Ц. Ли
Д. Джадж
А. Зельнер

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ ЗА РУБЕЖОМ
12345
67890

ОЦЕНИВАНИЕ ПАРАМЕТРОВ МАРКОВСКИХ МОДЕЛЕЙ ПО АГРЕГИРОВАННЫМ ВРЕМЕННЫМ РЯДАМ

Ц. Ли, Д. Джадж, А. Зельнер

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ ЗА РУБЕЖОМ

ESTIMATING THE PARAMETERS OF THE MARKOV PROBABILITY MODEL FROM AGGREGATE TIME SERIES DATA

T. C. Lee
University of Connecticut
G. G. Judge
University of Illinois
A. Zellner
University of Chicago

North-Holland
Publishing Company
Amsterdam — London
1970

Ц. Ли, Д. Джадж, А. Зельнер

ОЦЕНИВАНИЕ ПАРАМЕТРОВ МАРКОВСКИХ МОДЕЛЕЙ ПО АГРЕГИРОВАННЫМ ВРЕМЕННЫМ РЯДАМ

Перевод с английского
А. Д. КАСАВИНА, В. А. ЛОТОЦКОГО, А. С. МАНДЕЛЯ
Под редакцией и с предисловием
Н. С. РАЙБМАНА

Москва
«Статистика»
1977

517.8
Л55

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ
МЕТОДЫ ЗА РУБЕЖОМ

ГОТОВИТСЯ К ПЕЧАТИ

1. Райфа Г., Шлейфер Р. Прикладная теория статистических решений.

Л $\frac{10805^{1}\text{-}033}{008(01)\text{-}77}$ 43-77

[1] Второй индекс 20204.

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Насущные потребности теории и практики систем различной природы (в том числе экономических) в решении задач анализа, управления, прогнозирования и т. п. привели к развитию многих подходов и методов для получения количественных закономерностей и соотношений, характеризующих эти системы. Модели, применявшиеся для решения подобных задач, должны были описывать количественные соотношения функционирования систем и не ограничиваться только качественными подходами. Количественные модели давали возможность численно оценивать и сравнивать результаты анализа, получать конкретные значения прогнозируемого состояния, управляющего действия, оптимального в определенном смысле плана, спроса или потребности, уровня производства и т. д. Однако при потребности развития систем еще не было возможности для количественного подхода во многих областях и в том числе экономической, так как построение и применение количественных моделей требует накопления и обработки чрезвычайно больших объемов информации. Такие возможности появились только в результате создания современных средств электронной вычислительной техники, включающей периферийные устройства для съема информации и ввода ее в электронную вычислительную машину (ЭВМ) и собственно ЭВМ. Привлечение только части этой техники обычно требует более длительного времени для получения количественной модели (для больших систем, какими являются экономические системы, это могут быть годы). Поскольку первые ЭВМ и области их применения были связаны с техническими системами, это содействовало бурному развитию теории и практики управления этими системами и в значительной мере способствовало появлению общих постановок задач и методов их решения при анализе экономических систем и возникновению нового направления — экономической кибернетики [3*, 12*, 15*, 18*, 19*, 31*, 37*, 38*][1]. Одновременно эти

[1] Здесь и далее номера ссылок, отмеченные звездочкой, относятся к дополнительному списку литературы, включенному редактором перевода.

методы ускоренными темпами развивались также и в биологии, медицине, экологии, в области организационных систем и т. д. Применение методов кибернетики в экономических системах у нас в стране в значительной мере было подготовлено работами академиков Л. В. Канторовича и В. С. Немчинова по созданию количественных подходов и методов построения экономико-математических моделей [12*, 14*, 19*].

Задача получения адекватной модели оказалась одной из основных в современной теории управления, так как выбор структуры и параметров системы управления, закона управления может быть сделан только при наличии модели объекта управления.

Не меньшую роль играют модели и в области анализа, где от выбранной модели и ее соответствия реальному объекту зависят и результаты анализа, выводы, которые будут сделаны по данному объекту, системе, явлению. Таким образом, от качества модели, от степени ее изоморфности изучаемой системе во многом зависят все дальнейшие результаты анализа, выбор значимых переменных, определяющих поведение системы в динамике, точность прогнозирования развития системы. Особенно возросла роль модели при применении методов экономической кибернетики, предусматривающих моделирование для решения задач анализа и синтеза системы управления экономической системой. Следует указать, что не только в области экономической кибернетики и не только для экономических систем, но и для систем другой физической природы, в том числе для технических, биологических, медицинских, проблема получения модели реальной системы стояла очень остро. В области технической кибернетики в 60-х годах возникло новое направление, разрабатывающее общие подходы и методы построения математической модели. Это направление получило название идентификации, и в настоящее время оно представляет самостоятельную и бурно развивающуюся отрасль теории и практики управления [20*, 24*, 29*, 33*]. Несмотря на то, что это направление появилось при решении проблемы описания технических систем, в последние годы оно нашло широкое распространение почти во всех областях науки и практики, где применяются методы моделирования для решения задач анализа, управления, прогнозирования динамических систем [24*, 33*, 39*].

Задача идентификации состоит в определении наилучшей в каком-то смысле оценки оператора по данным, полученным в условиях реального функционирования системы, т. е. по статистическим данным.

Предлагаемый перевод книги американских ученых посвящен решению задачи идентификации в узком смысле, т. е. оценке параметров при заданной структуре модели для обобщенных, преобразованных агрегированных временных рядов. В качестве общей модели выбрана вероятностная модель — марковская цепь первого порядка. В современной теории и практике динамических систем в качестве стохастической модели их широко применяются марковские процессы. Благодаря работам академика А. Н. Колмогорова и его школы в области теории марковских процессов в течение последних десятилетий были

получены основные результаты. Широкое использование именно марковских процессов в качестве моделей динамических систем объясняется тем, что они давали возможность получить адекватное, изоморфное описание широкого класса систем, а также относительной простотой, так как для этих процессов будущее при известном настоящем не зависит от прошлого. Пусть случайный процесс $X(t)$ таков, что в моменты времени $t \in T_X$ принимает значения s из некоторого множества S ($s \in S$). Множество S называют пространством состояний, конкретное значение s из множества S, которое случайный процесс $X(t)$ принимает в момент времени t, называется состоянием процесса в момент t. Достаточно полной характеристикой случайного процесса является условная вероятность $P\{X(t) \in G \mid X(\sigma), \sigma \leqslant \tau\}$ того, что состояние процесса в момент времени t принадлежит множеству $G \subset S$, если известна траектория процесса до момента времени $\tau \leqslant t$. Для марковских процессов эта вероятность зависит только от состояния процесса в текущий момент времени τ, т. е.

$$P\{X(t) \in G \mid X(\sigma), \quad \sigma \leqslant \tau\} = P\{X(t) \in G \mid X(\tau)\}. \qquad (1)$$

Если множества T_X и S конечные или счетные, то $X(t)$ называется дискретным в пространстве состояний S и во времени T_X. В этом случае случайный процесс $X(t)$ рассматривается в дискретном ряде фиксированных точек $t_1, t_2, \ldots, t_n, \ldots$, и значения $X(t)$ образуют случайную последовательность $X_i = X(t_i)$, $i = 1, 2, \ldots, n$. Случайную последовательность $\{X_i\}$ можно рассматривать как случайную функцию номера, т. е. целочисленного аргумента i. Если структура случайного процесса $X(t)$ такова, что вероятность того, что $X_{m+1} = s_{m+1}$ при условии всех предыдущих значений X_i, зависит только от X_m и не зависит от всех X_i, т. е.

$$P\{X_{m+1} = s_{m+1} \mid X_m = s_m, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0\} = \\ = P\{X_{m+1} = s_{m+1} \mid X_m = s_m\}, \qquad (2)$$

то говорят, что случайный процесс $X(t)$ обладает марковским свойством.

Дискретные случайные процессы, обладающие марковским свойством, называются цепями Маркова. Случайный процесс, для которого выполняется условие (2), называется также простой марковской цепью, или цепью Маркова первого порядка. Сложной цепью Маркова порядка l называется случайный процесс, для которого условное распределение случайной величины $X_{m+1} = s_{m+1}$ относительно всех предыдущих значений $X_i = s_i$ зависит только от l значений X_i, которые непосредственно предшествуют $X_{m+1} = s_{m+1}$, т. е.

$$P\{X_{m+1} = s_{m+1} \mid X_m = s_m, X_{m-1} = s_{m-1}, \ldots, X_0 = s_0\} = P\{X_{m+1} = \\ = s_{m+1} \mid X_m = s_m, X_{m-1} = s_{m-1}, \ldots, X_{m-l+1} = s_{m-l+1}\}. \qquad (3)$$

Для простой марковской цепи условную вероятность того, что в момент $m+1$ система находится в состоянии j, если в момент m она находилась в состоянии i, обозначим через $p_{ij}(m)$:

$$P\{X_{m+1} = j \mid X_m = i\} = p_{ij}(m).$$

Когда эта условная вероятность зависит только от состояний и не зависит от времени m, то p_{ij} называются стационарными вероятностями перехода из состояния i в состояние j, а цепь Маркова называется однородной. Последняя полностью определяется множествами T_X и S, распределением начальных состояний $p(i) = P\{X_0 = i\}$ и матрицей вероятностей перехода

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} \cdots \\ p_{10} & p_{11} & p_{12} \cdots \\ p_{20} & p_{21} & p_{22} \cdots \\ \cdot & \cdot & \cdot \quad \cdot \quad \cdot \end{bmatrix}.$$

В теории марковских цепей кроме вероятностей перехода p_{ij} из состояния i в состояние j за один шаг рассматриваются также вероятности перехода за r шагов — $p_{ij}^{(r)}$. Если обозначить через $p_{ik}^{(l)}$ вероятность перехода из состояния i в промежуточное состояние k за l шагов, а через $p_{kj}^{(r)}$ веро ятность перехода из промежуточного состояния k в состояние j за r шагов, то справедливо следующее уравнение Колмогорова — Чепмена:

$$p_{ij}^{(l+r)} = \sum_k p_{ik}^{(l)} p_{kj}^{(r)}.$$

Эти краткие справочные сведения из теории марковских процессов нам представляются достаточными для ознакомления с основной терминологией в этой области. В перечне предлагаемой литературы читатель найдет необходимую информацию по теории марковских процессов. Для освоения материала книги читатель должен располагать знаниями основ теории вероятностей и математической статистики в объеме курсов для факультетов экономической кибернетики вузов.

Небезынтересно отметить, что книга вышла в свет как 65-й выпуск серии, издающейся под рубрикой «Вклад в экономический анализ».

Книга представляет значительный интерес в связи с тем, что в ней воедино собраны почти все современные результаты по методам оценивания параметров для систем, описываемых цепями Маркова.

Среди множества работ по марковским процессам книга Ли, Джаджа, Зельнера занимает особое место. Она представляет собой, образно говоря, практический путеводитель по алгоритмам и программам обработки статистических данных при построении марковской модели. Это, естественно, обеспечивает расширение области применения этого класса моделей. Приведенные сквозные примеры хотя и иллюстрируют применимость методов к решению экономических задач, однако не представляют самостоятельного интереса. Перечень (по словам авторов, далеко не полный) результатов исследований по применению марковских моделей для решения таких экономических задач, как распределение прибылей и заработной платы, определение размеров фирм, анализ потребления, поведения потребителей и т. д., который приводится в конце введения к первой главе, свидетельствует о широком применении этого класса моделей в экономических исследованиях. Такому ши-

рокому применению, бесспорно, во многом содействовало развитие методов оценивания параметров марковских моделей, рассмотренных в этой книге, так как на базе современных методов оптимизации удалось разработать практически легко реализуемые алгоритмы. Основные результаты приведены для безразмерных показателей—относительных частот, или пропорций, по терминологии авторов, поэтому они легко воспринимаются и исследователи во многих областях могут легко интерпретировать их при решении различных задач. Значительный интерес представляют поэтому приведенные в книге программы для ЭВМ, применение которых даст возможность ускорить процесс использования разработанных методов для исследований в различных областях науки и техники и в том числе при решении экономических задач.

Несколько слов об авторах книги. Профессор Цунг-Чао Ли — специалист по эконометрии и экономике сельского хозяйства, работает на факультете сельскохозяйственной экономики университета Коннектикута. Опубликованные результаты в основном связаны с исследованиями по оценкам вероятностей перехода марковских процессов, теорией принятия решений и прогнозирования, математическим программированием и его применением в экономике сельского хозяйства.

Специальность профессора Джорджа Г. Джаджа — экономика и статистика, он работает на факультете экономики университета штата Иллинойс. Результаты его исследований в основном относятся к методам получения количественных закономерностей в экономических процессах, построению моделей ценообразования и размещения.

Профессор Арнольд Зельнер работает на экономическом факультете в университете Чикаго, специальность — эконометрия, статистика. Основные результаты его исследований относятся к методам регрессии, регионального анализа, оценки параметров.

Авторы книги — члены ряда американских эконометрических обществ, редакторы журналов, преподаватели отделений усовершенствования экономистов при соответствующих университетах. Многие из полученных авторами результатов приведены в данной книге.

При переводе книги на русский язык был исправлен ряд опечаток и сохранены обозначения, принятые в оригинале.

Результаты предлагаемой книги представляют интерес для широкого круга специалистов, занимающихся вопросами управления и идентификации в различных областях науки и техники, и особенно для специалистов в области экономики, которым в первую очередь адресована эта книга. Научные работники как экономических, так и технических исследовательских институтов и лабораторий, математики — алгоритмисты и программисты, преподаватели высших учебных заведений найдут в книге новые результаты и приложения, которыми они смогут воспользоваться в своей работе. Книга будет также полезна аспирантам и студентам при изучении курсов теории управления и идентификации, вычислительной математики, программирования.

Н. РАЙБМАН

ПРЕДИСЛОВИЕ

Марковская цепь представляет собой вероятностную модель, пригодную для описания последовательностей наблюдений, которые в каждый момент времени характеризуют состояние наблюдаемой величины или класс, к которому она принадлежит. Несмотря на относительно недавнее происхождение, марковские модели нашли применение при описании временных рядов в самых разных областях научной и практической деятельности. Давно известны методы статистического анализа параметров стохастических систем по временным рядам микронаблюдений, применение этих методов нашло широкое отражение в литературе. Однако первое обсуждение задачи оценивания переходных вероятностей марковских цепей первого порядка связано с началом 50-х годов, а основные результаты в этой области получены в последние пять лет.

Основные задачи этой книги состоят в том, чтобы: (1) подытожить и дать оценку тем разрозненным результатам, которые можно найти в литературе, (2) дать способы оценки макрохарактеристик вероятностных моделей и предложить соответствующие машинные процедуры и (3) сравнить влияние конечного объема выборки на качество различных оценок по ограниченному статистическому эксперименту. Хотя излагаемые результаты в основном применимы к агрегированным данным, которые описываются стационарной марковской цепью первого порядка, в приложениях рассматривается распространение результатов как на случай, когда переходные вероятности зависят от времени, так и на общий случай задачи оценивания по относительным частотам состояний. Содержание этой книги формировалось под влиянием многих исследователей. Статьи Миллера [79], Гудмана [41], Маданского [72], Телсера [110] определили основное направление исследований и дали ценное интуитивное понимание того, как могут решаться задачи оценивания.

