + S_3)}{S_2 + S_3} \times \frac{n + m - 2p}{p}$$

имеет F-распределение с p и $(n + m - 2p)$ степенями свободы, где

S_1 — сумма квадратов отклонений от регрессии, рассчитанной по объединенной совокупности $n + m$ наблюдений;

S_2 — сумма квадратов отклонений от регрессии, рассчитанной по выборке из n наблюдений;

S_3 — сумма квадратов отклонений от регрессии, рассчитанной по m наблюдениям;

p — число оцениваемых параметров регрессии.

Полученное значение F сравнивается с соответствующим табличным значением $F^{\alpha}_{p,n+m-2p}$ для фиксированного уровня значимости α и числа степеней свободы p и $(n + m - 2p)$. Если $F > F^{\alpha}_{p,n+m-2p}$, то гипотеза об идентичности регрессии для двух групп наблюдений отвергается.

Метод имеет тот недостаток, что в случае, если нулевая гипотеза отвергается, то неизвестно, чем обусловлено различие уравнений регрессии — их сдвигом или различием коэффициентов регрессии и каких именно.

В случае временных рядов проверка гипотез сводится к образованию непрерывных сегментов и проверке идентичности уравнений, построенных по этим сегментам. Поскольку

[2] Индексы (1) и (2) при коэффициентах означают принадлежность к первой и второй выборкам соответственно.

ку минимальная величина сегмента m определяется числом параметров и равна $p + 1$, то точка перелома зависимости при больших p находится в пределах некоторого достаточно протяженного временного отрезка. Для ее локализации может быть использован прием, основанный на смещении точек начала отсчета интервалов. Однако имеется дополнительная трудность, связанная с произвольностью выбора краевых точек отсчета для первоначального интервала. Кроме того, выбрав первоначальный интервал, мы априорно предполагаем отсутствие на нем точек перелома зависимости.

Метод фиктивных переменных [Gujarati, 1970] свободен от некоторых из указанных выше недостатков и состоит в построении уравнения регрессии по объединенной совокупности двух выборок. Общий вид регрессионного уравнения для объединения совокупности следующий:

$$y^* = a_0 + Da_1 + \sum_{j=1}^{p} c_j x_j + \sum_{j=1}^{p} Db_j x_j$$

($D = 1$, если наблюдение принадлежит первой совокупности, и $D = 0$ — для второй). Если a_1 статистически значим, то величина сдвига для первой выборки равна $a_0 + a_1$, для второй — a_0. Если a_1 статистически незначим, то a_0 дает оценку общего для обеих совокупностей сдвига. Если b_j статистически значим, то значение соответствующего коэффициента для первой выборки равно $b_j + c_j$, а для второй — c_j. Если b_j статистически незначим, то c_j дает оценку общего для обеих выборок значения коэффициентов.

При помощи рассматриваемого метода можно установить, по каким параметрам различаются две регрессионные модели.

Приведем пример использования дисперсионного критерия для проверки гипотезы о наличии структурного изменения в экономических показателях.

На рис. 3.5 приведена динамика добычи нефти на двух месторождениях за период с начала эксплуатации. Как видно, по месторождению № 1 можно предположить наличие двух точек перелома зависимости ($t = 3$ и $t = 8$).

Исходная информация для проверки нулевой гипотезы по F-критерию приведена в табл. 3.4. Рассчитаем F-критерий для точки перелома $t = 3$. Сумма квадратов отклонений от регрессии, рассчитанной по первым трем точкам, $S_2 = 80782{,}41$. Сумма квадратов отклонений от регрессии, рассчитанной по последующим пяти точкам, $S_3 = 16935{,}53$. Сумма квадратов отклонений от регрессии, рассчитанной по объединенной выборке из 8 наблюдений, $S_1 = 451239{,}93$.

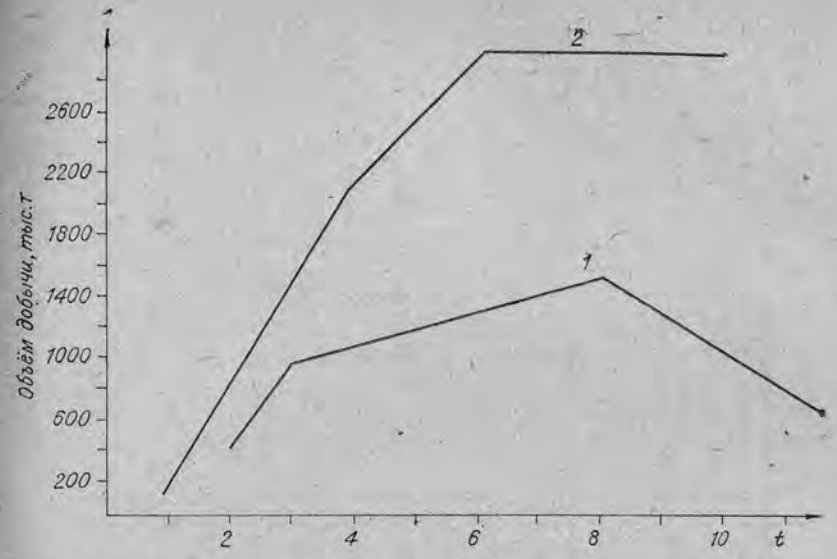


Рис. 3.5. Динамика добычи нефти: *1* — месторождение № 1, *2* — месторождение № 2.

Величина F-критерия составляет

$$F = \frac{451239{,}93 - (80782{,}41 + 16935{,}53)}{80782{,}41 + 16935{,}53} \times \frac{(3+5) - (2\times 2)}{2} = 7{,}24.$$

Соответствующее табличное значение F-критерия для $\alpha = 0{,}05$ равно $F_{2,4}^{0,05} = 6{,}94$.

Сравнивая F с $F_{2,4}^{0,05}$, видим, что $F > F_{2,4}^{0,05}$. Следовательно, есть основание утверждать, что в точке $t = 3$ существует перелом зависимости.

Аналогичные расчеты для точки перелома $t = 8$ приводят к следующему значению F-критерия:

$$F = \frac{430937{,}21 - (16935{,}53 + 12867{,}49)}{16935{,}53 + 12867{,}49} \times \frac{(5+4) - (2\times 2)}{2} = 33{,}65;$$

$$F_{2,5}^{0,05} = 5{,}79;\ F \gg F_{2,5}^{0,05}.$$

Следовательно, и в точке $t = 8$ есть перелом зависимости.

Таким образом, в результате проверки отвергнута нулевая гипотеза об отсутствии структурных изменений в обеих точках — кандидатах.

Кусочно-линейное уравнение, описывающее изменение динамики нефтедобычи на месторождении № 1, имеет следу-

Таблица 3.4

Проверка гипотезы о наличии структурных изменений (месторождение № 1)

Год	t	Объем добычи, тыс. т	Точка перелома $t=3$				Точка перелома $t=8$			
			Кусочно-линейное уравнение		Общее уравнение		Кусочно-линейное уравнение		Общее уравнение	
			$y_{\text{расч}}$	Δy	$y_{\text{расч}}$	Δy	$y_{\text{расч}}$	Δy	$y_{\text{расч}}$	Δy
1964	1	15,7	—100,30	116,03	238,23	—222,53				
1965	2	150,1	382,17	—232,07	450,49	—300,39				
1966	3	980,7	864,67	116,03	662,76	317,94				
1967	4	1266,4	1188,10	78,30	875,03	391,37	1188,10	78,30	1463,78	—197,38
1968	5	1202,4	1264,33	—61,93	1087,30	115,10	1264,33	—61,93	1388,64	—186,24
1969	6	1290,5	1340,56	—50,06	1299,56	—9,06	1340,56	—50,06	1313,50	—23,00
1970	7	1389,5	1416,79	—27,29	1511,83	—122,33	1416,79	—27,29	1288,36	151,14
1971	8	1554,0	1493,02	60,98	1724,10	—170,10	1493,02	60,98	1163,22	390,78
1972	9	1386,4					1325,73	60,67	1088,08	298,32
1973	10	998,7					1069,61	—70,91	1012,94	—14,24
1974	11	773,8					813,49	—40,19	937,80	—164,50
1975	12	607,8					557,37	50,43	862,66	—254,86

ющий вид:

$$y^*=\begin{cases}-582{,}83+482{,}50t, & 0<t\leqslant 3;\\ R^2=0{,}85; \quad \sigma_{\text{ост}}=284{,}22;\\ 883{,}18+76{,}23t, & 3<t\leqslant 8;\\ R^2=0{,}77; \quad \sigma_{\text{ост}}=75{,}14;\\ 2862{,}45-256{,}12t, & 8<t\leqslant 12;\\ R^2=0{,}96; \quad \sigma_{\text{ост}}=80{,}20.\end{cases}$$

Поскольку в данном примере нет априорной информации о разрывах зависимости в точках структурных изменений, то представляет интерес построение кусочно-линейного сплайна с узлами в установленных точках и сравнение его характеристик с характеристиками кусочно-линейной регрессии (табл. 3.5)

$$y^*=\begin{cases}-665{,}48+544{,}49t, & 0<t\leqslant 3;\\ -665{,}48+544{,}49t-427{,}33\,(t-3){,}3 & <t\leqslant 8;\\ -665{,}48+544{,}49t-427{,}33(t-3)-363{,}95(t-8);8<t\leqslant 12;\end{cases}$$

$$R^2=0{,}95; \quad \sigma_{\text{ост}}=133{,}66.$$

Как видно из табл. 3.5, в данном конкретном примере качество аппроксимации зависимости кусочно-линейной регрессией и сплайн-функцией примерно совпадают. Однако при отсутствии разрывов и скачков зависимости использование непрерывной аппроксимации при помощи сплайнов предпочтительнее, поскольку при его построении используется вся выборочная информация.

Следующий пример иллюстрирует это утверждение.

Анализ динамики добычи нефти на месторождении № 2 (см. рис. 3.5) позволяет выдвинуть гипотезу о наличии изменений зависимости в узлах, соответствующих $t=4$ и

Таблица 3.5

Распределение отклонений

Вид функции	Абсолютные отклонения				
	До 80	80—160	160—240	240—320	Свыше 320
Кусочно-линейная регрессия	9	2	1	0	0
Кусочно-линейный сплайн	10	1	0	1	0

Таблица 3.6

Проверка гипотезы о наличии структурных изменений (месторождение № 2)

Год	t	Объем добычи, тыс. т	Точка перелома $t=4$			
			Кусочно-линейное уравнение		Общее уравнение	
1966	1	188,4	149,77	38,63	304,06	—115,66
1967	2	811,9	859,14	—47,24	875,06	—63,15
1968	3	1547,1	1678,51	—21,41	1446,05	101,05
1969	4	20307,9	2277,88	30,02	2017,05	290,85
1970	5	2534,1	2613,97	—79,93	2588,05	—53,95
1971	6	2999,9	2959,97	39,93	3159,04	—159,14

$t = 6$. Проверим справедливость этого предположения для $t = 4$. Построение кусочно-линейной регрессии и проверка гипотезы о наличии структурного изменения зависимости приводят к результатам, показанным в табл. 3.6. Поступая так же, как и в случае месторождения № 1, определим F, исходя из рассчитанных значений $S_1 = 140407{,}29$; $S_2 = 5083{,}48$; $S_3 = 7983{,}21$:

$$F = \frac{140407{,}29 - (5083{,}48 + 7983{,}21)}{5083{,}48 + 7983{,}21} \times \frac{(4+2)-(2\times 2)}{2} = 9{,}75;$$

$$F_{2,2}^{0,05} = 19{,}00;\ F < F_{2,2}^{0,05}.$$

Сравнивая F с $F_{2,2}^{0,05}$, отмечаем, что результаты проверки не позволяют отвергнуть нулевую гипотезу об отсутствии перелома зависимости в точке $t = 4$.

Рассмотрим характеристики кусочно-линейной регрессии и сплайн-функции.

Кусочно-линейная регрессия:

$$y^* = \begin{cases} -559{,}6 + 709{,}37t, & 0 < t \leqslant 4; \\ R^2 = 0{,}99; \quad \sigma_{ост} = 50{,}41; \\ 883{,}97 + 346{,}00t, & 4 < t \leqslant 6; \end{cases}$$

$$R^2 = 0{,}96; \quad \sigma_{ост} = 97{,}82.$$

Кусочно-линейный сплайн:

$$y^* = \begin{cases} \underset{(24{,}53)}{-535{,}71 + 695{,}04t}, & 0 < t \leqslant 4; \\ \underset{(5{,}22)}{-535{,}71 + 695{,}04t - 309{,}83\,(t-4)}, & 4 < t \leqslant 6; \end{cases}$$

$$R^2 = 0{,}99; \quad \sigma_{ост} = 66{,}22.$$

Характеристики кусочно-линейного сплайна говорят о существенности перелома в точке $t = 4$.

В данном примере мы сталкиваемся с ситуацией, когда проверка на наличие перелома зависимости с помощью F-критерия и оценки значимости β-коэффициентов кусочно-линейного сплайна оказалась противоречивой.

Выше мы отмечали, что изложенный классический подход к проверке гипотез о наличии структурных изменений не лишен ряда недостатков. К одному из них, в частности, относится недоучет того факта, что в распоряжении исследователя имеется некоторая (не обязательно точная) информация о наличии структурного изменения. В этой ситуации полезно рассмотреть другой подход (байесовский), в рамках которого такая информация может быть учтена в вероятностном виде. *Байесовские методы* применимы как в непрерывных случаях, так и при наличии разрывов и скачков зависимости.

Байесовскому анализу структурных изменений посвящен ряд исследований [Пуарье, 1981; Ferreira, 1975; Halpern, 1973].

Следуя Гальперну [Halpern, 1973], рассмотрим базовую регрессионную модель с линейными сплайнами:

$$y_i = \beta_0 \omega_{i0} + \beta_1 \omega_{i1} + \ldots + \beta_k \omega_{ik} + \varepsilon_i,$$

где $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m) \sim N(0,\ \sigma^2 I)$ и $\omega_0 = 1$; $\omega_1 = x_1$;

$$\omega_{ij} = \begin{cases} x_i - x_{j-1}, & \text{если } x_i > \bar{x}_{j-1}; \\ 0, & \text{если } x_i \leqslant x_{j-1}, \end{cases}$$

с $x_1 < \bar{x}_2 \ldots < \bar{x}_{k-1}$ возможными узлами.

Пусть $S_1, S_2, \ldots, S_l$ есть $l = 2^{k-1}$ подмножеств множества $\{1, 2, \ldots, k-1\}$ и m_i — число элементов множества S_i. Обозначим расположенные в порядке возрастания элементы S_i через $\tau(i, 1), \tau(i, 2), \ldots, \tau(i, m_i)$. Пусть $H_i (i = \overline{1, l})$ обозначает нулевую гипотезу, в соответствии с которой реальные узлы находятся в точках $\bar{x}_{\tau(i,1)}; \bar{x}_{\tau(i,2)}, \ldots, \bar{x}_{\tau}$. Гипотезу H_i можно записать следующим образом: $\beta_{j+1} \neq 0$ лишь для тех j, при которых $j \in S_i$.

В случае $i = 0$ гипотеза H_0 описывает ситуацию, когда сплайн состоит всего из одного куска. Запишем:

$$\beta_i = \begin{cases} [\beta_0,\ \beta_1]', & \text{если } m_i = 0; \\ [\beta_0, \beta_1, \beta_{\tau(i,1)}, \ldots, \beta_{\tau(i,m_i)}], & \text{если } m_i > 0. \end{cases}$$

Тогда при условии справедливости гипотезы H_i β_i будет ненулевым вектором.

Пусть W — матрица $m \times (k+1)$ вида

$$W = \begin{pmatrix} \omega_{10}, \omega_{11}, \ldots, \omega_{1k} \\ \omega_{20}, \omega_{21}, \ldots, \omega_{2k} \\ \omega_{m0}, \omega_{m1}, \ldots, \omega_{mk} \end{pmatrix},$$

где ω_{ij} были определены выше. Обозначим через $Z_i (i = \overline{1, k})$ матрицу размерности $(k+2) \times (m_i + 2)$, общий элемент которой

$$z^i_{iq} = \begin{cases} 1, \text{ если } j = q = 1,\ j = q = 2 \text{ или если } j = \tau(i, q-2)+1; \\ 0, \text{ в другом случае.} \end{cases}$$

Вводя $y = (y_1, y_2, \ldots, y_m)$, можно сказать, что модель соответствует гипотезе H_i, если

$$y \sim N(Wz_i\beta_i,\ \sigma^2 I).$$

Априорную неопределенность в значениях β_i будем представлять с помощью равномерного распределения.

Гальперну удалось доказать теорему, в которой содержится все необходимое для анализа общей линейной сплайновой регрессии, в которой неизвестно ни число узлов, ни их расположение. В теореме дается аналитическое выражение для апостериорного распределения β_i, а также апостериорная вероятность того, что истинное расположение узлов описывается гипотезой H_i.

Апостериорное распределение $f(H_i/y)$ можно использовать при выборе наиболее вероятной модели, а с помощью апостериорного распределения $f(\beta_i/y)$ можно сделать некоторые выводы о коэффициентах регрессии.

Сравнивая классический и байесовский подходы к проблеме оценивания параметров модели с неизвестными точками структурных изменений, следует отметить сложность идентификации точек переключения для классических методов, что может привести к неудаче в их применении.

Использование же байесовских методов может быть успешным при условии, что исходная модель идентифицирована с помощью информативных априорных распределений. Кроме того, байесовский анализ позволяет учитывать априорную информацию о положении точек структурных изменений в вероятностных терминах.

Далее, приведенные конкретные результаты по одной из теоретических моделей (например, у Федера [Feder, 1975]) носят асимптотический характер. При этом они дают плохое приближение для относительно малых выборок, используемых в экономическом анализе. С другой стороны, хотя байесовские апостериорные распределения почти никогда не имеют простого известного вида, они являются точными результатами для конечных выборок.

Остановимся еще на одном способе оценивания структурных сдвигов с помощью моделей с неизменными параметрами. В излагаемом ниже подходе применяется аппарат математического программирования к семейству кусочно-регрессионных моделей. В нем предполагается, что точки, в которых происходят изменения параметров, неизвестны. Рассмотрим линейную модель с двумя режимами (см. [Ginsburg a. o., 1980]):

$$y^{*t} = \begin{cases} a_1' x_1^t + \varepsilon^t, \text{ если } t \in I_1; \\ a_2' x_2^t + \varepsilon^t, \text{ если } t \in I_2, \end{cases} \tag{3.1}$$

где $a_1 \in R^{n_1}$ и $a \in R^{n_2}$ — векторы параметров, которые могут быть зависимыми или независимыми друг от друга;

$x_1^t \in R^{n_1}$ и $x_2^t \in R^{n_2}$ — векторы независимых переменных;

ε^t — случайная ошибка;

I_1, I_2 — множества двух различных режимов.

Линейность и непрерывность в независимых переменных и параметрах позволяют переписать (3.1) в виде

$$y^{*t} = \max\{a_1' x_1^t, a_2' x_2^t\} + \varepsilon^t \tag{3.2}$$

или

$$y^{*t} = \min\{a_1' x_1^t, a_2' x_2^t\} + \varepsilon^t. \tag{3.3}$$

Заметим, что y^{*t} выпукла в (3.2) и вогнута в (3.3). Обе модели — кусочно-линейные. Совершенно ясно, что операторы max (min) дают недифференцируемые модели и делают задачу оценивания чрезвычайно трудной.

Альтернативный способ записи

$$y^{*t} = \begin{cases} a_1' x_1^t + \varepsilon^t, \text{ если } a_1' x_1^t > a_2' x_2^t; \\ a_2' x_2^t + \varepsilon^t, \text{ если } a_1' x_1^t \leqslant a_2' x_2^t. \end{cases} \tag{3.4}$$

Предполагая, что $\varepsilon^t \sim N(0, \sigma^2)$, оценку метода максимального правдоподобия для (3.2) — (3.4) можно найти из

$$\min \sum_{t=1}^{\omega} (\varepsilon^t)^2, \tag{3.5}$$

где ω — число наблюдений и минимизация проводится по a_1 и a_2. Мы уже упоминали, что модели (3.2) — (3.4) не явля-

ются непрерывными и поэтому (3.5) имеет не непрерывную первую производную. Следовательно, прямое решение, основанное на классических методах оптимизации, невозможно.

Другое «свойство», которое усложняет метод,— невыпуклость. И поэтому, как всегда при невыпуклой оптимизации, трудно найти глобальный минимум.

Для того чтобы найти решение (3.4), можно проверить все возможные распределения наблюдений между режимами, т. е. для всех номеров множеств I_1 и I_2, удовлетворяющих

$$I_1 \cap I_2 = \varnothing \tag{3.6}$$

и

$$I_1 \cup I_2 = \{1, 2, \ldots, \omega\}, \tag{3.7}$$

мы можем оценить параметры a_1 и a_2 модели

$$y^{*t} = \delta^t a_1' x_1^t + (1 - \delta^t) a_2' x_2^t + \varepsilon^t,$$

где $\delta^t = 1$, если $t \in I_1$, и $\delta^t = 0$ — в противном случае, и сохранить наилучшее решение. Однако, проведя 2^ω возможных перестановок номеров, удовлетворяющих (3.6) и (3.7), мы получим в результате огромное число регрессионных проблем, и метод нельзя будет реализовать даже при наилучших вычислительных возможностях.

В работе [Ginsburg e. a., 1980] для решения задачи (3.4) предлагается простой комбинированный подход. Однако, как замечают сами авторы, он осуществим, если $n \leqslant 3$, т. е. для задач малой размерности. Предложенные алгоритмы, однако, обладают замечательным свойством: они находят глобальное решение задачи невыпуклого оценивания за конечное число шагов. И именно с этой точки зрения они являются наилучшими из классических методов.

Рассмотрим один из алгоритмов.

Решим (3.2) или (3.3) при $n_1 = n_2$. В этом случае выражение, соответствующее методу наименьших квадратов (3.5), эквивалентно следующей нелинейной целочисленной оптимизационной проблеме:

$$\min \sum_{t=1}^{\omega} [y^t - \delta^t a_1' x_1^t - (1 - \delta^t) a_2' x^t]^2 \tag{3.8}$$

при

$$\delta^t (a_1' x_1^t - a_2 x_2^t) \geqslant 0;$$
$$(1 - \delta_2^t)(-a_1' x_1^t + a_2' x_2^t) \geqslant 0;$$
$$\delta^t \in \{0, 1\};\ t = 1, 2, \ldots, \omega,$$

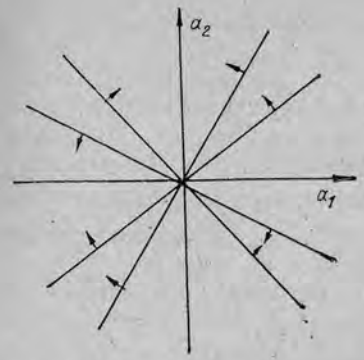


Рис. 3.6. Допустимая точка (0, 0).

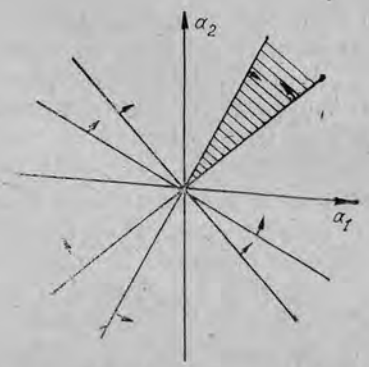


Рис. 3.7. Допустимый конус.

где минимизация происходит по a_1, a_2 и δ^t, $t = 1, \ldots, \omega$. Если зафиксировать δ^t, то мы сталкиваемся с ω линейными ограничениями; каждое ограничение представляет собой луч, делящий пространство значений (a_1, a_2) на два полупространства. Если принять очень слабое ограничение, что $x_1^i/x_2^i \neq x_1^j/x_2^j$ для всех $i \neq j$, то возникают две возможности: либо допустимой точкой является (0, 0), так как ограничения направлены в противоположные направления, либо допустимое множество — конус, определенный двумя ограничениями, все другие ограничения направлены в «допустимом» направлении. Эти две ситуации показаны соответственно на рис. 3.6 и 3.7.

Это замечание приводит к следующему простому алгоритму:

Шаг 0. Отправляясь от

$$S^* = \sum_{i=1}^{\omega} (y^t)^2,\ a_1^* = a_2^* = 0,$$

переупорядочим наблюдения в порядке убывания величины x_1^t/x_2^t. Если $x_2^t = 0$, то для некоторого t положим, что отношение равно некоторому достаточно большому числу.

Шаг 1. ($j = 1, \ldots, \omega$).

а) Возьмем (переупорядоченные вновь) наблюдения j и $j + 1$ (если $j = \omega$, возьмем первое наблюдение как $j + 1$) как связанные ограничения; зафиксируем все δ^t так, чтобы все ограничения указывали верное направление.

б) Вычислим значение

$$S = \min \sum_{i=1}^{\omega} [y^t - \delta^t a' x_1^t - (1 - \sigma^t) a_2' x_2^t]^2$$

с зафиксированными ранее δ^t, удовлетворяющее только двум ограничениям, полученным на шаге а).

Поскольку эта задача минимизации двумерна, выпукла и квадратична, с двумя линейными ограничениями, то решение ее есть решение системы Куна — Таккера. Проверим, является ли $S < S^*$. Если да, то положим $S^* = S$ и используем новые значения a_1^* и a_1^*, полученные из решения (3.8).

в) Установим $\delta^t = 1 - \delta^t (t = 1, 2, \ldots, \omega)$; это изменит все ограничения. Повторим шаг б) с новыми δ^t.

Завершение. Если $j = \omega + 1$, то конец; решение — S^*, a_1^* и a_2^*.

Основная особенность алгоритма, несмотря на его простоту, в том, что после ω шагов он дает глобальное решение (3.8).

§ 3. МНОГОМЕРНЫЕ ВРЕМЕННЫЕ РЯДЫ

Если можно считать, что параметры моделей, построенных по многомерным временным рядам, изменяются сравнительно медленно, то для описания этого процесса пригодна множественная регрессионная модель с переключениями. Сам процесс изменения параметров моделей представляет собой неслучайную функцию времени и на определенных интервалах временной оси эти параметры стабильны.

В тех случаях, когда такие участки стабильности априори могут быть выделены (известные точки перелома), то задача сводится к проверке существенности различий уравнений, построенных на разных сегментах временной оси. Следует подчеркнуть, что в многомерном случае применение сплайнов затруднено из-за громоздкости и сложности вычисления и трудности обоснования гипотезы о непрерывности функции в точках переключения. Поэтому основной тип модели в многомерном случае — кусочно-линейный. Сама же процедура проверки существенности различий уравнений, построенных на разных сегментах, аналогична описанной в предшествующем параграфе. Наиболее распространенным является дисперсионный подход. Его использование покажем на наглядном примере.

По одному из тракторных заводов страны имеются помесячные данные об объемах производства и себестоимости продукции за три года (1978—1980). Выдвинута гипотеза о

Таблица 3.7

Оценка существенности различий годовых уравнений регрессии

Год	Параметры уравнений				
	$\overline{R}^2$	$\sigma_{ост}$	a_0	b	$\Sigma (\Delta y)^2$
1978	0,83	82,14	7986,78	—0,81	67400,45
1979	0,63	177,48	7930,73	—0,78	314761,16
1978—1979	0,67	136,64	7799,75	—0,75	410497,21
1980	0,54	171,57	6921,58	—0,52	294261,94
1978—1980	0,62	145,02	7449,79	—0,66	714412,93

том, что погодовые зависимости себестоимости от объемов выпуска различны. Основанием для выдвижения этой гипотезы явилось то, что рассматриваемое предприятие в данный период находилось на стадии освоения производственных мощностей. Расчеты приведены в табл. 3.7. Для 1978—1979 гг. значение F-критерия рассчитывается следующим образом:

$$F = \frac{410597{,}21 - (67400{,}45 + 314761{,}16)}{67400{,}45 + 314761{,}16} \times \frac{(12+12) - (2\times 2)}{2} = 0{,}74;$$

$$F_{2,20}^{0,05} = 3{,}49;\ F < F_{2,20}^{0,05}.$$

Аналогичный расчет для 1978—1980 гг.:

$$F = \frac{714412{,}93 - (410497{,}21 + 294261{,}94)}{410497{,}21 + 294261{,}94} \times \frac{(24+12) - (2\times 2)}{2} = 16{,}22;$$

$$F_{2,32}^{0,05} = 3{,}32;\ F > F_{2,32}^{0,05}.$$

Проверка нулевой гипотезы показывает, что для 1978—1979 гг. может быть построено общее уравнение регрессии, а зависимость для 1980 г. существенно отличается, и для ее описания должно быть использовано другое уравнение регрессии

$$y = \begin{cases} 7799{,}75 - 0{,}75x, & \text{для 1978 — 1979 гг.;} \\ 6921{,}58 - 0{,}52x, & \text{для 1980 г.,} \end{cases}$$

где y — себестоимость продукции, руб./шт;

x — объем производства, шт.

Анализ модели свидетельствует о существенном снижении влияния объема производства по мере освоения производственной мощности.

Однако случаи априорно известных точек структурных изменений для многомерных временных рядов сравнительно редки и возникает задача поиска этих точек. Для многомер-

ного случая решение задачи осложняется двумя обстоятельствами. Во-первых, исключена возможность визуального анализа графиков, и поэтому сложно выдвинуть гипотезу о вероятном местоположении точки структурного изменения. Во-вторых, имеет место естественная упорядоченность наблюдений во времени, которую нельзя нарушать. Вследствие этого прямое использование методов многомерной классификации может привести к нарушению хронологической последовательности в выделенных группах.

Для выдвижения априорных гипотез о наличии структурных переломов можно предложить процедуру исследования характера поведения отклонений расчетных значений моделируемого показателя от фактических во времени. Если зависимость y от вектора x может быть описана единым уравнением регрессии, то расположение отклонений $\Delta y = y_{\text{расч}} - y_{\text{ср}}$ во времени не должно носить систематического характера, а разброс точек носит случайный характер и определяется точностью аппроксимации. Гипотеза о том, что совокупность отклонений Δy представляет собой упорядоченный ряд, может быть проверена соответствующими статистическими методами (метод серий, расчет коэффициентов автокорреляций, построение зависимости $\Delta y = f(t)$ и т. п.). Если такая гипотеза подтверждается, то это свидетельствует о вероятном структурном изменении и могут быть ориентировочно намечены области изменения структуры. Если отклонения группируются сериями постоянного знака, то область перехода от серии одного знака к серии другого знака и есть возможная область изменения структуры.

Рассмотрим пример исследования зависимости технико экономических показателей работы угольных карьеров Кузбасса. Анализировались месячные данные за пятилетний период (60 наблюдений) и устанавливались зависимости пяти экономических показателей (себестоимость 1 т угля; производительность труда; объем работ по добыче угля по вскрыше, полные затраты на добычу угля) от 23 факторов.

Приведем в качестве иллюстрации уравнение регрессии и анализ отклонений по показателю «Производительность труда» для Томусинского карьера.

$$y = \exp\{0{,}50 x_2^{0,07} \cdot x_4^{-0,04} \cdot x_5^{0,01} \cdot x_8^{0,07} \cdot x_{12}^{-0,01} \cdot x_{13}^{-0,04} \cdot x_{14}^{-0,01} \cdot x_{16}^{0,01} \cdot x_{17}^{0,01} \cdot x_{18}^{0,95} \cdot x_{19}^{0,01}\}; \; R^2 = 0{,}98,$$

где y — производительность труда;

x_2 — условная производительность экскаваторов по добыче;
x_4 — объем вскрытия запасов угля;
x_5 — объем готовых к выемке запасов угля;
x_8 — протяженность фронта работ по вскрыше;
x_{12} — среднее расстояние перевозки угля автотранспортом;
x_{13} — среднее расстояние перевозки вскрыши автотранспортом;
x_{14} — общая численность промышленно-производственного персонала;
x_{16} — число условных экскаваторов, занятых добычей угля;
x_{17} — число условных экскаваторов, занятых вскрышей;
x_{18} — количество угля, вывезенного автотранспортом;
x_{19} — количество породы, вывезенной автотранспортом.

Распределение отклонений и их накопленной суммы во времени представлено на рис. 3.8.