Ц. ЛИ,
Д. ДЖАДЖ,
А. ЗЕЛЬНЕР

Глава 1 ● ВВЕДЕНИЕ

Одна из целей количественной экономики состоит в том, чтобы объяснить или интерпретировать статистические данные, описывающие какие-либо экономические процессы или системы. Такое знание желательно, так как обычно предполагается, что, понимая механизмы явлений, лежащих в основе наблюдаемых экономических характеристик, можно предсказывать будущие значения соответствующих переменных и (или) управлять ими. Кроме того, такое знание может подсказать, как изменить экономическую структуру или как выбрать значения переменных состояний или управлений, если должны быть достигнуты определенные экономические цели.

В рамках экономической теории, по словам Маршака [73], числовые данные обычно рассматриваются как решения систем уравнений, которые в общем случае являются взаимозависимыми, динамическими и случайными. В другой работе [74] Маршак замечает, что экономические наблюдения почти никогда не бывают безошибочными, а уравнения экономики почти всегда подвержены действию возмущений. Таким образом, величины, которые мы хотим предсказывать, случайные, и прогноз состоит в назначении вероятностей, иными словами, в установлении распределений вероятностей. В теории вероятностей случайный процесс [24] понимается как процесс, развитие которого во времени и (или) пространстве подчиняется вероятностным законам. Следовательно, будущие значения процесса не являются детерминированными и можно только приписать различным будущим значениям или состояниям их вероятности. Для таких вероятностных моделей стохастические и динамические свойства экономических данных заставили многих экономистов утверждать, что некоторые экономические временные ряды можно интерпретировать как случайные процессы. В этом смысле проблема предсказания траектории изменения экономических переменных сводится к прогнозированию траектории случайного процесса, связанного с конкретной экономической структурой и предполагаемыми управляющими воздействиями.

Во многих современных теоретических и прикладных исследованиях основной вероятностной моделью является марковская цепь с дискретным временем, использованная при анализе временных рядов экономических показателей, когда в любой момент наблюдения фиксируется состояние, в котором находится наблюдаемая величина.

Неполный перечень исследований по применению марковских моделей включает работы [16] и [102] по распределению прибылей и заработной платы, [42] и [88] по исследованию социальной подвижности, [1], [44], [89] и [104] по определению размеров фирм и концентрации деловой активности, [103] по анализу потребления продуктов питания, [39] и [101], посвященные проблемам международной и межрегиональной торговли, [108] — [110] по анализу поведения потребителей, а также работу [22] и ряд других исследований по изменению рыночных цен. Если говорить о применении марковских моделей в других общественных науках, то можно назвать работы [15] по стохастическим моделям обучения, [13] и [76] по изучению миграций населения и текучести кадров, [20], в которой исследуются процессы распространения информации, и [3] по исследованию мнений избирателей. В дополнение следует заметить, что марковские процессы принятия решений представляют существенный класс оптимальных задач. Примерами исследований в этой области могут служить работы [5], [51], [99].

1.1. МАРКОВСКИЕ ЦЕПИ

Рассматриваемый тип стохастических процессов характеризуется тем, что (1) пространство элементарных событий определяется бесконечными последовательностями вида $\{x_0, x_1, x_2, \ldots\}$, в которых случайная величина x_t может принимать конечное число значений s_i, $i = 1, 2, \ldots, r$, в равноотстоящие моменты времени $t = 0, 1, 2, \ldots, T$. Распределение вероятностей различных состояний или исходов в каждый из этих моментов зависит только от состояния в предыдущий момент времени, т. е.

$$\Pr(x_t \mid x_{t-1}, x_{t-2}, \ldots) = \Pr(x_t \mid x_{t-1}) \text{ для всех } t, \qquad (1.1.1)$$

где через $\Pr(x_t \mid \ldots)$ обозначено условное распределение вероятностей x_t. Вероятность упорядоченного набора состояний определяется по закону умножения условных вероятностей:

$$\Pr(x_0, x_1, \ldots, x_T) = \Pr(x_0) \Pr(x_1 \mid x_0) \Pr((x_2 \mid x_0)x_1) \ldots \qquad (1.1.2)$$

Для марковских процессов эту формулу можно переписать в виде

$$\Pr(x_0, x_1, \ldots, x_T) = \Pr(x_0) \prod_{t=1}^{T} \Pr(x_t \mid x_{t-1}). \qquad (1.1.3)$$

Такой механизм функционирования полностью определяется начальным распределением вероятностей $\Pr(x_0)$ и условными вероятностями $\Pr(x_t \mid x_{t-1})$. Далее, пусть $x_{t-1} = s_i$ и $x_t = s_j$, тогда

$$\Pr(x_t = s_j \mid x_{t-1} = s_i) = p_{ij}(t) = p_{ij} \text{ для всех } t, \qquad (1.1.4)$$

где p_{ij} — не зависящая от времени вероятность перехода из состояния s_i в состояние s_j. Эти вероятности p_{ij} могут быть записаны в виде $(r \times r)$-матрицы переходных вероятностей $P = [p_{ij}]$ для всевозможных пар состояний s_i, s_j $(i, j = 1, 2, \ldots, r)$. Переходные вероятности обладают следующими свойствами:

$$0 \leqslant p_{ij} \leqslant 1 \tag{1.1.5}$$

и

$$\sum_j p_{ij} = 1, \quad i = 1, 2, \ldots, r. \tag{1.1.6}$$

Поскольку распределение вероятностей начального состояния x_0 и матрица переходных вероятностей P полностью определяют вероятностное поведение марковской цепи, то при заданных x_0 и P возникает задача установления распределения вероятностей для каждого x_t и, возможно, предельного распределения значений случайной величины x_t при $t \to \infty$, если такое распределение существует. Если марковская цепь неприводимая и апериодическая, а следовательно, и эргодическая, то существует единственный вектор-строка $\pi = (\pi_1, \pi_2, \ldots, \pi_r)$ — вектор стационарных вероятностей, такой, что

$$\lim_{t \to \infty} p_{ij}^{(t)} = \pi_j, \quad i, j = 1, 2, \ldots, r, \tag{1.1.7}$$

где $p_{ij}^{(t)}$ представляет собой (i, j)-й элемент матрицы P^t,

$$0 \leqslant \pi_j \leqslant 1, \tag{1.1.8}$$

$$\sum_j \pi_j = 1 \tag{1.1.9}$$

и

$$\pi = \pi P. \tag{1.1.10}$$

В этом разделе дано определение только основных понятий, используемых в следующих главах. Более полное изложение теории конечных цепей Маркова можно найти в [7], [24] и [60].

1.2. ЗАДАЧИ ОЦЕНИВАНИЯ

Когда имеются последовательность повторяющихся наблюдений и данные об изменении во времени состояний микрообъектов, то, как показано в работе [4], формула (1.1.3) определяет функцию правдоподобия, которая служит для вычисления оценок переходных вероятностей и для проверки различных гипотез о принимаемых ими значениях.

К сожалению, столь детальная информация часто недоступна, обходится слишком дорого или является неполной, поэтому приходится работать с агрегированными данными. Так, например, при переписи определяется только распределение числа индивидуальных объектов по всевозможным состояниям за каждый год переписи, но остается неизвестной траектория движения отдельного объекта. Возможно

также, что общее число микрообъектов может изменяться во времени, поэтому доступная информация характеризует лишь относительные частоты* каждого состояния.

Столкнувшись с тем, что во многих случаях доступны только агрегированные или частотные данные, естественно спросить, достаточно ли такой агрегированной информации для оценки матрицы переходных вероятностей P, которая отражает или определяет поведение микрообъектов. Именно этому вопросу посвящена настоящая книга, в связи с чем основные задачи исследования состоят в том, чтобы (1) построить оценки элементов матрицы переходных вероятностей P по агрегированным частотным данным и (2) изучить распределения и качества различных способов оценивания по выборочным результатам экспериментов.

1.3. ПЛАН КНИГИ

Глава 2 посвящена обычным и байесовским оценкам переходных вероятностей в случаях, когда имеются все необходимые выборочные данные по микрообъектам (микроданные). В главе 3 прослеживается путь развития оценок переходных вероятностей по временным рядам агрегированных данных. В частности, вводится линейная статистическая модель и исследуются свойства оценок, полученных обычным методом наименьших квадратов и с ограничениями. В главе 4 формулируется более общая модель выборочных экспериментов, которая рассматривается сначала при определении качества оценок в главах 2 и 3.

В главах 5—10 выводятся оценки, полученные методом взвешенных наименьших квадратов с ограничениями, оценки по критерию χ^2, оценки максимального правдоподобия, байесовские оценки и оценки по методу наименьших модулей, которые применяются к агрегированным выборочным данным и к решению двух примеров, основанных на фактических данных. В главах 11 и 12 рассматриваются возможности использования оценок для прогнозирования и их сравнительные характеристики. Глава 13 включает интерпретацию полученных результатов и комментарий по поводу направления будущих исследований.

Применение псевдообратных матриц к задаче оценивания переходных вероятностей по обобщенному методу наименьших квадратов рассматривается в приложении *А*. В приложении *Б* обсуждается линейная вероятностная модель в связи с общей задачей оценивания по данным об относительных частотах и приводится несколько специальных оценок. В приложении *В* рассматривается оценивание переходных вероятностей неоднородных марковских цепей. И наконец, приложение *Г* содержит машинные программы, реализующие все включенные в книгу процедуры оценивания.

---

* В советской литературе относительную частоту иногда называют частостью или просто частотой, в оригинале употребляется термин пропорция (*proportion*). —*Прим. пер.*

Глава 2 ● ОЦЕНИВАНИЕ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ ПО МИКРОДАННЫМ

После краткого введения в теорию конечных марковских цепей (параграф 1.1) рассмотрим методы статистического анализа параметров марковской модели при большом числе наблюдений за состоянием микрообъектов и не зависящих от времени вероятностей перехода. Будут получены оценки максимального правдоподобия и байесовские оценки переходных вероятностей в предположении, что имеются выборка микроданных и последовательность повторяющихся наблюдений за состоянием марковской цепи.

2.1. МИКРООЦЕНКА МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ [1]

Придерживаясь определений и обозначений, введенных в параграфе 1.1, предположим, что имеется последовательность наблюдений за состоянием эргодической марковской цепи. Допустим также, что в момент времени $t = 0$ в состоянии i находится $n_i(0)$ микрообъектов и пусть в процессе наблюдения фиксируется последовательное изменение состояний микрообъектов в моменты времени $t = 1, 2, \ldots, T$. В соответствии с параграфом 1.1 определим вероятность упорядоченного набора состояний однородной марковской цепи по формуле

$$\Pr(x_0, x_1, x_2, \ldots, x_T) = \Pr(x_0) \prod_t \Pr(x_t \mid x_{t-1}). \qquad (2.1.1)$$

Пусть через $n_{ij}(t)$ обозначено число микрообъектов, для которых $x_{t-1} = s_i$ и $x_t = s_j$, и пусть

$$n_{ij} = \sum_t n_{ij}(t). \qquad (2.1.2)$$

По формуле (2.1.1) вероятность фиксированной совокупности наборов состояний n микрообъектов можно определить как

$$\Pr(x_0, x_1, \ldots, x_T \mid \mathrm{n}) \propto \Pr(x_0) \prod_{i,j} p_{ij}^{n_{ij}} \qquad (2.1.3)$$

[1] В этом параграфе изложение следует работе [4].

(здесь $\propto$ — знак пропорциональности), и, как показано в [4], совокупность n_{ij} образует множество достаточных статистик.

Распределение вероятностей $n_{ij}(t)$ можно получить, рассматривая $n_i(t-1) = \sum_j n_{ij}(t)$ наблюдений значений величин, распределенных по полиномиальному закону с вероятностями p_{ij}. Тогда плотность вероятности (пв) для $n_{ij}(t)$ запишется как

$$\Pr(n_{11}(t), n_{12}(t), \ldots \mid \mathbf{n}(0)' \, p_{11}, \ldots) = \prod_i \left(\prod_i \left[\frac{n_i(t-1)!}{\prod_j n_{ij}(t)!} \times \right.\right.$$

$$\left.\left. \times \prod_j p_{ij}^{n_{ij}(t)} \right]\right) = \left[\prod_{i,i} \frac{n_i(t-1)!}{\prod_j n_{ij}(t)!} \right] \left[\prod_{i,j} p_{ij}^{\sum_i n_{ij}(t)} \right] =$$

$$= \left[\prod_{i,i} \frac{n_i(t-1)!}{\prod_j n_{ij}(t)!} \right] \left[\prod_{i,j} p_{ij}^{n_{ij}} \right], \qquad (2.1.4)$$

где $\mathbf{n}(0)' = [n_1(0), n_2(0), \ldots, n_r(0)]$ — вектор, элементы которого равны числу микрообъектов в различных состояниях при $t = 0$.

Для фиксированных наблюдений $n_{ij}(t)$ можно получить оценки не зависящих от времени вероятностей перехода p_{ij}, максимизируя функцию правдоподобия (2.1.4) по p_{ij} при учете ограничения (1.1.6). Заменяя выражение (2.1.3) его логарифмом, получим функцию Лагранжа:

$$\log \Pr(x_0, x_1, \ldots, x_T \mid n) - \sum_i \lambda_i \left(\sum_j p_{ij} - 1 \right) \qquad (2.1.5)$$

или

$$\log \left(\prod_{ij} p_{ij}^{n_{ij}} \right) - \sum_i \lambda_i \left(\sum_j p_{ij} - 1 \right) + \text{const}.$$

Поскольку $n_i(0)$ считается неслучайным, можно записать следующие необходимые условия максимума функции (2.1.5):

$$[\partial/\partial p_{ij}] \left[\sum_i \sum_j n_{ij} \log p_{ij} - \sum_i \lambda_i \left(\sum_j p_{ij} - 1 \right) \right] = n_{ij}/p_{ij} - \lambda_i = 0 \qquad (2.1.6)$$

и

$$[\partial/\partial \lambda_i] \left[\sum_i \sum_j n_{ij} \log p_{ij} - \sum_i \lambda_i \left(\sum_j p_{ij} - 1 \right) \right] = \sum_j p_{ij} - 1 = 0. \qquad (2.1.7)$$

Из (2.1.6) следует, что $n_{ij} = \lambda_i p_{ij}$. Суммируя это выражение по j, из (2.1.7) находим, что

$$\lambda_i = \sum_j n_{ij}. \qquad (2.1.8)$$

Подставляя (2.1.8) в (2.1.6) и решая полученное уравнение относительно p_{ij}, определим оценку максимального правдоподобия (ОМП):

$$\dot{p}_{ij} = n_{ij} / \sum_j n_{ij} \geqslant 0. \qquad (2.1.9)$$

В силу неотрицательности n_{ij} для ОМП (2.1.9) всегда выполняется ограничение (1.1.5).

Как показано в [62], ОМП (2.1.9) состоятельна, но, вообще, она не является несмещенной. Однако в [62] отмечается, что с увеличением объема выборки смещение стремится к нулю, и показано, что эта оценка асимптотически нормальная (см. также [4]). С помощью результатов, полученных в [10] и [49], удалось доказать [4], что при $T \to \infty$ асимптотические свойства полученной оценки такие же, как и при $n = 1$. Этот результат сохраняется и тогда, когда компоненты $\mathbf{n}$ (0) — случайные величины.

Для ОМП в условиях только что описанного выборочного эксперимента в работе [4] построены таблицы сопряженности признаков по критериям максимального правдоподобия и критерию χ^2 для проверки следующих гипотез:

(1) переходные вероятности марковской цепи первого порядка не зависят от времени;

(2) переходные вероятности равны заданным числам;

(3) рассматриваемый процесс представляет собой марковскую цепь заданного порядка.

2.2. БАЙЕСОВСКИЙ АНАЛИЗ МИКРОМОДЕЛИ

Вывод оценки, построенной в предыдущем параграфе, был основан на знании выборочных значений n_{ij} и на том, что вероятности перехода (1) не могут быть отрицательными, (2) не должны превышать единицу и (3) их сумма по строке равна единице.

Во многих случаях исследователь располагает априорной информацией о структуре отдельных элементов матрицы переходных вероятностей P. Такая информация может быть связана, например, с некими содержательными соображениями и (или) с опытом предварительных экспериментов. В этом случае может быть желательным учет подобной информации при оценивании переходных вероятностей по выборочным данным. Хорошо известно, что методом, позволяющим сочетать выборочные и априорные данные, является байесовский подход, к рассмотрению которого мы сейчас приступим.

а) Теорема Байеса и статистический анализ. В байесовском подходе предполагается, что степень неопределенности в знании вектора параметров θ может быть выражена априорной плотностью вероятности Pr (0), $\theta \in \Omega$, где Ω — область допустимых значений в пространстве параметров[1]. Для определения апостериорной плотности вероятности

[1] В обозначении априорной плотности вероятности Pr (0) не фигурирует явно зависимость от тех параметров, выбор которых находится в распоряжении исследователя.

после того, как получен вектор выборочных наблюдений $\mathbf{y}$, по теореме Байеса необходимо знание плотности условного распределения вероятностей $\Pr(\mathbf{y}\,|\,\theta)$, $\mathbf{y} \in R_{\mathbf{y}}$, где $R_{\mathbf{y}}$ — выборочное пространство. Апостериорная плотность вероятности, которая содержит всю априорную и выборочную информацию, может быть полезна при оценивании неизвестных параметров. Далее, если в распоряжении исследователя имеется функция, отражающая потери вследствие неточного оценивания, то, вообще говоря, можно получить оценку $\hat{\theta}$, которая минимизирует апостериорное математическое ожидание функции потерь. В широком диапазоне условий $\hat{\theta}$ минимизирует также средний риск. В этом случае $\hat{\theta}$ называют байесовской оценкой по отношению к данной функции потерь и выбранному априорному распределению. Известно, что байесовские оценки допустимы и составляют полный класс (см., например, [36]).

Прежде чем перейти к байесовскому оцениванию марковских вероятностей перехода, сформулируем сначала теорему Байеса. Пусть $\Pr(\mathbf{y}, \theta)$ — совместная плотность вероятности вектора наблюдений $\mathbf{y}$ и вектора параметров θ, $\mathbf{y} \in R_{\mathbf{y}}$, $\theta \in \Omega$. Справедливы обычные соотношения

$$\Pr(\mathbf{y}, \theta) = \Pr(\mathbf{y}\,|\,\theta)\Pr(\theta) = \Pr(\theta\,|\,\mathbf{y})\Pr(\mathbf{y}) \tag{2.2.1}$$

и

$$\Pr(\theta\,|\,\mathbf{y})\Pr(\mathbf{y}) = \Pr(\mathbf{y}\,|\,\theta)\Pr(\theta). \tag{2.2.2}$$

Таким образом, для апостериорной плотности вероятности при фиксированном векторе наблюдений можно записать:

$$\Pr(\theta\,|\,\mathbf{y}) = [\Pr(\theta)\Pr(\mathbf{y}\,|\,\theta)]/\Pr(\mathbf{y}) \propto \Pr(\theta)\Pr(\mathbf{y}\,|\,\theta),\ \theta \in \Omega. \tag{2.2.3}$$

Здесь $\propto$ — знак пропорциональности. Отсюда

$$\Pr(\theta\,|\,\mathbf{y}) \propto \Pr(\theta)\, l(\theta\,|\,\mathbf{y}),\ \theta \in \Omega, \tag{2.2.4}$$

где $\Pr(\mathbf{y}\,|\,\theta)$, рассматриваемая как функция θ, есть функция правдоподобия $l(\theta\,|\,\mathbf{y})$. Уравнение (2.2.4) выражает теорему Байеса. Отметим, что апостериорная плотность вероятности сочетает априорную и выборочную информацию, в дальнейшем она используется при выводе байесовских оценок.

б) Априорная плотность вероятности. В случае, когда оцениваемыми параметрами являются переходные вероятности, имеется по меньшей мере три специфических ограничения: (1) параметры не могут быть отрицательными, (2) параметры не могут быть больше единицы и (3) сумма вероятностей полной системы непересекающихся событий должна быть равна единице.

Кроме того, часто из содержательных соображений и (или) после предварительных экспериментов становятся известными приближенные значения определенных параметров. Для использования всей этой информации в качестве априорного распределения будет применяться

многомерное бета-распределение. Семейство бета-распределений достаточно полно для того, чтобы во многих случаях из него можно было выбрать распределение, адекватно отражающее имеющуюся априорную информацию.

Таким образом, следуя [75] и [77], запишем априорную плотность вероятности для элементов i-й строки матрицы переходных вероятностей в виде пв многомерного бета-распределения:

$$f(p_{i1}, p_{i2}, \ldots, p_{ir}) = \frac{\Gamma(a_{i1}+a_{i2}+\ldots+a_{ir})}{\Gamma(a_{i1})\,\Gamma(a_{i2})\ldots\Gamma(a_{ir})} \cdot (p_{i1}^{a_{i1}-1}\, p_{i2}^{a_{i2}-1} \ldots p_{ir}^{a_{ir}-1}), \tag{2.2.5}$$

где $0 \leqslant p_{ij} \leqslant 1$, $j = 1, 2, \ldots, r$ и $\sum_j p_{ij} = 1$, Γ обозначает гамма-функцию, a — положительные параметры, значения которых должны быть выбраны на основе имеющейся априорной информации. Маргинальная пв для одной из p_{ij} будет, конечно, обычной пв одномерного бета-распределения.

Матрица переходных вероятностей P содержит r строк, поэтому необходимо иметь r априорных распределений. Если априори предполагается, что столбцы матрицы P распределены независимо и в соответствии с формулой (2.2.5), то пв совместного распределения всех p_{ij} запишется как

$$\prod_i f(p_{i1}, p_{i2}, \ldots, p_{ir}) = \prod_i \left(\frac{\Gamma\left(\sum_j a_{ij}\right)}{\prod_j \Gamma(a_{ij})} \prod_j p_{ij}^{a_{ij}-1} \right), \tag{2.2.6}$$

для $0 \leqslant p_{ij} \leqslant 1$, $j = 1, 2, \ldots, r$, что представляет собой произведение многомерных бета-распределений. Эта пв может быть названа частным случаем матричного бета-распределения [75].

Для отображения имеющейся априорной информации необходимо выбрать определенные значения параметров a бета-распределения, задаваемого формулой (2.2.6). Это можно сделать по априорной информации о средних значениях и дисперсиях p_{ij}, если иметь в виду, что

$$E(p_{ij}) = a_{ij} \Big/ \sum_{j=1}^{r} a_{ij} = a_{ij}/a_i; \tag{2.2.7}$$

$$\mathrm{Var}(p_{ij}) = \frac{a_{ij}(a_i - a_{ij})}{a_i^2(a_i+1)} = \frac{E(p_{ij})(1 - E(p_{ij}))}{(a_i+1)}. \tag{2.2.8}$$

Таким образом, при заданных $E(p_{ij})$ и $\mathrm{Var}(p_{ij})$ по формуле (2.2.8) можно определить a_i, с помощью (2.2.7) можно найти a_{ij}. Как показано в [75], при заданных $E(p_{ij})$ и $E(p_{ih})$ и определенном a_i априорное значение взаимной ковариации будет отрицательным[1].

[1] Как показано в [75], отказ от взаимной независимости строк приводит к усложнению формул для моментов распределения и т. п.

в) **Апостериорная плотность вероятности.** При известной функции правдоподобия переходных вероятностей p_{ij} (2.1.3) и априорном распределении, заданном в форме матричного бета-распределения, можно применить теорему Байеса. При фиксированных выборочных данных о микрообъектах n_{ij} функция правдоподобия запишется так:

$$l(P \mid \mathbf{n}) \propto \prod_{i,j} p_{ij}^{n_{ij}} \propto \prod_{i=1}^{r} \prod_{j=1}^{r-1} p_{ij}^{n_{ij}} \left(1 - \sum_{j=1}^{r-1} p_{ij}\right)^{n_{ir}}, \qquad (2.2.9)$$

где $0 \leqslant p_{ij} \leqslant 1$ и $\sum_j p_{ij} = 1$, а $\mathbf{n}$ — вектор с элементами n_{ij}, $i, j = 1, 2, \ldots, r$. Тогда по теореме Байеса имеем

$$\Pr(P \mid \mathbf{n}) \propto \Pr(P)\, l(P \mid \mathbf{n}) \propto \prod_{i,j} p_{ij}^{n_{ij}+a_{ij}-1} \propto$$

$$\propto \prod_{i=1}^{r} \prod_{j=1}^{r-1} p_{ij}^{n_{ij}+a_{ij}-1} \left(1 - \sum_{j=1}^{r-1} p_{ij}\right)^{n_{ir}+a_{ir}-1}. \qquad (2.2.10)$$

Таким образом, апостериорная плотность вероятности также принадлежит к классу произведений многомерных бета-распределений с параметрами $n_{ij} + a_{ij}$:

$$\Pr(P \mid \mathbf{n}) = \prod_{i=1}^{r} \left(\frac{\Gamma\left(\sum_{j=1}^{r} n_{ij} + a_{ij}\right)}{\prod_{j}^{r} \Gamma(n_{ij} + a_{ij})} \prod_{j}^{r} p_{ij}^{n_{ij}+a_{ij}-1} \right) =$$

$$= k \prod_{i=1}^{r} \prod_{j=1}^{r-1} p_{ij}^{n_{ij}+a_{ij}-1} \left(1 - \sum_{j=1}^{r-1} p_{ij}\right)^{n_{ir}+a_{ir}-1}. \qquad (2.2.11)$$

Как показано в [75], маргинальная апостериорная пв подматриц матрицы P будет также пв матричного бета-распределения. Маргинальная апостериорная пв распределения любой из переходных вероятностей является пв обычного бета-распределения со средним

$$E(p_{ij} \mid \mathbf{n}) = \frac{n_{ij} + a_{ij}}{\sum_{j=1}^{r} (n_{ij} + a_{ij})} = \frac{n_{ij} + a_{ij}}{c_i} \qquad (2.2.7а)$$

и дисперсией

$$\operatorname{Var}(p_{ij} \mid \mathbf{n}) = \frac{(n_{ij} + a_{ij})(c_i - n_{ij} - a_{ij})}{c_i^2 (c_i + 1)} = \frac{E(p_{ij} \mid \mathbf{n})(1 - E(p_{ij} \mid \mathbf{n}))}{c_i + 1}. \qquad (2.2.8а)$$

г) Байесовская оценка. Апостериорная пв, объединяющая априорную и выборочную информацию, может быть использована для оценки значений переходных вероятностей. Из теории оценок известно, что среднее значение апостериорного распределения представляет собой байесовскую оценку относительно модуля отклонения оценки от истинного значения. Кроме того, отметим, что с помощью моды апостериорного распределения получаются следующие оценки переходных вероятностей:

$$\ddot{p}_{ij} = \frac{n_{ij}+a_{ij}-1}{\sum\limits_j n_{ij}+\sum\limits_j a_{ij}-r}; \quad i=1, 2, \ldots, r; \quad j=1, 2, \ldots, r-1. \quad (2.2.12)$$

Эту формулу можно получить, максимизируя логарифм (2.2.11) по совокупности p_{ij}. Значения последних r вероятностей p_{ir}, $i=1, 2, \ldots, r$, можно определить из условия $\sum\limits_{j=1}^{r} p_{ij}=1$. Функция потерь, минимизация которой приводит к байесовской оценке в виде моды апостериорного распределения, получена в работе [12].

Глава 3 ОЦЕНИВАНИЕ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ ПО МАКРОДАННЫМ

В предыдущей главе были выведены оценки максимального правдоподобия и байесовские оценки в предположении, что имеются упорядоченные во времени последовательности наблюдений за состоянием микрообъектов (микроданные). В этой и последующих главах строятся оценки переходных вероятностей для случая, когда вместо информации о траекториях движения отдельных микрообъектов $n_{ij}(t)$ имеется только агрегированная информация в виде относительных частот состояний в каждый из моментов времени t.

3.1. СООТНОШЕНИЯ ДЛЯ МАКРОДАННЫХ

Если вместо выборочных значений $n_{ij}(t)$ наблюдаются только агрегированные данные $n_j(t) = \sum_i n_{ij}(t)$, то для того, чтобы использовать наблюдаемые относительные частоты при оценке вероятностей перехода, необходимо записать следующее соотношение:

$$\Pr(x_{t-1} = s_i, x_t = s_j) = \Pr(x_{t-1} = s_i) \Pr(x_t = s_j \mid x_{t-1} = s_i). \tag{3.1.1}$$

Тогда по формуле полной вероятности имеем:

$$\Pr(x_t = s_j) = \sum_i \Pr(x_{t-1} = s_i) \Pr(x_t = s_j \mid x_{t-1} = s_i) \tag{3.1.2}$$

или

$$q_j(t) = \sum_i q_i(t-1)\, p_{ij}, \tag{3.1.3}$$

где $q_j(t)$ и $q_i(t-1)$ представляют собой безусловные вероятности $\Pr(x_t = s_j)$ и $\Pr(x_{t-1} = s_i)$ соответственно. Если безусловные вероятности $q_j(t)$ и $q_i(t-1)$ в (3.1.3) заменить фактическими значениями наблюдаемых частот $y_j(t)$ и $y_i(t-1)$, то не найдется таких оценок вероятностей перехода, которые с вероятностью, равной единице, удовлетворяли бы соотношению (3.1.3). Таким образом, если признать, что истинное значение относительной частоты и ее оценка $y_j(t)$ не совпадают, то, обозначая через $u_j(t)$

величину ошибки в правой части уравнения (3.1.3), можно записать, что выборочные наблюдения удовлетворяют следующему стохастическому уравнению:

$$y_j(t) = \sum_i y_i(t-1)\, p_{ij} + u_j(t). \tag{3.1.4}$$

В работе [79] это уравнение предлагается как линейная статистическая модель для оценивания переходных вероятностей. Именно к ее анализу мы и приступим.