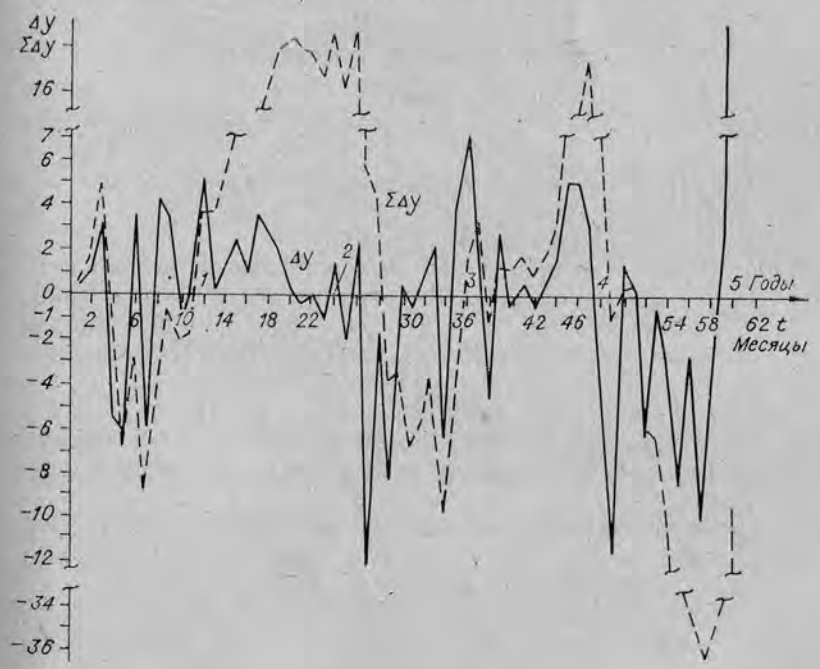


Рис. 3.8. Распределение отклонений (Δy) и их накопленной суммы $\left(\sum \Delta y\right)$.

Анализируя график визуально, можно прийти к следующему выводу: несмотря на высокий коэффициент детерминации ($R^2 = 0{,}98$), существуют ярко выраженные серии отклонений одинакового знака. Так, например, с 10-го по 24-й месяц все отклонения положительны, а их накопленная сумма переходит через 0 только в 11-м и 28-м месяцах, т. е. почти в течение года расчет по модели показывал завышенные значения производительности труда. Также можно выделить две, не столь явно выраженные серии отрицательных отклонений (с 26-го по 34-й и с 47-го по 58-й месяцы). Такой характер графика отклонений позволяет предположить наличие структурных изменений в исследуемом периоде времени.

Для проверки гипотезы H_0 о том, что обнаруженная несимметричность не является случайной, воспользуемся критерием количества серий [Большев, Смирнов, 1965, с. 146]. Серией называется часть последовательности отклонений, каждая из которых состоит из отклонений одного знака. Гипотеза H_0 принимается, если удовлетворяется следующее неравенство:

$$g(Q, m, n) < \gamma < G(Q, m, n),$$

где γ — число серий;
$g(Q, m, n)$ — нижнее критическое значение количества серий;
$G(Q, m, n)$ — верхнее критическое значение количества серий;
Q — уровень значимости;
m — количество положительных отклонений;
n — количество отрицательных отклонений;

В нашем примере $\gamma = 27$; $m = 35$; $n = 25$. Выбрав $Q = 0{,}05$ по таблице [Большев, Смирнов, 1965], определяем g и G [3]:

$g(0{,}05, 25, 25) = 18$; $G(0{,}05, 25, 25) = 33$; $18 < 27 < 33$.

Следовательно, гипотеза H_0 принимается, а выявленная асимметрия свидетельствует о наличии структурных изменений.

После подтверждения гипотезы о наличии структурных изменений встает задача локализации точек перелома. Поскольку многомерные временные ряды представляют собой строго упорядоченные хронологические последовательности, то мы считаем, что для решения задачи весьма эффективным может оказаться алгоритм многомерной группировки, предложенный Г. С. Кильдишевым и Ю. И. Аболенцевым [1978] и описанный нами в § 2 гл. 2.

[3] Поскольку таблицы для $m \neq n$ составлены лишь для $m = 20$, то воспользовались таблицей $m = n$.

Таблица 3.8

Поиск точки структурного изменения методом $U(\rho^2)$-группировки

t	Месяц	Объем производства x	Себестоимость продукции y	Число точек		Значение критерия $U(\rho^2)$
				l	$n-l$	
1	Апрель	4,0	4,9	1	11	0,060
2	Май	3,9	4,9	2	10	0,003
3	Июнь	3,9	4,5	3	9	0,930
4	Июль	3,4	5,3	4	8	1,870
5	Август	3,6	5,1	5	7	4,560
6	Сентябрь	3,8	5,2	6	6	5,690
7	Октябрь	3,9	5,0	7	5	7,950
8	Ноябрь	4,2	4,9	8	4	5,800
9	Декабрь	4,3	4,4	9	3	1,610
10	Январь	4,1	5,1	10	2	1,640
11	Февраль	4,1	4,9	11	1	1,710
12	Март	4,1	4,6			

Действительно, этот алгоритм позволяет решить две задачи: 1) получить статистическую оценку однородности (неоднородности) совокупности и 2) определить оптимальную границу (скачок) между группами.

Для пояснения техники поиска точки перелома приведем иллюстративный пример, относящийся к исследованию зависимости себестоимости продукции от объемов производства. Исходные данные и расчет $U(\rho^2)$ критерия для первой итерации приведен в табл. 3.8.

Из таблицы видно, что не для всех значений критерия однородности выполняется условие

$$U(\rho^2) \leqslant \chi^2_{0{,}05;\,2} = 5{,}99.$$

В частности, это имеет место для октября (№ 7). Следовательно, нулевая гипотеза должна быть отклонена, т. е. совокупность является неоднородной и должна быть разбита на два класса по максимальному значению $U(\rho^2) > \chi^2_{\text{табл}}$. В данном случае это значение 7,95. Поскольку в данном примере нет других значений $U(\rho^2)$, превосходящих критическое значение χ^2, и число наблюдений невелико, то можно ограничиться одной итерацией для локализации точек перелома и описать зависимость двумя кусочно-линейными функциями:

$$y = \begin{cases} g_1(x), & \text{для } 1 < t < 7; \\ g_2(x), & \text{для } t > 7. \end{cases}$$

При этом можно отметить такой характерный момент: найденная точка перелома делит календарный год на два периода (зимний и летний).

Остановимся еще на одном способе поиска точек структурных изменений, предложенном в работе [Mac Gee, Cartetion, 1970]. Его сущность заключается в последовательном образовании совокупностей минимального размера (если число оцениваемых параметров уравнения p, то минимальный размер совокупности $p + 1$), оценивании параметров соответствующих уравнений, в проверке гипотезы о возможности присоединения к фиксированной группе дополнительных точек. Данная процедура является многоуровневой и многошаговой и включает следующие действия.

На нулевом уровне все изменяющиеся наблюдения (n) рассматриваются как самостоятельные совокупности. На первом шаге процесса образуются последовательные группы из $p + 1$ наблюдения. Каждая группа сдвигается на одно наблюдение и общее число образованных групп равно $k = n - p$. Пусть, например, $n = 10$, $p = 2$ и есть последовательность наблюдений $y_1, y_2, \ldots, y_{10}; x_1, x_2, \ldots, x_{10}$, тогда соответствующие группы запишем так:

Номер группы	Состав
1	$y_1, y_2, y_3; x_1, x_2, x_3$
2	$y_2, y_3, y_4; x_2, x_3, x_4$
r	. . .
8	$y_8, y_9, y_{10}; x_8, x_9, x_{10}$

Общее число групп — 8. По каждой из них строится уравнение регрессий и рассчитываются значения отклонений $\Delta y_i = y_{i\text{ф}} - y_{i\text{расч}}$.

Вычисляется мера сходства [4]

$$\varepsilon_1 = \sum_{i=1}^{p} \varepsilon_i^2;\ \varepsilon_2 = \sum_{i=2}^{p+1} \varepsilon_i^2;\ \ldots,\ \varepsilon_r = \sum_{i=r}^{r+l} \varepsilon_i^2;\ \ldots,\ \varepsilon_k = \sum_{i=k}^{n} \varepsilon_i^2,$$

где

$$\varepsilon_i^2 = (\Delta y_i)^2.$$

[4] Приведена мера сходства для групп минимального размера. В общем случае, если размер группы $l > p + 1$, то мера сходства рассчитывается по формуле $\varepsilon_r = \sum_{i=r}^{r+l} \varepsilon_i^2/(l - p);\ (r = 1, 2, \ldots, k)$.

В качестве первоначального «центра кристаллизации» выбирается подсовокупность, у которой соответствующее ε_i минимально.

На следующем шаге делается попытка присоединить к этой группе одну точку (предшествующую или последующую). Возможность присоединения проверяется по F-критерию, рассчитанному следующим образом:

$$F_{1,l-p} = \frac{(y_0 - \hat{y}_0^*)^2 + \sum_{i=r}^{r+l} (\hat{y}_i - \hat{y}_i^*)^2}{\varepsilon_r},$$

где y_0 — фактическое значение дополнительного наблюдения;

$\hat{y}_0^*$ — расчетное значение для дополнительной точки по уравнению, построенному по $l + 1$ последовательному наблюдению;

$\hat{y}_i$ — расчетное значение для фиксированной совокупности длины l;

$\hat{y}_i^{*\prime}$ — расчетные значения для соответствующих точек по уравнению, построенному но $l + 1$ наблюдению.

Пусть для рассматриваемого выше случая из 10 наблюдений минимальное значение меры сходства оказалось у второй группы $\begin{pmatrix} y_2, & y_3, & y_4 \\ x_2, & x_3, & x_4 \end{pmatrix}$. Следует попытаться присоединить к ней первую или пятую точки. Для случая присоединения первой точки

$$F = \frac{(y_1 - \hat{y}_1)^2 + \sum_{i=2}^{4} \left((\hat{y}_i - \hat{y}_i^*)^2 \right.}{\varepsilon_2}.$$

Если рассчитанное значение F меньше (равно) табличного значения F-критерия для фиксированного уровня значимости и числа степеней свободы $1, l - p$, то точку можно включить в «центр кристаллизации».

На следующем шаге процедуры точки, включенные в найденный «центр кристаллизации», исключаются из рассмотрения, а описанная выше последовательность действий повторяется для оставшихся точек. Процесс продолжается до тех пор, пока не будут найдены два примыкающих друг к другу «центра кристаллизации» (r и $r + 1$). В этом случае по дисперсионному критерию, описанному ранее (§ 2 гл. 3), проверяется идентичность двух уравнений регрессии. Если различия между уравнениями существенны, то одна из точек

перелома найдена. В противном случае «центры кристаллизации» объединяются и процедура присоединения точек продолжается.

В результате описанной процедуры могут быть следующие ситуации:

все наблюдения в конечном итоге объединяются в одну фиксированную совокупность; следовательно, исследуемый многомерный временной ряд не имеет структурных изменений и может быть описан одним уравнением регрессии;

выделяются несколько примыкающих друг к другу «центров кристаллизации»; границы между ними соответствуют точкам перелома и в найденных сегментах строятся различные регрессионные модели;

выделяются несколько «центров кристаллизации», между которыми имеются аномальные точки, ряд аппроксимируется разрывной кусочно-линейной функцией.

Глава 4

МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ В ПРОСТРАНСТВЕННЫХ И ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ВЫБОРКАХ

Как было показано ранее, особенностью пространственных выборок является скачкообразность изменения зависимости при переходе от одного типа к другому. Поэтому естественной формой описания зависимости выступают методы дискретного и дискретно-непрерывного моделирования. Что касается смешанных пространственно-временных выборок, то здесь, как правило, имеет место сочетание обоих типов изменчивости структуры влияния. Общий подход к описанию таких зависимостей состоит в сочетании методов многомерной классификации (распознавания образов) и корреляционно-регрессионного анализа. Теоретические основы использования упомянутого подхода подробно рассматривались в главе 2.

Однако в реальных ситуациях общий подход к моделированию претерпевает целый ряд специфических изменений. Тип модели и подход к ее построению определяется в основном следующим:

типом исходной совокупности (пространственные или пространственно-временные выборки);

характером набора используемых признаков (номинальные, ранговые, количественные) и степенью их вариации в классах;

типом решаемой задачи (оценивание показателя или управление); характером реакции моделируемого показателя на воздействие факторов производства (непрерывный или пороговый);

способом разбиения совокупности на классы (непересекающиеся, пересекающиеся, непрерывная аппроксимация разбиения и др.).

Сочетания (комбинации) перечисленных обстоятельств определяют наличие большого числа типов моделей и способов их построения.

Ниже на конкретных примерах мы рассмотрим некоторые наиболее часто встречающиеся типы моделей.

§ 1. ДИСКРЕТНЫЕ (КЛАССИФИКАЦИОННЫЕ) МОДЕЛИ

В общем случае многофакторная дискретная модель экономических показателей представляет собой две классификации объектов в пространстве факторов и показателей производства и описание связи этих классификаций. Эта модель позволяет — при условии, что известен класс объектов в одном признаковом пространстве,— с той или иной степенью точности определить принадлежность объекта к классу в другом признаковом пространстве.

Многофакторная дискретная модель — сравнительно новый в практике экономического анализа тип экономико-статистической модели. Возможность построения адекватных дискретных моделей появилась только после включения в арсенал экономико-статистического исследования теории распознавания образов. Содержание процесса моделирования при построении дискретной модели сводится к следующему. Тем или иным методом распознавания образов производится классификация элементов системы в пространствах факторов и результатов производства. Заключительный этап моделирования состоит в установлении наличия связи между двумя полученными классификациями. Для этой цели привлекаются приемы статистики качественных признаков. Каждая классификация трактуется как многомерный качественный признак, строится матрица сопряженности между ними, по которой рассчитывается коэффициент связи между классификациями. Если величина коэффициентов связи статистически значима, можно говорить о существовании модели зависимости результативных показателей от факторов производства в виде дискретной функции.

Дискретные модели экономических показателей, построенные методами распознавания образов, имеют достаточно широкую сферу применения в экономических исследованиях. В целом ряде случаев они оказываются более эффективными, чем непрерывные регрессионные модели. Целесообразность применения в той или иной ситуации дискретной экономико-статистической модели может быть обусловлена как характером исследуемой совокупности, так и содержанием решаемой экономической задачи.

С точки зрения характера исследуемой совокупности наибольшее значение имеет факт ее структурной неоднородности и наличие существенных качественных признаков в описании объектов.

Что касается связи характера экономической задачи и типа модели, то здесь необходимо отметить следующие соображения.

Во-первых, целесообразность использования дискретных моделей может быть обусловлена пороговым характером реакции экономических систем на те или иные воздействия. Существенное изменение экономических показателей происходит не при любом изменении производственных факторов, а лишь при определенном, превосходящем пороговую величину. В свою очередь, и для значений экономических показателей важно не любое изменение, а лишь такое, которое существенно скажется на результатах работы зависящих от данного объекта элементов экономической системы (последующие производственные переделы и т. п.). Например, нормативы экономического стимулирования устанавливаются не для каждого предприятия в отдельности, а для групп предприятий с «близкими» объективными условиями производства. Различного рода системы премирования и поощрения имеют ступенчатые шкалы (например, премиальная надбавка за качество стали выплачивается при превышении определенных значений механических свойств последней).

Во-вторых, в ряде случаев сама экономическая задача требует представления данных в дискретной форме, например, тогда, когда результаты экономико-статистического моделирования объектов предполагается использовать для подготовки информации для дискретных моделей оптимального отраслевого планирования.

Наконец, в-третьих, дискретные модели — эффективный инструмент анализа и прогнозирования комплекса взаимосвязанных экономических показателей. При использовании непрерывных регрессионных моделей учет комплексности, системности результативных характеристик работы предприятия связан с известными трудностями — необходимостью использования сложного аппарата для оценки параметров системы структурных уравнений. Между тем функционирование промышленного объекта — столь многогранное явление, что для его описания необходимо привлекать, как правило, комплекс показателей. Необходимость в системном подходе возникает и при включении экономико-статистической модели в состав многоуровневого комплекса. Следовательно, дискретные модели — один из эффективных путей описания зависимости вектора экономических показателей от вектора производственных факторов.

Пофакторный подход, когда отдельно устанавливается регрессионная связь каждого из компонент выходного показателя со всем набором производственных факторов, искусственно разъединяет вектор результирующих показателей

на отдельные компоненты, причем изменение каждого из компонентов рассматривается изолированно.

Резюмируя, можно сказать, что наиболее часто дискретные модели применяются в тех случаях, когда либо сама экономическая задача требует группировки объектов, либо для измерения уровня выходного показателя используются качественные признаки (например, марка или сорт продукции), либо при структурной неоднородности исследуемой совокупности, когда число объектов, представляющих отдельный тип, мало́, либо требуется моделирование зависимости взаимосвязанного комплекса показателей от вектора факторов производства.

Построить дискретную модель можно различными способами в зависимости от решаемой экономической задачи. Нами выделяются три способа: метод «принудительной» классификации, метод «свободной» классификации и их комбинация.

Метод «принудительной» классификации

При этом способе независимой (первичной) является одна классификация, вторая подстраивается под нее, зависима от первой. Взаимосвязь классификаций предусматривается самим способом построения, число классов в обоих признаковых пространствах одинаково. Наиболее просто такой способ моделирования реализуется тогда, когда результаты функционирования объекта описываются одним экономическим показателем. В этом случае достаточно провести таксономию в пространстве факторов производства X, затем для каждого полученного класса (таксона) определить среднее значение экономического показателя y и его доверительные интервалы.

В общем случае, когда оба признаковых пространства многомерные, схема построения модели этого типа предполагает последовательное решение двух задач теории распознавания образов:

1) таксономии (самообучения) в независимом признаковом пространстве, результатом которой является построение активной (обучающей) классификации;

2) построения вторичной классификации в другом признаковом пространстве, т. е. обучение распознаванию первичной классификации.

Выбор «активной» классификации (пространства, в котором производится первичное разбиение) зависит от содержательной постановки задачи. Если ставится задача пассивного характера (оценивания, нормирования, пассивного прогноза), то в качестве первичной выступает классификация в пространстве факторов производства X. Решением задачи является оценка значений результативного показателя (показателей) в выделенных классах. Содержательно решение отвечает на вопрос, какие изменения произойдут в результатах производства при изменении типологии объекта по неуправляемым условиям производства (природные условия, внешние факторы и т. п.).

Если стоит задача активного управления процессом, то в качестве первичной выступает классификация в пространстве результативных показателей. Решением задачи в данной ситуации выступает описание типологии в пространстве Y в терминах пространства факторов производства. Оно по существу нацелено на получение ответа на вопрос, какие изменения нужно произвести в пространстве управляемых факторов, чтобы изменить принадлежность объекта к классу по результатам производства.

Критерием качества модели может выступать либо коэффициент связи классификаций при раздельном распознавании объектов в обоих признаковых пространствах по сформулированным в ходе обучения решающим правилам, либо суммарное число ошибок распознавания принадлежности объектов к классу первичной классификации по вторичной.

К достоинствам этого способа следует отнести заведомо тесную связь между двумя классификациями. Существенный недостаток способа «принудительной» классификации — возможность нарушения естественно сложившейся классификации в зависимом признаковом пространстве. При принудительном переходе от одной классификации к другой группы (типы) могут быть «разрезаны», нарушены.

Наибольший интерес в этом классе моделей представляют модели активного управления. Приведем два примера. Первый является более общим по содержанию, а второй иллюстрирует технику построения такого типа моделей.

Пример 1. Суть заключается в оценке необходимых изменений входных характеристик предприятий отрасли для перевода их в более высокий класс по эффективности работы [Беккер, 1979].

Объект управления — предприятие — характеризуется рядом входных и выходных параметров, а процесс развития объекта — их изменением. В пространстве этих параметров можно наблюдать определенные сочетания их значений. Другими словами, если рассматривать множество объектов управления, то можно выделить определенные комбинации

параметров, которые объединяют объекты в образы (типы). Под образом в данном случае следует понимать производственную ситуацию, характеризующую такое состояние объекта управления, при котором комбинация регулирующих воздействий предопределяет его выходное состояние.

Будем считать, что в базисный период совокупность предприятий отрасли распадается на k классов по результирующим показателям и эти классы уже упорядочены. Обозначим их через V_1, V_2, V_3, . . ., V_k. Следовательно, развитие объекта можно представить как его движение по классам V_i, $i = \overline{1, k}$. При этом переход через несколько классов — довольно редкое явление. Обычно это связано с резким изменением регулирующих воздействий и не всегда возможно.

Заметим, что переход объекта из класса в класс в настоящее время не является полностью управляемым процессом, и в этом смысле развитие объекта частично происходит пассивно.

Рассматривается метод активного управления процессом перехода из класса в класс. Задача состоит в нахождении набора значений регулирующих параметров (x) некоторого объекта из класса V_j таким образом, чтобы этот объект перешел в класс V_{j+l} по выходному показателю, где l — число, выбираемое исследователем. Переход из класса с низким «качеством» в класс с высоким «качеством» обеспечить трудно, поэтому число l не может быть очень большим.

Таким образом, считаем, что у нас имеется всего два класса — V_1 и V_2, и будем отыскивать значения регулирующих параметров для каждого из объектов класса V_1, чтобы эти объекты перешли в класс V_2.

Поставленная задача решалась для случая детерминированного разделения классов V_1 и V_2. Когда же эти классы не разделены строго, т. е. классификация практически состоит в указании для каждого объекта вероятностей его принадлежности к классам V_1, и V_2, постановка задачи должна сводиться к нахождению преобразования, увеличивающего вероятности отнесения объектов к классу V_2.

В качестве объекта анализа рассматривалась отрасль приборостроения. Было исследовано 168 предприятий. Несмотря на принадлежность исследуемой совокупности предприятий к одной отрасли, все они имеют свои особенности и представляют набор объектов, существенно различающихся по уровню специализации, масштабам производства, видам используемого оборудования и степени его прогрессивности, профессиональному уровню и структуре кадров.

Для всестороннего отражения вариации исследуемых объектов и определения резервов повышения производительности труда первоначально были выбраны 45 признаков, характеризующих различные аспекты функционирования промышленных объектов — рабочую силу, материально-вещественные элементы производства (орудия, предметы и продукты труда), технологию и организацию производства.

В результате применения алгоритма диагонализации корреляционной матрицы связей признаков друг с другом [Экономико-статистическое..., 1977] отобран сокращенный набор признаков x:

среднегодовая заработная плата рабочего, руб. (x_1);

коэффициент сменности оборудования (x_2);

среднегодовой объем основных производственных фондов, тыс. руб. (x_3);

отношение численности рабочих к численности промышленно-производственного персонала (x_4);

отношение стоимости измерительных и регулирующих приборов и устройств лабораторного оборудования к стоимости рабочих машин и оборудования (x_5);

удельный вес расходов на подготовку и освоение производства в полной себестоимости продукции, % (x_6);

удельный вес материальных затрат в полной себестоимости продукции, % (x_7);

удельный вес оборотных средств в производственных фондах, % (x_8);

удельный вес покупных изделий, полуфабрикатов и услуг кооперативных предприятий в сумме сырья и материалов, % (x_9);

удельный вес оборудования в среднегодовых основных производственных фондах, % (x_{10});

удельный вес численности рабочих в численности промышленно-производственного персонала, % (x_{11});

фондовооруженность рабочих, тыс. руб./чел. (x_{12});

удельный вес вновь введенных основных производственных фондов в стоимости среднегодовых производственных фондов, % (x_{13}).

В качестве характеристики уровня развития предприятия (y) был выбран показатель производительности труда, по которому все предприятия были классифицированы на два класса: V_1 — предприятия с низкой производительностью труда и класс V_2 — с высокой.

Далее рассматривались эти два класса с позиции производственных факторов. Здесь полученные два класса могут пересекаться, так как изучаются совсем в другом простран-

Таблица 4.2

Результаты вторичной классификации

Технический параметр	Среднее значение признака			Обратная величина дисперсии		
	I сорт	II сорт	III сорт	I сорт	II сорт	III сорт
Содержание влаги в коксе, % (x_1)	0,783	0,69	0,72	1,657	2,063	1,439
Содержание недопала в шихте, % (x_2)	2,604	3,456	2,389	0,107	0,071	0,095
Содержание недопала в корректировочной извести, % (x_3)	2,150	2,596	3,329	0,648	0,216	0,149
Удельный вес фракции кокса, %						
более 40 мм (x_4)	0,737	1,151	2,076	6,774	0,25	0,29
40—25 мм (x_5)	65,618	67,73	67,13	0,011	0,012	0,016
менее 25 мм (x_6)	33,281	30,76	30,52	0,012	0,012	0,015
Удельный вес фракции извести, %						
более 40 мм (x_7)	42,129	41,375	43,41	0,084	0,004	0,004
40—25 мм (x_8)	50,742	52,391	51,09	0,039	0,004	0,004
менее 25 мм (x_9)	7,128	5,571	5,15	0,100	0,09	0,041
Коэффициент использования мощности (x_{10})	0,843	0,846	0,824	1656,0	1941,5	692,0
Количество извести шихтовой, т (x_{11})	66,5	69,83	65,88	0,0048	0,003	0,003
Количество извести корректировочной, т (x_{12})	13,73	16,90	15,64	0,014	0,01	0,018
Сила тока						
I фаза (x_{13})	97,629	107,3	99,5	0,002	0,01	0,002
II фаза (x_{14})	108,96	115,0	44,6	0,0043	0,0196	0,010
III фаза (x_{15})	103,1	110,4	109,89	0,0015	0,031	0,013
Напряжение трансформатора						
I фаза (x_{16})	3,1	4,07	7,568	0,1165	0,098	0,078
II фаза (x_{17})	10,1	10,99	13,44	0,042	0,1867	0,125
III фаза (x_{18})	4,35	3,273	6,48	0,050	0,1953	0,029
Общая мощность печи, кВт (x_{19})	40,25	41,23	39,09	0,073	0,1263	0,021

задании «цены» ошибок — проводился по тому же алгоритму случайного поиска с адаптацией.

Наиболее значимыми по влиянию на сортность карбида оказались признаки: x_{16}, x_4 и x_7.

Описанная классификационная модель оценки качества продукции по значению технологических признаков может быть применена для разных целей. Одно из таких направлений, как видно из примера,— выявление основных технологических причин колеблемости качества продукции. Так как в число существенных входят сырьевые факторы, то полученные данные могут быть также полезны при обосновании стратегии закупок. Они позволяют провести сопоставление выигрыша, связанного с улучшением качества продукции, и проигрыша из-за использования более дорогого сырья.

Выявление важнейших технологических причин колебаний качественных показателей — лишь одно из приложений рассматриваемого типа моделей к анализу производства. Другим важным приложением его является применение результатов расчета для эффективного управления производственным процессом. Зная существенные факторы и границы их варьирования для класса, соответствующего I сорту, можно использовать эти сведения для поддержания такого технологического режима, который обеспечивает наивысшее качество продукта. Наиболее интересны данные признаков x_{16}, x_4, x_7: они свидетельствуют, что для получения карбида повышенного качества требуется на первой фазе держать высокое напряжение (что равносильно подаче на эту фазу большей мощности) и уменьшать крупные фракции кокса и извести.

Классификационные модели подобного типа могут также быть применены для прогнозирования качественных показателей продукции по параметрам технологического режима. Это важно для таких производств, где получение фактического значения качественного показателя связано с длительными лабораторными или другими испытаниями, например, в цементной промышленности, где активность цемента становится известной лишь через двое суток после его производства, или в металлургической промышленности, где сваренную сталь подвергают длительным испытаниям, от которых зависит характер ее дальнейшей переработки. Прогнозирование в этих случаях способствовало бы уменьшению числа ошибочных решений в процессе оперативного управления производством.

Метод «свободной» классификации

Обе классификации одних и тех же объектов (в пространствах факторов и показателей производства) строятся независимо друг от друга, а затем оценивается взаимосвязь между ними. Связь классификаций при этом подходе основана только на соответствующем выборе признаковых пространств и характера разбиения в них.

Таким образом, в общем случае построение моделей этого типа включает две самостоятельные операции таксономии, а также описание и измерение связи между полученными независимыми классификациями объектов. В ряде частных случаев необходимость в выполнении отдельных этапов может отпасть. Так, одно или оба разбиения могут оказаться заданными — это бывает тогда, когда признаковые пространства сформированы из небольшого числа качественных признаков.

Поскольку при данном подходе разбиения в каждом признаковом пространстве выполняются независимо друг от друга, то сначала оптимальность каждой классификации оценивается по ее внутренним критериям. При таксономии информация об индивидуальных свойствах объектов группы заменяется описанием средних свойств этой группы объектов, и чем меньше информации мы теряем при такой операции, тем при прочих равных условиях лучше качество таксономии. В литературе предложен ряд критериев качества таксономии при отсутствии «суперцели» [Загоруйко, 1972]. Для построения двух независимых классификаций модели можно воспользоваться одним из них.

Следующий этап построения модели — изучение связи между двумя выбранными вариантами классификации. Здесь естественно ориентироваться на характер и силу связи двух классификаций, составляющих дискретную модель, т. е. это можно трактовать как изучение связи двух многомерных качественных признаков.

Описание связи классификаций удобно начинать с построения матрицы взаимной сопряженности.

Очевидно, что дискретная модель тем лучше описывает изучаемое явление, чем ближе построенная матрица к диагональной. Для последней имеет место взаимно однозначное соответствие каждого класса по выходным параметрам определенному классу по входным параметрам. Модель с диагональной матрицей обеспечивает максимальные аналитические возможности и гарантирует высокую точность прогноза значений моделируемого комплекса показателей. Поэтому при дискретном моделировании желательно провести серию экспериментальных расчетов, цель которых состоит в выявлении параметров, дающих лучшую матрицу сопряженности. При этом варьируются следующие элементы модели:

набор факторов производства;
состав комплекса результативных показателей;

число классов в обеих классификациях;
методы классификации;
априорные веса признаков и др.

Количественная оценка качества полученной матрицы сопряженности, ее близости к оптимальной может проводиться с использованием приемов оценки взаимосвязи качественных признаков. В частности, для проверки гипотезы о независимости классификаций (нуль-гипотезы) используется критерий χ^2 Пирсона [Андерсон, 1963].

Однако критерий χ^2 позволяет лишь ответить на вопрос, значима ли статистически взаимосвязь классификаций. Этот критерий не может выступать в качестве стандартной меры силы связи, так как его значения непормированны — с увеличением числа объектов m изменяется и верхняя граница χ^2. Поэтому для измерения степени тесноты связи используются нормализованные коэффициенты взаимной сопряженности Пирсона (c) и Чупрова (k), рассчитываемые на основе χ^2:

$$c = \sqrt{\frac{\chi^2}{\chi^2 + m}};$$

$$k^2 = \frac{\varphi^2}{(m-1)(n-1)}, \quad \varphi^2 = \frac{\chi^2}{m},$$

а также другие коэффициенты связи качественных признаков. Коэффициент среднеквадратичной сопряженности Пирсона (c) имеет один серьезный недостаток. Хотя он возрастает по мере того, как увеличивается расхождение между теоретическими и эмпирическими численностями, он все же никогда не достигает предела, к которому стремится. Величина c зависит от размерности таблицы взаимной сопряженности, что, строго говоря, делает несравнимыми значения коэффициента для различных классификаций одной и той же совокупности. Более предпочтителен коэффициент взаимной сопряженности Чупрова, свободный от указанного недостатка коэффициента c.

Таким образом, одним из критериев качества дискретной модели является показатель силы связи между двумя классификациями: чем больше его значения, тем лучше модель.

Второй, более важный критерий качества дискретной модели — это величина ошибки прогноза, получаемая при проведении «экзамена», который предполагает использование дискретной модели, полученной на обучающем материа-

ле, для прогноза значений комплекса показателей по новым, «экзаменуемым» объектам.

В процессе доводки модели до окончательного вида «экзамен» может преследовать две взаимосвязанные цели:

1. Выбор оптимальной дискретной модели при наличии множества ее вариантов.

2. Оценка качества выбранного окончательного варианта модели (степени ее адекватности моделируемому явлению).

Процедура расчетов в обоих случаях одинакова, только при решении первой задачи она повторяется столько раз, сколько вариантов модели проверяется.

После того как «оптимальные» по внутренним критериям классификации в пространстве показателей и факторов производства построены и определена степень их взаимосвязи, необходимо организовать процесс изменения одной или обеих классификаций, направленный на возможное увеличение коэффициента связи между ними.

Число классов — одна из важнейших характеристик дискретной модели. Увеличение числа классов при ограниченном объеме исходной совокупности оказывает двойственное влияние на точность дискретной модели: с одной стороны, повышается однородность классов (групп), и, следовательно, уменьшается внутригрупповая вариация моделируемых показателей, повышается представительность среднего значения для всех объектов класса, с другой — в связи с уменьшением численности классов возрастает ошибка выборки.

В связи с этим для каждой конкретной ситуации существует некоторое предельное, максимально допустимое число классов (и соответственно минимально допустимый размер отдельного класса). Минимально допустимый размер класса должен определяться исходя из цели исследования и характера информации. При построении моделей экономических показателей (в тех случаях, когда показатели представлены числами) минимально допустимый размер классов следует устанавливать исходя из достоверности полученных групповых средних. Тот или иной класс можно считать самостоятельным только тогда, когда даваемое им среднее значение экономических показателей существенно отлично от средних по другим классам.