3.2. ОЦЕНКА ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ БЕЗ ОГРАНИЧЕНИЙ

Развивая подход, предложенный в [79], для оценивания переходных вероятностей по выборочным значениям частот с помощью (3.1.4) перепишем это уравнение в матричной форме:

$$\mathbf{y}_j = X_j \mathbf{p}_j + \mathbf{u}_j, \tag{3.2.1}$$

где $\mathbf{y}_j = \{y_j(t)\}$ — это $(T \times 1)$-вектор выборочных значений, $\mathbf{p}_j' = (p_{1j}, p_{2j}, \ldots, p_{rj})$ — $(r \times 1)$-вектор неизвестных параметров (оцениваемых переходных вероятностей), $\mathbf{u}_j$ — $(T \times 1)$-вектор случайных ошибок и X_j — следующая $(T \times r)$-матрица:

$$X_j = \begin{bmatrix} y_1(0) & y_2(0) & \ldots & y_r(0) \\ \vdots & \vdots & & \vdots \\ y_1(t-1) & y_2(t-1) & \ldots & y_r(t-1) \\ \vdots & \vdots & & \vdots \\ y_1(T-1) & y_2(T-1) & \ldots & y_r(T-1) \end{bmatrix}. \tag{3.2.2}$$

Предполагается, что ранг матрицы X_j равен r. Что касается вектора ошибок $\mathbf{u}_j$, то

$$E(\mathbf{u}_j) = 0 \tag{3.2.3}$$

и

$$E(\mathbf{u}_j \mathbf{u}_j') = \sigma_j \omega_{jj}, \tag{3.2.4}$$

где ω_{jj} — положительно-определенная диагональная $(T \times T)$-матрица[1].

Уравнения (3.1.4) и (3.2.1) представляют собой часть расширенной системы уравнений:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_r \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_r \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_r \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_r \end{bmatrix} \tag{3.2.5}$$

---

[1] Возможности такого описания вектора невязок рассмотрены в [72].

или

$$\mathbf{y} = X\mathbf{p} + \mathbf{u}, \tag{3.2.6}$$

где $\mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_r')$, $\mathbf{p}' = (\mathbf{p}_1', \mathbf{p}_2', \ldots, \mathbf{p}_r')$, $\mathbf{u}' = (\mathbf{u}_1', \mathbf{u}_2', \ldots, \mathbf{u}_r')$, а X — блочно-диагональная матрица, в которой $X_1 = X_2 = \ldots = X_r$,

$$E(\mathbf{u}) = 0 \tag{3.2.7}$$

и

$$E(\mathbf{u}\mathbf{u}') = \Sigma. \tag{3.2.8}$$

Здесь Σ — недиагональная вырожденная матрица[1] размера $Tr \times Tr$.

В случае линейной статистической модели в форме (3.2.5) или (3.2.6) при $T > r$ в работе [79] было предложено для оценки переходных вероятностей применять обычный метод наименьших квадратов. Иными словами, задача оценивания рассматривалась как задача определения такой оценки $\tilde{\mathbf{p}}$, которая минимизирует положительно определенную квадратичную форму

$$\varphi = \mathbf{u}'\mathbf{u} = (\mathbf{y} - X\mathbf{p})'(\mathbf{y} - X\mathbf{p}). \tag{3.2.9}$$

Решая эту экстремальную задачу, находим, что

$$\tilde{\mathbf{p}} = (X'X)^{-1} X'\mathbf{y} \tag{3.2.10}$$

при условии, что $X'X$ не является вырожденной. Но это не так, поскольку $X'X$ представляет собой блочно-диагональную матрицу с матрицами $X_j'X_j$ на диагонали, которые не вырождены, так как, по предположению, матрицы X_j, $j = 1, 2, \ldots, r$, имеют ранг r. В силу положительной определенности (а также симметричности) матрицы $X'X$ для вектора $\mathbf{p}$ выполняются не только необходимые, но и достаточные условия минимума функционала (3.2.9).

Несмотря на то что уравнения системы (3.2.5) «связаны помехой», в силу условия $X_1 = X_2 = \ldots = X_r$ r уравнений этой системы могут оцениваться как вместе, так и порознь, с одними и теми же результатами [121]. Таким образом, в соответствии с (3.2.10) «обычная» оценка метода наименьших квадратов $\mathbf{p}_j$ (подвектора $\tilde{\mathbf{p}}$) запишется как

$$\tilde{\mathbf{p}}_j = (X_j'X_j)^{-1} X_j'\mathbf{y}_j, \quad j = 1, 2, \ldots, r. \tag{3.2.11}$$

Возникает вопрос, удовлетворяют ли полученные оценки условию неотрицательности и нормировки, а именно

$$0 \leqslant p_{ij} \leqslant 1 \tag{3.2.12}$$

и

$$\sum_j p_{ij} = 1 \text{ для всех } i. \tag{3.2.13}$$

---

[1] Вырожденность Σ следует из того, что $\sum_j \mathbf{y}_j - \sum_j X_j \mathbf{p}_j = \eta_T - X_j \eta_r = \eta_T - \eta_T = 0$, где η_T и η_r представляют собой векторы с единичными элементами размера T и r соответственно.

а) **Условие нормировки.** Для того чтобы в постановке задачи учесть условие нормировки (3.2.13), сформулируем ее как задачу минимизации суммы квадратов ошибок:

$$\varphi = \mathbf{u}'\mathbf{u} = (\mathbf{y} - X\mathbf{p})' \ (\mathbf{y} - X\mathbf{p}) \tag{3.2.9}$$

при условии

$$G\mathbf{p} = \boldsymbol{\eta}_r \tag{3.2.14}$$

или

$$\sum_j p_{ij} = 1 \text{ для всех } i, \tag{3.2.13}$$

где G — $(r \times r^2)$-матрица вида $[I_1, I_2, \ldots, I_r]$, а каждая матрица I_i представляет собой единичную матрицу размера $r \times r$, и $\boldsymbol{\eta}_r$ — $(r \times 1)$-вектор-столбец, составленный из одних единиц[1].

Не требуя неотрицательности переменных

$$\mathbf{p} \geqslant 0, \tag{3.2.15}$$

можно учесть условие нормировки (3.2.14) и решить задачу оценивания с одним ограничением, сформировав следующий лагранжиан:

$$\varphi(\mathbf{p}, \boldsymbol{\lambda}) = (\mathbf{y} - X\mathbf{p})' (\mathbf{y} - X\mathbf{p}) - 2\boldsymbol{\lambda}' (G\mathbf{p} - \boldsymbol{\eta}_r), \tag{3.2.16}$$

где $\boldsymbol{\lambda}$ — $(r \times 1)$-вектор множителей Лагранжа, и определив стационарную точку $(\mathbf{p}^*, \boldsymbol{\lambda}^*)$ функционала $\varphi(\mathbf{p}, \boldsymbol{\lambda})$:

$$\mathbf{p}^* = (X'X)^{-1} X'\mathbf{y} + (X'X)^{-1} G' (G (X'X)^{-1} G')^{-1} \times (\boldsymbol{\eta}_r - G (X'X)^{-1} X'\mathbf{y}) \tag{3.2.17}$$

и

$$\boldsymbol{\lambda}^* = (G (X'X)^{-1} X'G')^{-1} (\boldsymbol{\eta}_r - G (X'X)^{-1} X'\mathbf{y}). \tag{3.2.18}$$

С помощью (3.2.10) можно переписать (3.2.17) и (3.2.18):

$$\mathbf{p}^* = \tilde{\mathbf{p}} + (X'X)^{-1} G' (G (X'X)^{-1} G')^{-1} (\boldsymbol{\eta}_r - G\tilde{\mathbf{p}}), \tag{3.2.19}$$

$$\boldsymbol{\lambda}^* = (G (X'X)^{-1} X'G')^{-1} (\boldsymbol{\eta}_r - G\tilde{\mathbf{p}}). \tag{3.2.20}$$

Для того чтобы показать, что условие нормировки выполняется автоматически, достаточно доказать одно из следующих эквивалентных утверждений:

$$\mathbf{p}^* = \tilde{\mathbf{p}}, \tag{3.2.21}$$

и

$$\boldsymbol{\lambda}^* = 0 \tag{3.2.22}$$

$$G\tilde{\mathbf{p}} - \boldsymbol{\eta}_r = 0. \tag{3.2.23}$$

---

[1] Оценивание по методу наименьших квадратов при наличии ограничений в форме линейных уравнений рассматривается в работах [40] и [112]. Более полное изложение теорем, представленных в этом и следующих параграфах, можно найти в [107].

Действительно, из (3.2.23) следует (3.2.22), из (3.2.22) следует (3.2.21) и, наконец, из (3.2.21) следует (3.2.23). Таким образом, для того чтобы заключить, что $\mathbf{p}^* = \widetilde{\mathbf{p}}$, достаточно доказать (3.2.23) или (3.2.22). Докажем сначала (3.2.23).

Теорема 1. $G\widetilde{\mathbf{p}} = \boldsymbol{\eta}_r$.

Доказательство. Поскольку $(r \times r^2)$-матрица $G = (I\,I \ldots I)$, мы имеем:

$$G\widetilde{\mathbf{p}} = [I\,I \ldots I]\,(X'X)^{-1}X'\,\mathbf{y} = \tag{3.2.24}$$

$$= [I\,I \ldots I]\begin{bmatrix} (X_1'X_1)^{-1} & & & \\ & (X_2'X_2)^{-1} & & \\ & & \ddots & \\ & & & (X_r'X_r)^{-1} \end{bmatrix}\begin{bmatrix} X_1' & & & \\ & X_2' & & \\ & & \ddots & \\ & & & X_r' \end{bmatrix}\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_r \end{bmatrix} = \tag{3.2.25}$$

$$= ((X_1'X_1)^{-1}X_1',\ (X_2'X_2)^{-1}X_2', \ldots,\ (X_r'X_r)^{-1}X_r')\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_r \end{bmatrix} = \tag{3.2.26}$$

$$= (X_1'X_1)^{-1}X_1'(\mathbf{y}_1 + \mathbf{y}_2 + \ldots + \mathbf{y}_r) = \tag{3.2.27}$$

$$= (X'X)^{-1}X'\boldsymbol{\eta}_T. \tag{3.2.28}$$

А в силу того, что $\boldsymbol{\eta}_T = X_1\boldsymbol{\eta}_r$, последнее равенство сводится к следующему:

$$G\widetilde{\mathbf{p}} = (X_1'X_1)^{-1}\,X_1'X_1\boldsymbol{\eta}_r = \boldsymbol{\eta}_r. \tag{3.2.29}$$

Таким образом, мы получаем известный результат Гудмана [41] о том, что $G\widetilde{\mathbf{p}} = \boldsymbol{\eta}_r$, т. е. оценка переходных вероятностей по методу наименьших квадратов автоматически удовлетворяет условию нормировки[1].

б) Условие неотрицательности. Как отмечалось, для того чтобы полученные оценки могли быть оценками переходных вероятностей, должны быть выполнены условия (3.2.12) и (3.2.13). Следуя Гудману [41], мы показали, что условие нормировки (3.2.13) выполняется автоматически. Однако при применении обычного метода наименьших квадратов (без учета ограничений) оценки переходных вероятностей могут быть отрицательными[2].

---

[1] Нетрудно непосредственно доказать формулу (3.2.22) [110], а именно: *Теорема 2.* $\boldsymbol{\lambda}^* = (G\,(X'X)^{-1}\,G')^{-1}\,(\boldsymbol{\eta}_r - G\widetilde{\mathbf{p}}) = 0$.

Доказательство. Поскольку

$$G\,(X'X)^{-1}\,G' = r\,(X_1'X_1)^{-1},\ (G\,(X'X)^{-1}\,G')^{-1} = (1/r)\,X_1'X_1,$$

то

$$\boldsymbol{\lambda}^* = (G\,(X'X)^{-1}\,G')^{-1}\,(\boldsymbol{\eta}_r - G\widetilde{\mathbf{p}}) = (1/r)\,(X_1'X_1)\,(\boldsymbol{\eta}_r - (X_1'X_1)^{-1}\times$$
$$\times X_1'\boldsymbol{\eta}_r) = (1/r)\,(X_1'X_1\boldsymbol{\eta}_r - X_1'\boldsymbol{\eta}_r) = (1/r)\,(X_1'\boldsymbol{\eta}_T - X_1'\boldsymbol{\eta}_T) = (1/r)\,(\boldsymbol{\eta}_r - \boldsymbol{\eta}_r) = 0.$$

[2] Это, естественно, означает, что ограничение сверху может также не выполняться, т. е. некоторые оценки могут превышать единицу.

Для того чтобы показать это, запишем оценки, полученные методом наименьших квадратов без учета ограничений (3.2.10), в виде следующей системы уравнений:

$$(X'X)\tilde{\mathbf{p}} = X'\mathbf{y}. \tag{3.2.10а}$$

В этой системе

$$X'X = \begin{bmatrix} X_1'X_1 & & & \\ & X_2'X_2 & & \\ & & \ddots & \\ & & & X_r'X_r \end{bmatrix} \tag{3.2.30}$$

представляет собой $(r^2 \times r^2)$-матрицу, а $X_l'X_l$ — $(r \times r)$-матрицу, составленную из неотрицательных элементов, и

$$X'\mathbf{y} = \begin{bmatrix} X_1\mathbf{y}_1 \\ X_2\mathbf{y}_2 \\ \vdots \\ X_r\mathbf{y}_r \end{bmatrix} \tag{3.2.31}$$

есть $(r^2 \times 1)$-вектор из неотрицательных элементов. Таким образом, правая часть и матрица системы (3.2.10а) неотрицательны.

Для того чтобы исследовать неотрицательность решений системы (3.2.10а), положим

$$(X'X)\tilde{\mathbf{p}} = d'(I - A)\,d\tilde{\mathbf{p}}, \tag{3.2.10б}$$

где d — невырожденная диагональная матрица из положительных элементов. Тогда (3.2.10а) можно переписать:

$$(I - A)\,d\tilde{\mathbf{p}} = [d]^{-1}X'\mathbf{y} \geqslant 0 \tag{3.2.10в}$$

или

$$(I - A)\,\tilde{\mathbf{w}} = \mathbf{c} \geqslant 0. \tag{3.2.10г}$$

Если A — неотрицательная блочно-диагональная матрица, у которой $0 \leqslant a_{ij} \leqslant 1$, $\sum_{i=1}^{r} a_{ij} < 1$, $\sum_{j=1}^{r} a_{ij} < 1$ для всех i и j, то справедлива следующая теорема.

Теорема 3. $(I - A)\,\tilde{\mathbf{w}} = B\,\tilde{\mathbf{w}} = \mathbf{c} \geqslant 0$ имеет неотрицательное решение $\tilde{\mathbf{w}}$ тогда и только тогда, когда

$$b_{11} > 0;\ \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} > 0, \ldots, \begin{vmatrix} b_{11} \ldots b_{1n} \\ \vdots \quad\ \vdots \\ b_{n1} \ldots b_{nn} \end{vmatrix} > 0. \tag{3.2.32}$$

Это так называемые условия Хоукинса — Саймона [46], которые эквивалентны утверждению того, что вектор с должен содержаться в выпуклом конусе, натянутом на вектор-столбцы матрицы B или $(I - A)$. В силу того что $X'X$ — положительно-определенная матрица с неотрицательными элементами, для нее условия (3.2.32) выполнены, однако требования, накладываемые в теореме 3 на элементы матрицы A, не выполняются. Отсюда следует, что в общем случае оценки по методу наименьших квадратов без учета ограничений могут получиться отрицательными или больше единицы.

в) Некоторые свойства оценок $\tilde{\mathbf{p}}$. Как показано в работе [72], u_j некоррелирована с X_j и, таким образом, обычная оценка переходных вероятностей по методу наименьших квадратов сходится по вероятности к $\mathbf{p}$ при $T \to \infty$. Кроме того, в [72] показано, что при фиксированном T оценка $\tilde{\mathbf{p}}$ сходится по вероятности к $\mathbf{p}$ при $n \to \infty$, т. е. доказана состоятельность оценки[1].