Строгой процедуры установления оптимального числа классов пока нет. На основе принятого критерия качества модели необходимое число классов можно определить путем последовательного приближения, апробируя различные варианты разбиения.

Таблица 4.3

Таблица сопряженности классификаций

Класс по экономическим показателям (A)	Число объектов в классах по производственным факторам (B)		Итого
	I	II	
I	20(40,18)	48(27,82)	68
II	58(37,82)	6(26,18)	64
Всего . . .	78	54	132

Примечание. В скобках проставлены выравнивающие частоты. Например, $a_{11}=\frac{68\times 78}{132}=40{,}18$.

Рассмотрим конкретный пример. Для мартеновского производства (132 объекта) строилась дискретная модель зависимости комплекса результативных показателей (расходы по переделу и себестоимость стали) от шести производственных факторов (годовой объем производства, средний вес плавки, доля чугуна в садке, доля разливки в изложницы на тележках, калорийность топлива, удельный расход кислорода). Классификация в обоих признаковых пространствах — показателей (A) и факторов (B) — проводилась по алгоритму и программе, реализующей метод потенциальных функций [Айзерман и др., 1970], при этом опробованы классификации на 2, 3, 4 и 5 классов, т. е. получено 16 вариантов модели. Взаимная сопряженность классификаций A и B для простейшего варианта (22) показана в табл. 4.3. Величина χ^2 составит

$$\frac{(20-40{,}18)^2}{40{,}18}+\frac{(58-37{,}82)^2}{37{,}82}+\frac{(48-27{,}82)^2}{27{,}82}+\frac{(6-26{,}18)^2}{26{,}18}=51{,}106.$$

При $\nu=1\,[(2-1)(2-1)]$ $p_{\chi^2}=0{,}01$, т. е. весьма мала. Следовательно, нуль-гипотезу о независимости классификаций A и B необходимо отвергнуть. Данные таблицы сопряженности свидетельствуют о статистически значимой связи между классификациями по экономическим показателям и производственным факторам.

Значение коэффициента взаимной сопряженности Чупрова рассчитывается следующим образом:

$$\varphi^2=\frac{\chi^2}{m}=\frac{51{,}106}{132}=0{,}387,$$

$$k^2=0{,}387;\quad k=0{,}622.$$

Таблица 4.4

Изменение коэффициента связи классификаций при изменении числа классов

Классификация по экономическим показателям (число классов)	Классификация по производственным факторам (число классов)			
	2	3	4	5
2	0,622	0,473	0,407	0,388
3	0,600	0,483	0,401	0,392
4	0,552	0,280	0,400	0,391
5	0,442	0,468	0,309	0,396

Влияние числа классов на величину коэффициента связи между элементами дискретной модели можно проследить по табл. 4.4.

Как видно из приведенных данных, для одних и тех же признаковых пространств и алгоритма классификации с изменением числа классов существенно меняется величина коэффициента взаимной сопряженности (от 0,280 до 0,622).

Из приведенной краткой характеристики способа «свободной» классификации ясны его достоинства и недостатки. Он учитывает реально сложившуюся концентрацию объектов в признаковых пространствах, но в то же время оказывается очень трудоемким процесс получения достаточно тесно связанных классификаций, необходимых для прогнозирования экономических показателей.

Комбинированный метод

Попытка преодолеть недостатки методов «принудительной» и «свободной» классификации привела к разработке комбинированной многоэтапной схемы построения дискретных моделей. Основное содержание этого метода заключается в том, что до выполнения операции свободного разбиения пространства факторов производства выполняется его преобразование (размерности и метрики) на основе изучения информативности факторов для разбиения по показателям. Действительно, в пространстве факторов производства может быть получено множество классификаций, каждая из которых условна в том смысле, что классификация, удовлетворяющая одному критерию, одной цели, может оказаться непригодной для другой цели. При построении модели экономических показателей естественным является требование получить такую классификацию объектов в пространстве факторов производства, которая была бы близка к классификации по экономическим показателям. Это может быть достигнуто соответствующим преобразованием признакового пространства.

Комбинированная схема построения дискретной модели включает следующие этапы:

1. Свободная классификация (таксономия) в пространстве экономических показателей Y и построение на этой основе обучающей выборки (к таксономии в пространстве экономических показателей относятся все соображения о максимально допустимом числе классов, которые были ранее высказаны при описании методики построения моделей свободного типа, т. е. доверительные интервалы групповых оценок показателей не должны перекрываться).

2. Обучение распознаванию этих классов в пространстве факторов производства X.

Первый и второй этапы по существу представляют собой построение модели способом «принудительной» классификации, где в качестве «активной» выступает классификация по выходным экономическим показателям.

3. Преобразование пространства факторов производства по результатам обучения (переход от X к X'). Это преобразование заключается в минимизации описания и придании при необходимости «весов» отдельным признакам.

4. Свободная классификация в преобразованном пространстве факторов производства X'.

5. Описание и измерение связи классификаций в пространствах Y и X'.

Остановимся несколько подробнее на третьем этапе схемы. Известна большая зависимость результатов таксономии от размерности и метрики признаковых пространств.

Очевидно, неправомерно рассчитывать на то, что диапазоны значений каждого из факторов производства (или их комбинаций), полученные в результате описания классификации выходных показателей в пространстве X, окажутся непересекающимися. Естественно считать наиболее информативными для построения типологии те факторы производства, разница в законах распределения которых оказывается наибольшей при переходе от одного класса в пространстве к другому. Эта идея лежит в основе построения большинства методов выбора информативных подсистем признаков.

Что касается метрики пространства описания, то для обеспечения сопоставимости признаков в настоящее время общепринята нормировка исходных данных по величине среднеквадратического отклонения. Однако, поскольку

целью моделирования является прогнозирование уровня экономических показателей объекта, то часто оказывается целесообразным дополнительно ввести веса («относительную значимость») для факторов производства. Эти веса могут быть ориентировочно установлены, например, при минимизации описания по алгоритму случайного поиска с адаптацией [Лбов, 1981]. Сокращение размерности пространства и предварительное взвешивание факторов производства обеспечивает при последующей независимой таксономии объектов в этом пространстве учет основной цели моделирования.

Описанная многоэтапная схема гарантирует сравнительно быстрое и целенаправленное получение двух классификаций с достаточно тесной связью между ними (во всяком случае со связью, близкой к максимально достижимой для данных признаковых пространств). При этом сохраняется возможность выделения типов в каждом пространстве и проведения четких и простых границ между ними. Реализация схемы сводится к последовательному решению типичных задач распознавания образов (самообучения, обучения с учителем, минимизации описания) по стандартным алгоритмам и программам.

Все сказанное выше относилось к статическому случаю, т. е. к пространственным выборкам. В случае пространственно-временных выборок (динамический случай) возникает дополнительная проблема проверки устойчивости классификаций во времени и разработки способов их динамизации. Конечной целью этого этапа является прогноз тех преобразований классификационной структуры совокупности, которые должны произойти с течением времени. На этом этапе могут быть использованы различные методы и алгоритмы. Подробнее этот вопрос будет рассмотрен в § 3 настоящей главы, специально посвященном проблемам динамизации дискретно-непрерывных моделей. Однако некоторые подходы к проверке устойчивости классификаций будут показаны ниже на примере *задачи использования дискретных моделей для разработки системы групповых нормативов экономического стимулирования* для предприятий отрасли [Розин и др., 1969].

Условием эффективности какой-либо системы материального стимулирования является достаточно полный учет объективных возможностей предприятий. Их коллективам должны быть созданы равные условия для достижения директивного уровня стимулируемого показателя.

Достигнутый уровень экономических показателей производства на предприятии обусловлен воздействием регу-

лируемых, нерегулируемых и случайных факторов. Объективная оценка работы предприятий может быть получена только тогда, когда система распределения заданий учитывает различия в объективных, неуправляемых на уровне предприятия условиях и факторах производства. Решение этой задачи достигается построением системы дифференцированных групповых нормативов материального стимулирования.

Одним из важнейших этапов построения системы групповых нормативов является формирование групп предприятий с близкими условиями производства. Принципиальная возможность такой группировки основывается на том, что предприятия отрасли обычно распадаются на ряд априори существующих групп (классов) по комплексу объективных условий производства.

Сложность задачи заключается в том, что группы должны формироваться по большому числу факторов производства, влияние которых на показатели работы предприятий носят стохастический характер. Предприятия отраслей промышленности различаются по качеству используемого сырья, местоположению, размерам, структуре и техническому уровню производственных фондов, степени концентрации производства, номенклатуре продукции и т. д. Для построения групп все важнейшие объективные условия производства должны быть учтены. Это требует, как правило, использования 10—15, а во многих случаях и большего числа параметров группировки. С этой точки зрения традиционные методы комбинационной группировки оказываются неприемлемыми, так как они позволяют использовать только 2—3, максимум 4 группировочных признака.

Перейдем к формальной постановке задачи. Пусть имеется некоторая совокупность объектов (предприятий), характеризуемая двумя наборами признаков — входными и выходными показателями (факторы производства и показатели экономической эффективности).

Основная цель задачи состоит в определении влияния некоторого набора (части) входных показателей — объективных условий производства — на выходные показатели, т. е. показатели экономической эффективности производства.

Формально каждому объекту (предприятию) ставятся в соответствие два вектора $x = (x_1, \ldots, x_n)$ и $y = (y_1, \ldots, y_r)$. Координаты вектора x есть входные характеристики объекта — набор факторов производства; координаты вектора y есть выходные характеристики объекта, т. е. показатели экономической эффективности (себестоимость, рента-

бельность и др.). Координаты вектора x можно разделить на две группы.

В экономической интерпретации первая группа — значения неуправляемых, вторая— управляемых факторов производства. Обозначим первую группу объективных условий через x_{I}, вторую — через x_{II}. Тогда вектор x можно представить в следующем виде: $x = (x_{\text{I}}, x_{\text{II}})$.

Вектор y есть многомерная случайная величина, зависящая от вектора $x = (x_{\text{I}}, x_{\text{II}})$ и некоторой случайной величины ξ. Величина ξ ассоциируется с влиянием группы неуправляемых факторов производства.

$$y = g(x, \xi) = g(x_{\text{I}}, x_{\text{II}}, \xi).$$

В такой терминологии основная цель задачи состоит в нахождении влияния неуправляемых факторов производства x_{I} на выходной показатель y.

Обозначим через $\hat{y}_t$ значения выходного показателя эффективности предприятий при фиксированных значениях управляемых факторов производства на некотором уровне t (ξ не учитывается), т. е. когда различия в показателях эффективности объясняются лишь различием объективных условий работы предприятий. В этом случае для решения задачи достаточно произвести группировку по выходному показателю $\hat{y}_t$. Однако на практике значения $\hat{y}_t$ неизвестны, так как невозможно учесть влияние случайных факторов ξ. Поэтому метод группировки предприятий по выходному показателю неприемлем.

Для решения задачи применяется комбинированный способ построения модели.

Результаты решения сформулированной задачи покажем на примере лесозаготовительной промышленности Урала и Сибири (176 леспромхозов, функционирующих четыре года).

Весь комплекс факторов, оказывающих влияние на экономические результаты деятельности предприятий, может быть подразделен на три группы: 1 — местоположение предприятий, 2 — качество используемых природных ресурсов, 3 — техническая вооруженность труда.

Такая группировка факторов применима для всех отраслей промышленности. В лесозаготовительной отрасли наибольшее значение имеет местоположение предприятия и качество природных ресурсов.

Анализ условий и результатов деятельности лесозаготовительных предприятий позволил включить в исходное описание объектов факторы, перечисленные в табл. 4.5

Таблица 4.5

Набор факторов производства

Наименование группы и фактор	Измеритель (условное обозначение)
1. Местоположение предприятия	
Географическое положение	Градус северной широты (x_1), градус восточной долготы (x_2)
Рельеф	Высота над уровнем моря, м (x_3)
Снежность зим района расположения ЛПХ	Глубина снежного покрова, см (x_4)
Удаленность нижних складов от лесосырьевых ресурсов	Среднее расстояние вывозки, км (x_{11})
Характер примыкания нижних складов	Удельный вес вывозки к сплаву (x_{13}), к пунктам потребления (x_{14}), к железной дороге (x_{15}), %
2. Качество используемых природных ресурсов	
Состав пород лесонасаждений	Удельный вес сосны (x_5); ели, пихты (x_6), лиственницы (x_7); березы, осины (x_8), %
Средний объем хлыста	(x_9), м^3
Средний запас древесины на 1 га	(x_{10}), м^3/га
3. Техническая вооруженность труда	
Виды используемого транспорта	Удельный вес вывозки автомашинами (x_{16}), тракторами (x_{17}), %
Стоимость производственных фондов	(x_{18}), тыс. руб.

($x_1 - x_{10}$ — объективно неуправляемые, а x_{11}, $x_{13} - x_{18}$ — условно неуправляемые факторы производства). Кроме того, учитывались объем производства (x_{12}) и удельный вес лесозаготовок в общем объеме производства (x_{19}). Эти признаки регламентируются централизованно и поэтому относятся к условно неуправляемым, т. е. они неуправляемы на уровне предприятия.

Первичные группы — эталоны формировались по одному выходному показателю — себестоимости продукции. При этом предполагалось, что если группа предприятий из года в год сохраняет свой относительный уровень себестоимости продукции, то с определенной вероятностью можно утверждать, что эти предприятия имеют близкие объективные условия производства. Следовательно, обязательным элементом работы по формированию первичных групп является проверка их поведения во времени.

Существует ряд методов выделения первичных групп — эталонов. Рассмотрим некоторые из них.

Оценка неизвестных параметров некоторого смешанного распределения. Основная идея решения таких задач сводится к нахождению апостериорных плотностей вероятностей с помощью априорных плотностей и имеющихся наблюдений $(y_1, y_2, \ldots, y_m)$, где $y_i = (y_{i1}, y_{i2}, \ldots, y_{ir})$, $i = \overline{1, m}$ (выборка).

Можно считать, что обучающая выборка $(y_1, \ldots, y_m)$ извлекается в соответствии с некоторым смешанным законом распределения

$$f(y/\alpha, q) = \sum_{i=1}^{k} f(y/i, \alpha_i)\, q_i,$$

где q_i $(i = \overline{1, k})$ — неизвестные априорные вероятности классов;

α_i $(i = \overline{1, k})$ — неизвестный параметр (возможно, многомерный), связанный с распределением класса V_i;

$\alpha = (\alpha_1, \ldots, \alpha_k)$;

$q = (q_1, \ldots, q_k)$.

а. Нахождение оценки неизвестных параметров методом минимума χ^2 [Миленький, 1967]. В предложенной ситуации выходной показатель — одномерная величина ($r = 1$), а частные распределения классов полагаем нормальными. Поэтому

$$f(y/\alpha, q) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{k} \frac{q_i}{\sigma_i} e^{-\frac{(y-\mu_i)^2}{2\sigma_i^2}},$$

где $\alpha = \{q_i, \mu_i, \sigma_i\}$ — неизвестные параметры. Вся область, в которую попадают реализации, разбивается на s интервалов; точки деления ξ_i выбираем как выборочные квантили порядка i/s:

$$\frac{m}{5} \geqslant s \geqslant R_k - 2,$$

где R_k — число неизвестных параметров. Подсчитывается число реализаций, попавших в i-ю область. Пусть это будет число β_i. Тогда искомыми параметрами будут те, которые

минимизируют выражение

$$\chi^2(k) = m \sum_{i=1}^{s} \frac{\left(\frac{\beta_i}{m} - \nu_i(\alpha)\right)^2}{\nu_i(\alpha)},$$

где $\nu_i(\alpha) = \int_{\xi_{i-1}}^{\xi_i} f(y/\alpha, q_i)\, dy$, т. е. $\nu_i(\alpha)$ — вероятность попадания в i-й интервал.

Все описанное выше выполняется для $m = 1, 2, 3, \ldots$ последовательно. Необходимо остановиться на минимальном m, при котором $\chi^2(m) < \chi_p^2$; здесь χ_p^2 есть значение χ^2 с $(s - R_n - 1)$ степенями свободы на уровне p. Число p задается из условия, что при $m = m_0$, где m_0 — априори существующее число классов с вероятностью $1 - p$, должно выполняться $\chi^2(m) > \chi_p^2$, а при $m < m_0$ всегда $\chi^2(m) > \chi_p^2$.

Изложенный метод представляется наиболее полным и точным. Однако его практическое применение требует разработки специальных программ на ЭВМ. Поэтому в ряде случаев можно прибегнуть к приближенным способам декомпозиции общего распределения на частные.

б. Простой метод разложения распределения на гауссовские компоненты [Bhattacharya, 1967]. Пусть $\beta(y)$ означает частоту попадания в класс — интервал со средней точкой y и длиной h.

Рассматривается множество точек с координатами $\{y, \log(\beta(y + h)/\beta(y))\}$. Это множество разделяется на группы следующих друг за другом (по y) точек, которые приближенно могут быть соединены отрезком прямой с отрицательным наклоном. Число таких групп будет оценкой числа k компонент смеси.

Пусть отрезок η пересекает ось y в точке y_r и образует с отрицательным направлением оси y угол Θ_r.

В качестве оценок параметров r компоненты смеси гауссовских распределений берутся следующие статистики:

$$\bar{y}_r = y_r + h/r, \quad \sigma_r^2 = (dh \operatorname{ctg} \Theta_r/b) - (h^2/r)$$

(d и b — единицы масштаба на осях x и $\Delta \log \beta(y)$ соответственно), где $\Delta \log \beta(y) = \log(\beta(y + h)/\beta(y))$. В качестве оценок априорных вероятностей классов q_i берутся частоты попадания точек на i-ю прямую.

Рассмотрим применение последнего метода для группировки предприятий по величине себестоимости. При $h = 15$

Таблица 4.6

Характеристики групп предприятий

Параметры	Группы предприятий					
	I	II	III	IV	V	VI
Среднее значение	499	630	682	725	815	884
Среднеквадратическое отклонение	40	45	28	28	22	37

совокупность разбивалась на 50 интервалов, для решения задачи использовались только 30. Те интервалы, для которых одна из величин $\beta(y + h)$ или $\beta(y)$ равнялась нулю, выпадали из рассмотрения.

По уровню себестоимости предприятия могут быть объединены в шесть групп (табл. 4.6).

Рассмотренный метод выполняется по данным за все годы. Представителями класса считаются предприятия, попавшие в один и тот же класс за все или большинство анализируемых лет. Полученные в итоге группы предприятий можно считать первичными классами — эталонами.

Группировка наблюдений (предприятий) с применением методов таксономии. При определении первичных групп — эталонов по алгоритму [Елкина, Загоруйко, 1966] были использованы значения выходного показателя (себестоимости) за четыре года. При этом рассматривались два подхода. Один из них состоял в том, что все значения выходного показателя 176 леспромхозов за четыре года рассматривались как одна совокупность, т. е. подлежала разбиению общая совокупность, включающая 704 наблюдения (176×4).

При другом подходе выходная характеристика показателя эффективности представлена многомерной величиной, т. е. значения себестоимости по годам служили в качестве признаков.

Представляет интерес сравнение результатов группировки, полученных обоими способами. Соответствующие данные приведены в табл. 4.7. Из сопоставления видно, что результаты расчетов по предложенным способам хорошо совпадают и можно пользоваться любым из них.

Для следующего анализа окончательно были сформированы группы — эталоны, представленные в табл. 4.8.

После выделения классов — эталонов по значению выходного показателя y осуществляется переход от описания по выходному показателю к описанию по неуправляемым факторам производства x_I. Необходимость такого перехода

Таблица 4.7

Результаты группировок, полученных разными способами

Номер группы	Первый способ (радиус 66)			Второй способ (радиус 52)						
	Координаты центров групп	Число предприятий	Номера предприятий, включенных в группу	Координаты центров групп по годам				Среднее значение центров	Число предприятий	Номера предприятий, включенных в группу
				1-й	2-й	3-й	4-й			
1	422	4	37, 54, 60, 87	420	405	429	437	425	6	37, 54, 60, 87, 95, 106
2	516	36	10, 27, 31, 33, 34, 36, 53, 55, 56, 57, 58, 59, 61, 62, 65, 70, 71, 73, 77, 81, 84, 85, 86, 94, 99, 100, 101, 102, 104, 107, 118, 122, 124, 125, 171, 173	507	487	518	508	505	33	10, 27, 31, 33, 36, 53, 55, 56, 57, 58, 59, 61, 62, 65, 71, 73, 75 77, 81, 82, 84, 85, 86, 94, 99, 100, 102, 104, 107, 122, 124, 125, 171
3	642	53	5, 6, 13, 15, 20, 25, 32, 38, 40, 42, 43, 44, 45, 46, 47, 49, 50, 66, 67, 68, 69, 74, 76, 78, 79, 88, 92, 93, 96, 97, 98, 103, 110, 111, 114, 119, 129, 135, 138, 140, 150, 154, 155, 156, 157, 158, 160, 161, 162, 163, 164, 165, 168	613	594	637	626	618	37	5, 6, 13, 32, 34, 39, 40, 42, 43, 46, 47, 48, 49, 50, 70, 76, 78, 79, 80, 93, 97, 98, 101, 103, 110, 111, 114, 121, 150, 154, 156, 157, 160, 161, 163, 164, 168
4	—	—	—	709	678	727	715	707	32	2, 16, 20, 24, 29, 38, 44, 45, 51, 52, 64, 66, 96, 112, 113, 116, 120, 126, 129, 131, 132, 134, 140, 141, 142, 145, 148, 158, 159, 162, 165, 167
5	766	28	12, 17, 19, 23, 24, 28, 29, 35, 52, 83, 90, 91, 117, 128, 130, 131, 132, 133, 136, 137, 139, 141, 144, 146, 148, 149, 152, 159	785	752	807	804	787	18	4, 12, 17, 19, 35, 83, 90, 91, 117, 128, 133, 136, 137, 139, 144, 146, 149, 152

Таблица 4.8

Характеристика групп — эталонов

Номер группы	Число предприятий	Координаты центров групп по годам				Среднее значение центров
		1-й	2-й	3-й	4-й	
1	6	420	405	429	437	425
2	33	507	487	518	508	505
3	37	613	594	637	626	618
4	32	709	678	727	715	707
5	18	785	752	807	804	787
6	3	942	873	965	920	925

обусловлена тем, что выходной показатель был использован только для получения первичных групп предприятий. Для получения окончательных групп предприятий необходимо учитывать влияние неуправляемых факторов производства (основная цель задачи), что и оправдывает формирование описаний классов — эталонов по этим факторам. Вследствие такого перехода в пространстве X получаем соответствующие классы — эталоны, для которых необходимо задать описание. Описание класса — эталона есть некоторая функция от элементов, входящих в этот класс, и необходимо для формирования окончательных групп; функция — описание класса — эталона есть мера близости, показывающая близость любого произвольного элемента с данным классом — эталоном. При решении данной задачи для описания классов — эталонов использовался алгоритм Г. С. Лбова [1981].

После получения на предшествующем этапе исследования описания классов — эталонов в пространстве X появляется возможность формирования нормативных групп. Это осуществляется путем переклассификации всех объектов по полученному в процессе обучения описанию. При переклассификации для некоторых объектов (предприятий) может быть изменен номер группы. С точки зрения решаемой задачи этот факт представляет значительный интерес для анализа качества работы коллектива предприятия. Если при исследовании был учтен достаточно полный набор неуправляемых факторов, то переход из группы в группу можно интерпретировать как отклонение качества работы коллектива предприятия от среднего уровня. Здесь возможны три варианта. Первый — предприятие не изменяет своего относительного положения (номера групп в пространстве Y и X совпадают). Это означает, что качество работы коллектива предприятия

находится на среднем уровне. Второй — предприятие переходит из группы с низким показателем y в группу, соответствующую более высокому показателю эффективности. Можно сделать вывод, что коллектив предприятия работает на уровне ниже своих объективных возможностей. Третий случай противоположен второму. При этом степень отклонения от средних условий работы пропорциональна числу перешагиваемых групп. В качестве примера предлагается таблица переходов (табл. 4.9).

Рассматриваются шесть групп, расположенных в порядке возрастания уровня себестоимости продукции. Общее число переходных объектов составило 57. При этом 33 предприятия повысили номер группы (качество работы коллективов этих предприятий выше среднего уровня). Большинство переходов происходит в соседние группы. Лишь десять предприятий переходят через одну и больше групп. Сформированные по полному набору неуправляемых факторов производства группы можно считать нормативными и устанавливать всем предприятиям, попавшим в одну группу, единый стимулирующий показатель. Однако использование полного набора факторов из-за громоздкости практически неудобно, поэтому на последнем этапе исследования необходимо минимизировать описания нормативных групп в пространстве X.

При решении данной задачи критерием значимости признаков будет считаться число ошибок при распознавании.

Таблица 4.9

Матрица переходов объектов при переклассификации

Первичная группа—эталон	Результаты переклассификации в пространстве X					
	I	II	III	IV	V	VI
I	—	54, 60, 95, 106	37	—	—	—
II	61, 81, 104	—	27, 31, 33, 36, 59, 171	124	57	
III	—	70, 78, 93, 97, 101	—	5, 6, 76, 79, 98, 103, 114, 121, 150, 154	156, 160, 164	—
IV	—	66	38, 44, 45, 52, 96, 162, 165, 167,	—	2, 51, 131, 141, 158, 159	—
V	—	83	35, 91	4, 12, 90, 117	—	
VI	—	—	—	—	—	

Таблица 4.10

Матрица потерь

Группа i	Группа j					
	I	II	III	IV	V	VI
I	0	1	2	3	4	5
II	1	0	1	2	3	4
III	2	1	0	1	2	3
IV	3	2	1	0	1	2
V	4	3	2	1	0	1
VI	5	4	3	2	1	0

Можно ввести матрицу потерь $||c_{ij}||$. Коэффициент c_{ij} пропорционален расстоянию между центрами i-й и j-й групп (табл. 4.10).

Таким образом можно вычислить функцию риска

$$R = \sum_{i=1}^{6} \sum_{j=1}^{6} c_{ij} m(j/i),$$

где $m(j/i)$ — число объектов i-го класса, попавших в j-й класс. По алгоритму [Елкина, Загоруйко, 1966] из девятнадцати неуправляемых факторов, оказывающих влияние на эффективность производства, выбраны восемь. Выбранная подсистема включает следующие факторы: географическое расположение леспромхоза (долгота) — x_2; глубина снежного покрова — x_4; удельный вес в составе насаждений сосны — x_5; лиственницы, осины, березы — x_7, x_8; средний запас древесины на 1 га — x_{10}; характер примыкания складов к пунктам потребления — x_{14}; удельный вес вывозки древесины автотранспортом — x_{16}.

§ 2. СТАТИЧЕСКИЕ ДИСКРЕТНО-НЕПРЕРЫВНЫЕ (КУСОЧНО-ПРОСТЫЕ) МОДЕЛИ

Как было показано в главе 2, дискретно-непрерывная модель (ДНМ) представляет собой комбинацию моделей двух типов — дискретной и непрерывной. При ее использовании совокупность объектов моделируется на двух уровнях: верхнем (уровень классов) и нижнем (внутриклассовая модель).

Различие между дискретно-непрерывной и непрерывной моделью определяется характером совокупности. Дискретно-непрерывные модели целесообразно использовать тогда, когда совокупность неоднородна, распадается на отдельные классы, которые различаются по характеру связи между факторами и показателями. Ясно, что понятие «отдельные» классы является условным. В случае вероятностного распознавания таковыми следует считать совокупности, в которых вероятность принадлежности каждого из объектов совокупности (или большинства их) к одному из классов высока и соответственно низка вероятность принадлежности ко всем остальным классам (ее величина меньше заданного порога). Использование дискретно-непрерывной модели в описанных случаях диктуется требованием соблюдения принципа структурного подобия модели и изучаемого явления. Здесь можно утверждать, что исследуемая совокупность по существу состоит из ряда изолированных подсовокупностей со своими специфическими закономерностями.

Дискретно-непрерывные модели в ряде случаев могут быть использованы и для неоднородной совокупности с пересекающимися, «смазанными» классами, если исследуемая зависимость носит сложный нелинейный характер. Здесь дискретно-непрерывная модель выступает как средство упрощения зависимостей для соблюдения принципа максимально возможной простоты моделей.

Дискретно-непрерывные модели — эффективный способ соизмерения качественных и количественных признаков в тех случаях, когда производственные факторы описываются двумя независимыми наборами признаков разной природы и каждый из этих наборов содержит существенную информацию о моделируемом процессе. При этом влияние качественных и близких к ним количественных признаков учитывается дискретной частью модели, остальных количественных — внутриклассовыми регрессиями.

Помимо разбиения совокупности на изолированные классы к условиям целесообразности применения «чистой» ДНМ следует также отнести достаточную численность полученных классов, существенную внутриклассовую вариацию моделируемого показателя, различный характер зависимостей в классах.

Приведем пример, иллюстрирующий бóльшую целесообразность и эффективность применения ДНМ по сравнению с непрерывной моделью.

В исследовании, посвященном анализу и планированию деятельности механизированных колонн [Оганесов и др.,

Таблица 4.11

Сравнение точности моделей различных типов

Тип модели	Точность аппроксимации, σ_a^2			Точность прогноза, σ_n^2		
	y_1	y_2	y_3	y_1	y_2	y_3
Линейная непрерывная	1,7	0,005	1,29	2,1	0,022	2,63
Квадратичная непрерывная	0,99	0,003	0,90	1,88	0,019	1,92
Кусочно-линейная (первого типа)	0,43	0,0008	0,37	1,04	0,008	1,09
Кусочно-линейная (второго типа)	0,15	0,0004	0,12	0,53	0,008	0,74

Примечание. Для аппроксимации использовалось 54 реализации, а для прогноза (экзамена) — 27.

1976], изучалась зависимость производительности труда (y_1), себестоимости работ (y_2) и выработки грунта на одного рабочего (y_3) от производственных факторов для 27 объектов Урала и Сибири за три года.

Анализ деятельности механизированных колонн позволил выдвинуть гипотезу о наличии условий, определяющих эффективность построения ЭСМ с переменной структурой. Это, во-первых, наличие групп объектов, различающихся географическими условиями, мощностью и технической вооруженностью труда (вид строительства, стадия строительства, климатические и инженерные условия, объем строительно-монтажных работ, мощность экскаватора, грузоподъемность автомашин и т. д.); во-вторых, наличие в описании объектов количественных и качественных признаков; в-третьих, существенная вариация показателей и факторов производства с резко выраженной многовершинностью и асимметрией распределений.

Для оценки эффективности использования ДНМ сравнивались модели четырех типов: линейная непрерывная, квадратичная непрерывная, кусочно-линейная, построенная методом предварительного кластерного анализа, и кусочно-линейная, построенная методом ФИЭР. Сравнение точности моделей при аппроксимации (остаточная дисперсия σ_a^2) и прогнозировании (σ_n^2) приведено в табл. 4.11.

Приведенные данные свидетельствуют, во-первых, о резком повышении точности аппроксимации и прогноза при применении кусочно-линейных моделей и, во-вторых, о существенном влиянии способа построения ДНМ на ее точность.

Основными классификационными признаками, определяющими тип модели и способ ее построения, являются характер разбиения совокупности на классы и метод определения параметров моделей.

По первому признаку следует различать совокупности, *а*) распадающиеся на изолированные, «далекие» классы (строгое выполнение гипотезы компактности), *б*) состоящие из пересекающихся «смазанных» классов и включающие переходные типы объектов.

В случае *а*) целесообразно строить дискретно-непрерывную модель с детерминированным разбиением на классы (нерандомизированные модели), в случае *б*) предпочтительнее прибегать к вероятностному разбиению совокупности и строить дискретно-непрерывные модели с учетом «весов» принадлежности объектов к классам (рандомизированные модели).

Ясно, что понятия «далекие», «близкие», «изолированные» — условны. Возникает необходимость применения определенных процедур, позволяющих уточнить эти понятия. В общем случае, если мы говорим, что совокупность состоит из k классов, то это означает, что каждому объекту соответствуют k действительных, неотрицательных чисел $p_i^1, p_i^2, \ldots$ $\ldots, p_i^k$, где p_i^l означает вероятность того, что объект z_i относится к l-му классу. Естественно, что $\sum_l p_i^l = 1$ в случае, если z_i-й объект относится только к одному из k классов, и $\sum_l p_i^l < 1$ — в противном случае.

Для получения вероятностей p_i^j отнесения каждого объекта к каждому классу может быть использован, например, подход, описанный в работе [Беккер и др., 1974].

Когда совокупность состоит из нескольких однородных классов объектов и эти классы в пространстве признаков расположены «далеко» друг от друга (случай нерандомизированных весов: $p_i^l = 1$ для объекта z_i, попавшего в класс с номером l, и $p_i^l = 0$ для всех других классов), следует строить дискретно-непрерывную модель в форме набора автономных уравнений регрессии для разных классов. Такой же тип модели можно использовать и тогда, когда моделируемый показатель является количественным признаком, а факторы производства описываются признаками двух типов (качественными и количественными).