---

[1] Для того чтобы исследовать некоторые свойства выборочных данных и показать несмещенность оценки $\tilde{\mathbf{p}}$, введем $q_l(t)$ — вероятность пребывания микрообъекта в состоянии l в момент времени t. Пусть генеральная совокупность достаточно велика (или рассматривается выборка с возвращением), чтобы вероятность обнаружить микрообъект в состоянии l не зависела от наблюдений за состоянием других микрообъектов и была равна $q_l(t)$. Пусть $y_l(t)$ — относительная частота состояния l в момент времени t, а $N(t)$ — объем выборки. Тогда $N(t)\, y_l(t)$ равно числу микрообъектов выборки, которые в момент времени t находятся в состоянии l, а $N(t)\,(1 - y_l(t))$ — числу микрообъектов, которые в момент времени t находятся в других состояниях. Математическое ожидание $y_l(t)$ по всем возможным выборкам равно $q_l(t)$, таким образом, оценка будет несмещенной. Если $\bar{X}$ и $\bar{\mathbf{y}}$ — соответствующие данные о значениях относительных частот во все моменты наблюдений по генеральной совокупности, а X_s и $\mathbf{y}_s$ — такие же данные об относительных частотах, но по выборке объема $N(t)$, то оценки с применением метода наименьших квадратов по генеральной совокупности и по выборке запишутся как

$$\mathbf{p} = (\bar{X}'\,\bar{X})^{-1}\,\bar{X}'\,\bar{\mathbf{y}}$$

и

$$\tilde{\mathbf{p}}_s = (X_s'\,X_s)^{-1}\,X_s'\,\mathbf{y}_s.$$

Поскольку

$$E(y_l(t)) = q_l(t),$$

то

$$E(X_s) = \bar{X} \quad \text{и} \quad E(\mathbf{y}_s) = \bar{\mathbf{y}}.$$

Однако $\tilde{\mathbf{p}}_s$ не будет несмещенной оценкой $\mathbf{p}$, так как

$$E(\tilde{\mathbf{p}}_s) = E\,[(X_s'\,X_s)^{-1}\,X_s'\,\mathbf{y}_s] \neq (EX_s'\,EX_s)^{-1}\,EX_s'\,E\mathbf{y}_s = (\bar{X}'\,\bar{X})^{-1}\,\bar{X}'\,\bar{\mathbf{y}} = \mathbf{p}$$

в силу взаимозависимости X_s, X_s' и $\mathbf{y}_s$. В то же время $y_l(t)$ есть состоятельная оценка $q_l(t)$, так как

$$\operatorname{plim} y_l(t) = \operatorname{plim} n_l(t)/N(t) = q_l(t).$$

Следовательно,

$$\operatorname{plim} X_s = \bar{X} \quad \text{и} \quad \operatorname{plim} \mathbf{y}_s = \bar{\mathbf{y}}.$$

(Здесь символ plim означает переход к пределу по вероятности, см., например, [7*, 9*, 21*]. — *Прим. пер.*)

Однако поскольку оценивание ведется по относительным частотам, имеет место гетероскедастичность. Иными словами, ковариационная матрица ошибок (3.2.4) не равна положительному скаляру, умноженному на единичную матрицу. А так как обычная оценка по методу наименьших квадратов не учитывает гетероскедастичности, то она не эффективна[1].

3.3. ОЦЕНИВАНИЕ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ ПРИ УЧЕТЕ ОГРАНИЧЕНИЙ

Из теорем 1 и 3 следует, что для обычных оценок по методу наименьших квадратов выполняется условие нормировки (3.2.13), однако условие неотрицательности оценок (3.2.15) может быть нарушено, что приведет к неприемлемости оценок переходных вероятностей. Эту проблему рассматривает Гудман в работе [41]. Он пишет (с. 247): «Задача состоит в том, чтобы оценить $(a \times a)$ параметров, удовлетворяющих линейным ограничениям. Мы будем заниматься минимизацией $\Sigma\, c_i c_i$ с тем, чтобы получить оценку, обеспечивающую наименьшее значение суммы средних квадратов. Другими словами, будут оцениваться параметры, представляющие собой точку a $(a - 1)$-мерного пространства. Поскольку $t_{ij} \geqslant 0$, значения параметров будут лежать в подмножестве этого пространства. Если потребовать, чтобы значения оценок $\hat{T}$ принадлежали этому же подмножеству, то такой подход оказывается слишком сложным. Сначала нужно получить обычную оценку по методу наименьших квадратов $\hat{T}$. Если $\hat{T}$ принадлежит требуемому подмножеству, то она и служит оценкой параметров T. Если $\hat{T}$ не принадлежит множеству допустимых значений, то искомая оценка будет лежать на границе множества. Тогда нужно будет пользоваться оценкой, которая принадлежит границе множества допустимых значений и минимизирует $\Sigma c_i c_i \ldots$. (В наших обозначениях $a = r$, $t_{ij} = p_{ij}$, $T = P$, $\Sigma\, c_i c_i = \mathbf{u}'_j \mathbf{u}_j$.)

Таким образом, в [41] было сделано признание, что обычные оценки по методу наименьших квадратов могут не удовлетворять условию $0 \leqslant p_{ij} \leqslant 1$. В этом случае предложено пользоваться оценками, которые принадлежат границе множества допустимых значений параметров и на множестве точек границы доставляют минимальное значение квадратичной форме (3.2.9). Идея, сформулированная в работах [41] и [79], была реализована в работе [110], в которой предлагается эвристическая процедура корректировки значений оценок переходных вероятностей, если они не принадлежат интервалу [0, 1]. Основываясь на [110], авторы работ [68] и [114] построили оценку, которая удовлетворяет условиям Гудмана. В этом случае отыскивается оценка, минимизирующая

$$\mathbf{u}'\mathbf{u} = (\mathbf{y} - X\mathbf{p})' \ (\mathbf{y} - X\mathbf{p}) \qquad (3.3.1)$$

[1] Проблема гетероскедастичности применительно к оцениванию переходных вероятностей обсуждается в [72], описание методов оценивания по данным об относительных частотах можно найти в [123].

при ограничениях

$$Gp = \eta_r, \tag{3.3.2}$$

$$p \geqslant 0. \tag{3.3.3}$$

Поскольку (3.3.1) — квадратичная форма, а ограничения линейны, эта задача относится к типичным задачам квадратичного программирования. Следуя [55] и опираясь на теорему Куна — Таккера [66] для задач нелинейного программирования и на принцип двойственности Дорна [30] для квадратичного программирования, можно свести задачу к следующей линейной постановке: найти $\widetilde{p}^c$, максимизирующее

$$(X'y - X'X\widetilde{p}^c)'\,p \tag{3.3.4}$$

при ограничениях

$$Gp \leqslant \eta_r, \tag{3.3.5}$$

$$-Gp \leqslant -\eta_r \tag{3.3.6}$$

и

$$p \geqslant 0, \tag{3.3.7}$$

где $\widetilde{p}^c$ — оптимальная оценка по методу наименьших квадратов при учете ограничений. В соответствии с принципом двойственности для задач линейного программирования [27] двойственная задача состоит в том, чтобы минимизировать

$$[\lambda_1' \; \lambda_2'] \begin{bmatrix} \eta_r \\ -\eta_r \end{bmatrix} \tag{3.3.8}$$

при ограничениях

$$[G' \; -G'] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \geqslant X'y - X'X\widetilde{p}^c \tag{3.3.9}$$

и

$$\lambda_1, \lambda_2 \geqslant 0, \tag{3.3.10}$$

где λ_1 и λ_2 — $(r \times 1)$-векторы двойственных переменных.

Чтобы построить алгоритм решения, заменим в прямой и двойственной задачах $\widetilde{p}^c$ на p и запишем задачу в следующей комбинированной формулировке, которая позволяет одновременно получить решение прямой и двойственной задач: максимизировать

$$(X'y - X'Xp)'\,p - \lambda_1'\eta_r + \lambda_2'\eta_r = -\lambda_1'\alpha_1 - \lambda_2'\alpha_2 - \beta'p \leqslant 0 \tag{3.3.11}$$

при ограничениях

$$Gp = \eta_r, \tag{3.3.12}$$

$$G'\lambda_1 - G'\lambda_2 + (X'X)\,p - \beta = X'y, \tag{3.3.13}$$

$$p,\ \lambda_1,\ \lambda_2,\ \alpha_1,\ \alpha_2,\ \beta \geqslant 0, \tag{3.3.14}$$

где α_1, α_2 и β — векторы дополнительных переменных прямой и двойственной задач.

Задача (3.3.11—14) может быть легко решена с помощью стандартного симплекс-метода для задач квадратичного программирования [119]. Соответствующие параметры задачи и алгоритма приводятся в табл. 3.1.

Таблица 3.1

Симплекс-таблица задачи квадратичного программирования для классической оценки по методу наименьших квадратов при учете ограничений

B_0	λ_1	λ_2	p	α_1	α_2	β
η_r			G	I		
$-\eta_r$			$-G$		I	
$X'y$	G'	$-G'$	$X'X$			$-I$

а) Выборочные свойства оценки, учитывающей ограничения. Выборочные свойства таких оценок при ограничениях, заданных в форме равенства (3.3.2), исследованы в работе [111]. Если имеются ограничения в форме неравенства (1.1.5) или (3.2.15) и для отыскания оценок решается задача квадратичного программирования, то можно рекомендовать работу [120], в которой изучены выборочные свойства таких оценок при одном параметре. В ней рассматривается случай, когда предполагается, что вектор u имеет многомерное нормальное распределение с нулевым средним и ковариацией $\sigma^2 I$. Как хорошо известно, обычная оценка параметра β по методу наименьших квадратов без учета ограничений распределена нормально со средним β и дисперсией $\sigma^2 (X'X)^{-1}$.

Как отмечено в [120], в отличие от обычных оценок оценки, полученные при учете ограничений, имеют усеченное нормальное распределение[1] с дискретными компонентами на концах интервала. При этом непрерывная часть распределения в случае одной переменной имеет форму усеченного закона нормального распределения внутри интервала [0, 1]. Для простейших задач регрессии можно подсчитать моменты распределения оценок, построенных при учете ограничений в форме

[1] Усеченный закон нормального распределения рассмотрен в работе [23].

неравенств. Однако для многомерных задач аналитическое исследование моментов и выборочных характеристик таких оценок становится затруднительным. Трудность состоит в том, что влияние того или иного ограничения на вид оценок зависит от конкретной выборки и должно рассматриваться как случайный вектор. Таким образом, оценка по методу наименьших квадратов при учете ограничений в форме неравенств для стандартной линейной задачи представляет собой сумму произведений случайных величин (набора активных ограничений и y), а не линейную комбинацию случайных величин, как это было в случае обычных оценок или при учете ограничений в форме равенств. Некоторые свойства оценок, учитывающих ограничения в форме неравенств, рассматривались в [45], точное распределение найдено в [48]. Однако, как отмечено в [48], сложность записи этого распределения ограничивает его применение, поскольку в общем случае достаточно сложно определить моменты распределения оценок.

Глава 4 ● ВЫБОРОЧНЫЙ ЭКСПЕРИМЕНТ. ПЕРВЫЕ РЕЗУЛЬТАТЫ

Трудности определения выборочных свойств оценок переходных вероятностей по методу наименьших квадратов при учете ограничений в форме неравенств и то обстоятельство, что для оценок других типов имеются лишь асимптотические результаты, определяют содержание этой главы. В ней описывается вероятностная модель для анализа свойств различных оценок по методу Монте-Карло при ограниченном объеме выборки, рассматривается схема, позволяющая оценивать качество различных оценок, приводятся результаты выборочных экспериментов для микро- и макрооценок, представленных в главах 2 и 3.

При интерпретации результатов этой и последующих глав хотелось бы иметь полную ясность относительно ограничений, связанных с выборочным методом экспериментального исследования оценок. В [115] и других работах отмечено, что выводы относительно свойств оценок могут зависеть от выбранных для эксперимента значений параметров. Отсюда следует, что вместо сопоставления свойств оценок в одной или нескольких изолированных точках необходимо ввести в пространстве параметров некоторую функцию риска. К сожалению, результаты, которые приводятся в этой и следующих главах, относятся к выборочному эксперименту только для одной точки пространства параметров, поэтому возможности их обобщения весьма ограничены. Тем не менее эти результаты все же обнаруживают различие в свойствах оценок в рассматриваемой точке и дают представление о скорости сходимости при увеличении объема выборки. К тому же хорошо известно, что для заданной функции потерь и априорного распределения байесовская оценка, если она существует, минимизирует средний риск. Следует также оговорить, что основой при анализе эффективности различных оценок служили «повторяемые выборочные эксперименты». Естественно, можно сомневаться в том, насколько это разумно применительно к общественным дисциплинам, в рамках которых повторная выборка обычно недоступна.

4.1. ИМИТАЦИОННАЯ МОДЕЛЬ И ЕЕ СВОЙСТВА

Для выборочных экспериментов используется однородная марковская цепь первого порядка со следующей матрицей переходных вероятностей:

$$P = [p_{ij}] = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} 0{,}60 & 0{,}40 & 0{,}00 & 0{,}00 \\ 0{,}10 & 0{,}50 & 0{,}40 & 0{,}00 \\ 0{,}00 & 0{,}10 & 0{,}70 & 0{,}20 \\ 0{,}00 & 0{,}00 & 0{,}10 & 0{,}90 \end{bmatrix} \end{array}. \qquad (4.1.1)$$

Траектории движения, определяемого такой матрицей вероятностей перехода, обнаруживают ярко выраженную тенденцию микрообъектов оставаться в прежних состояниях, а при изменении состояния — переходить только в соседние. Представляется, что такая модель хорошо согласуется со многими процессами, которые наблюдаются в экономике. Вектор стационарных вероятностей, соответствующих матрице (4.1.1), имеет вид

$$\pi = (0{,}0189 \;\; 0{,}0755 \;\; 0{,}3020 \;\; 0{,}6036). \qquad (4.1.2)$$

Это означает, что по мере увеличения t каждая строка матрицы P^t стремится к (4.1.2), или в общем виде

$$P^t \to \eta_r \pi \quad \text{при} \quad t \to \infty, \qquad (4.1.3)$$

где η_r — вектор-столбец, все r элементов которого равны единице. Если процесс с матрицей (4.1.1) начинается из первого состояния, то он приближается к вектору стационарных вероятностей примерно за 40 шагов.

Для того чтобы полностью определить марковскую цепь при фиксированной матрице вероятностей переходов (4.1.1), нужно задать начальное распределение вероятностей. Если обозначить через $\mathbf{y}(0)$ $(1 \times r)$-вектор начальных значений частот, а через $\mathbf{y}(t)$ $(1 \times r)$-вектор вероятностей (агрегированных частот) в момент времени t, то по определению марковской цепи

$$\mathbf{y}(t) = \mathbf{y}(0)\, P^t. \qquad (4.1.4)$$

Очевидно, что разные начальные векторы $\mathbf{y}(0)$ приводят к разным траекториям агрегированных данных $\mathbf{y}(t)$. В модели (4.1.1) имеются четыре разных начальных состояния микрообъектов. Траектории изменения агрегированных данных, соответствующих каждому из этих начальных состояний, показаны на рис. 4.1. При выходе из состояния 1 обнаруживаются сильные изменения траекторий на начальном участке с последующим экспоненциальным приближением к положению равновесия.

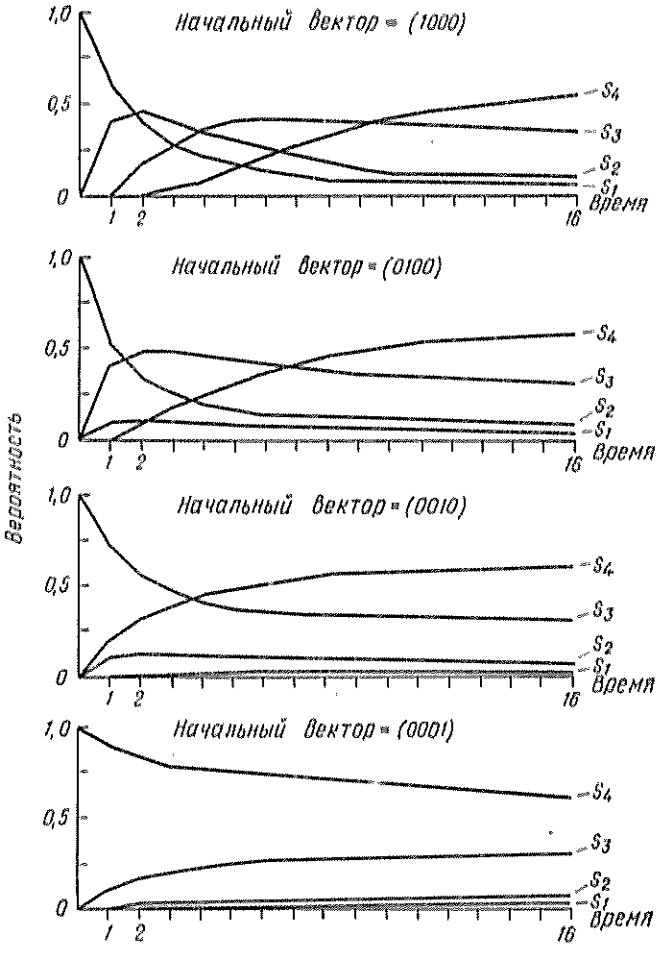

Рис. 4.1. Траектория агрегированных данных для разных начальных состояний

4.2. ПРОЦЕДУРА МОДЕЛИРОВАНИЯ

Модель (4.1.1) применялась для исследования свойств оценок переходных вероятностей по методу наименьших квадратов при учете ограничений. Для этого на машине IBM-7094 моделировалась система из 1000 микрообъектов, которые в начальный момент времени находились в состоянии 1. В соответствии с матрицей (4.1.1) рассматривались случайные изменения 1000 чисел, соответствующих состояниям этих микрообъектов на интервале в 24 временных такта. Так, например, если известно, что микрообъект k, $k = 1, 2, \ldots, 1000$, на t-м шаге находится в состоянии i, где $i = 1, 2, 3$ или 4, и если t-е значение псевдослучайного числа $< 10\ p_{i1}$, то к следующему шагу k-й микрообъект попадает в состояние 1; если псевдослучайное число оказалось заклю-

ченным между числами 10 p_{i1} и 10 (p_{i1} + p_{i2}), то следующим состоянием микрообъекта будет состояние 2, и т. д. Процедура повторяется для каждого из 1000 микрообъектов. Таким образом, в результате генерируется (1000 × 24)-матрица данных, которая показывает, в каком состоянии находится каждый из микрообъектов в каждый момент времени.

4.3. МОДЕЛИРУЕМАЯ СОВОКУПНОСТЬ И ВЫБОРКА

Полученные агрегированные частоты для 1000 микрообъектов на интервале в 20 шагов приведены в табл. 4.1.

Из совокупности в 1000 микрообъектов производится отбор 25, 50 и 100 микрообъектов. Для каждой выборки проводятся 50 повторных экспериментов. В качестве базисного временного интервала берутся отрезки с 1-го по 16-й шаг или со 2-го по 14-й. Таким образом, для каждой выборки получаются 50 оценок и сравниваются средние значения и дисперсии этих оценок. Поскольку длительность интервала времени и число повторений эксперимента фиксированы, единственным переменным фактором будет объем выборки[1], и интересно выяснить, как он влияет на свойства оценок.

Таблица 4.1

Агрегированные частоты при моделировании системы из 1000 микрообъектов

Номер шага	S_1	S_2	S_3	S_4
0	1,000	0,000	0,000	0,000
1	0,600	0,400	0,000	0,000
2	0,406	0,428	0,166	0,000
3	0,289	0,385	0,296	0,030
4	0,206	0,325	0,393	0,076
5	0,152	0,290	0,413	0,145
6	0,115	0,262	0,403	0,220
7	0,098	0,217	0,399	0,286
8	0,088	0,170	0,399	0,343
9	0,063	0,175	0,375	0,387
10	0,059	0,143	0,366	0,432
11	0,048	0,118	0,366	0,468
12	0,040	0,113	0,355	0,492
13	0,042	0,104	0,336	0,518
14	0,033	0,105	0,333	0,529
15	0,031	0,104	0,315	0,550
16	0,032	0,103	0,293	0,572
17	0,025	0,091	0,311	0,573
18	0,025	0,094	0,304	0,577
19	0,029	0,083	0,298	0,590
20	0,027	0,087	0,295	0,591

[1] Еще раз подчеркнем, что объем выборки относится к числу микрообъектов, а не к длительности интервала и числу повторений эксперимента.

4.4. ВЫБОРОЧНЫЕ ОТНОСИТЕЛЬНЫЕ ЧАСТОТЫ КАК ОЦЕНКИ ИСТИННЫХ ОТНОСИТЕЛЬНЫХ ЧАСТОТ

После отбора микрообъектов случайные траектории их движения агрегируются с тем, чтобы получить значения относительных частот в каждом состоянии в любой из моментов времени. Эти частоты, в дальнейшем называемые наблюдаемыми относительными частотами $y_j(t)$, $j = 1, 2, \ldots, r$, представляют собой оценки истинных частот $q_j(t)$, $j = 1, 2, \ldots, r$. Иными словами, предполагается, что при заданных $y_i(t-1)$, $i = 1, 2, \ldots, r$,

$$q_j(t) = \sum_i y_i(t-1)\, p_{ij}, \qquad i = 1, 2, \ldots, r; \qquad t = 1, 2, \ldots, T. \tag{4.4.1}$$

В этом разделе будет показано, что наблюдаемые частоты являются несмещенными оценками истинных частот, т. е. величина ошибки

$$u_j(t) = y_j(t) - q_j(t) \tag{4.4.2}$$

обладает свойством, заключающимся в том, что $Eu_j(t) = 0$.

Рассмотрим случайную выборку из совокупности, каждый элемент которой может находиться в состоянии i или в состоянии, отличном от состояния i. Если объем выборки $N(t)$ и $y_i(t)$ — величина относительной частоты микрообъектов выборки в i-м состоянии, то $N(t)\, y_i(t)$ — число микрообъектов выборки, которые находятся в состоянии i, а $N(t)\,(1 - y_i(t))$ — число микрообъектов выборки, которые находятся в состояниях, отличных от состояния i. Предполагается, что генеральная совокупность достаточно велика или выборка проводится с возвращением, поэтому вероятность $q_i(t)$ обнаружить микрообъект в состоянии i не зависит от предыдущих извлечений. Вероятность обнаружить $N(t)\, y_i(t)$ микрообъектов в состоянии i и $(1 - y_i(t))\, N(t)$ — в состоянии «не i» в момент времени t равна[1]

$$\binom{N(t)}{N(t)\, y_i(t)}\, q_i(t)^{N(t)\, y_i(t)}\,(1 - q_i(t))^{N(t)\,(1 - y_i(t))}. \tag{4.4.3}$$

Тогда функция правдоподобия запишется следующим образом:

$$L = K \prod_t \prod_i q_i(t)^{N(t)\, y_i(t)}\,(1 - q_i(t))^{N(t)\,(1 - y_i(t))}, \tag{4.4.4}$$

---

[1] Если число состояний больше двух, то биномиальное распределение получается из полиномиального в результате объединения всех состояний, отличных от состояния i, в состояние «не i».

где K — константа. Для получения оценки максимального правдоподобия $q_i(t)$ можно максимизировать $\log L$ по $q_i(t)$. В результате максимизации (4.4.4) мы приходим к следующему необходимому условию

$$\frac{N(t)\,y_i(t)}{q_i(t)} - \frac{N(t)\,(1-y_i(t))}{1-q_i(t)} = 0 \qquad (4.4.5)$$

или после преобразований

$$\hat{q}_i(t) = y_i(t). \qquad (4.4.6)$$

Таким образом, в каждый момент времени наблюдаемые частоты будут оценками максимального правдоподобия истинных частот. Несмещенность оценок максимального правдоподобия показывается непосредственно. Можно воспользоваться следующим известным приемом. Прежде всего справедлива формула

$$E(y_i(t)) = (1/N(t))\,E(N(t)\,y_i(t)) = (1/N(t))\,E(n_i(t)), \qquad (4.4.7)$$

где $n_i(t) = N(t)\,y_i(t)$. По определению

$$E(N(t)\,y_i(t)) = \sum_{n_i(t)} n_i(t) \binom{N(t)}{n_i(t)} q_i(t)^{n_i(t)} (1-q_i(t))^{N(t)-n_i(t)} \qquad (4.4.8)$$

и сумма в правой части равна $N(t)\,q_i(t)$. Объединяя этот результат и формулу (4.4.7), имеем

$$E(y_i(t)) = q_i(t). \qquad (4.4.9)$$

Следовательно, наблюдаемая частота и есть несмещенная оценка. Кроме того, как известно, эта оценка будет состоятельной, т. е.

$$\operatorname{plim} y_i(t) = \operatorname{plim} n_i(t)/N(t) = q_i(t). \qquad (4.4.10)$$

Дисперсии выборочных значений частот равны

$$q_i(t)\,(1 - q_i(t)/N(t)),\ i = 1, 2, \ldots, r;\quad t = 1, 2, \ldots, T, \qquad (4.4.11)$$

а ковариации

$$-q_i(t)\,q_j(t)/N(t),\ i \neq j;\ t = 1, 2, \ldots, T. \qquad (4.4.12)$$

Состоятельные оценки этих дисперсий и ковариаций можно записать как

$$y_i(t)\,(1 - y_i(t))/N(t),\ i = 1, 2, \ldots, r;\ t = 1, 2, \ldots, T \qquad (4.4.13)$$

и

$$-y_i(t)\,y_j(t)/N(t),\ i \neq j;\ t = 1, 2, \ldots, T. \qquad (4.4.14)$$

Эти оценки дисперсий и ковариаций будут рассматриваться в следующих главах в связи с оценками по обобщенному методу наименьших квадратов, оценками по критерию χ^2 и оценками максимального правдоподобия.

4.5. КРИТЕРИИ ДЛЯ ИССЛЕДОВАНИЯ КАЧЕСТВА ОЦЕНОК

Одним из показателей, определяющих качество оценок, является среднеквадратическая ошибка (СКО) $\left(\sum_{k=1}^{N}(\hat{p}_{ijk}-p_{ij})^2/N\right)^{1/2}$, где $\hat{p}_{ijk}$ — k-я выборочная оценка, p_{ij} — истинное значение, а N — число выборочных оценок. В соответствии с этим критерием потери, связанные с отличием оценки от истинного значения исследуемого параметра, пропорциональны корню квадратному из средней величины квадратов отклонений. С точки зрения априорной информации критерий СКО как мера разброса мало отличается от среднего абсолютного отклонения (САО). Одно из преимуществ критерия СКО в простоте его аналитической записи как функции смещения и дисперсии оценок.

С другой стороны, величина абсолютного отклонения оценок от истинных значений переходных вероятностей также использовалась при анализе качества различных оценок. На этом показателе основаны непараметрические критерии попарного сравнения и ранговые критерии обработки данных [100]. В работе [62] степень ранговой корреляции измеряется с помощью коэффициента согласия Кендэла W.

Значимость величины разброса оценок относительно истинных значений параметров проверялась по критерию совпадения средних абсолютных отклонений. В этом случае проверяемая гипотеза состоит в том, что

$$\Pr(|\hat{p}_{ij}-p_{ij}|<|\tilde{p}_{ij}-p_{ij}|)=\frac{1}{2}, \qquad (4.5.1)$$

где $\hat{p}_{ij}$ и $\tilde{p}_{ij}$ — разные оценки неизвестного параметра, истинное значение которого равно p_{ij}. При проверке этого критерия одновременно проверялась гипотеза биномиальности распределения.

Кроме того, использовался также знако-ранговый критерий попарного сравнения Уилкоксона. Для этого критерия ранги отклонений выступают в качестве весовых коэффициентов, а процедура состоит в том, что для 50 пар оценок вычисляются величины

$$d_k=|\hat{p}_{ijk}-p_{ij}|-|\tilde{p}_{ijk}-p_{ij}|, \quad k=1,\ 2,\ldots,\ 50, \qquad (4.5.2)$$

которые ранжируются безотносительно к их знакам. Если $d_k=0$, то это отклонение не ранжируется. Всем ненулевым значениям d_k приписываются ранги: ранг 1 — наименьшему по модулю, ранг 2 — следующему и т. д. Если два или большее число d_k принимают одинаковые значения, то им приписывается среднее значение ранга. После

этого каждому рангу приписывается знак соответствующего d_h и определяется сумма рангов T, которой присваивается знак, противоположный преобладающему знаку в совокупности d_h. Получающаяся статистика T имеет распределение, близкое к нормальному закону со средним $\mu_T = (N(N+1))/4$ и стандартным отклонением $\sigma_T = \sqrt{[(N(N+1)(2N+1))/24]}$. Таким образом, можно определить соответствующую нормализованную величину $z = (T - \mu_T)/\sigma_T$ и сравнить ее с табличными данными для стандартного нормального распределения с тем, чтобы принять или отвергнуть гипотезу о совпадении средних значений по выборке двух разных оценок.

4.6. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ ДЛЯ МИКРООЦЕНОК МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

В нашем распоряжении имелись полные последовательности данных для 1000 микрообъектов, поэтому для получения основания для сравнений по этим данным были вычислены оценки максимального правдоподобия всех переходных вероятностей. Как уже говорилось и как показано в работе [61], микрооценки максимального правдоподобия состоятельны, но в общем случае они могут быть смещенными. Однако по мере увеличения объема выборки величина смещения стремится к нулю, а оценка имеет асимптотически нормальное распределение.

Если данные, описывающие траектории движения 1000 микрообъектов, рассматривать как выборку из бесконечной генеральной совокупности, то соответствующие выборочные оценки переходных вероятностей p_{ij} по методу максимального правдоподобия на интервале $t = 1, 2, \ldots, 16$ равны:

$$\dot{P} = [\dot{p}_{ij}] = \left[n_{ij} \Big/ \sum_j n_{ij}\right] = \begin{bmatrix} 0{,}6044 & 0{,}3956 & 0{,}0000 & 0{,}0000 \\ 0{,}0988 & 0{,}4930 & 0{,}4082 & 0{,}0000 \\ 0{,}0000 & 0{,}1013 & 0{,}6948 & 0{,}2039 \\ 0{,}0000 & 0{,}0000 & 0{,}0961 & 0{,}9039 \end{bmatrix}. \tag{4.6.1}$$

Как и следовало ожидать, эти оценки находятся в хорошем соответствии с истинными значениями вероятностей перехода (4.1.1), с помощью которых моделировалась эволюция микрообъектов. Проверка гипотезы стационарности процесса по критерию χ^2 дает значение критерия, равное 57,92, что гораздо ниже пороговой величины 191,75, которая соответствует 10%-ному уровню значимости при 168 степенях свободы[1].