Рандомизированные дискретно-непрерывные модели могут оказаться полезными в тех случаях, когда выявлены не-

однородность совокупности и наличие классов в ней, но эти классы расположены близко друг от друга в признаковом пространстве. В такой ситуации этот тип модели дает значительный выигрыш в использовании информации, так как оценка параметров модели производится не по каждому классу в отдельности, а по всей совокупности в целом.

Частным случаем рандомизированной модели является ситуация, когда каждое наблюдение рассматривается как самостоятельный класс и в каждой точке выборочной совокупности оцениваются параметры регрессии с учетом всех остальных наблюдений. По существу, такая операция представляет собой непараметрическое оценивание функции регрессии, позволяющее учесть непрерывную структурную изменчивость моделируемой совокупности. Эти модели обычно используются в неоднородных совокупностях ограниченного объема.

Как было уже сказано в главе 2, к числу параметров дискретно-непрерывных моделей относятся параметры разбиения выборки и параметры внутриклассовых моделей. Напомним, что поиск этих параметров может проводиться тремя различными способами. Первый состоит из двух четко разделенных последовательных этапов: нахождение наилучшего разбиения выборки по выбранному критерию таксономии и поиска внутри фиксированного разбиения оптимальных параметров внутриклассовых регрессий.

Второй способ использует итеративный процесс поочередного корректирования параметров разбиения и параметров регрессий.

Третий заключается в единовременном поиске глобального оптимума поставленной задачи оценивания параметров.

Наиболее распространены и практически реализуемы первые два способа. Несмотря на то, что они не всегда гарантируют нахождение глобального оптимума, их широкое применение объясняется простотой реализации, а также тем, что они обычно обеспечивают необходимую для использования моделей точность. Последнее обусловлено, во-первых, уровнем точности исходной информации и, во-вторых, необходимой точностью экономических расчетов.

Методику построения ДНМ покажем на примере реализации раздельного последовательного поиска параметров классификации и регрессии, как наиболее наглядном. В этом случае построение ДНМ распадается на следующие этапы.

1. Построение дискретной части модели, включающее выбор способа классификации и набора признаков, по которым должно производиться разбиение совокупности на однород-

ные классы; формирование математического описания классов и правила отнесения нового объекта к тому или иному классу, оценку вероятности принадлежности объектов к сформированным классам.

2. Построение непрерывной части модели, содержащее все операции, присущие регрессионному анализу, но выполняемые для каждого класса в отдельности. В случае построения «взвешенной» ДНМ содержание второго этапа несколько модифицируется — параметры внутриклассовых регрессий рассчитываются с учетом вероятностей отнесения всех объектов совокупности к полученным на первом этапе классам.

3. Оценка аппроксимирующей и прогнозирующей способности модели.

Описанная трехэтапная схема построения ДНМ во многом (особенно при выборе «рациональной» классификации) базируется на содержательных и интуитивных соображениях. Покажем принципиальную возможность объединения всех трех этапов в единый алгоритм построения кусочно-линейной функции регрессии, рассматриваемой как задача многомерной классификации с единой целевой функцией.

Построение дискретной части ДНМ. Этот этап часто смыкается с исследованием совокупности на однородность (см. § 4 гл. 1), особенно в тех случаях, когда для обоснования гипотезы неоднородности совокупности используются методы таксономии. Однако и в последнем случае имеет смысл выделить самостоятельный этап построения дискретной части ДНМ, так как он имеет свои специфические особенности и задачи (выбор рационального пространства признаков, числа классов, способа разбиения и т. д.). Специфические требования к «наилучшей» классификации в случае построения ДНМ — это получение возможно более однородных (в смысле близости к нормальному распределению) и в то же время «далеких» друг от друга классов.

При построении дискретной части ДНМ прежде всего определяют, проводить ли разбиение только в пространстве факторов производства (X), или в объединенном пространстве показателей и факторов (Y, X). Если ставится цель построения ДНМ с линейными уравнениями регрессии (а это наиболее распространенная постановка задачи), то это равносильно требованию формирования классов с многомерным распределением показателей и факторов производства, близким к нормальному, а следовательно, и необходимости проведения разбиения в объединенном пространстве (Y, X). Правда, здесь могут возникнуть определенные трудности с

последующим определением класса нового объекта, для которого известны лишь производственные факторы и неизвестно значение моделируемого показателя. Следует заметить, что в большинстве практических ситуаций, когда набор производственных факторов достаточно полно аппроксимирует Y, классификации в пространствах (X) и (Y, X) близки, и проблема отнесения нового объекта к тому или иному классу решается легко. Получение существенно различных классификаций в этих пространствах является сигналом о неполном учете исходными данными производственных факторов, влияющих на моделируемый показатель, и о необходимости уточнения исходных посылок исследования.

Далее, при формировании рабочего признакового пространства необходимо установить, надо ли ориентироваться на комбинацию качественных и количественных признаков, или можно обойтись одним из этих комплексов, что значительно упрощает процесс моделирования.

Во всех случаях в ходе построения дискретной части ДНМ возникает проблема отбора из общего набора признаков тех, которые являются возможной причиной неоднородности совокупности. Здесь возникает дилемма определения рационального числа признаков, образующих рабочее пространство классификации. С одной стороны, чем большее число признаков положено в основу классификации, тем устойчивее таксоны. При малом числе признаков конструкция таксонов может оказаться чрезмерно чувствительной, учет дополнительных признаков в этом случае может привести к нарушению системы классификации, к получению совершенно других таксонов. С другой стороны, включение в набор признаков, не вызывающих явной неоднородности совокупности, распределения которых по разным группам объектов близки, ведет к «смазыванию» границ между классами, к нарушению четкости таксономии. Точного решения проблемы выбора рационального набора признаков пока нет; на практике прибегают к методу последовательных приближений, пробуя разные варианты. Останавливаются на том варианте, который дает устойчивую классификацию и в то же время обеспечивает получение четких границ таксонов.

В число признаков — кандидатов для формирования рабочего признакового пространства в первую очередь включаются существенные качественные признаки. Отбор «претендентов» из количественных признаков требует предварительного исследования характера их распределений и связей с моделируемым показателем. Здесь следует обратить

внимание на те количественные признаки, значения которых располагаются только в определенных интервалах числовой оси, а между интервалами имеются разрывы. Такой характер распределения является свидетельством неоднородности совокупности по данному признаку. В число количественных признаков — кандидатов для классификации следует включать также те, которые имеют многовершинное распределение. В отдельных случаях разбиение признаков описания объекта на используемые для классификации и учитываемые внутриклассовыми моделями определяется самим содержанием экономической задачи. Например, если ДНМ предназначена для измерения технического уровня производства или темпа его изменения, то логично классификацию проводить по прочим (не входящим в характеристику технического уровня) признакам, влияющим на экономические показатели, с тем, чтобы элиминировать их влияние. Измерение темпов технического прогресса тогда можно проводить на основе внутриклассовых регрессий.

Существенную информацию для отбора признаков — кандидатов дает также анализ полей корреляции связи признаков с моделируемым показателем. Обнаружение резких изменений характера связи в интервале изменения того или иного признака говорит о проявлении неоднородности совокупности по данному признаку и необходимости включения последнего в набор. Известно, что исследование парных связей часто не позволяет выявить истинный характер зависимости, поскольку за ними может скрываться влияние других факторов. Но в данном случае при изучении неоднородности совокупности это обстоятельство не мешает, наоборот оно может оказаться полезным, поскольку способствует разбиению совокупности по комплексу взаимосвязанных признаков.

После того как набор признаков для классификации сформирован, можно переходить к выбору способа разбиения совокупности объектов. Следует различать два принципиально различных способа классификации — метод последовательных логических делений по каждому признаку в отдельности и метод формирования групп на основе набора признаков. Ни один из этих методов не является необходимым или достаточным условием принадлежности объекта к данной группе (см. гл. 2).

Первый способ реализуется комбинационной группировкой или ее модификацией — методом последовательных разбиений (см. § 2 гл. 2). При втором способе используются различные алгоритмы таксономии.

Целесообразность применения того или иного способа построения дискретной части модели зависит от характера исходного материала. Если набор признаков — кандидатов включает в основном качественные признаки и близкие к ним, то наиболее целесообразен многошаговый процесс последовательных разбиений. В том случае, когда набор состоит из большого числа количественных признаков, более целесообразно для классификации использовать алгоритмы таксономиц. Здесь в свою очередь встает проблема выбора из множества существенных алгоритмов более подходящего для условий данного конкретного исследования.

При использовании любого из методов разбиения один из важных вопросов построения дискретной части модели — определение рационального числа классов. С этой точки зрения (способа определения числа классов) все алгоритмы таксономии можно подразделить на две группы: алгоритмы, где число классов заранее задается исследователем, и алгоритмы, где число классов в явной форме не задается, а находится в процессе разбиения. Следует отметить, что жесткой границы между этими двумя группами алгоритмов нет. И во второй группе алгоритмов число классов, по существу, определяется исследователем, но не прямо, а опосредованно — либо путем задания шага изменения радиуса описывающей гиперсферы [Загоруйко, 1972], либо путем задания величины порога [Беккер и др., 1974], либо другими аналогичными способами.

Критерием выбора рационального числа классов в идеале должен выступать показатель общей точности модели (подробнее см. гл. 2, § 3). Однако полный перебор всех вариантов классификации и внутриклассовых регрессий в них связан с очень большими объемами расчетов. Поэтому на практике определение рационального числа классов проводят в две стадии: на первой стадии рациональное число классов определяется исходя из частного критерия качества таксономии, на второй стадии осуществляется доводка модели по общему критерию точности всей ДНМ. Форма частного критерия для выбора рациональной группировки определяется целью исследования (получение наиболее адекватной модели, объясняющей большую долю вариации моделируемого показателя). Критерием может, например, служить отношение

$$F = \frac{\sigma^2_{\text{общ}}/(m-1)}{\sigma^2_{\text{ср.взв}}/(m-k-1)},$$

где $\sigma^2_{\text{общ}}$ — дисперсия моделируемого показателя по всей совокупности объектов;

$\sigma^2_{\text{ср.взв}} = \dfrac{\sum_i \sigma_i^2 m_i}{m}$ — средневзвешенная дисперсия моделируемого показателя внутри каждого класса;

m — число объектов в общей совокупности;

m_i — число объектов в i-м классе;

k — число классов.

Классификацию, для которой отношение F максимально, можно считать лучшей из сравниваемых вариантов классификаций.

Таким образом, задача построения рациональной дискретной части модели имеет целый ряд степеней свободы: характер разбиения общего набора признаков на классификационные и учитываемые внутриклассовыми регрессиями, способ и алгоритм классификации, число классов. Все это приводит к многовариантности решения задачи первого этапа построения ДНМ и обусловливает необходимость специального анализа полученных вариантов классификации. Главным направлением такого анализа является проверка полученных классов на однородность (гл. 2, § 2).

Возможны два подхода к выработке критерия, позволяющего остановиться на определенной классификации:

1) строгий подход, основанный на применении статистических критериев;

2) эвристический подход, основанный на изучении тесноты связи между объектами, попавшими в один класс и объектами из разных классов.

В работе Л. А. Верховской [1971] введен информационный статистический критерий для проверки гипотезы об однородности средних значений и корреляционных матриц нескольких независимых нормальных многомерных выборок. Найдена аппроксимация закона распределения информационного статистического критерия. Суть критерия в следующем.

Предположим, что имеется k независимых выборок по m_i реализаций каждая ($i = \overline{1, k}$) из совокупности нормальных популяций. Проверяется гипотеза H_0

$$\mu_1 = \mu_2, \ldots, = \mu_k; \quad A_1 = A_2 = \ldots = A_k.$$

Конкурирующая гипотеза H_1

$\mu_1 \neq \mu_2, \ldots, \neq \mu_k; A_1 \neq A_2, \ldots, \neq A_k$, (т. е. не все равны).

Здесь $\mu_1, \mu_2, \ldots, \mu_k$ — средние значения для $1, 2, \ldots, k$ классов, $A_1, A_2, \ldots, A_k$ — корреляционные матрицы для классов $1, 2, \ldots, k$.

В основу построения критерия положена статистика минимума различающей информации $(* : H)$

$$p\{I(* : H_0) - I(* : H_1) \geqslant c/H_0\} \leqslant \alpha,$$

где α — ошибка первого рода; c — заданный порог.

Информационная статистика для проверки гипотезы об однородности средних и корреляционных матриц имеет вид [Верховская, 1971].

$$Q_{m,k} = I(*: H_0) - I(*: H_1) = \frac{1}{2}\sum_{i=1}^{k}(m_i - 1)\ln\left(\frac{|\widehat{\Sigma}_0|}{|\widehat{\Sigma}_i|}\right).$$

Если справедлива гипотеза H_0, то $Q_{m,k}$ асимптотически распределено как нецентральное χ^2-распределение с $\frac{(k-1)m(m+3)}{2}$ степенями свободы и параметром нецентральности

$$\beta^2 = \sum_{i=1}^{k}\left(\frac{1}{m_i - 1} - \frac{1}{m-k}\right)\frac{m(2m^2+3m-1)}{12} + \frac{m(k-1)(m-k-2)}{2(m-k)}.$$

В качестве эвристического подхода к выработке критерия неоднородности предложен подход, основанный на изучении матрицы вероятностей отнесения каждой из точек к полученным классам [Экономико-статистическое..., 1977].

Будем считать, что каждый объект может принадлежать к любому из полученных классов, но с некоторой вероятностью. Для проверки отдаленности или близости классов поступаем следующим образом:

а) пусть число объектов в обучающей выборке равно m, число классов равно k, тогда построим матрицу L размерности $k \times k$ таким образом: на пересечении i-й строки и j-го столбца этой матрицы будет стоять число m_{ij}, равное числу объектов из класса V_i, принадлежащих классу V_j, которое подсчитывается по формуле

$$m_{ij} = \sum_{z_q \in V_i} p_{\mathbf{z}_{qj}};$$

Из этой матрицы получим матрицу L^+ путем нормировки каждой строки, т. е.

$$L^+ = \|m_{ij}^*\|, \text{ где } m_{ij}^* = \frac{m_{ij}}{\sum_j m_{ij}};$$

б) вычисляются характеристики H_i каждой строки матрицы по формуле

$$H_i = -\sum_j m_{ij}^* \lg m_{ij}^*, \quad i = \overline{1,k};$$

в) вычисляется общая характеристика H по формуле

$$H = \sum_{i=1}^{k} H_i p_i,$$

где H — априорная вероятность появления произвольной точки для класса.

Если H велико, то можно сказать, что классы в пространстве близки, если же H мало, то классы далеки друг от друга.

Приведем результаты применения каждого из рассмотренных критериев в анализе классификаций угольных шахт. Для разбиения шахт в 10-мерном пространстве признаков использовался алгоритм классификации «Рельеф». Проанализируем с точки зрения принятых критериев два варианта разбиения: трехклассовый и двухклассовый. Для трехклассового разбиения получены следующие результаты. Информационная статистика для проверки гипотезы об однородности средних значений и корреляционных матриц имеет значение $Q_{m,k} = 5{,}96$. Следовательно, если верна гипотеза H_0, т. е. общая совокупность однородна в смысле этой группировки, то $Q_{m,k}$ асимптотически распределено как нецентральное χ^2-распределение со 154 степенями свободы и параметром нецентральности $\beta^2 = 18{,}78$. Получая $\chi_0^2 < \chi_{n,p,\beta^2}^2$, делаем вывод о том, что каждый из трех рассматриваемых классов не есть однородная подсовокупность общей совокупности объектов.

Этот вывод подтверждает и результаты проверки по второму (эвристическому) подходу. По изучаемой трехклассовой группировке была получена матрица вероятностей принадлежности каждого объекта к классам. Рассчитанная по ней матрица L размерности 3×3 имеет вид

$$L = \begin{pmatrix} 16{,}4 & 3{,}56 & 6{,}04 \\ 3{,}75 & 18{,}73 & 8{,}52 \\ 6{,}16 & 8{,}12 & 23{,}72 \end{pmatrix}$$

Из этой матрицы получили матрицу L^*

$$L^* = \begin{pmatrix} 0{,}63 & 0{,}14 & 0{,}23 \\ 0{,}12 & 0{,}60 & 0{,}28 \\ 0{,}16 & 0{,}21 & 0{,}63 \end{pmatrix}$$

и следующие значения для H:

$$H_1 = -0{,}63(-0{,}201) + 0{,}14(-0{,}854) + 0{,}23(-0{,}638) = 0{,}394;$$

$$H_2 = -0{,}12(-0{,}921) + 0{,}60(-0{,}222) + 0{,}28(-0{,}553) = 0{,}399;$$

$$H_3 = -0{,}16(-0{,}796) + 0{,}21(-0{,}678) + 0{,}63(-0{,}201) = 0{,}396.$$

Отсюда получаем, что $H = 0{,}395$, т. е. взаимопроникновение полученных групп друг в друга достаточно велико.

Результаты проверки для двухклассовой группировки следующие. По первому критерию при значении параметра нецентральности $\beta^2 = 20{,}68$ получаем $\chi_0^2 > \chi^2_{n,p,\beta^2}$ ($p = 0{,}05$), т. е. гипотеза H_0 отвергается и исследуемые два класса являются однородными подсовокупностями данной совокупности.

Это разбиение удовлетворяет требованиям кусочно-линейной аппроксимации и может служить базой построения ДНМ. Распределение объектов по классам в данном случае произошло с вероятностями, близкими к единице (табл. 4.12). Следовательно, в исследуемой совокупности можно выбрать два непересекающихся класса для построения линейных функций регрессии.

Построение внутриклассовых регрессий. Как уже указывалось ранее, при реализации второго этапа построения ДНМ следует различать две ситуации:

1. Случай хорошей разделимости классов (выполняется гипотеза компактности) — внутриклассовые регрессии строятся без учета вероятностей принадлежности объектов к полученным классам. Оценка параметров регрессий производится с помощью классического метода наименьших квадратов.

2. Классы «размазаны», имеют переходные формы — при оценке параметров внутриклассовых регрессий необходимо учесть неопределенность, возникающую для объектов, находящихся вблизи границ классов. Критерием выбора параметров внутриклассовых регрессий является минимизация функционала (остаточной дисперсии)

$$Q_l = \sum_{j=1}^{m} \| y_j - g_l(x_j) \|^2 p_j^l, \qquad (4.1)$$

где Q_l — величина функционала (остаточной дисперсии) для l-го класса ($l = \overline{1, k}$);

Таблица 4.12

Распределение объектов по вероятности принадлежности к классам

Класс	Значение вероятности					
	До 0,95	0,95—0,96	0,96—0,97	0,97—0,98	0,98—0,99	Свыше 0,99
I	—	2	—	7	40	5
II	2	2	—	5	3	—

y_j — фактическое значение моделируемого показателя;
$g_l(x_j)$ — функция регрессии l-го класса;
p_j^l — вероятность принадлежности j-го объекта к l-му классу;
m — общее число единиц в совокупности.

Из выражения (4.1) следует, что в рассматриваемом случае (случай рандомизированных весов) оценка параметров для каждой внутриклассовой модели производится методом наименьших квадратов со взвешенными наблюдениями, т. е. по существу производится объединение внутриклассовых регрессий в единую модель.

Дело в том, что если некоторый выделенный каким-либо алгоритмом классификации класс объектов достаточно «далек» от остальных классов, то функция регрессии, построенная для него, достаточно хорошо аппроксимирует значения показателей объектов, попавших в этот класс, и значения показателей для новых точек, которые относятся к этому классу по входным параметрам x. Таким образом, мы решаем здесь задачу «чистого» дискретно-непрерывного моделирования (ситуация 1).

Если же классы близки друг к другу, то отнесение нового объекта к этим классам во входном пространстве признаков X может быть неоднозначным. Другими словами, мы предполагаем, что для каждой точки, находящейся близко к границе соседних классов, существуют ненулевые вероятности отнесения ее к каждому из этих классов. При стыковке моделей, построенных для таких классов, необходимо несколько изменить эти полученные внутриклассовые уравнения регрессии таким образом, чтобы учесть неопределенность отнесения точек вблизи границ соседних классов. Учет такой неопределенности можно производить, используя вероятности отнесения каждой выборочной точки к определенному классу. Приведем пример построения ДНМ обоими описанными способами.

Моделировались показатели себестоимости 1 м³ древесины и комплексной выработки на одного рабочего лесозаготовительной промышленности по информации 74 предприятий Урала и Сибири. Вся совокупность предприятий рассматривалась за семь лет.

Методами многомерной классификации были получены три класса (A, B, C), для которых строились как «чистые», так и взвешенные ДНМ. Один из вариантов этих моделей для показателя комплексной выработки приводится ниже.

«Чистая» ДНМ

$A.\ y = 489{,}42 - 0{,}86x_1 - 4{,}23x_4 - 2{,}92x_8 + 0{,}290x_9 + 0{,}380x_{11} + 0{,}080x_{13} + 1{,}42x_{14}$;

$B.\ y = 115{,}0 + 0{,}40x_2 - 0{,}220x_8 + 0{,}26x_9 + 0{,}29x_{14} + 0{,}4x_{16}$;

$C.\ y = 361{,}13 - 1{,}95x_1 + 1{,}31x_2 - 6{,}06x_4 - 1{,}05x_5 + 0{,}364x_6 - 3{,}108x_8 + 0{,}40x_9$.

«Взвешенная» ДНМ

$A.\ y = 552{,}0 - 1{,}934x_1 - 2{,}956x_4 - 2{,}451x_8 + 0{,}356x_9 - 0{,}362x_{11} - 0{,}235x_{13} + 2{,}308x_{14}$;

$B.\ y = 391{,}1 + 1{,}870x_2 - 3{,}178x_8 + 0{,}247x_9 + 1{,}423x_{14} + 0{,}708x_{16}$;

$C.\ y = 210{,}4 + 2{,}847x_2 + 346{,}9x_6 - 1{,}344x_8 + 0{,}164x_9 - 0{,}550x_{10} + 1{,}300x_{13}$

Учитываемые факторы

x_1 — глубина снежного покрова;
x_2 — удельный вес сосны в древостоях;
x_4 — удельный вес лиственницы;
x_5 — удельный вес березы и осины;
x_6 — средний объем хлыста;
x_8 — среднее расстояние вывозки;
x_9 — объем вывозки;
x_{10} — удельный вес примыкания к сплаву;
x_{11} — удельный вес примыкания складов к пунктам потребления;
x_{13} — удельный вес вывозки автотранспортом;
x_{14} — удельный вес вывозки тракторами;
x_{16} — удельный вес лесозаготовок.

Сравнение обоих типов моделей приводится в табл. 4.13. Таким образом, во всех классах использование весов приводит к повышению точности аппроксимации.

В обеих рассмотренных ситуациях второй этап построения ДНМ включает все стадии, обычные для многофакторного регрессионного анализа. Специфика задачи проявляется лишь в особой важности стадии отбора существенных переменных, что связано с уменьшением численности выборки при разбиении совокупности на классы.

Таблица 4.13

Значение множественного коэффициента корреляции

Класс	Себестоимость		Комплексная выработка	
	Невзвешенная модель	Взвешенная модель	Невзвешенная модель	Взвешенная модель
A	0,53	0,61	0,71	0,72
B	0,72	0,85	0,78	0,90
C	0,69	0,78	0,78	0,86

Обычно число количественных переменных, включенных в анализ и имеющих значительную внутриклассовую вариацию, достаточно велико (15—20 и более). Включение всех этих переменных во внутриклассовые регрессионные модели при ограниченном объеме подсовокупностей практически невозможно, так как, во-первых, ведет к чрезмерному расширению доверительных областей оценок, во-вторых, снижает точность оценок из-за работы с приближенными данными и, в-третьих, затрудняет интерпретацию.

Поэтому при построении ДНМ задача отыскания небольшого набора признаков, которые объяснили бы основную часть внутриклассовой вариации моделируемого показателя, особенно важна. Начальным этапом этого процесса является анализ структуры матриц парных коэффициентов и выделение связанных групп признаков. После исключения дублирующих и тесно коррелированных переменных производится выбор существенных переменных для построения частных регрессий, для чего можно использовать разные способы. Предварительно следует проанализировать уровень внутриклассовой вариации переменных после разбиения совокупности. Наиболее распространенные формальные способы отыскания существенных переменных базируются на последовательном исключении (добавлении) переменных, дающих наименьшее (наибольшее) изменение достаточной дисперсии по сравнению с исходным набором.

Когда ДНМ применяется для прогнозирования показателей новых объектов, в ряде случаев оказывается целесо-

образным использовать не стабильные уравнения, а специфические частные уравнения регрессии, учитывающие особенности характеристик данного объекта. Отбор признаков—кандидатов для внутриклассового уравнения регрессии при этом производится следующим образом. Сначала определяют принадлежность нового объекта к тому или иному классу, затем его количественные характеристики сравниваются со средними значениями характеристик данного класса, в качестве признаков — кандидатов для частной регрессии выступают те количественные переменные, по которым обнаруживается существенное расхождение среднегрупповых и индивидуальных значений. Таким образом, здесь вместо одной частной регрессионной модели для каждого класса используется набор специфических уравнений, структура которых определяется особенностями исследуемых объектов.

Отметим, что для объектов совокупности, выделившихся с единичными вероятностными в малочисленные самостоятельные классы, целесообразно строить дискретные модели. Здесь в качестве расчетного значения моделируемого показателя выступает среднее значение $\bar{y}_l$, соответствующее определенному классу l.

Что касается этапа оценки аппроксимирующей и прогнозирующей способности ДНМ, то он подробнее будет рассмотрен в главе 5.

Покажем описанную выше технику построения ДНМ на простом примере.

Решалась задача изучения влияния социально-демографических характеристик рабочего (факторов) на уровень заработной платы y (показатель) на заводе железобетонных изделий. Обследованием было охвачено 230 рабочих цехов основного производства. Для каждого из них фиксировались три группы факторов: демографические (возраст —x_1, состав семьи — x_2), функциональные качества рабочего (образование — x_3, общий трудовой стаж — x_4, стаж работы на данном предприятии — x_5, разряд — x_6, число дней невыходов по болезни — x_7, число дней невыходов по другим причинам — x_8, длительность отпуска — x_9), производственные условия (дополнительный отпуск — x_{10}).

В качестве рабочей гипотезы было выдвинуто положение, что совокупность обследуемых рабочих неоднородна и распадается на ряд подсовокупностей со специфическим для каждой подсовокупности характером зависимости моделируемого показателя от социально-демографических факторов. Причиной этому служат различия в уровне функциональных качеств рабочих. Кроме того, производственный

Таблица 4.14

Типология постоянных рабочих основного производства

Фактор, показатель	Обозначение	Средние значения факторов и показателей по классам			
		I	II	III	IV
Возраст, лет	x_1	47,1	44,2	38,5	35,1
Состав семьи, чел.	x_2	3,5	3,6	3,4	2,7
Образование, классов	x_3	6,73	7,02	7,76	8,02
Общий трудовой стаж, лет	x_4	26,5	24,0	19,1	14,9
Стаж работы на предприятии, лет	x_5	19,4	13,3	8,2	3,3
Разряд	x_6	3,8	3,6	3,4	3,2
Число невыходов по болезни в году, дней	x_7	10,5	14,2	13,7	15,6
Число невыходов по другим причинам, дней	x_8	0,05	0,08	0,16	0,24
Длительность годового отпуска, дней	x_9	27,6	28,3	27,6	25,4
Дополнительный отпуск, дней	x_{10}	6,32	6,57	6,32	6,07
Размер годового пособия по временной нетрудоспособности, руб.	x_{11}	98,4	133,2	117,4	109,6
Среднемесячная заработная плата, руб.	y	211,9	208,1	199,8	179,6

коллектив состоит из представителей разных поколений, отличающихся свойственными только им характеристиками.

С помощью метода многомерной классификации, основанного на непараметрическом оценивании функции плотности распределения выборочных объектов [Беккер и др., 1974], выборка была разбита на четыре класса (средние значения факторов по классам приведены в табл. 4.14).

Классы характеризуются следующими содержательными признаками:

I класс (56 рабочих) — средний возраст 47 лет, 63% среди рабочих составляют женщины. Стаж работы на предприятии около 20 лет. Образование 6 классов, средний разряд — 4-й, среднемесячная зарплата 212 руб. В выборочной совокупности численность класса — 24,3%.

II класс (53 рабочих) — средний возраст 44 года. Стаж работы на предприятии более 14 лет. Образование 6 классов. Среднемесячная зарплата 208 руб., разряд 4-й. В выборочной совокупности численность класса — 23,1%.

III класс (37 рабочих) — средний возраст 38 лет. Стаж работы на предприятии около 8 лет, образование около 8 классов, разряд 3-й, среднемесячная зарплата 200 руб. В выборочной совокупности численность класса — 16,1%.

IV класс (84 рабочих) — средний возраст 35 лет. Стаж работы на предприятии 3 года. Образование 8 классов, разряд 3-й, среднемесячная зарплата 180 руб. В выборочной совокупности численность класса — 36,5%.

Наиболее сильным классификационным признаком оказался стаж работы на предприятии. Таким образом, выделились классы рабочих, длительно работающих на предприятии (от 14 до 20 лет), группа рабочих, в основном адаптировавшаяся на предприятии (стаж работы около 8 лет) и самая многочисленная группа молодых рабочих, стаж работы которых на предприятии несколько выше трех лет. Оказалось — и в этом одна из положительных сторон проведенной классификации,— что различия между классами проявились почти по всем использованным переменным. По ряду переменных (возраст, заболеваемость, квалификация) изменения идут строго по вертикали от первого класса к четвертому (возраст, квалификация). В дискретной модели ярко представлен социальный процесс — смена поколений рабочих: в I класс вошла группа рабочих, большинство которых в ближайшие 5—8 лет получат право на пенсию по возрасту (женщины 55 лет) или на льготных основаниях (женщины — 50, мужчины — 55 лет). В IV класс включены рабочие, стаж трудовой деятельности которых до получения права на пенсионное обеспечение составляет для женщин 20 лет, для мужчин 25 лет (соответственно по льготной пенсии — на 5 лет меньше).

Проведенная классификация позволила выделить устойчивые типологии рабочих основного производства.

Второй этап построения ДНМ заключался в построении и анализе зависимости заработной платы рабочих от определяющих ее социально-демографических характеристик для каждого выделенного класса. В табл. 4.15 приведены характеристики моделей (в натуральном и стандартизованном масштабах) для полученных четырех классов.

Как видно из таблицы, типологические группы отличаются не только величиной параметров уравнений, но также набором существенных переменных.

Сравним, например, модели I и IV классов. Из переменных, входящих в модели, пересекаются только две: число дней невыходов по болезни и размер годового пособия по временной нетрудоспособности.

Характерным для модели группы старшего поколения (I класс) является сильное влияние квалификации и стажа работы на предприятии на уровень заработной платы. Для группы молодых рабочих (IV класс) факторами, оказываю-

Таблица 4.15

Характеристики моделей заработной платы рабочих

Класс	Модели-руемый показа-тель	Коэффициенты регрессии													
		R^2	$\sigma_{ост}$	a_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
I	y	0,39	47,23	−103,27					7,78 0,153	15,83 0,377	−1,09 −0,407		3,65 0,288		0,16 0,428
II	y	0,43	42,80	92,02			4,22 0,161			13,81 0,401	−0,88 −0,331			3,45 0,275	0,20 0,546
III	y	0,52	43,46	56,76						20,34 0,504		−61,23 −0,435	3,58 0,295	−2,72 −0,172	
IV	y	0,26	51,47	86,97			8,95 0,355	1,88 0,295			−2,18 −0,950				0,24 0,837

Примечание. b_i/β_i: в числителе — оценка соответствующего коэффициента регрессии в натуральном масштабе, в знаменателе — оценка коэффициента в стандартизованном масштабе. Относительно небольшое значение коэффициента детерминации связано с тем, что в исследовании ставилась цель определить влияние, оказываемое на моделируемый показатель лишь групповой социально-демографических факторов, и не учитывалось влияние других факторов.

щими существенное влияние на уровень заработной платы, являются общее образование и общий стаж работы.

Отсюда ясна невозможность описания моделируемого показателя единой непрерывной моделью.

§ 3. ДИНАМИЧЕСКИЕ ДИСКРЕТНО-НЕПРЕРЫВНЫЕ МОДЕЛИ (ПРОСТРАНСТВЕННО-ВРЕМЕННЫЕ ВЫБОРКИ)

Эффективность применения той или иной концепции о структуре модели экономической системы существенно зависит от наличия соответствующей статистической информации. В частности, для реализации идеи построения моделей с переменной структурой (или, что то же, описания экономической динамики моделью управляемого случайного процесса [1]) необходимо иметь семейство реализаций экономического процесса. Принципиальное отличие экономических процессов, например, от физических состоит прежде всего в практической невозможности многократного их воспроизведения с достаточной для исследования точностью. Таким образом возникает важная методологическая проблема получения семейства реализаций значений изучаемого экономического процесса на одном и том же фиксированном отрезке времени. Решение этой проблемы связано с использованием в теории и практике статистического моделирования аппарата многомерной классификации. Дело в том, что до появления методов многомерной классификации единственно доступной статистической информацией о поведении во времени экономического объекта была единственная реализация (один одномерный или многомерный временной ряд). Применение методов многомерной классификации позволяет реализовать идею рассмотрения множества пространственных реализаций однотипных экономических систем на фиксированном (сравнительно коротком) отрезке времени как различные варианты развития в динамике одной системы, т. е. идею о том, что типы объектов, выделенные в одном или нескольких пространственных сечениях процесса в то же время являются ступенями развития объекта в динамике (эргодичность). Принадлежность к определенному типу — внутренняя характеристика объекта. В формальном отношении это позволяет рассматривать случайные процессы, описывающие выходные показатели однотипных систем как семейство реализаций, представляющих один управляемый случайный процесс.

[1] В самом общем виде управляемый случайный процесс определяется как семейство случайных процессов, с каждым элементом которого ассоциирован параметр, называемый параметром управления.

Во всех случаях динамическая модель, построенная на базе анализа данных по совокупности предприятий, имеет ряд преимуществ перед динамическими моделями отдельного объекта (см. гл. 3). Как правило, на информации по отдельному объекту затруднительно реализовать идею построения полной модели с переменной структурой. Дело в том, что объект зачастую не реализует большинства возможных состояний структуры связей. Даже для такого предельного случая, когда история развития предприятия завершена и фактически его модель уже не имеет практического значения, не будут реализованы многие возможные состояния, представляющие интерес для прогнозирования, планирования и управления.

Кроме того, наличие большого числа реализаций позволяет повысить надежность и устойчивость оценок параметров моделей. Динамические модели с переменной структурой, построенные на базе информации по совокупности объектов, обладают максимально возможной для статистических моделей областью приложения, ибо они описывают большинство реализованных состояний. Более того, анализ этих моделей позволяет конструировать и новые, потенциально возможные состояния объектов.

При использовании данных пространственно-временных выборок следует различать три случая, для которых методы построения моделей с переменной структурой различны.

1. Сравнительно большое число объектов с короткими временными реализациями ($m \gg 1,\ k \geqslant 1,\ p \geqslant 1,\ t \geqslant 1$).

2. Большое число объектов со сравнительно длинными реализациями ($m \gg 1,\ k \geqslant 1,\ p \geqslant 1,\ t \gg 1$).

3. Небольшое число объектов со сравнительно длинными реализациями ($m > 1,\ k \geqslant 1,\ p \geqslant 1,\ t \gg 1$).

Если исходная информация охватывает большое количество объектов, но протяженность динамических рядов невелика (случай 1), то общую динамическую ЭСМ нужно строить на базе частных статических моделей. Иными словами, частные статические (пространственные) модели должны быть отправным пунктом конструирования общей динамической модели.

Прием конструирования динамической ЭСМ на базе частных динамических моделей, построенных для каждого объекта совокупности, применим в том случае, когда протяжен-

ность многомерных динамических рядов достаточно велика (случай 2). Частные динамические модели (временное сечение процесса) при этом могут быть довольно разнообразны. Это и модели одномерные, отражающие только изменение моделируемого показателя во времени и в связи со своей предысторией, но не учитывающие взаимозависимостей его с факторами производства и многомерные, учитывающие наряду с вышесказанным зависимости моделируемых показателей от факторов производства в текущий и некоторые предшествующие моменты времени.

Анализ частных динамических моделей позволяет выявить определенные закономерности и различия развития объектов в динамике. Пространственный анализ полученных параметров частных динамических моделей с помощью различных методов математической статистики и теории распознавания образов позволит выделить группы предприятий, однородных в смысле происходящих в них динамических изменений, и для каждой из них построить общую динамическую модель. Недостаток такого подхода — громоздкость вычислений.

Третья группа способов конструирования динамической ЭСМ характеризуется тем, что строится одна модель для общей совокупности, в которую каждое предприятие входит несколько раз, т. е. в качестве объекта наблюдения выступают «заводо-годы» (случай 3).

Подходы, отнесенные нами к первым двум группам — построение динамической ЭСМ на базе частных моделей (динамических или статических)— предполагают наличие информации о работе достаточно большого числа предприятий ($m > 30$) в течение определенного периода времени. В экономической практике в ряде случаев возникают ситуации, при которых необходимо проанализировать и построить динамическую модель (с возможностью использования ее в качестве прогнозной) для небольшого числа промышленных объектов. В этом случае обычно и применяют так называемый прием «заводо-годы», при котором анализируется одна общая совокупность без дифференциации во времени. Каждый объект при этом выступает в качестве самостоятельного наблюдения несколько раз, увеличивая тем самым объем совокупности.

Нет необходимости доказывать, что такой прием не всегда правомочен, так как в этом случае исследуется фактически не одна совокупность, а смесь нескольких совокупностей с характеристиками и взаимозависимостями, изменяющимися во времени. В каждом отдельном случае необходимо либо

показать, что полученную совокупность «заводо-лет» можно без особого ущерба считать однородной, либо выделить действительно однородную подсовокупность устойчивых во времени предприятий и для нее строить модель, либо привести, в крайнем случае, какие-либо рассуждения на эту тему и пользоваться моделью с определенными оговорками.

Построение динамических моделей на основе информации по совокупности объектов (большое число объектов со сравнительно короткими временными рядами, $m > 30$, $t \geqslant 1$). Данный случай наиболее типичен для большинства экономических задач, поскольку, как правило, затруднительно получение сопоставимой информации о функционировании большого числа экономических объектов за длительный период времени. Отраслевая система включает «разновозрастные» предприятия, общее число которых ограниченно (несколько десятков, в крайнем случае, сотен).

Построение моделей с переменной структурой связано с исследованием изменения типологии совокупности во времени. При данном подходе она включает набор пространственных сечений (состояние совокупности объектов в фиксированные последовательные моменты времени $t_0, t_1, \ldots, t_\omega$). Каждое пространственное сечение в неявном виде включает элементы динамики: объекты, входящие в него, находятся на разных точках траектории своего развития, имеют различную предысторию, неодинаковый технический уровень.

Типологическую структуру каждого пространственного сечения можно характеризовать тремя элементами: составом классов; количественными характеристиками объектов, входящих в выделенный класс; числом классов. В общем случае, все три элемента типологической структуры изменяются во времени.

Изменения двух элементов имеют систематический плавный характер, но скорость их протекания переменная. В отдельные периоды эти процессы протекают более интенсивно, в другие — менее. Третий элементарный процесс — изменение набора классов — имеет дискретный периодический характер.

Изменение базисной типологической структуры совокупности (t_0) происходит за счет двух источников — развития действующих на период t_0 предприятий и ввода в эксплуатацию новых предприятий.

Чтобы более наглядно проиллюстрировать существо задачи построения и использования динамических типологий при конструировании моделей с переменной структурой, целесообразно начать изложение с примера, когда изменение

структуры связей можно отнести в основном за счет различий в предыстории развития объектов. Такая ситуация имеет место в отраслях с устоявшейся технологией при изучении сравнительно небольших календарных периодов (5—10 лет). В частности, такая ситуация имела место при изучении работы лесозаготовительных предприятий Красноярского края [Кулешов и др., 1972] и нефтяных месторождений Западной Сибири [Сергеева, Ягольницер, 1979]. В последнем случае совокупность включала небольшое число действующих месторождений (15), находящихся либо на стадии освоения, либо на стадии стабилизации добычи нефти, так как нефтегазовые ресурсы Западной Сибири начали интенсивно осваиваться в 70-е годы.

Исследовалось влияние на объем добычи нефти на месторождениях геолого-промысловых (проницаемость коллекторов, вязкость нефти, внутрипластовое давление и др.) и технико-технологических факторов (плотность сетки скважин, объем закачиваемой в пласт воды, фонд нагнетательных скважин и др.). С целью учета положения предприятий на траектории развития была проведена их классификация по «возрасту» и для каждого «возрастного» слоя строились модели. В качестве примера приведем варианты моделей для месторождений первого (y^1) и пятого (y^5) годов эксплуатации.

$$y^1 = 29{,}26 + 91{,}98x_2^1 - 1{,}29x_5^1 - 6{,}36x_8^1;$$

$$y^5 = -7284{,}5 + 81{,}65x_6^5 + 93{,}37x_{12}^5,$$

где y^1, y^5 — объемы добычи нефти для первого и пятого годов эксплуатации, тыс. т;

x_2^1 — проницаемость коллектора, дарси;

x_5^1 — относительная вязкость;

x_8^1 — коэффициент расчлененности;

x_6^5 — давление насыщения, кгс/см^2;

x_{12}^5 — фонд нагнетательных скважин, шт.

Таким образом, модели, полученные для разных точек траектории развития месторождений, подтверждают гипотезу о неодинаковой структуре связей переменных на различных этапах развития предприятий. На этапе освоения месторождений основная роль в определении добычи нефти принадлежит геолого-промысловым факторам. По мере перехода к этапу стабилизации добычи резко возрастает значение технико-технологических факторов, а также той части геолого-промысловых, которые связаны с технологией добычи нефти.

Общий случай выделения пространственно-временных блоков требует учета всех причин изменения типологии совокупности (предыстория, технический уровень производства, природные условия, эволюционные факторы). Здесь могут быть использованы различные методические подходы. Их основой являются два направления формирования динамических типологий. Первое основано на построении независимых пространственных типологий для каждого фиксированного периода времени (год, квартал, месяц и т. д.) с последующим изучением взаимосвязей между ними. Второе заключается в построении пространственно-временных типологий с фиксированными во времени границами путем объединения информации разных периодов. Рассмотрим подробнее каждое из указанных направлений.

При использовании однородных групп за каждый год приходится устанавливать устойчивость распределения однородных групп объектов во времени, и некоторое соответствие между группами объектов, полученных в разные годы. После разбиения статистической совокупности на однородные группы по годам нужно определить некоторый критерий, указывающий степень устойчивости распределения классов в рассматриваемом периоде. В качестве одного из приемов анализа устойчивости можно использовать построение последовательности таблиц взаимной сопряженности классификаций двух периодов и расчет на их основе некоторых статистик взаимосвязи (χ^2, коэффициенты взаимной сопряженности Пирсона и Чупрова и др.). На основе этих коэффициентов может быть построена квадратная матрица Λ, характеризующая взаимосвязи классификаций разных периодов, имеющая размерность $\omega - 1$, где ω — число временных отсчетов. Исследование этой матрицы позволяет сделать некоторые выводы о взаимосвязях классификаций. Например, если заполненность матрицы небольшая и ее элементы концентрируются вокруг главной диагонали, то можно утверждать, что наблюдаются лишь существенные взаимосвязи по последовательным периодам. Используя матрицу Λ в качестве характеристики взаимосвязи классификаций, можно организовать итеративный процесс поиска набора разбиений, наиболее удовлетворительного с позиции построения динамических моделей. Наиболее удовлетворительным следует считать такой набор пространственных разбиений, при котором число классов по пространственным сечениям либо стабильно, либо изменяется незначительно; в одноименных классах разных сечений должно присутствовать устойчивое «ядро», включающее достаточную долю объектов класса.

Для выбора представителей «ядра» классов и оценки общей устойчивости «ядра» можно прибегнуть к расчету информационной меры устойчивости объектов в классе.

Если число сечений обозначить через ω, а номер класса через l ($l = \overline{1, k}$), то каждому объекту i можно поставить в соответствие ω-мерный вектор $D_l^i = (D_{l1}^i, D_{l2}^i, \ldots, D_{l\omega}^i)$, $i = \overline{1, n}$.

$$D_{lj}^i = \begin{cases} 0, & \text{если } i\text{-й объект не попадает в класс } l \text{ в период } j; \\ 1, & \text{если } i\text{-й объект попадает в класс } l \text{ в период } j. \end{cases}$$

Понятно, что $\sum_{j=1}^{\omega} \sum_{l=1}^{k} D_{lj}^i = \omega$.

Мера неопределенности, возникающая при отнесении объекта к типу устойчивых, может быть рассчитана для каждого объекта по формуле

$$H^i = -\sum_{j=1}^{\omega} \sum_{l=1}^{k} \frac{D_{lj}^i}{\omega} \log \frac{D_{lj}^i}{\omega}.$$

В этом случае показателем устойчивости может служить величина μ^i, $\mu^i = 1 - H^i$.

Для формирования устойчивого «ядра» определенного класса отбираются объекты, у которых μ^i превышает некоторый заданный порог (например, 0,7—0,8). Эта же мера может быть использована для построения общей характеристики устойчивости типологии. Для этого необходимо сгруппировать объекты по величине μ^i и проанализировать это распределение. Приведем пример такого анализа устойчивости для 176 леспромхозов и четырех пространственных сечений совокупности (табл. 4.16).

Понятно, что первая ситуация (см. табл. 4.16) характеризует полностью устойчивые предприятия, к устойчивым можно отнести и вторую ситуацию ($\mu^i = 0{,}6$). Эти ситуации реализуются по 128 объектам, что составляет 61,4% общего числа наблюдений.

Разумеется, выделение или невыделение групп всегда связывается с некоторым порогом, задание которого производится по-разному в каждой из решаемых конкретных задач.

Разбиение всего периода на несколько подпериодов позволяет рассматривать каждый из подпериодов отдельно как устойчивую по годам выборку. В общем случае разбиение может быть и вероятностным. Другими словами, будем счи-

Таблица 4.16

Значения показателей неопределенности и устойчивости для группы леспромхозов

Возможная ситуация	Мера неопределенности, H^i	Мера устойчивости, μ^i	Число предприятий
1. Все годовые реализации попали в один таксон	0,0	1,0	66
2. Три годовых реализации попали в один таксон, а одна — в другой	0,4	0,6	62
3. Две годовых реализации в одном таксоне и две — в другом	0,5	0,5	33
4. Две годовых реализации в одном таксоне, а две распределились между двумя другими таксонами	0,75	0,25	15
5. Все годовые реализации попали в разные таксоны	1,0	0,0	0

тать, что имеется несколько классов с заданным набором вероятностей принадлежности объектов к ним. После получения классов по каждому из них необходимо построить частные модели. В самом благоприятном случае их параметры стабильны, в противном случае для оценки их дрейфа могут быть использованы динамизации многофакторных статических моделей.

Описанный выше этап представляет собой изучение динамики типологии совокупности. В качестве примера приведем анализ изменчивости пространственной типологии шахт Кузбасса в пяти временных сечениях, охватывающей десятилетний период. Кузнецкий угольный бассейн является удобным объектом для иллюстрации изменений типологии во времени, так как включает достаточное число шахт с большим разнообразием природных и технико-организационных условий добычи угля. Классификация осуществлялась в 11-мерном факторном пространстве: учитывались следующие факторы: газообильность (x_1), максимальная глубина разработки (x_2), водообильность (x_3), распределение добычи угля по углам падения ($x_4 - x_6$), распределение добычи угля по мощности пластов ($x_7 - x_{10}$), количество разрабатываемых пластов (x_{11}). Использовался непараметрический алгоритм многомерной классификации «Рельеф» (см. [Экономико-статистические..., 1977]) со стабильным параметром аппроксимации $h = 0{,}3$. Полученные результаты сведены в табл. 4.17 и 4.18.

Из табл. 4.17 видно, что, несмотря на существенное изменение объема выборочной совокупности и сравнительно

Таблица 4.17

Изменение числа и наполненности классов по сечениям

Год	Число шахт	Общее число классов	Численность объектов по неединичным классам							
			I	II	III	IV	V	VI	VII	VIII
1963	72	23	13	29	7	2	2	2	2	—
1966	77	25	10	35	4	4	2	2	2	2
1968	80	24	13	30	6	5	5	5	2	2
1971	69	20	21	26	2	2	2	2	2	—
1972	71	22	18	25	4	4	3	2	—	—

длительный период обследования (10 лет), основные параметры типологии сравнительно устойчивы: общее число классов меняется в пределах 20—25; число неединичных классов 6—8; примерно половина всех шахт (35—40) распределены между первыми двумя классами.

Из табл. 4.18, в которой приведены доли одинаковых объектов в представительных классах за два ближайших периода времени, ясно, что во II классе легко выделяется устойчивое во времени «ядро». Что касается I класса, то из анализа данных следует, что характер типологии изменился в период между 1968 и 1971 гг.

Следующий за выделением устойчивых во времени «ядер» классов этап исследования — динамизация внутриклассовых регрессий. Возможная процедура динамизации набора пространственных моделей для каждого класса состоит из двух этапов: построение уравнений регрессии, учитывающих лаги во влиянии основных факторов и тенденцию; выявление и учет динамики коэффициентов уравнений регрессии. Таким образом, получаются модели с различной степенью учета динамических закономерностей.

Результаты первого типа представляют собой уравнение регрессии, построенное по преобразованной информации. После получения промежуточных моделей, т. е. статических моделей, в которых учтены лаги и характер тенденции, переТаблица 4.18

Пересечения двух основных классов по сечениям

Класс	Процент совпадений в парах последовательных сечений			
	1963—1966	1966—1968	1968—1971	1971—1972
I	40,0	30,3	38,5	72,2
II	82,8	96,7	84,6	84,0

ходят к исследованию поведения параметров этих моделей во времени и их учету.

Применяемая схема динамизации пространственных (статических) многофакторных моделей совокупности объектов сводится к следующему: строится набор пространственных регрессий за каждый год исследуемого периода в устойчивых «ядрах»; на основе динамических моделей отдельных типичных объектов с длительной предысторией выявляются лаги и тенденции; с учетом этих характеристик производится преобразование исходной информации (сдвиг факторов во времени, введение новых переменных, введение разных форм фактора времени); осуществляется построение нового набора пространственных моделей по преобразованной информации и исследуется динамика параметров полученного набора пространственных уравнений регрессии.

Резюмируя, можно сказать, что моделирование процесса развития отраслевой системы при использовании указанных подходов включает следующие действия: выделение устойчивых «ядер» классов, оценка динамики характеристик классов и выделение пространственно-временных типов; построение динамических многофакторных внутриклассовых моделей.

Модель для случая линейной аппроксимации при использовании описанного выше подхода может быть представлена в виде

$$y^l(t) = \sum_{j=1}^{n} \sum_{\tau=1}^{t_j} a^l_{j\tau}(t)\, x_j(t-\tau) + \sum_{\tau=1}^{t_0} b^l_\tau(t)\, y(t-\tau) + c^l(t);$$

$$\{x(t),\ y(t)\} \in O_l,$$

где $y^l(t)$ — значение моделируемого показателя в момент t для объектов, принадлежащих классу O_l $(l = \overline{1, k})$;

$x(t)$ — вектор независимых переменных в момент t;

$a^l_{j\tau}(t)$, $b^l_\tau(t)$, $c^l t$ — параметры динамизированных внутриклассовых регрессий;

n — общее число факторов;

t_j — число временных сдвигов каждого фактора;

t_0 — число временных сдвигов моделируемого показателя.

Эта модель учитывает все виды изменения характера моделируемой зависимости во времени: скачкообразное —

путем изменения структуры модели в разных классах; эволюционное — путем представления изменения параметров модели в виде некоторой функции времени и введения авторегрессионной части; запаздывающие — путем введения лаговых переменных. Однако построение такого развернутого вида модели не всегда целесообразно и необходимо. В ряде случаев могут быть реализованы ситуации, когда возможно упрощенное представление зависимости.

Например, если выявленная структура связей для фиксированных классов устойчива во времени, отпадает необходимость в динамизации параметров модели. Противоположная ситуация — типологическая структура совокупности однородна (нет скачкообразных изменений зависимости), но существенны эволюционные изменения. В этом случае нет необходимости в разбиении выборки на классы, но нужна динамизация параметров модели.

Приведенный подход — использование независимых и пространственных типологий для конструирования динамической модели — позволяет получить достаточно полную информацию о поведении типологической структуры совокупности во времени, но в то же время имеет и ряд моментов, усложняющих процесс моделирования. К ним прежде всего относятся трудность идентификации классов разных лет; подвижность границ классов и изменение их характеристик во времени; трудность разложения общего изменения характеристик объекта во времени на изменение их, вызванное движением класса объектов как целого, и вызванное собственным движением объекта внутри класса (или между классами).

В этих обстоятельствах необходимо использовать в ряде случаев несколько иной подход к построению динамических дискретно-непрерывных моделей.

Его принципиальное отличие от описанного выше заключается в том, что объектом классификации выступает не пространственное сечение, а объединенный информационный массив разных временных периодов. Здесь каждая годовая реализация выступает как отдельное наблюдение. Границы классов при данном подходе стабильны во времени. Поэтому отпадает необходимость в идентификации классов объектов разных лет и в проверке устойчивости классификации. Кроме того, к достоинствам следует отнести возможность анализа движения объектов без искажающего влияния смещения характеристик классов. Параметры внутриклассовых моделей могут строиться методом «объекто-годы» с учетом лагов, тенденций и т. п.

Однако условием применимости этого подхода является инвариантность признакового пространства разбиения во времени — возможность использования стабильного во времени набора признаков. В связи с этим требуется предварительный этап проверки гипотезы о стабильности признакового пространства.

Для исследования движения отдельных объектов могут быть построены матрицы переходов и рассчитаны меры устойчивости объектов. С помощью этих характеристик можно выделить типы траекторий объектов по классам, учесть влияние базисного состояния на последующее движение, а также прогнозировать будущие переходы объекта.

Построение динамических моделей на основе информации по совокупности объектов (большое число объектов с длительной предысторией, $n > 30$, $t \gg 1$). Суть этого способа построения динамических моделей с переменной структурой заключается в выделении классов объектов (частей временных реализаций), однородных по типу динамики процесса с построением общей внутриклассовой модели. Теоретически этот подход перспективен, поскольку нацелен на прямое выявление различных типов динамики. Однако практическая его реализация наталкивается на серьезную трудность — необходимо иметь значительное число длинных временных реализаций по сопоставимым этапам развития экономических объектов. Последнее обусловливает разработку способов преодоления этого затруднения. Они сводятся, во-первых, к преобразованию исходной информации с целью достижения сопоставимости и, во-вторых, к использованию смешанных подходов, когда динамическая типология строится по временным рядам одного показателя (выходного), а различие классов устанавливается на базе использования сечений (одного или нескольких) в пространстве факторов производства. Поскольку опыт исследований свидетельствует о том, что на траекторию развития объекта большое влияние оказывает его базисное состояние, в качестве одного из сечений используется классификация по характеристикам базисного состояния.

Рассмотрим подробнее методику конструирования динамической модели для последнего случая.

Схема решения сформулированной задачи включает следующие этапы.

I. Построение индивидуальных кривых динамики показателя каждого объекта в сопоставимом периоде и формирование ансамбля реализаций выходного показателя. Во многих случаях на этом этапе нет необходимости рассматривать

непрерывные функции, можно ограничиться и заданием индивидуальных кривых в виде набора дискретных точек в пространстве (Y, T) («показатель — время»). Для обеспечения сопоставимости необходимо на кривых динамики показателей выделить характерные точки и участки [2].

Для выделения характерных точек могут быть использованы методы, подробно описанные в гл. 3. Одним из таких способов, например, является введение функционала отличия и анализ его поведения во времени. Функционал отличия может конструироваться как разница количественных характеристик кривой на соседних участках. Появление локальных экстремумов и превышение заранее заданного порога фиксирует наличие характерных точек. В качестве частного случая такого функционала можно рассматривать ошибки фильтрации (следящий сигнал) при выравнивании ряда методами скользящего среднего или экспоненциального сглаживания. В истории развития предприятия характерными точками обычно являются момент пуска в эксплуатацию, моменты полного освоения проектной мощности и проектных значений экономических показателей, моменты проведения реконструкции или внедрения новых технологических способов производства и т. д. Для обеспечения сопоставимости формирование ансамбля должно проводиться только из временных реализаций, относящихся к определенному характерному участку кривой. Кроме того, может потребоваться приведение отрезков к единой длине путем перехода на условные измерители.

II. Формирование обучающей выборки в пространстве выходных показателей (Y, T), для чего осуществляется многомерная классификация объектов по характеру кривых (вернее анализируемых отрезков кривых). При классификации индивидуальных отрезков кривых могут использоваться различные алгоритмы таксономии, при этом отдельные временные реализации могут задаваться как всеми точками, так и наиболее информативными. Результаты этого этапа — построение алфавита классов типичных кривых для определенного характерного участка траектории развития предприятия (например, для периода освоения) и выбор предприятий — представителей каждого класса.

III. Математическое описание полученной на I этапе системы классов — эталонов в новом признаковом пространстве, т. е. переход из пространства «показатель — время» (Y, T) в пространство «факторы производства — время» (X, T) и обучение распознаванию принадлежности его к классам.

IV. Преобразование по результатам обучения пространства (X, T) в пространстве (X', T). Это преобразование заключается в минимизации описания классов и при необходимости в придании весов отдельным признакам.

V. Формирование классов O_j ($j = \overline{1, m}$) в преобразованном пространстве (X', T) методами таксономии. Классификация, полученная на этом этапе, может быть названа «конечной», она выступает как дискретная часть дискретно-непрерывной модели показателя.

VI. Построение типичных кривых динамики показателя для всех реализаций, попавших в данный класс. Содержание этого этапа заключается в поиске наиболее подходящего способа описания внутриклассовой динамики показателя, заданного конечным числом реализаций. Для решения этой задачи могут использоваться методы регрессионного анализа, авторегрессионные схемы, методы теории сплайнов, а также их возможные комбинации.

В случае использования комбинации методов регрессионного анализа с авторегрессионной схемой общий вид модели может быть следующим:

$$y^*(t) = \begin{cases} g_1[t, y(t-1), y(t-2), \ldots, y(t-k)], & x \in O_1; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ g_i[t, y(t-1), y(t-2), \ldots, y(t-k)], & x \in O_i; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ g_m[t, y(t-1), y(t-2), \ldots, y(t-k)], & x \in O_m. \end{cases}$$

Во многих реальных ситуациях весьма эффективным инструментом описания динамики экономического показателя внутри выделенных классов являются сплайн-функции. Это обусловлено тем, что для процесса развития экономических объектов характерно изменение темпов экономических показателей на различных участках траектории.

Приведем пример построения модели с переменной структурой для случая относительно длинных временных реализаций. Содержание задачи сводится к прогнозированию значений параметров процесса освоения (построению динамической модели экономических показателей шахт в период освоения и использованию ее для прогноза) на базе статистической информации об изменениях показателей в период освоения.

Объектами исследования послужили 230 угольных шахт, по которым имелись данные об экономических показателях

[2] Характерные точки — моменты существенного изменения структурных свойств обрабатываемой реализации. Характерные участки представляют собой отрезки кривых, заключенные между соседними характерными точками.

не менее чем за 6 лет с начала пуска в эксплуатацию. По каждому производственному объекту исследовались два типа характеристик: параметры собственно процесса освоения; комплекс априорно известных характеристик шахт, по которым можно прогнозировать процесс освоения для вновь построенных предприятий.

Параметром процесса освоения выступал процент освоения производственной мощности в каждом из первых шести лет с момента пуска шахты в эксплуатацию. Комплекс априорно известных характеристик угледобывающих предприятий включал 32 признака, которые можно подразделить на три группы: природно-географические, технико-технологические и организационно-экономические. К первой относятся признаки, характеризующие местоположение предприятия (широта, долгота и т. д.) и горно-геологические условия его функционирования (угол падения пласта, характер падения, марка угля, глубина разработки, категория по газу и др.). Ко второй группе принадлежат признаки, отражающие уровень технологии производства: количество разрабатываемых пластов, число лав, уровень механизации навалки, протяженность горных выработок и др. В третью группу входили следующие характеристики: проектная мощность шахты, среднесписочное число рабочих по добыче и др.

На этапе I исследования для каждого выбранного объекта был проведен расчет уровня освоения производственной мощности за 6 лет с момента пуска в эксплуатацию. Выбор такого периода основан на том, что большинство шахт изучаемой совокупности осваивают свою проектную мощность в течение указанного времени.

На этапе II проводилась многомерная классификация угольных шахт по относительному уровню (в %) освоения проектной мощности за первые шесть лет эксплуатации (230 объектов в шестимерном пространстве). Результатом этапа явилось выделение четырех типов предприятий, отличающихся динамикой освоения проектной мощности.

Выделенные типы предприятий представляют собой обучающую выборку для формирования классов объектов на этапе III решения задачи на основе характеристик условий его функционирования (природно-географических, технико-технологических, организационно-экономических). Решение задачи обучения осуществлялось с помощью алгоритма и программы [Елкина, Загоруйко, 1966]. Результатом явилось получение в каждом из четырех классов решающего правила, в основе построения которого лежит вектор средних значений признаков $\{\bar{x}_1^i, \bar{x}_2^i, \ldots, \bar{x}_{32}^i\}$ $(i = \overline{1, 4})$, а также

величин, обратных дисперсиям признаков. Однако описание классов явно оказалось избыточным, и поэтому возникла необходимость сокращения исходного описания с целью уменьшения минимального набора информативных признаков. Эта процедура (минимизация описания) явилась этапом IV решения поставленной задачи.

Поиск информативного набора признаков осуществлялся по той же программе. Информативный набор включал 8 признаков (из 32 первичных). В него вошли: географическая широта, угол падения пласта, категория шахты по газу, максимальная глубина выработки, проектная мощность и др.

Сравнение классификаций по полному набору признаков и информативному поднабору показало, что при стабильной численности групп лишь небольшое число объектов изменяет свою групповую принадлежность.

На этапе V в пространстве восьми признаков осуществлялась классификация шахт с целью уточнения принадлежности объектов к классам. Последнее необходимо для определения четкого разбиения объектов в новом признаковом пространстве, поскольку оно выступает как дискретная часть динамической дискретно-непрерывной модели экономического показателя. В результате проведенной классификации получено разбиение на четыре класса, включавших 171 объект. Остальные объекты резко различались по своим характеристикам и выделились в малочисленные и единичные классы.

Дискретная часть модели, полученная на этом этапе, представлена в табл. 4.19. К первому классу с низкими темпами освоения производственной мощности относятся шахты со сравнительно большой проектной мощностью, разрабатывающие крутопадающие маломощные пласты и имеющие высокую загазованность. Самые высокие темпы освоения имеют шахты четвертого класса, характеризующиеся наибольшими промышленными запасами, мощными пластами, средней загазованностью и т. д.

На завершающем этапе VI осуществляется окончательное построение дискретно-непрерывной модели, т. е. выбор наиболее подходящей кривой освоения проектной мощности для каждого из выделенных четырех классов.

Для описания кривых было использовано два метода: аппроксимация нелинейной регрессионной моделью и линейными сплайн-функциями.

Выбор кусочно-линейных сплайнов для аппроксимации показателя уровня освоения производственной мощности

Таблица 4.19

Характеристики дискретной части модели по информативному набору признаков *

Признак	Обозначение	Классы							
		I(h_1=33)		II(h_2=23)		III(h_3=87)		IV(h_4=28)	
		$\bar{x}$	$1/\sigma^2$	$\bar{x}$	$1/\sigma^2$	$\bar{x}$	$1/\sigma^2$	$\bar{x}$	$1/\sigma^2$
Проектная мощность, тыс. т	x_1	530,3	$5 \cdot 10^{-6}$	360,6	$21 \cdot 10^{-6}$	420,2	$13 \cdot 10^{-6}$	527,2	$6 \cdot 10^{-6}$
Максимальная глубина разработки, м	x_2	248,2	$8 \cdot 10^{-5}$	158,8	$11 \cdot 10^{-5}$	243,1	$4 \cdot 10^{-5}$	241,6	$5 \cdot 10^{-5}$
Количество лав, шт.	x_3	4,64	$57 \cdot 10^{-3}$	4,09	$174 \cdot 10^{-3}$	5,03	$123 \cdot 10^{-3}$	5,04	$121 \cdot 10^{-3}$
Угол падения пласта, град	x_4	21,2	$6 \cdot 10^{-3}$	19,5	$2 \cdot 10^{-3}$	14,3	$5 \cdot 10^{-3}$	26,8	$2 \cdot 10^{-3}$
Категория по газу, %	x_5	11,9	$20 \cdot 10^{-3}$	6,65	$18 \cdot 10^{-3}$	6,92	$22 \cdot 10^{-3}$	8,41	$18 \cdot 10^{-3}$
Промышленные запасы угля, млн. т	x_6	17,4	$34 \cdot 10^{-4}$	10,1	$131 \cdot 10^{-4}$	14,5	$41 \cdot 10^{-4}$	27,4	$5 \cdot 10^{-4}$
Географическая широта, град	x_7	49,9	$69 \cdot 10^{-3}$	51,0	$96 \cdot 10^{-3}$	50,7	$61 \cdot 10^{-3}$	52,0	$53 \cdot 10^{-3}$
Средняя вынимаемая мощность пласта, м	x_8	1,31	2,9	1,54	2,1	1,42	2,7	1,79	1,4
Уровни освоения, % по годам:	y								
в первом		15,3	—	36,5	—	21,5	—	69,7	—
во втором		40,2	—	64,9	—	57,9	—	90,2	—
в третьем		47,2	—	73,9	—	78,1	—	96,8	—
в четвертом		50,6	—	76,5	—	93,0	—	105,0	—
в пятом		54,1	—	73,7	—	102,9	—	107,7	—
в шестом		58,9	—	76,0	—	107,7	—	112,7	—

* $\bar{x}$ — среднее значение; σ^2 — дисперсия.