[1] В этом случае критерий χ^2 записывается как

$$\chi^2_{r(r-1)(t-1)} = \sum_i \sum_t \sum_j n_i(t-1)(\dot{p}_{ij}(t) - \dot{p}_{ij})^2 / \dot{p}_{ij},$$

где $\dot{p}_{ij}(t)$ — оценка переходной вероятности по методу максимального правдоподобия в момент времени t, $n_i(t-1)$ — число микрообъектов, которые в момент времени $t-1$ находятся в состоянии i, а $\dot{p}_{ij}$ — элементы матрицы (4.6.1).

Применение критерия χ^2 для проверки гипотезы состоятельности оценок (4.6.1) дает значение критерия, равное 2,52, что меньше величины 18,55, которая равна значению 10%-ного уровня значимости при 12 степенях свободы, и, таким образом, гипотеза не отвергается[1].

Затем для рассматриваемой системы были найдены значения оценок максимального правдоподобия при 50-кратном повторении моделирования и разных объемах выборки. В табл. 4.2. приведены экспериментальные значения этих оценок для выборок объема 25, 50 и 100 при 50-кратном повторении эксперимента на интервале $t = 1, 2, \ldots, 16$.

Таблица 4.2

Средние значения и среднеквадратические ошибки для оценок максимального правдоподобия

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}5829 & 0{,}4171 & 0{,}0000 & 0{,}0000 \\ 0{,}0949 & 0{,}4998 & 0{,}4053 & 0{,}0000 \\ 0{,}0000 & 0{,}1038 & 0{,}6950 & 0{,}2012 \\ 0{,}0000 & 0{,}0000 & 0{,}1055 & 0{,}8945 \end{bmatrix}$	$\begin{bmatrix} 0{,}0529 & 0{,}0529 & 0{,}0000 & 0{,}0000 \\ 0{,}0283 & 0{,}0656 & 0{,}0458 & 0{,}0000 \\ 0{,}0000 & 0{,}0245 & 0{,}0332 & 0{,}0361 \\ 0{,}0000 & 0{,}0000 & 0{,}0316 & 0{,}0316 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5934 & 0{,}4066 & 0{,}0000 & 0{,}0000 \\ 0{,}0999 & 0{,}4934 & 0{,}4067 & 0{,}0000 \\ 0{,}0000 & 0{,}1034 & 0{,}6966 & 0{,}2000 \\ 0{,}0000 & 0{,}0000 & 0{,}1060 & 0{,}8940 \end{bmatrix}$	$\begin{bmatrix} 0{,}0361 & 0{,}0361 & 0{,}0000 & 0{,}0000 \\ 0{,}0265 & 0{,}0566 & 0{,}0374 & 0{,}0000 \\ 0{,}0000 & 0{,}0173 & 0{,}0300 & 0{,}0265 \\ 0{,}0000 & 0{,}0000 & 0{,}0245 & 0{,}0245 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}6077 & 0{,}3993 & 0{,}0000 & 0{,}0000 \\ 0{,}0981 & 0{,}4926 & 0{,}4093 & 0{,}0000 \\ 0{,}0000 & 0{,}1041 & 0{,}6975 & 0{,}1984 \\ 0{,}0000 & 0{,}0000 & 0{,}1019 & 0{,}8981 \end{bmatrix}$	$\begin{bmatrix} 0{,}0224 & 0{,}0224 & 0{,}0000 & 0{,}0000 \\ 0{,}0141 & 0{,}0490 & 0{,}0224 & 0{,}0000 \\ 0{,}0000 & 0{,}0141 & 0{,}0200 & 0{,}0200 \\ 0{,}0000 & 0{,}0000 & 0{,}0173 & 0{,}0173 \end{bmatrix}$

Как видно из этой таблицы, средние значения оценок для каждой из выборок близки к истинным значениям (4.6.1) в генеральной совокупности так же, как и к элементам порождающей процесс матрицы вероятностей перехода (4.1.1). Величина суммарного расхождения значений оценок относительно элементов матрицы (4.1.1), определяемая величиной суммы по всем элементам матрицы модулей отклонений оценок от истинных значений, равна 0,0658, 0,0486 и 0,0320 для выборок объема 25, 50 и 100 соответственно. Для выборок объема 25, 50 и 100 суммарная среднеквадратическая ошибка, определяемая суммой по всем элементам матрицы квадратов отклонений оценок от соответ-

[1] В этом случае критерий χ^2 записывается как

$$\chi^2_{r(r-1)} = \sum_i \sum_j \left(\sum_l n_l(l-1)\right)(\dot{p}_{ij} - p_{ij})^2 / p_{ij},$$

где $\dot{p}_{ij}$ и p_{ij} — элементы матриц (4.6.1) и (4.1.1) соответственно.

ствующих истинных значений приблизительно равна 0,0007, 0,0003 и 0,0002 соответственно, а сумма среднеквадратических ошибок всех оценок для трех разных выборок составляет 0,4025, 0,3155 и 0,2190. В общем случае, как и следовало ожидать, по мере увеличения объема выборки оценки стремятся к истинным значениям параметров.

По критерию согласия χ^2 проверялось соответствие выборочных распределений оценок отличных от нуля переходных вероятностей нормальному распределению вероятностей со средним и дисперсией, которые равны соответствующим среднему и дисперсии выборочной оценки. При построении выборочных распределений область определения функции распределения в соответствии со средним значением и дисперсией разбивалась на пять равных интервалов так, чтобы в каждый интервал попадало не менее пяти точек[1]. Соответствующие значения критерия χ^2 сведены в табл. 4.3.

Таблица 4.3

Значения критерия χ^2 при проверке гипотезы о нормальном распределении оценок переходных вероятностей по методу максимального правдоподобия

Элементы матрицы вероятностей перехода p_{ij}	Объем выборки		
	25	50	100
p_{11}	1,4	3,8	2,6
p_{12}	1,4	3,8	2,6
p_{21}	0,4	0,8	1,6
p_{22}	4,6	1,4	1,2
p_{23}	1,4	2,2	2,6
p_{32}	3,4	2,4	1,0
p_{33}	1,8	1,6	2,6
p_{34}	1,6	1,4	7,4
p_{43}	3,8	0,4	0,8
p_{44}	3,0	0,6	0,6

Ни одно из приведенных в таблице значений не является статистически значимым отклонением от нормального распределения при 10%-ном уровне значимости[2].

4.7. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ ПОСЛЕДОВАТЕЛЬНОСТИ ВЕКТОРОВ ВЕРОЯТНОСТЕЙ СОСТОЯНИЙ

Как и раньше, допустим, что в начальный момент времени все элементы системы находятся в одном из четырех марковских состояний, а для следующих моментов времени $t = 1, 2, \ldots, 20$ определяются векторы вероятностей состояний для каждого из начальных состояний (рис. 4.1). Если траектории, представленные на рисунке, рассматри-

[1] Для применимости критерия χ^2 по грубым оценкам в каждом интервале должно быть не менее 5 точек [6].

[2] Табличное значение критерия χ^2 при 4 степенях свободы для 10%-ного уровня значимости равно 7,78.

ваются как четыре выборочные последовательности данных о величине относительных частот и если выбрать $t = 2, 3, \ldots, 16$, то соответствующие оценки по методу наименьших квадратов без учета ограничений (3.2.10) близки к истинным значениям вероятностей перехода, задаваемым матрицей (4.1.1). Однако для выбранной матрицы (4.1.1) при выходе из состояния 1 элементы вектора вероятностей $y_i(t)$ обнаруживают сильные изменения в начальный период времени, а затем экспоненциально приближаются к стационарному распределению. Поэтому возникает вопрос о влиянии выбора длительности и расположения интервала наблюдения на значения оценок по методу наименьших квадратов без учета ограничений. Для того чтобы получить представление об интересующих нас свойствах оценок, для построения оценок были взяты разные интервалы наблюдений. Результаты приводятся в табл. 4.4.

Как видно из табл. 4.4, для интервала наблюдения $t = 2, 3, \ldots, 6$ формула (3.2.10) приводит к оценкам, которые весьма близки к истинным значениям параметров. Однако при наблюдениях на интервале $t = 8, 9, \ldots, 12$ оценки по методу наименьших квадратов без учета ограничений приводят к просчетам и заметному нарушению условий неотрицательности вероятностей перехода. Это связано с высокой степенью коллинеарности X-ов и тем, что матрица $X'X$ является почти вырожденной (это характерно для временных рядов, описывающих экономические наблюдения).

4.8. РЕЗУЛЬТАТЫ ВЫБОРОЧНОГО ЭКСПЕРИМЕНТА ДЛЯ МАКРОДАННЫХ

а) Метод наименьших квадратов без учета ограничений. По выборке, извлекаемой из генеральной совокупности 1000 микрообъектов, для получения оценок переходных вероятностей по методу наименьших квадратов без учета ограничений определялись значения относительных частот в моменты времени с 1-го по 16-й включительно. Во всех случаях нарушалось условие неотрицательности оценок переходных вероятностей. Средние значения и среднеквадратические ошибки оценок для разных выборок при 50-кратном повторении эксперимента представлены в табл. 4.5.

Для каждой выборки три или четыре раза нарушалось условие неотрицательности оценок. Кроме того, при сравнении средних значений и среднеквадратических ошибок эти оценки оказались гораздо хуже микрооценок максимального правдоподобия. Как и следовало ожидать, с увеличением объема выборки среднеквадратическая ошибка уменьшается. Если сложить все среднеквадратические ошибки оценок разных переходных вероятностей, то для выборок объема 25, 50 и 100 соответствующие агрегированные ошибки равны 2,49, 2,05 и 1,95.

Для экспериментального изучения распределения вероятностей оценок было извлечено 90 выборок объема 100, и соответствующие эмпирические распределения каждой из 16 оценок сравнивались по критерию χ^2 с нормальным распределением, обладающим теми же средним значением и дисперсией. Из шестнадцати оценок только одна

Таблица 4.4

Оценки переходных вероятностей по методу наименьших квадратов для разных интервалов наблюдения за марковской цепью

Момент времени	Оценки без учета ограничений	Оценки при учете ограничений
$t=2, 3, \ldots, 6$	$\begin{bmatrix} 0{,}6000 & 0{,}3996 & -0{,}0006 & 0{,}0010 \\ 0{,}1000 & 0{,}5007 & 0{,}4009 & -0{,}0016 \\ 0{,}0000 & 0{,}0993 & 0{,}6991 & 0{,}2016 \\ -0{,}0003 & 0{,}0013 & 0{,}1015 & 0{,}8995 \end{bmatrix}$	$\begin{bmatrix} 0{,}6001 & 0{,}3999 & 0{,}0000 & 0{,}0000 \\ 0{,}0999 & 0{,}5001 & 0{,}4000 & 0{,}0000 \\ 0{,}0000 & 0{,}0999 & 0{,}6999 & 0{,}2002 \\ 0{,}0000 & 0{,}0004 & 0{,}1003 & 0{,}8993 \end{bmatrix}$
$t=3, 4, \ldots, 7$	$\begin{bmatrix} 0{,}5926 & 0{,}4003 & 0{,}0020 & 0{,}0051 \\ 0{,}1084 & 0{,}4997 & 0{,}3978 & -0{,}0059 \\ -0{,}0047 & 0{,}1000 & 0{,}7011 & 0{,}2036 \\ 0{,}0038 & 0{,}0004 & 0{,}0993 & 0{,}8965 \end{bmatrix}$	$\begin{bmatrix} 0{,}6006 & 0{,}3998 & 0{,}0005 & 0{,}0000 \\ 0{,}0994 & 0{,}5013 & 0{,}3993 & 0{,}0000 \\ 0{,}0000 & 0{,}0094 & 0{,}7004 & 0{,}2002 \\ 0{,}0005 & 0{,}0007 & 0{,}0997 & 0{,}8992 \end{bmatrix}$
$t=5, 6, \ldots, 9$	$\begin{bmatrix} 0{,}5858 & 0{,}3751 & -0{,}0562 & 0{,}1060 \\ 0{,}1119 & 0{,}5203 & 0{,}4488 & -0{,}0906 \\ -0{,}0333 & 0{,}0953 & 0{,}6859 & 0{,}2250 \\ 0{,}0007 & 0{,}0001 & 0{,}1031 & 0{,}8954 \end{bmatrix}$	$\begin{bmatrix} 0{,}5990 & 0{,}3775 & 0{,}0199 & 0{,}0036 \\ 0{,}1005 & 0{,}5182 & 0{,}3813 & 0{,}0000 \\ 0{,}0000 & 0{,}0956 & 0{,}7065 & 0{,}1977 \\ 0{,}0000 & 0{,}0000 & 0{,}0979 & 0{,}9021 \end{bmatrix}$
$t=8, 9, \ldots, 12$	$\begin{bmatrix} 0{,}3179 & 0{,}6159 & -0{,}4783 & 0{,}2704 \\ 0{,}3226 & 0{,}3260 & 0{,}7974 & -0{,}2206 \\ -0{,}0539 & 0{,}1434 & 0{,}5953 & 0{,}2567 \\ 0{,}0078 & -0{,}0065 & 0{,}1178 & 0{,}8907 \end{bmatrix}$	$\begin{bmatrix} 0{,}5945 & 0{,}4055 & 0{,}0000 & 0{,}0000 \\ 0{,}1023 & 0{,}4963 & 0{,}4014 & 0{,}0000 \\ 0{,}0000 & 0{,}1004 & 0{,}6993 & 0{,}2002 \\ 0{,}0000 & 0{,}0002 & 0{,}1000 & 0{,}8998 \end{bmatrix}$

(p_{19}) обнаружила существенное отклонение от нормальности при 5%-ном уровне значимости.

Таблица 4.5

Средние значения и среднеквадратические ошибки для оценок по методу наименьших квадратов без учета ограничений

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,5165 & 0,5089 & 0,0180 & -0,0434 \\ 0,1504 & 0,3866 & 0,4466 & 0,0164 \\ -0,0104 & 0,1408 & 0,6067 & 0,2629 \\ -0,0001 & 0,0007 & 0,1586 & 0,8408 \end{bmatrix}$	$\begin{bmatrix} 0,1583 & 0,2228 & 0,2134 & 0,1407 \\ 0,1437 & 0,2170 & 0,2541 & 0,1608 \\ 0,1127 & 0,1272 & 0,1795 & 0,1265 \\ 0,0802 & 0,0973 & 0,1503 & 0,1099 \end{bmatrix}$
50	$\begin{bmatrix} 0,5666 & 0,4547 & -0,0344 & 0,0131 \\ 0,1471 & 0,4067 & 0,4972 & -0,0410 \\ -0,0412 & 0,1701 & 0,6127 & 0,2584 \\ 0,0257 & -0,0310 & 0,1506 & 0,8547 \end{bmatrix}$	$\begin{bmatrix} 0,1041 & 0,1656 & 0,1793 & 0,1112 \\ 0,1179 & 0,1758 & 0,2095 & 0,1480 \\ 0,0974 & 0,1457 & 0,1434 & 0,1034 \\ 0,0621 & 0,1007 & 0,1054 & 0,0842 \end{bmatrix}$
100	$\begin{bmatrix} 0,5480 & 0,4857 & -0,0218 & -0,0119 \\ 0,1724 & 0,2553 & 0,4719 & 0,0004 \\ -0,0350 & 0,1806 & 0,6501 & 0,2043 \\ 0,0144 & -0,0260 & 0,1114 & 0,9002 \end{bmatrix}$	$\begin{bmatrix} 0,1115 & 0,1526 & 0,1845 & 0,1268 \\ 0,1494 & 0,2017 & 0,2295 & 0,1615 \\ 0,0824 & 0,1235 & 0,1137 & 0,0858 \\ 0,0411 & 0,0688 & 0,0655 & 0,0513 \end{bmatrix}$

Для проверки гипотезы о сходимости этих оценок к истинным значениям был взят t-критерий со статистикой вида [53]:

$$\frac{\tilde{p}_{ij}-p_{ij}}{\sqrt{[(\mathrm{SSE}/(Tr-r^2))\,a_{ii}]}}, \qquad (4.8.1)$$

где SSE обозначает сумму квадратов ошибок и a_{ii} — это i-й диагональный элемент матрицы $(X'X)^{-1}$. Ни в одном из случаев критерий не показал значащих отклонений оценок по методу наименьших квадратов без учета ограничений от истинных значений.