связан с его существенной нелинейностью и наличием точек перелома зависимости.

Модель показателя освоения производственной мощности имеет вид

а) для случая аппроксимации нелинейной регрессией

$$y^*(t) = 64{,}61 - \frac{49{,}53}{t} \text{ при } (x_1, x_2, \ldots, x_8) \in O_1;$$

$$R^2 = 0{,}99;\ \sigma_{ост} = 1{,}62;$$

$$y^*(t) = 86{,}74 - \frac{48{,}54}{t} \text{ при } (x_1, x_2, \ldots, x_8) \in O_2;$$

$$R^2 = 0{,}97;\ \sigma_{ост} = 3{,}23;$$

$$y^*(t) = 118{,}54 - \frac{101{,}97}{t} \text{ при } (x_1, x_2, \ldots, x_8) \in O_3;$$

$$R^2 = 0{,}96;\ \sigma_{ост} = 7{,}41;$$

$$y^*(t) = 117{,}02 - \frac{48{,}97}{t} \text{ при } (x_1, x_2, \ldots, x_8) \in O_4;$$

$$R^2 = 0{,}97;\ \sigma_{ост} = 3{,}10;$$

б) для случая аппроксимации кусочно-линейным сплайном

$$y^*(t) = 9{,}60 + 24{,}90t - 18{,}19(t-2) - 2{,}85(t-3)$$

при $(x_1, x_2, \ldots, x_8) \in O_1$;

$$R^2 = 0{,}99;\ \sigma_{ост} = 0{,}53;$$

$$y^*(t) = 8{,}10 + 28{,}40t - 18{,}80(t-2) - 9{,}25(t-3)$$

при $(x_1, x_2, \ldots, x_8) \in O_2$;

$$R^2 = 0{,}99,\ \sigma_{ост} = 1{,}66;$$

$$y^*(t) = -15{,}62 + 37{,}12t - 19{,}08(t-2) - 11{,}20(t-4)$$

при $(x_1, x_2, \ldots, x_8) \in O_3$;

$$R^2 = 0{,}99;\ \sigma_{ост} = 1{,}85;$$

$$y^*(t) = 49{,}41 + 20{,}29t - 13{,}07(t-2) - 3{,}27\ (t-4)$$

при $(x_1, x_2, \ldots, x_8) \in O_4$;

$$R^2 = 0{,}99;\ \sigma_{ост} = 0{,}76$$

где t — номер года, O_1, O_2, O_3, O_4 — классы шахт по признакам $(x_1, x_2, \ldots, x_8)$.

Использование модели для прогнозирования уровня освоения вновь вводимых шахт в t-м году включает два этапа:

Таблица 4.20

Относительные ошибки аппроксимации для шахт второго класса

Тип модели	Ошибки аппроксимации по годам освоения, %					
	1-й	2-й	3-й	4-й	5-й	6-й
Нелинейная регрессия	—4,4	3,8	4,7	2,5	—4,3	—3,3
Кусочно-линейный сплайн	0,30	0,00	—0,80	2,20	—1,99	0,60

Таблица 4.21

Темпы освоения производственной мощности для четырех классов угольных шахт, %

Период освоения	Классы			
	I	II	III	IV
Первый и второй	24,9	28,4	37,1	20,3
Третий	6,7	9,6	—	—
Переход от третьего к четвертому	—	—	18,0	7,1
Переход от четвертого к пятому и шестому	3,8	0,4	—	—
Переход от пятого к шестому	—	—	6,8	3,8

определение принадлежности вновь вводимого объекта к классу по априорно известным характеристикам $x_1, x_2, \ldots, x_8$;

расчет уровня освоения по соответствующему уравнению.

Представляет интерес сравнение качества аппроксимации нелинейной регрессией и сплайн-функцией (табл. 4.20). Сравнение ошибок аппроксимации свидетельствует о том, что кусочно-линейный сплайн более чувствителен к структурным изменениям в моделируемом динамическом ряду.

Помимо большей точности описания представление кривой освоения кусочно-линейным сплайном позволяет просто определять изменение темпов освоения производственной мощности как изменение коэффициентов кусочно-линейной функции (табл. 4.21).

Резюмируя, следует отметить, что характерным для приведенного примера является сочетание двух типов изменчивости структуры модели: дискретного (скачкообразного)— при переходе от класса к классу и непрерывного (плавного)— при моделировании динамики освоения внутри классов.

Глава 5

ОЦЕНКА КАЧЕСТВА МОДЕЛЕЙ И ИХ КОРРЕКЦИЯ

§ 1. АНАЛИЗ КАЧЕСТВА МОДЕЛЕЙ ПО РЕЗУЛЬТАТАМ ПРОГНОЗИРОВАНИЯ

Поскольку прогнозы имеют вероятностную природу, всегда следует ожидать расхождения между прогнозом и его реализацией. К основным причинам несовпадения прогноза по модели и реализации можно отнести следующие:

неоправдавшиеся общие предположения;

недостаточность или низкое качество исходной информации;

неадекватность принятой модели прогноза;

погрешность в математико-статистической обработке данных, относящаяся к выбору функции, к методике расчета коэффициентов уравнений и т. п.

Наиболее сложным является анализ адекватности модели прогноза. Нельзя дать общих правил для определения методических дефектов модели при несогласованности между прогнозом и действительностью, так как здесь требуется конкретный экономический анализ. Во всяком случае, прежде чем пересматривать структуру принятой модели, следует устранить по возможности все другие причины несовпадений. Если после устранения этих причин все же остаются расхождения, надлежит, во-первых, проверить модель с точки зрения логики ее построения, теоретической обоснованности принятых закономерностей и структурных связей; во-вторых, требуют исследования те экономические решения, которые были заложены в прогноз и могли не оправдаться в действительности. В-третьих, нужен анализ расхождения между прогнозами технического прогресса и его фактической реализацией, в особенности в том, что касается прогноза экономической эффективности новой техники. Практика показывает, что именно здесь отклонения встречаются особенно часто.

Переходя к рассмотрению статистических способов анализа точности прогноза, следует заметить, что все они в той или иной форме основаны на сопоставлении расчетной и фактической величины показателя.

С этой точки зрения любые оценки по статистической модели можно трактовать как прогнозные, так как по существу процесс расчета по модели есть условная оценка величины показателя при данных значениях факторов производства. Разделение расчетов по модели на оценочные и прогнозные производится исходя из принадлежности объектов к анализируемой совокупности. Если расчеты проводятся по объектам, входящим в исходную совокупность (находящимся в пределах ее пространственно-временных границ), то это операция оценки аппроксимации, в противном случае имеем дело с прогнозом — пространственным, временным, пространственно-временным (в зависимости от того, какие границы исходной совокупности нарушены). С этой точки зрения использование модели для оценки прошлых (уже реализованных) значений показателей, но не относящихся к анализируемой совокупности также представляет собой прогноз, но ретроспективный. В связи с изложенным может быть использован единый методический подход к анализу точности аппроксимации и прогнозирования. В основе этого подхода лежат исследования отклонений расчетного значения показателя (y_i^*) от фактического (y_i). Наиболее распространены показатели общей точности моделей (среднеквадратическая ошибка, средняя относительная ошибка и др. Подробнее см. § 2 настоящей главы). Эти общие характеристики точности модели могут выступать как критерии сравнения различных типов моделей. Однако для обоснованного выбора наилучшей модели прогноза и ее корректировки общие усредненные характеристики точности обязательно должны быть дополнены анализом характера распределения и расположения в пространстве и во времени индивидуальных ошибок.

Индивидуальное отклонение фактической реализации от расчетной (индивидуальная ошибка аппроксимации или прогноза) состоит из двух частей, одна из которых связана со случайной компонентой прогноза, другая — со случайной компонентой реализации. Это связано с тем, что как статистический прогноз, так и реализация содержат в себе случайную компоненту. Прогноз стохастичен потому, что строится на основе моделей, параметры которых оценены с определенной ошибкой. Ошибки также могут содержаться в оценке независимых переменных и всегда возникают из-за неучета влияния отдельных факторов. Прогнозируемую величину обычно представляют в виде комбинации нестохастической и стохастической частей. При этом основной акцент, естественно, делается на оценку значения нестохастической части. Что касается стохастической компоненты, то здесь нельзя ожидать особо надежных результатов. Таким образом, в наилучшем случае (при идеальном прогнозе) представленное значение будет равно нестохастической части реализации, а ошибки будут возникать из-за наличия искажения реализации ее случайной компонентой. Это минимально достижимая величина ошибки прогноза. В действительности ошибка прогноза всегда больше.

Исследование индивидуальных отклонений расчетных значений y_i^* от фактических y_i (ошибок аппроксимации и прогноза) позволяет сделать ряд выводов о точности прогноза и направлении корректировки модели. Такое исследование удобно начинать с графических методов анализа.

Один из основных видов графиков — это диаграмма зависимости реализованных (фактических) величин показателя от предсказанных по модели. Конечно, в качестве аргумента может выступать и фактическое значение показателя, но более логичен первый подход (представление реализации как функции прогноза). Действительно, прогнозы составляются раньше, чем появляются соответствующие им реализации. При выборе в качестве аргумента расчетного значения показателя диаграмма $y_i = g(y_i^*)$ позволяет ответить на вопрос, что можно сказать о реализациях, если известны результаты прогноза.

В таком виде диаграмму целесообразно строить также при анализе точности аппроксимации, так как расчетное значение показателя можно трактовать как комплексную характеристику положения объекта в пространстве факторов производства. Биссектриса координатного угла, проходящая через начало координат,— это линия совершенных прогнозов, а любая точка вне ее представляет прогноз с ненулевой ошибкой.

По характеру расположения точек относительно биссектрисы координатного угла можно сделать ряд заключений о характере прогнозов. В частности, расположение точек над линией совершенных прогнозов соответствует переоценке уровня показателя, под линией — недооценке уровня. Далее, если точки по ту или иную сторону линии совершенных прогнозов располагаются сериями, скоплениями и эти серии не случайны, то это означает зависимость ошибки прогноза от положения точек — объектов по оси расчетных по-

казателей. Такая ситуация свидетельствует о систематической погрешности моделей в определенных областях признакового пространства.

Таким образом, диаграмма прогноз-реализация может показать следующее:

а) отклонения от полученного уравнения регрессии носят систематический характер, отрицательные отклонения соответствуют одним интервалам значения y_i^*; положительные — другим; необходим переход к дискретно-непрерывным моделям либо корректировка прогноза на величину его смещения в определенной области;

б) дисперсия не постоянна, как это предполагается при использовании метода наименьших квадратов; можно прибегнуть к взвешенному методу наименьших квадратов или произвести преобразование перед выполнением регрессионного анализа.

Если вместо абсолютных значений строить диаграммы прогноз—реализация в относительных отклонениях от базисного периода, то на основе таких диаграмм могут быть выявлены не только ошибки в оценке уровня, но и ошибки в поворотных точках, что чаще встречается и труднее поддается исправлению, чем ошибки в оценке уровня [Тейл, 1970].

Для корректировки моделей большое значение имеет не только общая диаграмма прогноз — реализация, но и исследование зависимости ошибки прогноза (аппроксимации) от других факторов. Здесь часто оказывается полезным исследование связи отклонений со временем. В случае аппроксимации строится график расположения отклонений в той последовательности, в какой расположены сами наблюдения. В случае регулярных многократных прогнозов для фиксированного объекта последовательность отклонений соответствует хронологическому порядку прогнозов. В первом случае (аппроксимация) выявление закономерностей связи величины отклонения со временем свидетельствует о недоучете этого фактора в модели и требует ее соответствующей корректировки. В случае прогноза факт систематического изменения ошибок во времени также может быть использован либо для корректировки самого прогноза, либо корректировки модели. Корректировка прогноза может, например, осуществляться путем построения функции ошибки прогноза от времени (в виде явной функции, авторегрессионной схемы либо их комбинации) и введения соответствующей поправки в прогноз.

Можно также располагать отклонения (ошибки прогноза) в последовательность в соответствии со значениями отдельных независимых переменных. Факт закономерного изменения отклонений от величины независимых переменных тоже свидетельствует об ошибках в модели (например, о недоучете нелинейности зависимости) и необходимости соответствующей корректировки.

Выше были рассмотрены отдельные подходы к анализу индивидуальных ошибок прогноза и способы использования результатов этого анализа при корректировке модели. Эти подходы можно рассматривать как отдельные элементы общего итерационного процесса совершенствования статистической модели на основе анализа ошибок аппроксимации (прогноза).

Общая схема одного цикла итерационного процесса совершенствования модели, основанная на использовании идей распознавания образов, состоит из нескольких этапов.

1-й этап. Рассчитываются ошибки аппроксимации (прогноза) по каждой реализации моделируемого процесса (в пространстве, во времени).

2-й этап. Выполняется анализ распределения ошибок по основной диаграмме.

Если в результате анализа выявляются какие-либо систематические тенденции в величине ошибок и в их расположении относительно линии совершенных прогнозов, то переходят к следующему этапу схемы.

3-й этап. Производится группировка реализаций по характеру расположения относительно линии совершенных прогнозов, величине к знаку ошибки и т. п. Полученные группы объектов (реализаций) выступают в качестве обучающей выборки.

4-й этап. Выполняется обучение распознаванию групп объектов (реализаций), сформированных по характеру ошибок прогноза, в пространстве факторов производства. При этом в пространство факторов производства включаются не только независимые переменные модели, но и все известные характеристики объекта, которые предположительно могут быть связаны с величиной и характером ошибок прогноза (аппроксимации). Если получена значимая зависимость классификации объектов в пространстве факторов производства и в пространстве ошибок, то можно перейти к последнему этапу схемы.

5-й этап. Выявляются основные факторы производства и характеристики объектов, оказывающие влияние на характер и величину ошибок прогноза, т. е. осуществляется операция минимизации описания.

Результаты последнего этапа могут служить базой совер-

шенствования структуры исходной модели, в частности для принятия решения об изменении типа модели (например, перехода от непрерывной к дискретно-непрерывной модели), введения в модель дополнительных факторов (например, фактора времени); учета нелинейности и взаимодействия факторов и т. д.

Итак, анализ ошибок прогноза (аппроксимации) по совокупности пространственно-временных реализаций позволяет провести уточнение прогноза

а) путем уточнения результатов расчета на основе модели авторегрессии ошибок, т. е. корректировки самого прогноза;

б) путем совершенствования структуры исходной модели, т. е. корректировки методики прогноза.

Второй путь более трудоемок, требует большего времени, но вместе с тем обеспечивает радикальное решение проблемы повышения точности прогнозов.

§ 2. МЕТОДЫ ОЦЕНИВАНИЯ КАЧЕСТВА МОДЕЛЕЙ

Здесь будут рассмотрены некоторые часто используемые, достаточно универсальные методы оценивания ожидаемой ошибки модели, которые можно использовать в условиях априорной неопределенности относительно истинного закона статистического взаимодействия признаков.

Любой параметрической модели прогноза Y по X — $g(X;\ \Theta)$, построенной алгоритмом A на основе анализа выборки V объема m, полученной согласно функции распределения $F(y, x)$, можно поставить в соответствие среднюю ожидаемую ошибку прогноза на реализациях генеральной совокупности

$$\delta = \int\limits_{R_{yx}} (y - g(x; \Theta))^2\, dF(y, x). \qquad (5.1)$$

При этом алгоритм A выбирает модель $g(X; \Theta)$ в некотором параметрическом классе функциональных зависимостей G, характеризующемся определенной степенью статистического разнообразия зависимостей $\rho(G)$ при данной выборке V для алгоритма A. Класс G может включать различные полиномиальные модели, кусочно-линейные или иные.

Так как закон распределения $F(y, x)$ неизвестен, то вычисление δ согласно (5.1) невозможно и выбор наилучшей $g(X; \Theta)$ в классе G осуществляется путем минимизации некоторой оценки величины (5.1) — $\hat{\delta}$.

Величина $\hat{\delta}$ является случайной, так как зависит от случайной выборки V, на основе которой строится $g(X; \Theta)$. Как всякая случайная величина, $\hat{\delta}$ имеет свой закон распределения.

Существуют различные критерии «надежности» оценки того или иного параметра (в частности, δ) — ее состоятельность, несмещенность, эффективность и др.

Очевидно, надежная оценка $\hat{\delta}$ должна зависеть от качества $g(X; \Theta)$ на V — суммарной эмпирической ошибки

$$\delta_V = \sum_{i=1}^{m} (y_i - g(x_i; \Theta))^2 \qquad (5.2)$$

и величины $\rho(G)$.

Если не учитывать $\rho(G)$, то при достаточно богатом классе G и малом объеме выборки m существует большая вероятность выбора далеко не лучшей модели. Например, увеличивая число компонент (k) в кусочно-линейной модели, можно добиться $\delta_V = 0$. Эта вероятность обусловлена большим элементом случайности расположения небольшого количества m точек (y_i, x_i) в R_{yx} относительно возможных гиперповерхностей из G. Определение же значения $\rho(G)$ весьма сложная и практически неразрешимая задача, так как $\rho(G)$ зависит от $F(y, x)$, а вид $F(y, x)$ нам неизвестен. В таком положении ничего не остается делать при аналитических исследованиях, как в качестве верхней оценки $\rho(G)$ брать число эмпирически различных на выборке объема m проверяемых моделей — $|G|$ [Вапник, 1979].

Поэтому надежные аналитические оценки в математической статистике получены лишь в простейших случаях, когда известен либо вид $F(y, x)$, либо класс проверяемых зависимостей весьма узок (например, гиперплоскость с неизвестными параметрами).

Мы же рассматриваем класс реальных задач со следующими особенностями:

относительно вида $F(y, x)$ существует априорная неопределенность;

алгоритмы синтеза КЛМ — сложные оптимизационные процедуры с весьма широким классом параметрических возможностей G (выбор информативных признаков $X' \subset X$, определение числа линейных частных моделей);

объем выборки m мал и соизмерим с размерностью n пространства R_x.

Поэтому особо остро ставится вопрос о получении надежных оценок $\widehat{\delta}$ для моделей $g(X;\Theta)$ из фиксированного класса G.

Среди аналитических исследований, учитывающих специфику вышеупомянутого класса реальных задач, стоит отметить работу В. Н. Вапника [1979]. Заметим, что вторая особенность класса задач обусловлена тем, что мы рассматриваем так называемые параметрические модели прогноза, представляющие собой функциональные преобразования $X \to Y$, зависящие от значений сравнительно небольшого количества параметров. В связи с этим вне нашего рассмотрения находятся так называемые непараметрические модели и методы оценки их качества [Тарасенко, 1976]. Однако нужно заметить, что современные методы непараметрической статистики хотя и не требуют предположений относительно вида закона распределения $F(y, x)$, но еще не соответствуют «духу» реальных, со сложной структурой, методов построения непараметрических моделей, оптимальных (по δ) на конечной выборке.

Из нашего рассмотрения выпадают также аналитические оценки, полученные без учета вероятностной природы задачи прогнозирования [Журавлев, 1978].

Работа [Вапник, 1979] базируется на результатах задачи равномерной сходимости частот $\{\widehat{p}_1, \ldots, \widehat{p}_s\}$ к вероятностям $\{p_1, \ldots, p_s\}$ несовместных событий $\left(\sum_{i=1}^{s} p_i = 1\right)$

$$p\{\sup_i |\widehat{p}_i - p_i| > \varepsilon\} \leqslant 1 - \eta, \tag{5.3}$$

где ε есть функция от (m, s, η). На основе проведенных теоретических исследований получены оценки типа

$$\delta_{b_1} = \frac{\delta_V/m}{h_1(|G|, \eta)} \text{ и } \delta_{b_2} = \delta_V/m + h_2(|G|, \eta) \tag{5.4}$$

для задач соответственно регрессионного и дискриминантного анализов. Для полученных оценок $\delta_в$ имеет место $p(\delta \leqslant \leqslant \delta_в) \geqslant 1 - \eta$.

Напомним (см. гл. 2) конкретный вид оценки δ_b для КЛМ:

$$\delta_b = \sum_{\nu=1}^{k} \frac{\delta_{V_\nu}/m}{1 - \sqrt{\dfrac{n_\nu \cdot k_2\left(\ln \dfrac{m}{n_\nu k_2} + 1\right) + \ln c_n^{n_\nu} - \ln \eta}{m}}}. \tag{5.5}$$

Оценки типа (5.4) обычно завышены. Это в основном обусловлено минимаксным («равномерным») аспектом постановки

задачи и заменой $\rho(G)$ на $|G|$. Кроме того, эффективность их использования зависит от величины η, для априорного задания значения которой у исследователя обычно немного информации. Однако нужно заметить, что от любой оценки $\widehat{\delta}$ практически требуется не столько минимальная ее смещенность относительно истинного значения, сколько способность правильно упорядочивать по качеству проверяемые модели в заданном классе G. Исследований же в этой области явно недостаточно.

Аналитическим оценкам можно противопоставить «алгоритмические», не представляющие из себя аналитического выражения для определения $\widehat{\delta}$ по $\{g(X;\Theta), V, \rho(G)\}$, но содержащие алгоритм вычисления $\widehat{\delta}$. Преимуществами алгоритмических оценок перед аналитическими являются более реальный учет $\rho(G)$ и универсальность формулировки по отношению к различным классам G (к различным алгоритмам A). Недостаток их — сложность использования в трудоемких оптимизационных алгоритмах.

Рассмотрим несколько различных оценок алгоритмического типа.

Часто рекомендуется оценка, полученная методом деления исходной выборки V на обучающую подвыборку V^o объемом m^o и экзаменационную V^e объемом m^e. При этом общий алгоритм синтеза $g(X;\Theta)$ таков. Сначала выбирается наилучшая структура $g(X;\Theta)$ из класса G путем построения различных моделей на основе V^o и оценивания их качества на V^e в виде

$$\delta_{oe} = \frac{1}{m^e} \sum_{(y_i, x_i) \in V^e} (y_i - g(x_i;\Theta))^2. \tag{5.6}$$

После того как по минимуму δ_{oe} выбрана наилучшая структура $g(X;\Theta)$, ее параметры Θ могут быть уточнены на основе V.

Оценка δ_{oe} просто вычисляется и может быть использована в сложных алгоритмах синтеза $g(X;\Theta)$. Однако надежность ее невелика. Статистические качества ее неплохи при сравнительно большой величине m^e, но при этом сама модель $g(X;\Theta)$, построенная на очень малой выборке V^o, далека от оптимальной в классе G. На различных этапах последовательной структуры алгоритма целесообразно перераспределять реализации V^o и V^e.

Другой $\widehat{\delta}$ является оценка, полученная методом «хаотизации» [Пинскнер, 1979]. Рассмотрим несколько возможных

вариантов ее вычисления. Обозначим через V_1 псевдовыборку, полученную из V путем случайной перестановки значений y_i у всех исходных реализаций. Пусть аналогично получена из V и псевдовыборка V_2. Через $\delta(V_g, V_j)$ обозначим суммарную эмпирическую ошибку модели $g(X; \Theta; V_g)$, построенной тем же самым алгоритмом A на основе выборки V_q, на реализациях выборки V_j

$$\delta(V_q, V_j) = \sum_{(y_i, x_i) \in V_j} (y_i - g(x_i; \Theta; V_q))^2. \qquad (5.7)$$

Тогда в качестве оценки «хаотизации» δ_x величины δ для модели $g(X; \Theta)$ может быть взято одно из следующих выражений:

$$\delta_{x_1} = (\delta_V + \delta(V_1, V_2) - \delta(V_1, V_1))/m;$$
$$\delta_{x_2} = (\delta_V/m)\,\delta(V_1, V_2)/\delta(V_1, V_1). \qquad (5.8)$$

Аддитивная (δ_{x_1}) и мультипликативная (δ_{x_2}) оценки могут быть улучшены путем усреднения их значений по различным вариантам псевдовыборок V_j. Моделирование на ЭВМ подтвердило положительные качества оценок (5.8) [Пипскнер, 1979].

Часто используется оценка «скользящего контроля» [Вапник, 1979]. Схема ее вычисления такова. Пусть $V_i = V \setminus (y_i, x_i)$ (из V удалена реализация (y_i, x_i)). Обозначим через $g(X; \Theta_i)$ модель, построенную алгоритмом A на основе V_i. Тогда имеем такую оценку

$$\delta_{ск} = \frac{1}{m} \sum_{i=1}^{m} (y_i - g(x_i; \Theta_i))^2. \qquad (5.9)$$

Величина $\delta_{ск}$ аналитически вычислима лишь для простейших моделей — линейной регрессионной и дискриминантной [Вапник, 1979]. Оценка (5.9) является несмещенной для $g(X; \Theta)$, построенной на любых $(m - 1)$ реализациях. Для модели $g(X; \Theta)$, построенной на основе V, она завышена, но незначительно. Вычислить же дисперсию $\delta_{ск}$ в общем случае не удается, но, как показывают экспериментальные исследования, она при малом m значительна. Кроме того, смещение $\delta_{ск}$ резко увеличивается при нарушении предпосылки независимости наблюдений из V. Оценка (5.9) относится к разряду сложно вычисляемых алгоритмических оценок.

Другой из рассматриваемых алгоритмических оценок является оценка «повторного моделирования» [Котюков, 1978]. Ее сложно вычислять, но она достаточно надежна

и сравнительно устойчива к нарушению условия независимости $(y_i, x_i) \in V$ (метод «повторного моделирования» будет рассмотрен нами более подробно ниже).

Аналитические и алгоритмические оценки в значительной степени дополняют друг друга. Поэтому целесообразно на разных стадиях работы реальных алгоритмов, имеющих сложную оптимизационную структуру, использовать и те и другие оценки. Например, в алгоритмах синтеза кусочно-линейных моделей на этапах выполнения сложных оптимизационных расчетов при фиксированном значении k целесообразно использовать оценку аналитического типа, а на окончательном этапе выбора наилучшей модели при различных k целесообразно использовать оценку другого типа.

Метод повторного моделирования

Для удобства верхним индексом o будем обозначать все оценки, полученные на основе анализа исходной выборки,

$$V^o = (y_i^o, x_i^o); \quad i = \overline{1, m}.$$

Часто построенная модель прогноза $g(X; \Theta^o)$ либо использует в явном виде некоторую оценку функции распределения $F^o(y, x)$ (или плотности $f^o(x, x)$), либо неявно предполагает наличие такой оценки. Например, регрессионная КЛМ, чьи частные модели $g(X; \alpha_\nu^o)$ были получены путем решения системы нормальных уравнений, подразумевают аппроксимацию $f(y, x)$ смесью k нормальных законов:

$$f^o(y, x) = \sum_{\nu=1}^{k} \pi_\nu^o N(\mu_\nu^o, \Sigma_\nu^o),$$

где μ_ν^o — оценка вектора математического ожидания величины (Y, X), а Σ_ν^o — оценка матрицы ковариаций признаков (Y, X) в области O_ν.

В качестве π_ν^o можно взять байесовские оценки, соответствующие априорному равномерному закону распределения соответствующих вероятностей

$$\pi_\nu^o = \frac{m_\nu + 1}{m + k}; \quad \nu = \overline{1, k}. \qquad (5.10)$$

Уточним одно обстоятельство. Допустим, модель $g(X; \alpha_\nu^o)$ включает лишь $n_\nu < n$ исходных информативных признаков. Однако вектор μ_ν^o и матрица Σ_ν^o должны соответствовать полному вектору (Y, X), так как алгоритм A, выби-

рая информативные $X^* \subset X$, анализировал полный вектор μ_v^o и полную матрицу Σ_v^o.

Таким образом, будем считать, что сопоставление $g(X; \Theta^o) \to F^o(y, x)$ возможно. Под $F^o(y, x)$ подразумевается функция распределения, параметры которой были использованы алгоритмом A при построении $g(X; \Theta^o)$.

Имея оценку $F^o(y, x)$, (или $f^o(y, x)$), мы можем в соответствии с ней генерировать (например, на ЭВМ) повторную многомерную выборку любого объема.

При этом «новые реализации» (y_i, x_i) должны удовлетворять системе параметрических ограничений детерминированной или стохастической природы

$$B(y, x) \leqslant b, \tag{5.11}$$

отражающей специальные знания в конкретной области, интерпретацию V^o и требования построенной модели $g(X; \Theta^o)$. Так, например, в случае кусочно-линейной модели при моделировании реализаций согласно $N(\mu_v^o, \Sigma_v^o)$ необходимо учитывать границы области O_v.

Сущность метода «повторного моделирования» заключается в следующем [Котюков, 1978].

Пусть $[y, x]_r$ — совокупность любых r статистически независимых реализаций, полученных согласно $F^o(y, x)$ и удовлетворяющих (5.11), а D — множество всевозможных различных таких совокупностей; $[y, x]_r \in D$. Обозначим через $g(X; \Theta)$ новую модель прогноза, построенную уже на основе анализа $[y, x]_r$ с помощью того же алгоритма A. Вектор Θ есть функция от $[y, x]_r$.

Тогда оценкой «повторного моделирования» значения δ для $g(X; \Theta^o)$ будет величина

$$\delta_{nм} = \int_D \left[\frac{1}{m} \sum_{i=1}^{m} (y_i^o - g(x_i^o; \Theta))^2 \right] dF^o([y, x]_r), \tag{5.12}$$

где $F^o(y_i, x_i) = F^o(y, x)$ для $\forall (y_i, x_i) \in [y, x]_r$; $F^o([y_i, x_i]_r) = \prod_{i=1}^{r} F^o(y_i, x_i)$ — функция распределения $[y, x]_r$. Под интегралом в (5.12) стоит средняя ошибка на реализациях исходной выборки V^o модели $g(X; \Theta)$, построенной по $[y, x]_r$.

Значение $\delta_{nм}$ можно приближенно вычислить методом Монте-Карло, т. е. с помощью многократного моделирования совокупностей $[y, x]_r$, построения по ним моделей $g(X; \Theta)$,

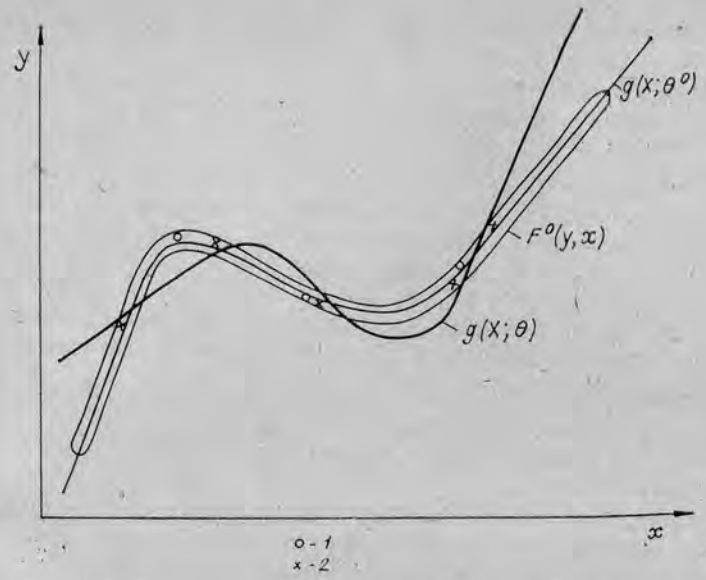


Рис. 5.1. Повторное моделирование выборки: *1* — исходные выборочные реализации; *2* — новые выборочные реализации.

оценивания их ошибки на исходной выборке V^o с последующим усреднением результата. Обычно уже десять повторных моделирований позволяют качественно определить величину $\delta_{nм}$.

Для оценки $\delta_{nм}$ при конкретных видах $F^o(y, x)$ можно обосновать сходимость: $\delta_{nм} \to \delta$ — при $m \to \infty$ и $F^o(y, x) \to F(y, x)$; $\delta_{nм} \to (\delta_V^o/m)$ при $r \to \infty$.

Здесь δ_V^o определяется для $g(X; \Theta)$ согласно (5.2). Если модели $g(X; \alpha_v^o)$ построены методом, минимизирующим δ_V^o, то справедливо соотношение $\delta_{nм} \geqslant (\delta_V^o/m)$ при любом r.

Очевидно, что при больших значениях r оценка $\delta_{nм}$ не обладает хорошими свойствами и практически совпадает с (δ_V^o/m).

Поэтому нами при использовании (5.12) предлагается принцип равномощности повторных выборок исходной $r = m$.

К сожалению, аналитическое исследование оценки $\delta_{nм}$ при конечной выборке затруднительно. В следующем параграфе будут приведены результаты моделирования поведения различных оценок, в том числе и $\delta_{nм}$. Пока же ограничимся

обсуждением «механизма действия» оценки $\delta_{nм}$ на содержательном уровне.

На рис. 5.1 показан принцип вычисления $\delta_{nм}$. Аппроксимация данных осуществлялась полиномом третьей степени. Специально рассмотрен случай почти вырожденного эллипсоида рассеивания $F^o(y, x)$. Видно, что ошибка $g(X; \Theta)$ на V^o может быть значительной, особенно в крайних по оси X точках. Очевидно, в данной ситуации (см. рис. 5.1) лучшим по оценке $\delta_{nм}$ был бы полином первой степени.

Влияние оценки $\delta_{nм}$ на выбор наилучшего значения k при построении КЛМ сказывается следующим образом. Для чересчур большого k и малого m при повторном моделировании для некоторых областей $\{O_v^o\}$ в выборке $[y, x]_m$ не будут вообще реализованы события $x \in O_v^o$. Это приведет к большой ошибке новой модели $g(X; \Theta)$ на реализациях $((y_i^o, x_i^o)\colon x_i^o \in O_v^o)$.

Метод повторного моделирования можно рассматривать и как вариант принципа «хаотизации» исходной выборки [Пинскнер, 1979]. Только здесь вид «хаотизации» не произволен, а следует из свойств $\{A, V^o\}$.

Сравнительный анализ критериев оценивания

Несмотря на наличие как аналитических, так и алгоритмических оценок качества моделей прогнозирования, некоторых теоретических и экспериментальных исследований каждой из них, данных о проведенном сравнительном анализе этих оценок с целью определения области предпочтения каждой из них явно недостаточно. Здесь сделана попытка несколько восполнить этот пробел в исследованиях. Сравнительный анализ критериев осуществлен при решении простейшей статистической задачи.

Для задачи распознавания двух классов по замеру дискретного признака X путем моделирования на ЭВМ было проведено сопоставление семи различных оценок вероятности ошибочной классификации δ. Интерес представлял анализ поведения («устойчивости») оценок при различных соотношениях объема обучающей выборки и числа оцениваемых параметров классификатора, а также при наличии коррелированных наблюдений.

Рассмотрим задачу распознавания двух классов $\{w_1, w_2\}$ на основе измерения случайной величины X, имеющей k возможных дискретных значений $\{X^1, \ldots, X^k\}$. Обозначим через

$p(j/y)$ вероятность события $x = X^j$ при условии $x \in w_y$ ($y = \overline{1,2}$).

Оптимальным классификатором для каждого возможного исхода $x = X^j$ является байесовское решающее правило

$$x \in w_1 \text{ при } \pi_1 p(j/1) \geqslant \pi_2 p(j/2);$$
$$x \in w_2 \text{ при } \pi_1 p(j/1) < \pi_2 p(j/2),$$

минимизирующее общую вероятность ошибочной классификации наблюдения $x \in R_x$, которая в этом случае равна

$$\delta_\infty = \sum_{j=1}^{k} \min\{\pi_1 p(j/1), \pi_2 p(j/2)\}. \tag{5.13}$$

Пусть априорные вероятности классов π_1, π_2 известны, а вероятности $p(j/y)$ неизвестны; обучающая выборка $V = V_1 \cup V_2$ объема m получена согласно $F(y, x) \equiv \{\pi_y, p(j/y)\}$; m_y — объем выборки класса w_y, а $m(j/y)$ — число событий $x \in X^j$ при условии $x \in w_y$.

Вычислим оценки вероятностей $\hat{p}(j/y) = m(j/y)/m_y$. Тогда оптимальным выборочным классификатором (ОВК) возможного исхода $x = X^j$ (для $j = \overline{1, k}$) является правило

$$\begin{aligned} x \in w_1 \quad &\text{при} \quad \pi_1 \hat{p}(j/1) \geqslant \pi_2 \hat{p}(j/2); \\ x \in w_2 \quad &\text{при} \quad \pi_1 \hat{p}(j/1) < \pi_2 \hat{p}(j/2), \end{aligned} \tag{5.14}$$

минимизирующее общую вероятность ошибочной классификации, которая для ОВК равна

$$\delta = \sum_{j=1}^{k} \sum_{y=1,2} \Psi(j/y)\,(\pi_y p(j/y)), \tag{5.15}$$

где

$$\Psi(j/1) = \begin{cases} 1, & \text{если } \pi_1 \hat{p}(j/1) < \pi_2 \hat{p}(j/2); \\ 0, & \text{если } \pi_1 \hat{p}(j/1) \geqslant \pi_2 \hat{p}(j/2); \end{cases}$$
$$\Psi(j/2) = 1 - \Psi(j/1).$$

Вычисление δ для ОВК невозможно, так как неизвестны $p(j/y)$. Нами рассматривались следующие различные оценки величины δ.

1. Средняя эмпирическая ошибка

$$\delta_V^* = \delta_V/m. \tag{5.16}$$

2. Оценка (5.6), полученная методом деления V на обучающую и экзаменационную подвыборки (δ_{oe}).

3. В качестве оценки $\widehat{\delta}$ можно взять математическое ожидание верхней границы δ:

$$\delta_{лм} = \sum_{j=1}^{k} \min\left\{\pi_1 \frac{m(j/1)+1}{m_1+1}, \pi_2 \frac{m(j/2)+1}{m_2+1}\right\}. \quad (5.17)$$

4. Оценка, полученная при использовании байесовских оценок для $p(j/y)$ при априорном равномерном распределении соответствующих вероятностей [Манохин, 1976]

$$\delta_м = \sum_{j=1}^{k} \min\{\pi_1 p_m(j/1), \pi_2 p_m(j/2)\}, \quad (5.18)$$

где

$$p_m(j/y) = (m(j/1) + 1)/(m_y + k). \quad (5.19)$$

5. Оценка, полученная на основе теории равномерной сходимости частот к вероятностям [Вапник, 1979],

$$\delta_b = \delta_V^* + \sqrt{\frac{\ln M - \ln \eta}{2m}}, \quad (5.20)$$

где M — число всевозможных различных решающих правил, которое для нашей задачи равно 2^k. Неравенство $\delta < \delta_b$ выполняется с вероятностью $(1 - \eta)$.

6. Оценка (5.9) «скользящего контроля» — $\delta_{ск}$.

7. Оценка (5.12) «повторного моделирования» $\delta_{пм}$, полученная при $F^o(y, x) \equiv \{\pi_y, p_m(j/y)\}$.

Оценки δ_V^*, $\delta_{лм}$, $\delta_м$, δ_b — аналитические, остальные — алгоритмические. Оценки $\delta_{лм}$, $\delta_м$ не являются универсальными — они имеют область применения, ограниченную схемой данной задачи.

Кратко опишем методику моделирования. Полагали $k = 12$ и варьировалась величина m. Имело место $\pi_1 = \pi_2 = 0{,}5$ и, кроме того, $m_1 = m_2$. Были рассмотрены три различные вероятностные схемы (табл. 5.1).

Исследования проводились для значений $m_y = 8, 12, 16, 20, 30$; $y = 1{,}2$. При каждом m_y осуществлялось 100 испытаний с последующим усреднением результата.

При расчете δ_b, согласно рекомендациям В. Н. Вапника, полагали $\eta = 0{,}5$. При расчете $\delta_{пм}$ методом Монте-Карло осуществлялось 20 повторных моделирований, что позволяло вычислить $\delta_{пм}$ с высокой точностью.

Моделирование проводилось на ЕС-ЭВМ при двух различных датчиках псевдослучайных чисел, равномерно распределенных в интервале (0,1). Соответствующие результаты моделирования при этом существенно не различались.

Таблица 5.1

Вероятностные схемы формирования выборки

Номер схемы	Вероятность ошибочной классификации, δ_∞	Условные вероятности для $j=\overline{1,6}$		Условные вероятности для $j=\overline{7,12}$	
		$p(j/1)$	$p(j/2)$	$p(j/1)$	$p(j/2)$
1	0,06	0,15666	0,01	0,01	0,15666
2	0,16	0,14	0,02666	0,02666	0,14
3	0,28	0,12	0,04666	0,04666	0,12

Дополнительно исследовалось поведение оценок при $k = 6$ (соседние по j пары значений X табл. 5.1 соединялись). Характер результатов моделирования с учетом отношения m и k при этом существенно не изменился.

Изучалось также поведение всех оценок при коррелированных наблюдениях. Был реализован упрощенный, «неклассический» принцип генерации зависимых наблюдений. Задавалась вероятность того, что каждое $(l + 1)$-е наблюдение равно предыдущему:

$$p_k = p((y_{l+1}, x_{l+1}) = (y_l, x_l)); \quad 1 \leqslant l \leqslant m_y.$$

При этом с вероятностью $(1 - p_k)$ каждое следующее наблюдение (y_{l+1}, x_{l+1}) определялось независимо и согласно $F(y, x)$. Это эквивалентно уменьшению объема обучающей выборки. Рассматривались три варианта: $p_k = 0$ (независимые наблюдения), $p_k = 0{,}3$; $p_k = 0{,}5$.

Приведем результаты моделирования. При каждом испытании для каждой из семи оценок $\widehat{\delta}_i$ вычислялась величина $\Delta_i = \widehat{\delta}_i - \delta$ $(i = \overline{1,7})$. На основе 100 испытаний в каждом варианте задачи для величин Δ_i вычислены оценки математических ожиданий $\widehat{M}(\Delta_i) = (\mu_V^*, \mu_{oe}, \mu_{лм}, \mu_в, \mu_{ск}, \mu_{пм})$ и их дисперсий $\widehat{D}(\Delta_i) = \{D_V^*, D_{oe}, D_{лм}, D_м, D_в, D_{ск}, D_{пм}\}$. Величина $\widehat{M}(\Delta_i)$ характеризует смещение оценки $\widehat{\delta}_i$ относительно истинного значения δ.

В табл. 5.2 и 5.3 приведены значения $\widehat{M}(\Delta_i)$ $(i = \overline{1,7})$ при различных значениях $\{m_y, p_k\}$ для первых двух схем моделирования ($p_\infty = 0{,}06$ и $p_\infty = 0{,}16$). Далее в табл. 5.4 и 5.5 приводятся соответствующие значения величин $\widehat{D}(\Delta_i)$.

Так как $\widehat{D}(\Delta_i)$ при возрастании p_k и δ_∞ увеличиваются незначительно, то соответствующие таблицы для них при $p_k = 0{,}5$ и $\delta_\infty = 0{,}28$ не приводятся.

Таблица 5.2

Смещение оценок при $\delta_\infty = 0{,}06$

p_k	m_y	μ_ν^*	μ_{oe}	$\mu_{лм}$	$\mu_м$	$\mu_в$	$\mu_{ск}$	$\mu_{пм}$
0,0	8	−0,145	0,109	0,515	0,156	0,386	0,019	0,064
	12	−0,093	0,102	0,363	0,146	0,340	0,011	0,052
	16	0,052	0,093	0,298	0,138	0,324	0,008	0,041
	20	0,043	0,042	0,241	0,127	0,293	0,004	0,037
	30	−0,009	0,039	0,182	0,116	0,183	0,001	0,018
0,3	8	−0,235	−0,002	0,429	0,109	0,296	−0,121	−0,017
	12	−0,154	0,063	0,305	0,096	0,279	−0,082	0,008
	16	−0,107	0,108	0,245	0,092	0,269	−0,047	0,012
	20	−0,083	0,091	0,211	0,089	0,253	−0,037	0,010
	30	−0,041	0,043	0,151	0,085	0,233	−0,022	0,006
0,5	8	−0,276	−0,010	0,376	0,084	0,255	−0,162	−0,046
	12	−0,217	−0,001	0,284	0,079	0,216	−0,119	−0,031
	16	−0,179	0,021	0,203	0,075	0,197	−0,098	−0,019
	20	−0,149	0,107	0,171	0,071	0,187	−0,056	−0,007
	30	−0,106	0,064	0,112	0,066	0,168	−0,034	0,002

Так как характер и соотношение изменения величин $\widehat{M}(\Delta_i)$ при увеличении δ_∞ сохраняются (исключение составляет $\mu_м$), то для $\delta_\infty = 0{,}28$ приводится лишь сокращенный вариант таблицы (табл. 5.6).

Кроме оценок $\widehat{M}(\Delta_i)$ и $\widehat{D}(\Delta_i)$ интерес представляют величины $\widehat{M}(\Delta_i^2)$. Их значения могут быть вычислены на основе приведенных по формуле $\widehat{M}(\Delta_i^2) = \widehat{D}(\Delta_i) + (\widehat{M}(\Delta_i))^2$ и поэтому здесь они не приводятся.

Анализ результатов моделирования позволяет сделать следующие основные выводы.

1. Величины $\widehat{D}(\Delta_i)$ позволяют утверждать, что оценки δ_{oe} (разделение V на обучение и «экзамен») и $\delta_{ск}$ (скользящий контроль) имеют большую дисперсию уклонения их значения от истинной величины δ. Дисперсии остальных оценок значительно меньше и примерно одинаковы.

2. При некоррелированных наблюдениях ($p_k = 0{,}0$) наилучшими свойствами обладают $\delta_{ск}$ (скользящий контроль) и $\delta_{пм}$ (повторное моделирование). У $\delta_{ск}$ по сравнению с $\delta_{пм}$ меньше смещение $\widehat{M}(\Delta_i)$, но больше дисперсия $\widehat{D}(\Delta_i)$.

3. При коррелированных наблюдениях лучшим качеством обладает оценка $\delta_{пм}$.

Таблица 5.3

Смещение оценок при $\delta_\infty = 0{,}16$

p_k	m_y	μ_ν^*	μ_{oe}	$\mu_{лм}$	$\mu_м$	$\mu_в$	$\mu_{ск}$	$\mu_{пм}$
0,0	8	−0,197	−0,049	0,456	0,060	0,332	0,020	0,029
	12	−0,120	0,082	0,331	0,054	0,313	0,014	0,021
	16	−0,113	0,067	0,232	0,050	0,263	0,009	0,014
	20	−0,093	0,046	0,187	0,045	0,243	0,005	0,008
	30	−0,040	0,022	0,149	0,041	0,234	0,002	0,004
0,3	8	−0,279	−0,004	0,378	−0,051	0,252	−0,129	−0,052
	12	−0,198	0,011	0,236	−0,039	0,235	−0,101	−0,041
	16	−0,148	0,083	0,193	−0,031	0,228	−0,075	−0,031
	20	−0,130	0,068	0,131	−0,024	0,206	−0,059	−0,020
	30	−0,079	0,035	0,110	−0,015	0,195	−0,041	−0,009
0,5	8	−0,312	−0,028	0,339	−0,081	0,219	−0,198	−0,078
	12	−0,261	−0,023	0,175	−0,063	0,172	−0,188	−0,061
	16	−0,223	−0,007	0,115	−0,056	0,153	−0,158	−0,050
	20	−0,184	−0,001	0,092	−0,038	0,152	−0,126	−0,036
	30	−0,129	0,021	0,061	−0,028	0,145	−0,092	−0,017

4. Сравнительно небольшое смещение и малую дисперсию имеет аналитическая оценка $\delta_м$, но она обладает медленной сходимостью при увеличении m_y и имеет несколько различный характер поведения при больших и малых величинах δ_∞.

Таблица 5.4

Дисперсия оценок при $\delta_\infty = 0{,}06$

p_k	m_y	D_ν^*	D_{oe}	$D_{лм}$	$D_м$	$D_в$	$D_{ск}$	$D_{пм}$
0,0	8	0,0062	0,0262	0,0058	0,0047	0,0062	0,0198	0,0073
	12	0,0046	0,0244	0,0044	0,0038	0,0046	0,0144	0,0057
	16	0,0036	0,0232	0,0035	0,0029	0,0036	0,0087	0,0045
	20	0,0025	0,0106	0,0024	0,0020	0,0025	0,0049	0,0029
	30	0,0017	0,0061	0,0017	0,0014	0,0017	0,0030	0,0008
0,3	8	0,0069	0,0326	0,0074	0,0062	0,0069	0,0215	0,0090
	12	0,0058	0,0304	0,0054	0,0058	0,0058	0,0174	0,0072
	16	0,0043	0,0292	0,0049	0,0043	0,0043	0,0137	0,0055
	20	0,0036	0,0237	0,0035	0,0039	0,0036	0,0090	0,0034
	30	0,0024	0,0076	0,0022	0,0018	0,0024	0,0043	0,0012

Таблица 5.5

Дисперсия оценок при $\delta_\infty = 0{,}16$

p_k	m_y	D^*_ν	D_{oe}	$D_{лм}$	$D_м$	$D_в$	$D_{ск}$	$D_{пм}$
0,0	8	0,0087	0,0264	0,0062	0,0052	0,0087	0,0243	0,0078
	12	0,0067	0,0245	0,0049	0,0044	0,0067	0,0178	0,0062
	16	0,0050	0,0235	0,0041	0,0037	0,0050	0,0130	0,0050
	20	0,0032	0,0163	0,0030	0,0029	0,0032	0,0095	0,0039
	30	0,0024	0,0103	0,0022	0,0023	0,0024	0,0044	0,0015
0,3	8	0,0094	0,0336	0,0084	0,0066	0,0094	0,0271	0,0111
	12	0,0069	0,0270	0,0076	0,0057	0,0069	0,0200	0,0092
	16	0,0062	0,0251	0,0061	0,0048	0,0062	0,0154	0,0069
	20	0,0050	0,0198	0,0048	0,0042	0,0050	0,0121	0,0048
	30	0,0039	0,0115	0,0029	0,0039	0,0039	0,0063	0,0031

Таблица 5.6

Смещение оценок при $\delta_\infty = 0{,}28$

p_k	m_y	μ^*_ν	μ_{oe}	$\mu_{лм}$	$\mu_м$	$\mu_в$	$\mu_{ск}$	$\mu_{пм}$
0,3	8	−0,334	−0,067	0,320	−0,138	0,197	−0,161	−0,079
	12	−0,269	−0,020	0,172	−0,129	0,164	−0,125	−0,062
	16	−0,236	0,038	0,088	−0,121	0,140	−0,106	−0,051
	20	−0,203	0,027	0,073	−0,116	0,133	−0,080	−0,040
	30	−0,163	0,009	0,024	−0,110	0,111	−0,063	−0,025

5. Характер соотношений величин $\widehat{M}(\Delta_i)$ и $\widehat{D}(\Delta_i)$ мало зависит от значения δ_∞.

Заметим, что изучение аналитических оценок в данном эксперименте было более простым по сравнению с алгоритмическими, так как рассмотренная вероятностная схема распознавания является самой простой с точки зрения возможностей ее аналитического исследования и получения соответствующих оценок.

Характер значений и закономерность их изменения для исследованных оценок соответствуют интуитивным представлениям о их возможностях. Эксперимент подтвердил хорошие робастные свойства метода повторного моделирования.

Некоторые дополнительные направления использования метода повторного моделирования

Метод повторного моделирования может быть использован не только для оценивания качества построенной на основе анализа выборки модели прогноза $g(X; \Theta)$. Класс решаемых с его помощью задач значительно шире. Проиллюстрируем это.

Рассмотрим задачу определения наилучшей оценки плотности распределения $f(x; \Theta)$ в некотором параметрическом классе плотностей G на основе анализа исходной выборки $V^o = \{x_i^o; i = \overline{1, m}\}$. В общем случае Θ — вектор. Информация об истинной плотности распределения отсутствует.

Допустим, что алгоритм A на основе анализа V^o построил оценку $f(x; \Theta)$. Заметим, что функция $f(x; \Theta^o)$ может относиться и к классу «непараметрических» [Тарасенко, 1976]. Требуется определить качество оценки $f(x; \Theta^o)$.

Когда тип $f(x; \Theta^o)$ априорно фиксирован и требуется определить лишь качество оценки Θ^o, часто используется критерий максимума правдоподобия Фишера

$$l = \prod_{i=1}^{m} f(x_i^o; \Theta^o). \tag{5.21}$$

Задача тогда заключается в выборе значения Θ^o, максимизирующего l.

Если же мы имеем дело с задачей не только определения наилучшего значения Θ^o, но и выбора наилучшего типа $f(x; \Theta)$ в некотором параметрическом классе плотностей G, то оценка (5.21) становится непригодной. Очевидно, что если класс G широк, то наилучшей оценкой в смысле (5.21) для $f(x; \Theta)$ является смесь из m δ-функций.

Такая постановка характерна для многих прикладных задач и для их решения наряду с подходом В. Н. Вапника [1979] может быть использован метод повторного моделирования. Рассмотрим возможность его применения [Котюков, 1981].

Через $[x]_r$ также обозначим совокупность любых r независимых реализаций, полученных согласно $f(x; \Theta^o)$, а через $f(x; \Theta)$ — оценку плотности, полученную тем же алгоритмом A на основе анализа $[x]_r$. Для оценки $f(x; \Theta)$ значе-

ние функции правдоподобия на исходной выборке V^o равно

$$l^o = \prod_{i=1}^{m} f(x_i^o; \Theta), \qquad (5.22)$$

Усредним величину (5.22) по множеству D всевозможных различных повторных выборок $[x]_r \in Q$. Тогда критерий «повторного максимального правдоподобия» имеет вид

$$l_{пм} = \int_D l^o f([x]_r)\, d([x]_r), \qquad (5.23)$$

где $f(x_j; \Theta^o) - f(x; \Theta^o)$, $\forall x_j \in [x]_r$; $f([x]_r) = \prod_{j=1}^{r} f(x_j; \Theta^o)$ — плотность распределения $[x]_r$.

Величина $l_{пм}$, так же как и (5.12), может быть приближенно вычислена методом Монте-Карло, т. е. путем многократного моделирования повторных выборок $[x]_r$ согласно $f(x; \Theta^o)$ определения по ним с помощью A оценок $f(x; \Theta)$, величин l^o с последующим усреднением результата.

Будем соблюдать принцип равномощности повторных выборок исходной — $r = m$.

При практическом использовании $l_{пм}$ вместо (5.22) целесообразно определять численно более устойчивую функцию $l^o = \prod_{i=1}^{m} (f(x_i^o; \Theta) + \varepsilon)$, где ε — некоторая малая величина. Значение (5.23) характеризует качество оценки $f(x; \Theta^o)$. Лучшей оценке $f(x; \Theta^o) \in G$ соответствует $l_{пм} = \max$.

Критерий $l_{пм}$ может быть использован при решении широкого класса задач, в том числе в процедурах кластерного анализа (разделения смесей).

Особо выделим следующую задачу. Первая интерпретация системы границ $\Gamma = \{\Gamma^1(X)_1, \ldots, \Gamma^s(X)\}$ метода ФИЭР позволяет использовать его при решении задачи аппроксимации истинной функции $f(x)$ многомерной гистограммой, границы блоков которой перпендикулярны исходным координатным осям x_j. При оптимизации структуры многомерных блоков с помощью метода ФИЭР в качестве минимизируемого критерия $\hat{\delta}$ при фиксированном значении k может быть взята величина $(-l)$, а при различных k — оценка $(-l_{пм})$.

Обсудим на содержательном уровне «механизм действия» критерия $l_{пм}$ при построении оценки $f(x; \Theta^o)$. На рис. 5.2

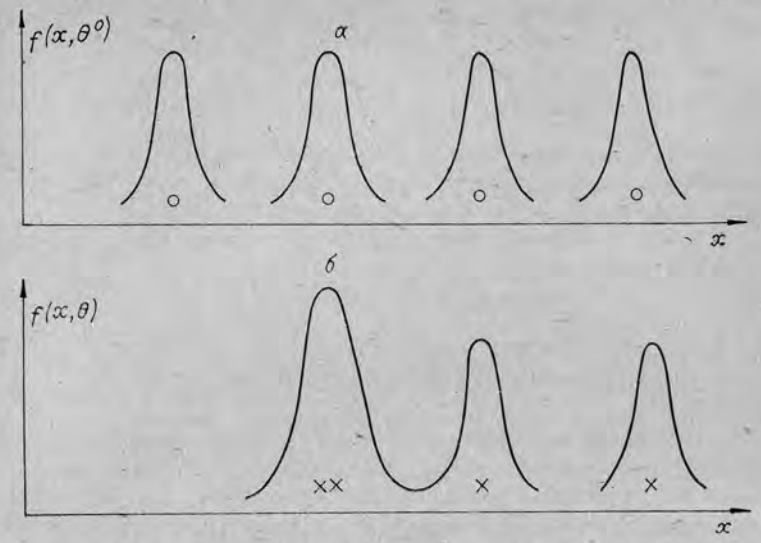


Рис. 5.2. Анализ выборки: *а* — исходной, *б* — повторной.

рассмотрена задача подбора «размаха ядер» в непараметрической оценке типа Розенблатта — Парзена [Тарасенко, 1976].

Алгоритм A определил для $f(x; \Theta^o)$ чересчур малый «размах ядер» при анализе $V^o = [x^o]_4$. Поэтому плотность $f(x; \Theta)$, построенная по повторной выборке, даст на V^o значение $l^o \approx \approx 0$.

Если множество R_x конечно ($R_x = \{S^{-1}, \ldots, S^k\}$), то $F(x; \Theta^o) \equiv \{\hat{p}_0(S^j);\ j = \overline{1, k}\}$, где $\{\hat{p}^o(S^j)\}$ — оценки вероятностей событий S^j $\left(\sum \hat{p}^o(S^j) = 1\right)$, полученные на основе анализа $V^o = \{x_1^o = S_1^o, \ldots, x_m^o = S_m^o;\ \forall S_i^o \in R_x\}$. В этом случае функция правдоподобия

$$l = \prod_{i=1}^{m} \hat{p}^o(S_i^o). \qquad (5.24)$$

Соответствующим образом трансформируется и критерий $l_{пм}$ оценки качества $F(x; \Theta^o)$. Имеем

$$l_{пм} = \int_D l^o dF([x]_r), \qquad (5.25)$$

где $F([x]_r) = \prod_{v=1}^{m} F(x; \Theta^o); \quad [x]_r \in D;$

$\widehat{p}(S^j)$ определяется на основе $[x]_r$; $l^o = \prod_{i=1}^{m} \widehat{p}(S_i^o)$.

Оценка типа (5.25) может быть использована при анализе случайных последовательностей с помощью цепей Маркова, при построении кусочно-марковских моделей.

Метод повторного моделирования может быть обобщен и на другие задачи.

§ 3. АДАПТАЦИЯ МОДЕЛЕЙ, ПОСТРОЕННЫХ ПО ВРЕМЕННЫМ РЯДАМ

Исходя из вероятностной природы модели всегда следует ожидать расхождения между результатами расчетов по моделям и фактическими реализациями процессов. Статистическое моделирование можно представить в виде разворачивающегося во времени процесса конструирования модели, ее апробации, корректировки и т. д. Анализ работы модели в реальных условиях дает возможность получить значительно больше информации о свойствах изучаемого явления, чем на первоначальном этапе конструирования модели. Итерационный процесс совершенствования модели с анализом ошибок на каждом этапе, вскрытием порождающих их причин и внесением в модель коррективов продолжается до тех пор, пока либо свойство ошибок не приблизится к характеристикам «белого шума», либо точность моделирования не достигнет заданной. Этот процесс можно назвать процессом адаптации (приспособлением) модели к условиям ее функционирования. Под адаптацией понимается «процесс изменения параметров и структуры системы, а возможно, и управляющих воздействий на основе текущей информации с целью достижения определенного, более оптимального, состояния при начальной неопределенности и изменяющихся условиях работы» [Цыпкин, 1968, с. 58]. Алгоритмы адаптации нашли широкое применение в задачах идентификации, оценивания и распознавания образов благодаря их достоинствам: простоте вычислений, универсальности и разнообразию алгоритмов, малым требованиям к объему памяти ЭВМ.

Характерная особенность адаптации состоит в последовательном накоплении и использовании текущей информации с целью устранения неопределенности, вызванной недостаточностью первоначальной информации. Адаптация требует

в определенном смысле совмещения этапов корректировки построенной модели и управления на ее основе.

В основу общего подхода к решению задачи адаптации положены итеративные методы, в частности методы стохастической аппроксимации [Невельсон, Хасьминский, 1972]. Ниже мы рассмотрим несколько часто используемых подходов к адаптации. Первый из них основан на сочетании методов регрессионного и дисперсионного анализа и включает следующие этапы:

1. Выбор интервала времени, в течение которого можно считать выявленную зависимость стабильной. Другими словами, определение промежутка времени, в течение которого параметры модели, найденные по прошлой информации, не изменились и могут быть использованы для управления.
2. Определение необходимого количества наблюдений на этом интервале времени.
3. Обновление полученной зависимости по данным текущей информации.
4. Определение точки разрыва (изменения) характера функции.

Этот подход по существу представляет собой модификацию метода построения модели с переменной структурой на основе дисперсионного критерия, описанного в § 2 гл. 3. Отличие заключается в том, что в данном подходе процедура многократно повторяется по мере поступления новых наблюдений. В связи с многократными пересчетами возникает проблема упрощения вычислений на основе рекуррентного оценивания [Невельсон, Хасьминский, 1972].

Другой подход к адаптации модели основан на идеях стохастической аппроксимации. Этот подход оказывается особенно полезным в том случае, когда нет априорной информации о статистических характеристиках изучаемого управляемого стохастического процесса.

В стохастических системах нельзя осуществлять оптимальное управление без достаточного хорошего знания вероятностных характеристик процесса. Уточнение этих характеристик процесса и его более полное изучение возможно только при функционировании системы, в процессе управления. Следовательно, в стохастических системах управление носит двойственный, противоречивый характер. С одной стороны, оно служит средством изучения системы, с другой — средством осуществления целей оптимального управления.

Решение задачи облегчается, если удается разделить процедуры обучения и управления, последовательно чередуя их. Для линейных стохастических систем с квадратичной целе-

вой функцией это удается осуществить. Именно с такой ситуацией мы сталкиваемся при оценивании параметров линейных регрессионных моделей методом наименьших квадратов.

Применение идей стохастической аппроксимации для построения адаптивных ЭСМ в этом случае сводится по существу к рекуррентному методу наименьших квадратов. Суть последнего заключается в том, что новая информация (под новой понимается либо вновь появившаяся информация, либо искусственно созданная в результате приложения модели к ретроспективным данным) используется для исправления имеющихся оценок параметров модели. Этот пересчет связан с систематическим преобразованием априорных ковариационных матриц в апостериорные. Последнее может осуществляться с использованием рекуррентных соотношений.

Идентификация линейных моделей, основанная на идеях стохастической аппроксимации, является рекуррентным градиентно-поисковым методом. Применение таких рекуррентных методов особенно удобно для оценивания коэффициентов динамических моделей, когда «наблюдения» появляются последовательно в процессе функционирования объектов.

Использование адаптационных алгоритмов превращает классическую регрессионную модель в обучающуюся.

Необходимо отметить, что в отличие от метода наименьших квадратов, который довольно чувствителен к соблюдению исходных посылок, адаптивные алгоритмы, основанные на идеях стохастической аппроксимации, работоспособны при менее жестких априорных ограничениях на исходную информацию. В частности, предположение о независимости и случайности входных величин не принципиально. Для нелинейного случая возможно получать оценки без предположения о монотонности зависимости. Однако несмотря на достаточно широкий класс адаптивных алгоритмов, их практическая реализация во многих случаях обнаружила, что некоторые из этих алгоритмов сходятся медленно, а попытка оптимизации путем выбора их параметров часто приводит к нестабильности. С целью преодоления этого недостатка в работе [Поляк, Цыпкин, 1979] предложен подход, основанный на учете имеющейся априорной информации, позволяющий осуществить выбор из разнообразия возможных алгоритмов такого, который в определенном смысле оптимален по отношению к этой априорной информации.

Задача построения ЭСМ заключается в отыскании неизвестных параметров α в соотношении $y = x^T\alpha + \varepsilon$ по совокупности наблюдений, где x — входные переменные; y — выходные переменные; ε — ошибки измерения.

Неизвестные параметры α можно отыскивать с помощью как оптимизационых (метод максимального правдоподобия, метод наименьших квадратов), так и рекуррентных (методы стохастической аппроксимации, рекуррентный метод наименьших квадратов) методов.

Общий подход к построению адаптивных алгоритмов заключается в следующем (см. [Поляк, Цыпкин, 1979]). Выбирается некоторая функция потерь $F(v)$ и строится функционал $J(c) = M\{F(c^T x - y)\}$, точка минимума которого при определенных условиях совпадает с α. Затем для определения α применяется стохастический градиентный метод вида

$$\alpha_n = \alpha_{n-1} - \Gamma_n x_n \Psi\left(x_n^T \alpha_{n-1} - y_n\right),$$

где $\Psi(v) = \Gamma'(v)$, а Γ_n — некоторая матрица. Различные адаптивные алгоритмы различаются способом выбора Γ_n и F, а следовательно Ψ. В работе Б. Т. Поляка и Я. З. Цыпкина [1979] Γ_n и Ψ предлагается выбирать исходя из имеющейся априорной информации о процессе.

Например, если ошибки измерения ε распределены нормально, то при использовании в качестве Γ_n выборочной матрицы ковариации и Ψ, пропорциональных дисперсии измерений, оптимальный адаптивный алгоритм совпадает с рекуррентным методом наименьших квадратов.

Если ошибки измерения не подчиняются нормальному закону, то метод наименьших квадратов в большинстве случаев вообще не дает состоятельных оценок, тогда как рекуррентный алгоритм работоспособен.

Рассмотрим (см. [Лукашин, 1979]) еще один, как нам представляется, интересный подход к адаптации модели, основанный на применении *байесовского принципа* к процессам, допускающим различные состояния, соответствующие наличию неустановившихся ошибок и скачкообразных изменений. При таком подходе события, имеющие случайный характер, получают ясное отражение в модели, а последовательно поступающие данные используются для подсчета апостериорных вероятностей и анализа ситуаций.

Рассмотрим базисную модель. Введем обозначения: d_t — текущее значение временного ряда: μ_t — текущее значение тренда; β_t — текущее значение коэффициента линейного роста; s_t — текущее значение сезонного фактора. В качестве модели используется процесс, определяемый уравнениями

$$d_t = \mu_t s_t + \varepsilon_t; \quad \varepsilon \sim N(O, V_\varepsilon); \tag{5.26}$$

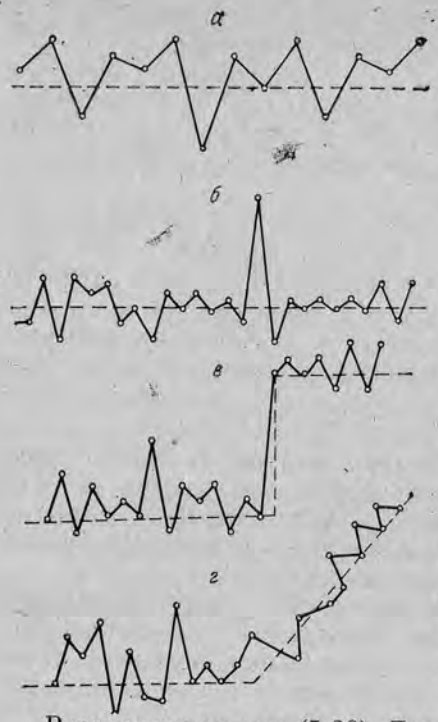


Рис. 5.3. Основные состояния процесса.

а — отсутствие изменений; *б* — случайный всплеск; *в* — ступенчатое изменение; *г* — изменение в коэффициенте линейного роста.

$$\mu_t = \mu_{t-1} + \beta_t + \gamma_t;$$
$$\gamma \sim N(O,\ V_\gamma); \qquad (5.26)$$
$$\beta_t = \beta_{t-1} + \delta_t;$$
$$\delta \sim N(O,\ V_\delta),$$

где ε_t — наблюдаемый шум; γ_t — возмущение тренда; δ_t — возмущение наклона; ε, γ, δ предполагаются независимыми нормально распределенными величинами с нулевыми средними и известными (не обязательно постоянными) дисперсиями V_ε, V_γ, V_δ.

Рассмотрим модель (5.26). Ясно, что шум ε_t влияет только на текущие наблюдения, в то время как возмущения γ_t, δ_t сказываются и в будущем. Очень большое значение γ_t может вызвать переход на новый уровень тренда, очень большое δ_t приводит к разрыву наклона, а очень большое ε_t вызывает всплеск и совершенно не сказывается на будущем.

Эти события обычно происходят в соответствии с вероятностным законом, их можно моделировать, предположив, что существует не одно, а несколько возможных распределений, порождающих текущие значения ε_t, γ_t, δ_t.

Запишем модель для этого случая. Пусть N — число состояний; π_j — вероятность наступления j-го состояния ($j = \overline{1, N}$). Если в момент t система находится в состоянии j, то случайные величины ε_t, γ_t, δ_t генерируются процессами

$$\varepsilon_t \sim N\left(0, V_\varepsilon^{(j)}\right);$$
$$\gamma_t \sim N\left(0, V_\gamma^{(j)}\right);$$
$$\delta_t \sim N\left(0, V_\delta^{(j)}\right).$$

Параметрами системы, таким образом, являются

$$\pi_j, V_\varepsilon^{(j)}, V_\gamma^{(j)}, V_\delta^{(j)}; \quad j = \overline{1, N}.$$

В дальнейшем различают четыре возможных изменения процесса: «отсутствие изменений», «случайный всплеск», «ступенчатое изменение», «изменение коэффициента линейного роста» (рис. 5.3).

Качественная характеристика дисперсий, соответствующих этим изменениям, приведена в табл. 5.7.

Естественно выражать μ_t и β_t через понятие распределения. Далее будет показано, как последовательные наблюдения модифицируют эти распределения вероятностей. Для простоты ограничимся рассмотрением двумерного нормального распределения.

Введем обозначение для первых и вторых моментов:

$$m = E(\mu);$$
$$b = E(\beta);$$
$$w_{\mu\mu} = E(\mu - m)^2;$$
$$w_{\mu\beta} = E(\mu - m)(\beta - b);$$
$$w_{\beta\beta} = E(\beta - b)^2.$$

Например, $E_\mu = \iint \mu p(\mu,\ \beta) d_\mu d_\beta$.

Обозначим вектор моментов:
$\Phi = \{m, b, v_{\mu\mu}, v_{\mu\beta}, v_{\beta\beta}\}$. Запись $(\mu\beta) \sim N(\Phi)$ означает, что пара переменных μ, β удовлетворяет закону двумерного нормального распределения с параметрами Φ; $(\mu, \beta) \sim \sim \sum_{i=1}^{M} \times \alpha^{(i)} N\left(\Phi^{(i)}\right)$ — смесь из m двумерных нормальных компонент с вероятностью $\alpha^{(i)}$ и параметрами $\Phi^{(i)}$ ($i = \overline{1, m}$).

Следует заметить, что распределение (μ_t, β_t) — байесовское апостериорное распределение вероятностей (по отноше-

Таблица 5.7

Характеристика дисперсий процесса

Тип изменения	V_ε	V_γ	V_δ
Отсутствие изменений	Нормальная	0	0
Случайный всплеск	Большая	0	0
Ступенчатое изменение	Нормальная	Большая	0
Изменения в коэффициенте линейного роста	»	0	Большая

нию к текущему наблюдению d_t). Иногда это будет записываться как $(\mu_t, \beta_t|d_t)$.

Перейдем к проблеме вычисления апостериорных вероятностей. Рассмотрим сначала простейший случай, когда генерирующий процесс имеет лишь одно состояние, т. е. дисперсии V_ε, V_γ, V_δ неизменны во времени. В этом случае, если совместное распределение μ_{t-1}, β_{t-1} в момент $t-1$ является двумерным нормальным, т. е. $(\mu_{t-1}, \beta_{t-1}|d_{t-1}) \sim N(\Phi_{t-1})$, то апостериорное распределение в момент t также является двумерным нормальным:

$$(\mu_t, \beta_t|d_t) \sim N(\Phi_t),$$

где компоненты Φ_t вычисляются следующим образом. Пусть

$$l_t = d_t - (m_{t-1} + b_{t-1})s_t \tag{5.27}$$

и

$$R = \begin{pmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{pmatrix}, \tag{5.28}$$

где

$$\begin{aligned} r_{11} &= v_{\mu\mu,t-1} + 2v_{\mu\beta,t-1} + v_{\beta\beta,t-1} + V_\gamma + V_\delta; \\ r_{12} &= v_{\mu\beta,t-1} + v_{\beta\beta,t-1} + V_\delta; \\ r_{22} &= v_{\beta\beta,t-1} + V_\delta. \end{aligned} \tag{5.29}$$

Далее пусть

$$V_l = s_t^2 r_{11} + V_\varepsilon; \tag{5.30}$$

$$A_1 = s_t r_{11}/V_l; \quad A_2 = s_t r_{12}/V_l.$$

Тогда параметры совместного апостериорного распределения в момент t будут следующими:

$$\begin{aligned} m_t &= m_{t-1} + b_{t-1} + A_1 l_t; \\ b_t &= b_{t-1} + A_2 l_t; \\ v_{\mu\mu,t} &= r_{11} - A_1^2 V_l; \\ v_{\mu\beta,t} &= r_{12} - A_1 A_2 V_l; \\ v_{\beta\beta,t} &= r_{22} = A_2^2 V_l. \end{aligned} \tag{5.31}$$

Преобразование, описываемое выражениями (5.27) — (5.31), называется B-преобразованием и обозначается

$$\Phi_t = B(\Phi_{t-1};\ V_s,\ V_\gamma,\ V_\delta).$$

Рассмотрим смешанное исходное распределение

$$(\mu_{t-1}, \beta_{t-1}) \sim \sum_{t=1}^{N} q_{t-1}^{(i)} N\left(\Phi_{t-1}^{(i)}\right),$$

где $q_{t-1}^{(i)}$ — вероятность (апостериорная по отношению к d_{t-1}) того, что процесс в момент $t-1$ находится в состоянии i. Применив B-преобразование, можно получить

$$(\mu_t, \beta_t|d_t) \sim \sum_{i,j} p_t^{(i,j)} N\left(\Phi_t^{(i,j)}\right), \tag{5.32}$$

где

$$\Phi_t^{(i,j)} = B\left(\Phi_{t-1}^{(i)}, V_s^{(i)}, V_\gamma^{(i)}, V_\delta^i\right),$$

и

$$\begin{aligned} p_t^{(i,j)} &= p\left(s_t = j, s_{t-1} = i \,|\, \vec{d}\,(t)\right) = kp\ \left(s_{t-1} = i\ |\vec{d}\,(t-1)\right) = \\ &= kL\left(d_t \,|\, s_t = j, s_{t-1} = i \,|\, d\,(t-n)\right) \pi_j q_{t-1}^{(i)} = \\ &= R\sqrt{\left(\frac{1}{2\pi V_l^{(ij)}}\right)} \exp\left\{-\frac{\left[d_t - s_t\left(m_{t-1}^{(i)} + b_{t-1}^{(i)}\right)\right]^2}{2V_l^{(ij)}}\right\} \pi_j q_{t-1}^{(i)}, \end{aligned}$$

где $\vec{d}_t$ обозначает все наблюдения включительно до момента t:

$$\begin{aligned} R &= 1/p(d_t/\vec{d}(t-1)); \\ V_l^{(i,j)} &= s_t^2 r_{11}^{(i,j)} + V_\varepsilon^{(j)}, \\ r_{11}^{(i,j)} &= v_{\mu\mu,t-1}^{(i)} + 2v_{\mu\beta,t-1}^{(i)} + v_{\beta\beta,t-1}^{(i)} + V_\gamma^{(j)} + V_\sigma^{(j)}; \end{aligned}$$

p_t^{ij} означает вероятность (апостериорную к d_t) того, что в момент $t-1$ процесс находится в состоянии i, а в момент t — в состоянии j. Полученный результат математически корректен, однако практически его применение весьма затруднено следующим обстоятельством: от исходного N-компонентного распределения мы перешли к N^2-компонентному распределению (5.32). Это усложняет вычислительную процедуру даже для небольших временных рядов.

Считается, что лучший выход из создавшейся ситуации такой: все компоненты распределения (5.32) с одним и тем же текущим состоянием j объединяются в одну группу. В результате

$$(\mu_t, \beta_t|d_t) \sim \sum_{j=1}^{N} q_t^{(i)} N\left(\Phi_t^{(j)}\right),$$

где

$$\begin{aligned} q_t^{(j)} &= \sum_i p_t^{(i,j)}; \\ m_t^{(j)} &= \sum_i p_t^{(i,j)} m_t^{(i,j)}/q_t^{(j)}; \\ b_t^{(j)} &= \sum_i p_t^{(i,j)} b_t^{(i,j)}/q_t^{(j)}. \end{aligned} \tag{5.33}$$

Выражения для вторых моментов несколько сложнее. К примеру,

$$v_{\mu\mu,t}^{(j)} = \sum_{i}^{(i,j)} p_t \left[v_{\mu\mu,t}^{(i,j)} + \left(m_t^{(i,j)} - m_t^{(j)} \right)^2 \right] / q_t^{(j)}. \qquad (5.34)$$

В результате этого сохраняется наиболее важная информация о процессе и апостериорное распределение $(\mu_t, \beta_t | d_t)$ опять состоит из N-компонент.

Сделаем краткий обзор метода по шагам цикла обновления.

Шаг 1. После получения наблюдения d_{t-1} предполагается, что апостериорное распределение $(\mu_{t-1}, \beta_{t-1} | d_t)$ представляет собой двумерное нормальное распределение $(\mu_{t-1}, \beta_{t-1} | d_{t-1}) \sim \sum_{i=1}^{N} q_{t-1}^{(i)} N\left(\Phi_{t-1}^{(i)}\right)$, где $q_{t-1}^{(i)} = p\{s_{t-1} = i | d_{(t-1)}\}$ и параметры $\Phi_{t-1}^{(i)}$ известны.

Шаг 2. В момент t до получения состояния d_t процесс находится в одном из возможных состояний, каждое из которых характеризуется параметрами

π_j — вероятность j-го состояния;

$V_\varepsilon^{(j)}, V_\gamma^{(j)}, V_\sigma^{(j)}$-дисперсии j-го состояния.

Шаг 3. Рассматривая все возможные переходы процесса из одного состояния в другое и учитывая вероятности этих переходов, получаем выражение для апостериорного распределения через смесь N^2 двумерных распределений $(\mu_t, \beta_t | d_t) = \sum_{i,j=1}^{N} p_t^{(i,j)} N\left(\Phi_t^{(i,j)}\right)$,

где $\Phi_t^{(i,j)} = B\left(\Phi_{t-1}^{(i)}, V_\varepsilon^{(j)}, V_\gamma^{(j)}, V_\delta^{(j)}\right)$

и $p_t^{(i,j)} = k \sqrt{\frac{1}{2\pi V_l^{(i,j)}}} \exp\left\{ - \frac{\left[d_t - s_t\left(m_{t-1}^{i} + b_{t-1}^{(i)} \right) \right]^2}{2 V_l^{(i,j)}} \right\} \pi_j q_{t-1}^{(i)}$.

Шаг 4. N^2-компонентное распределение, полученное на предыдущем шаге, приближенно преобразуется в эквивалентное распределение с N компонентами

$$(\mu_t, \beta_t | d_t) \simeq \sum_{j=1}^{N} q_t^{(j)} N\left(\Phi_t^{(j)}\right),$$

где $q_t^{(j)} = \sum_i p_t^{i,j}$ и параметры $\Phi_t^{(j)}$ представлены выражениями (5.32) и (5.33).

Шаг 5. Апостериорное распределение, полученное на шаге 4, сейчас соответствует форме, используемой на шаге 1, и следующая итерация проводится аналогичным образом.

ЛИТЕРАТУРА

Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений. М.: Статистика, 1974. 236 с.

Айзерман М. А., Браверман Э. М., Розоноэр А. И. Метод потенциальных функций в теории обучения машин. М.: Наука, 1970. 384 с.

Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее применение. М.: Мир, 1972. 316 с.

Александров В. В., Горский Н. Д., Поляков А. О. Пакет прикладных программ для классификации многомерных наблюдений.— В кн.: Автоматизация исследований и проектирования. М.: Наука, 1978, с. 27—34.

Андерсон Т. Введение в многомерный статистический анализ. М.: Наука, 1963. 520 с.

Бард Й. Нелинейное оценивание параметров. М.: Статистика, 1979. 350 с.

Беккер А. В. Определение резервов повышения эффективности производства предприятий отрасли.— В кн.: Моделирование и анализ экономических показателей промышленного производства. Новосибирск: Наука, 1979, с. 130—141.

Беккер А. В., Розин Б. Б., Ягольницер М. А. Методы классификации при статистическом моделировании экономических показателей.— В кн.: Многомерный статистический анализ в социально-экономических исследованиях. М.: Наука, 1974, с. 87—115.

Бобров С. Переменная корреляция.— В кн.: Математические методы в статистике. М.: Экономическая жизнь, 1927.

Бородкин Ф. М., Лукацкая М. Л. К вопросу о статистической оценке причинных связей.— В кн.: Статистические методы в экономическом анализе производства. Новосибирск: Наука, 1968, с. 3—9.

Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М.: Наука, 1965. 476 с.

Бронштейн И. Н., Семендяев К. А. Справочник по математике для инженеров и учащихся втузов. М.: Наука, 1964. 608 с.

Вапник В. Н. Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979. 447 с.

Василенко В. А. Теория сплайн-функций. Учебное пособие. Новосибирск: изд. Новосиб. гос. ун-та, 1978. 65 с.

Верховская Л. А. Однородность средних значений и корреляционных матриц многомерных выборок.— Изв. АН СССР. Сер. Техн. кибернетика, 1971, № 5, с. 171—176.

Вишнев С. М. Основы комплексного прогнозирования. М.: Наука, 1977. 287 с.

Дорофеюк А. А. Алгоритмы автоматической классификации (обзор литературы).— В кн.: Проблемы расширения возможностей автоматов. Вып. 1. М.: изд. Ин-та проблем управления, 1971, с. 5—41.

Дорофеюк А. А., Ибрагимли Ш. Д., Мовсумов В. Г. Использование критерия статистической эквивалентности моделей в задаче кусочной аппроксимации.— Автоматика и телемеханика, 1976, № 7, с. 94—99.

Дорофеюк А. А., Касавин А. Д., Торговицкий И. Ш. Применение методов автоматической классификации для построения статистической модели объектов.— Автоматика и телемеханика, 1970, № 1, с. 144—158.

Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Статистика, 1973. 392 с.

Елкина В. Н., Загоруйко А. Г. Об алфавите объектов распознавания.— В кн.: Вычислительные системы. Вып. 22. Новосибирск: изд. Ин-та математики СО АН СССР, 1966, с. 59—77.

Жидков Г. Н. Некоторые методы автоматической классификации (обзор).— В кн.: Структурные методы опознавания и автоматическое чтение. М.: изд. ВИНИТИ, 1970, с. 68—85.

Журавлев Ю. И. Об алгебраическом подходе к решению задач распознавания или классификации.— В кн.: Проблемы кибернетики. Вып. 33. М.: Наука, 1978, с. 4—50.

Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций. М.: Наука, 1980. 352 с.

Загоруйко Н. Г. Методы распознавания и их применение. М.: Советское радио, 1972. 206 с.

Загоруйко Н. Г. Методы обнаружения закономерностей. М.: Знание, 1981. 62 с.

Ивахненко А. Г., Зайченко Ю. П., Димитров В. Д. Принятие решений на основе самоорганизации. М.: Советское радио, 1976. 280 с.

Исаенко О. К., Урбах Б. Ю. Разделение смесей распределений вероятностей на их составляющие.— В кн.: Теория вероятностей. Математическая статистика. Теоретическая кибернетика. М.: изд. ВИНИТИ, 1976, с. 3—41.

Касавин А. Д. Адаптивные алгоритмы кусочной аппроксимации в задаче идентификации.— Автоматика и телемеханика, 1972, № 12, с. 101—105.

Кемени Дж., Снелл Дж. Конечные цепи Маркова. М.: Наука, 1970.

Кильдишев Г. С., Аболенцев Ю. П. Многомерные группировки. М.: Статистика, 1978. 160 с.

Котюков В. И. Синтез признаков распознавания в булевом пространстве.— В кн.: Структурные методы опознавания и автоматическое чтение. М.: изд. ВИНИТИ, 1970, с. 140—144.

Котюков В. И. Оптимизация решений некоторых задач автоматической классификации (таксономии).— В кн.: Информационные методы в системах управления, измерения и контроля. Т. 2. Владивосток: изд. ДВНЦ АН СССР, 1972, с. 144—151.

Котюков В. И. Оптимизация кусочного представления сложных моделей процессов и явлений.— В кн.: Методы представления и аппаратурный анализ случайных процессов и полей. Ч. I. Л.: изд. Всесоюз. НИИ электроизмерительных приборов, 1974.

Котюков В. И. О возможном подходе к непараметрическому оцениванию качества сложных моделей прогнозирования.— В кн.: Адаптивные системы и их приложения. Новосибирск: Наука, 1978, с. 130—135.

Котюков В. И., Буторин А. Е. Синтез иерархических многофакторных моделей и их адаптация.— В кн.: Эмпирическое предсказание и распознавание образов. Новосибирск: изд. Ин-та математики СО АН СССР, 1979, с. 100—112.

Котюков В. И. Построение кусочно-простых многофакторных моделей в некоторых задачах анализа данных. М.: изд. ВИНИТИ, 1981, № 557—82 Деп. 26 с.

Кулешов В. В., Лукацкая М. Л., Талышева Л. П. Исследование некоторых особенностей динамики функционирования лесозаготовительных предприятий.—В кн.: Распознавание образов и регрессионный анализ в экономических исследованиях. Новосибирск: изд. ИЭ и ОПП СО АН СССР, 1972, с. 171—178.

Лбов Г. С. Методы обработки разнотипных экспериментальных данных. Новосибирск: Наука, 1981. 158 с.

Лбов Г. С., Котюков В. И., Манохин А. Н. Об одном алгоритме распознавания в пространстве разнотипных признаков.— В кн.: Вычислительные системы. Вып. 55. Новосибирск: изд. Ин-та математики СО АН СССР, 1973, с. 108—110.

Ли Ц., Джадж Д., Зельнер А. Оценивание параметров марковских моделей по агрегированным временным рядам. М.: Статистика, 1977. 221 с.

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования. М.: Статистика, 1979. 254 с.

Маленво Э. Статистические методы эконометрии. М.: Статистика, 1976.

Манохин А. Н. Методы распознавания образов, основанные на логических решающих функциях.— В кн.: Вычислительные системы. Вып. 67. Новосибирск: изд. Ин-та математики СО АН СССР, 1976, с. 42—53.

Марчук Г. И. Методы вычислительной математики. М.: Наука, 1977.

Миленький А. В. Классификация сигналов в условиях неопределенности. М.: Советское радио, 1975. 328 с.

Моисеев Н. Н., Иванилов Ю. П., Столярова Е. П. Методы оптимизации. М.: Наука, 1978. 351 с.

Мюллер Д., Шусслер К. Статистические методы в социологии. Ч. II. М.: изд. Ин-та социальных проблем, 1968. 183 с.

Невельсон М. Б., Хасьминский Р. З. Стохастическая аппроксимация и рекуррентное оценивание. М.: Наука, 1972. 348 с.

Оганесов И. С., Ткаченко В. Я., Шатилов А. Г. Статистические методы в анализе и планировании строительно-монтажных работ. М.: Транспорт, 1976. 190 с.

Опойцев В. И. Идентификация многомерных статистических объектов.— Автоматика и телемеханика, 1970, № 9, с. 104—107.

Пасхавер Б. И. Проблема интервалов в группировках.— Вестник статистики, 1972, № 6, с. 24—28.

Пинскнер И. Ш. Принцип «хаотизации» и его применение при обработке наблюдений.— В кн.: Модели, алгоритмы, принятие решений. М.: Наука, 1979, с. 5—37.

Пляскина Н. И. Имитация основных этапов разработки газовых месторождений.— В кн.: Моделирование стохастических процессов развития отраслевых систем. Новосибирск: изд. ИЭ и ОПП СО АН СССР, 1980, с. 117—140.

Поляк Б. Т., Цыпкин Я. З. Адаптивные алгоритмы оценивания (сходимость, оптимальность, стабильность).— Автоматика и телемеханика, 1979, № 3, с. 71—84.

Пуарье Д. Эконометрия структурных изменений (с применением сплайн-функций). М.: Финансы и статистика, 1981. 183 с.
Розанов Г. В., Френкель А. А. Корреляционный и регрессионный анализ в экономических исследованиях.— Экономика и математические методы, 1967, т. III, вып. 3, с. 426—440.
Розин Б. Б., Сергеева Л. А. Прогнозирование сроков и кривых освоения производственной мощности угольных шахт.— В кн.: Вопросы экономико-статистического моделирования и прогнозирования в промышленности. Новосибирск: изд. ИЭ и ОПП СО АН СССР, 1970, с. 90—127.
Розин Б. Б. Учет влияния качественных признаков при моделировании экономических показателей. Там же, с. 168—201.
Розин Б. Б., Лбов Г. С., Журавель Н. М. Использование методов распознавания образов в технико-экономическом анализе производства.— В кн.: Проблемы экономико-статистического анализа и моделирования промышленного производства. Новосибирск: Наука, 1969, с. 239—253.
Розин Б. Б., Беккер А. В., Вотрина Н. В. Применение методов распознавания образов для внутриотраслевой группировки предприятий. Там же, с. 67—101.
Розин Б. Б. Теория распознавания образов в экономических исследованиях. М.: Статистика, 1973. 224 с.
Розин Б. Б., Журавель Н. М., Мучник И. Б., Берсенева Г. Т. Ступенчатая и кусочно-линейная аппроксимация экономических показателей мартеновского производства.— В кн.: Вопросы построения и применения статистических моделей экономических показателей предприятий. Новосибирск: изд. ИЭ и ОПП СО АН СССР, 1971, с. 141—169.
Розин Б. Б., Ягольницер М. А. Конструирование экономико-статистических моделей с заданными свойствами. Новосибирск: Наука, 1981. 172 с.
Себестиан Г. С. Процессы принятия решений при распознавании образов. Киев: Техника, 1965. 151 с.
Сергеева Л. А., Ягольницер М. А. Построение и анализ активных моделей добычи нефти.— В кн.: Моделирование и анализ экономических показателей промышленного производства. Новосибирск: Наука, 1979, с. 78—93.
Сиськов В. И. Об определении величины интервалов при группировках.— Вестник статистики, 1971, № 12, с. 3—11.
Сиськов В. И. Корреляционный анализ в экономических исследованиях. М.: Статистика, 1975. 168 с.
Тарасенко Ф. П. Непараметрическая статистика. Томск: Изд. Томск. ун-та, 1976. 292 с.
Тейл Г. Прикладное экономическое прогнозирование. М.: Прогресс, 1970. 510 с.
Фукунага К. Введение в статистическую теорию распознавания образов. М.: Наука, 1979. 367 с.
Хейс Д. Причинный анализ в статистических исследованиях. М.: Финансы и статистика, 1981. 255 с.
Цыпкин Я. З. Основы теории обучающихся систем. М.: Наука, 1970.
Цыпкин Я. З. Адаптация и обучение в автоматических системах. М.: Наука, 1968. 400 с.
Четвериков Н. С. Статистические и стохастические исследования. М.: Госстатиздат, 1963. 300 с.
Шлезингер М. И. Взаимосвязь обучения и самообучения в распознавании образов.— Кибернетика, 1968, № 2, с. 81—88.

Экономико-статистическое моделирование в промышленности (методологические и методические вопросы). Новосибирск: Наука, 1977. 240 с.
Ямпольский В. З., Макаров И. П. Решение задачи классификации при априорно незаданном числе классов.— В кн.: Кибернетика и вуз. Вып. 4. Томск: изд. Томск. ун-та, 1971, с. 12—24.
Янч Э. Прогнозирование научно-технического прогресса. М.: Прогресс, 1974. 586 с.
Ястремский Б. С. Избранные труды. М.: Статистика, 1964. 391 с.
Ball D., Hall T. Isodata, a new menhod of data analysis and pattern classification.— Technical Reports, Stanford Research Institute, 1965. 11 p.
Bellman R. On the approximation of curves by lines segments ising dynamic programming.— Communication of the Association for Computing Machinery, 1961, N 2, p. 34—39.
Bhattacharya C. G. A simple method of resolution of a distribution into gaussian components.— Biometrics, 1967, vol. 23, N 1, p. 115—135.
Boor C. E. de, Rice J. R. Least squares cubic spline approximation II — variable knots. Computer Science Department, Division of Mathematical Sciences, 1968, CSD TR 21, Purdue University.
Chester P. R. An adaptive nonparametric approach to the classification.— Technometrics, 1969, N 11, p. 24—31.
Chow G. C. Tests of equality between sets of coefficients in two linear regressions.— Econometrica, 1960, vol. 28, N 3, p. 591—605.
Feder P. I. On asymptotic distribution theory in segmented regression problems. Identified case.— Annals of Statistics, 1975a, vol. 3.
Feder P. I. The log likelihood ratio in segmented regression. — Annals of Statistics, 1975b, vol. 3, p. 84—97.
Ferreira P. E. A Bayesian analysis of a switching regression model: Known number of regims.— Journ. of the American Statistical Association, 1975, vol. 70, p. 370—374.
Ginsburg V., Tishler A., Zang I. Alternative estimation methods for two-regime models.— European Economic Review, 1980, vol. 13, N 2, p. 207—228.
Gujarati D. Use of dummy variables in testing for equality between sets of coefficients in two linear regressions.— Amer. Statist., 1970, vol. 24, N 1, p. 50—52.
Halpern E. F. Bayesian spline regression when the number of knots is unknown.— Jour. of Royal Statistical Society, 1973, vol. 35, ser. B, p. 347—360.
Hudson D. J. Fitting segmented curves whose joint points have to be estimated.— Journ. of the American Statistical Association, v. 61, 1966, p. 1097—1125.
Mac Gee V. E., Cartetion W. T. Piecewise regression.— Journ. of the American Statistical Association. 1970, vol. 65, N 331.
Parsen E. On the estimation of a probability function and mode.— Ann. Math. Stat., 1962, vol. 33, N 3, p. 1065—1076.
Powell M. J. D. On best L_2 spline approximations.— Numerische Mathematic, Differentialgleichungen, Approximationtheorie, 1968, N 9.
Robinson D. E. Estimates for the points of intersection of two polynomial regressions.— Journ. of the American Statistical Association, 1964, vol. 59, p. 214—224.
Sprent P. Some hypotheses concerning two-phase regresion lines.— Biometrics, 1961, vol. 17, p. 634—645.
Wilks S. S. Multidimensional statistical scatter.— In: Contribution to Probability and Statistics. Stanford Univ. press., 1960.

ОГЛАВЛЕНИЕ

Введение 3

Глава 1. Структурные изменения в экономических системах и их учет при моделировании 10

§ 1. Структура экономического процесса . . . . . —
§ 2. Вариация экономических показателей и типы статистических моделей 17
§ 3. Основные причины изменчивости абстрактной технологии. Структурная неоднородность отраслевых систем 22
§ 4. Изменчивость структур и неоднородность статистических совокупностей 34

Глава 2. Основные задачи и методы конструирования моделей с переменной структурой 40

§ 1. Формальная постановка задачи. Типы моделей с переменной структурой . . . . . —
§ 2. Методы анализа типологической структуры совокупности 51
§ 3. Методы построения кусочно-линейных моделей . . . 81

Глава 3. Методы построения моделей с переменной структурой для временных выборок 101

§ 1. Одномерные временные ряды 104
§ 2. Поиск точек перелома в одномерных временных рядах. 116
§ 3. Многомерные временные ряды 130

Глава 4. Модели с переменной структурой в пространственных и пространственно-временных выборках . . 139

§ 1. Дискретные (классификационные) модели 140
§ 2. Статические дискретно-непрерывные (кусочно-простые) модели 166
§ 3. Динамические дискретно-непрерывные модели (пространственно-временные выборки) 186

Глава 5. Оценка качества моделей и их коррекция 205

§ 1. Анализ качества моделей по результатам прогнозирования . . . . . —
§ 2. Методы оценивания качества моделей 210
§ 3. Адаптация моделей, построенных по временным рядам 228

Литература 237

Бенциан Борисович Розин
Владислав Игоревич Котюков
Мирон Аркадьевич Ягольницер

ЭКОНОМИКО-СТАТИСТИЧЕСКИЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

Ответственный редактор Виктор Григорьевич Соколов

Утверждено к печати
Институтом экономики и организации промышленного производства СО АН СССР

Редактор издательства *Л. И. Легкоступ*
Художник *В. В. Растегаев*
Художественный редактор *С. М. Кудрявцев*
Технический редактор *А. В. Сурганова*
Корректоры *С. М. Погудина, К. И. Сергеева*

ИБ № 23705

Сдано в набор 28.06.83. Подписано к печати 03.01.84. МН-0с898. Формат 84×108/32. Бумага типографская 3. Обыкновенная гарнитура. Высокая печать. Усл. печ. л. 12,6. Усл. кр.-отт. 12,8. Уч. изд. л. 14,5. Тираж 1550 экз. Заказ 685. Цена 2 р. 50 к.

Издательство «Наука», Сибирское отделение. 630099, Новосибирск, 99, Советская, 18.

4-я типография издательства «Наука». 630077, Новосибирск, 77, Станиславского, 25.