б) Метод наименьших квадратов при учете ограничений. Средние значения и среднеквадратические ошибки оценок по методу наименьших квадратов при учете ограничений (см. параграф 3.3) для 50-кратного повторения эксперимента представлены в табл. 4.6.

Оказывается, что эти оценки гораздо лучше оценок переходных вероятностей без учета ограничений из табл. 4.5. Прежде всего оценки, полученные при учете ограничений, неотрицательны, а их среднеквадратические ошибки меньше среднеквадратических ошибок соответствующих оценок из п. а).

При исследовании распределения оценок по методу наименьших квадратов при учете ограничений вычисляется величина D-статистики

Колмогорова — Смирнова [70] с целью проверки критерия согласия гипотезы о законе нормального распределения. D-статистика определяется как

$$Dn = \sup_x |Fn(x) - F(x)|, \tag{4.8.2}$$

где $Fn(x)$ — эмпирическое распределение, которое записывается как

$$Fn(x) = j/N, \quad X(j) < x < X(j+1), \quad j = 0, 1, \ldots, N, \tag{4.8.3}$$

а $F(x)$ — функция распределения предполагаемого нормального закона.

Таблица 4.6

Средние значения и среднеквадратические ошибки для оценок по методу наименьших квадратов при учете ограничений

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}4992 & 0{,}4272 & 0{,}0723 & 0{,}0013 \\ 0{,}1315 & 0{,}4593 & 0{,}3943 & 0{,}0149 \\ 0{,}0160 & 0{,}0992 & 0{,}6344 & 0{,}2504 \\ 0{,}0051 & 0{,}0262 & 0{,}1361 & 0{,}8326 \end{bmatrix}$	$\begin{bmatrix} 0{,}1521 & 0{,}1513 & 0{,}1369 & 0{,}0058 \\ 0{,}1011 & 0{,}1603 & 0{,}1406 & 0{,}0291 \\ 0{,}0307 & 0{,}0649 & 0{,}1324 & 0{,}0921 \\ 0{,}0143 & 0{,}0407 & 0{,}1123 & 0{,}1149 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5680 & 0{,}3933 & 0{,}0386 & 0{,}0001 \\ 0{,}1080 & 0{,}4806 & 0{,}4020 & 0{,}0094 \\ 0{,}0069 & 0{,}1031 & 0{,}6562 & 0{,}2338 \\ 0{,}0039 & 0{,}0174 & 0{,}1247 & 0{,}8540 \end{bmatrix}$	$\begin{bmatrix} 0{,}0881 & 0{,}1159 & 0{,}0842 & 0{,}0000 \\ 0{,}0696 & 0{,}1155 & 0{,}0888 & 0{,}0192 \\ 0{,}0168 & 0{,}0560 & 0{,}0965 & 0{,}0684 \\ 0{,}0086 & 0{,}0364 & 0{,}0744 & 0{,}0807 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}5657 & 0{,}3953 & 0{,}0390 & 0{,}0000 \\ 0{,}1224 & 0{,}4754 & 0{,}3959 & 0{,}0063 \\ 0{,}0043 & 0{,}1060 & 0{,}6875 & 0{,}2022 \\ 0{,}0017 & 0{,}0113 & 0{,}0929 & 0{,}8941 \end{bmatrix}$	$\begin{bmatrix} 0{,}0868 & 0{,}1147 & 0{,}0840 & 0{,}0000 \\ 0{,}0712 & 0{,}1199 & 0{,}0912 & 0{,}0145 \\ 0{,}0099 & 0{,}0393 & 0{,}0629 & 0{,}0404 \\ 0{,}0049 & 0{,}0232 & 0{,}0459 & 0{,}0432 \end{bmatrix}$

При проверке было взято 50 оценок для выборки объема 50. Величина D-статистики оценок отличных от нуля параметров равна

$$\begin{bmatrix} 0{,}0857 & 0{,}0671 & \ldots & \ldots \\ 0{,}1102 & 0{,}0093 & 0{,}0880 & \ldots \\ \ldots & 0{,}0663 & 0{,}1336 & 0{,}0759 \\ \ldots & \ldots & 0{,}0842 & 0{,}1108 \end{bmatrix}, \tag{4.8.4}$$

и ни в одном случае она не превышает табличного значения 0,17 для выборки объема 50 при 10%-ном уровне значимости. Таким образом, гипотеза о нормальности не может быть отвергнута. Что касается оценок параметров, истинные значения которых равны нулю, то их распределения в нуле являются усеченными. Соответствующие показатели приводятся в табл. 4.7.

Таблица 4.7

Некоторые характеристики оценок нулевых параметров по методу наименьших квадратов при учете ограничений

Параметр	Число нулевых оценок	Число положительных оценок	Максимальная величина	Среднее значение	Стандартное отклонение
p_{13}	30	20	0,4109	0,0386	0,0748
p_{14}	49	1	0,0049	0,0001	0,0007
p_{24}	35	15	0,0583	0,0084	0,0167
p_{31}	38	12	0,0611	0,0069	0,0153
p_{41}	35	15	0,0302	0,0039	0,0076
p_{42}	33	17	0,1448	0,0174	0,0320

Как видно из табл. 4.7, по меньшей мере 60% оценок точно совпадают с истинным значением оцениваемого параметра. За исключением оценок вероятностей p_{13} и p_{42} максимальные значения оценок невелики. Максимальное значение оценки вероятности p_{13} неожиданно велико, однако следующее за ним по величине равно всего 0,1909. Что касается оценок p_{14}, то они почти идеальны и только одна из них отлична от нуля и равна 0,0049.

4.9. ПРИМЕР

В дополнение к модельным экспериментам рассмотрим в качестве примера практического приложения оценок переходных вероятностей реальные данные об изменении рыночной конъюнктуры при продаже табачных изделий.

а) Задача Телсера. В работе [110] Телсер исследовал данные о потреблении курильщиками различных табачных изделий. По каждому конкретному потребителю данных не было. Имелись только ежегодные данные об объеме продажи сигарет (в миллиардах штук) трех ведущих сортов (см. табл. 4.8).

На основании этой таблицы Телсер предположил, что доли продажи определяются по фиксированной выборке курильщиков, а поведение курильщиков можно описать однородной марковской цепью первого порядка.

б) Результаты для ограниченных и неограниченных оценок. Применение оценок по методу наименьших квадратов (3.2.10), полученных без учета ограничений, привело к следующим значениям оценок переходных вероятностей:

$$\tilde{P} = [\tilde{p}_{ij}] = \begin{bmatrix} 0{,}5454 & 0{,}3574 & 0{,}0972 \\ -0{,}0744 & 0{,}9654 & 0{,}1090 \\ 0{,}6489 & -0{,}3949 & 0{,}7460 \end{bmatrix}, \qquad (4.9.1)$$

Две оценки отрицательны и не могут быть переходными вероятностями. При учете требования неотрицательности ограниченные оценки получаются с помощью решения задачи квадратичного программирования из параграфа 3.3. Они равны:

$$\widetilde{P}^c = [\widetilde{p}^c_{ij}] = \begin{bmatrix} 0{,}6686 & 0{,}1423 & 0{,}1891 \\ 0{,}0000 & 0{,}8683 & 0{,}1317 \\ 0{,}4019 & 0{,}0000 & 0{,}5981 \end{bmatrix}. \qquad (4.9.2)$$

Таблица 4.8

Доли продажи сигарет и общий объем продажи

Год	«Camel»	«Lucky Strike»	«Chesterfield»	Всего
1925	0,5056	0,2028	0,2916	67,9
26	0,4879	0,1899	0,3222	77,9
27	0,4504	0,2236	0,3260	85,3
28	0,4068	0,3039	0,2893	90,2
29	0,3637	0,3616	0,2747	102,0
1930	0,3365	0,4118	0,2517	104,0
31	0,3311	0,4425	0,2264	100,4
32	0,2936	0,4498	0,2566	81,5
33	0,2794	0,4008	0,3198	91,6
34	0,3418	0,3301	0,3281	97,6
1935	0,3867	0,3013	0,3120	101,9
36	0,4074	0,2906	0,3020	113,9
37	0,4084	0,2949	0,2967	116,9
38	0,3842	0,3195	0,2963	113,8
39	0,3746	0,3358	0,2896	114,2
1940	0,3708	0,3500	0,2792	120,0
41	0,3579	0,2653	0,2768	135,5
42	0,3527	0,3851	0,2622	154,5
43	0,3276	0,3875	0,2849	175,5

Поведение, задаваемое матрицей (4.9.2), отражает значительную привязанность курильщиков к любимому сорту сигарет. Оказывается, что, сменяя 2-й и 3-й сорта сигарет, курильщики переходят только на один из двух оставшихся сортов. Так, курившие «Lucky Strike» и на следующий год никогда не переходят на «Camel», а курившие «Chesterfield» — на «Lucky Strike».

Глава 5 ● ОЦЕНКИ ПО МЕТОДУ НАИМЕНЬШИХ ВЗВЕШЕННЫХ КВАДРАТОВ С ОГРАНИЧЕНИЯМИ В ФОРМЕ НЕРАВЕНСТВ

В главе 3 при изложении метода наименьших квадратов с ограничениями мы пренебрегли особенностями ковариационной матрицы помех Σ (3.2.8) и предположили, что Σ есть скаляр, умноженный на единичную матрицу размера $(Tr \times Tr)$. В работе [72] показана неадекватность такого предположения. Для моделей, основанных на относительных частотах заполнения состояний, характерна гетероскедастичность, и поэтому оценки по методу наименьших квадратов, не учитывающие гетероскедастичность, оказываются неэффективными. В этой ситуации можно применить обобщенный метод наименьших квадратов Эйткена [2] для получения следующей линейной несмещенной оценки с минимальной дисперсией:

$$\tilde{\tilde{p}} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y. \qquad (5.0.1)$$

Следует, однако, отметить, что для рассматриваемой нами модели матрица Σ вырожденная, так что Σ^{-1} не существует. Единственный способ учесть гетероскедастичность в случае вырожденной Σ состоит во введении неособенной весовой матрицы размера $(Tr \times Tr)$, заменяющей Σ в (5.0.1)[1].

В этой главе для построения оценок берутся разные весовые матрицы, и свойства этих оценок с помощью выборочных экспериментов сравниваются со свойствами оценок по методу наименьших (невзвешенных) квадратов. Таким образом удается получить некоторую информацию о степени чувствительности характеристик оценок к изменениям способа взвешивания.

---

[1] Для решения таких вырожденных задач Чипманом [18], Рао [92] и другими был предложен метод с применением псевдообратных матриц. Однако в нашем случае вырождена не только ковариационная матрица помех, но линейно-зависимы и оцениваемые уравнения. Метод псевдообращения в применении к этой задаче обсуждается в приложении *А*.

5.1. СТАТИСТИЧЕСКАЯ МОДЕЛЬ

Как говорилось в главе 3, статистическая модель имеет вид

$$\mathbf{y} = X\mathbf{p} + \mathbf{u}, \tag{5.1.1}$$

где

$$E\mathbf{u} = 0, \tag{5.1.2}$$

$$E\mathbf{u}\mathbf{u}' = \Sigma. \tag{5.1.3}$$

Здесь $\mathbf{y}$ и $\mathbf{u}$ — $(rT \times 1)$-векторы-столбцы, X — блочно-диагональная $(rT \times r^2)$-матрица, $\mathbf{p}$ — $(r^2 \times 1)$-вектор-столбец и Σ — недиагональная вырожденная $(rT \times rT)$-матрица.

Переформулируем модель в рамках обобщенного метода наименьших квадратов [2], учитывая вес каждого наблюдения (или каждого периода наблюдений) в предположении гетероскедастичности. Умножая (5.1.1) на весовую матрицу H размера $(rT \times rT)$, получим

$$H\mathbf{y} = HX\mathbf{p} + H\mathbf{u}, \tag{5.1.4}$$

тогда оценка $\mathbf{p}$ по методу наименьших взвешенных квадратов примет вид

$$\tilde{\tilde{\mathbf{p}}} = (X' H' HX)^{-1} X' H' H\mathbf{y}. \tag{5.1.5}$$

Имея в виду этот результат, перейдем к задаче отыскания подходящей матрицы H.

5.2. МЕТОД НАИМЕНЬШИХ ВЗВЕШЕННЫХ КВАДРАТОВ С ОГРАНИЧЕНИЯМИ

Поскольку Σ в (5.1.3) вырождена, не существует матрицы H, преобразующей гетероскедастичные ошибки в гомоскедастичные. Однако можно найти матрицы весов для наблюдений, определяющие оценки по методу наименьших взвешенных квадратов, которые могут оказаться более эффективными, чем обычные оценки по методу наименьших квадратов (3.2.10). Сначала мы укажем весовые матрицы, приписывающие постоянные веса каждому уравнению, различные для разных уравнений. Затем мы исследуем различные веса для отдельных наблюдений в каждом уравнении.

В случае, когда наблюдениям, входящим в одно уравнение, приписываются одинаковые веса, меняющиеся лишь от уравнения к уравнению, оценка по методу наименьших взвешенных квадратов совпадает с обычной оценкой по методу наименьших квадратов. Другими словами, если

$$H' H = \hat{\Sigma} \otimes I, \tag{5.2.1}$$

где $\hat{\Sigma}$ — $(r \times r)$-матрица, $\otimes$ — символ кронекеровского произведения, а I — единичная $(T \times T)$-матрица, то, поскольку X — блочно-диагональная матрица с одинаковыми блоками, имеем

$$(X'H'HX)^{-1}\ X'H'H\mathrm{y} = (X'X)^{-1}\ X'\mathrm{y}. \qquad (5.2.2)$$

Результат (5.2.2), казалось бы, свидетельствует о бесполезности взвешивания. Однако в случае, когда параметры стеснены ограничениями, оценки по методу наименьших взвешенных квадратов с ограничениями могут отличаться от оценок при отсутствии ограничений, поскольку локальный минимум квадратичной функции потерь зависит от весов.

В более общем случае, когда различные веса приписываются каждому наблюдению, использование оценок по методу наименьших взвешенных квадратов без ограничений может привести к нарушению условий неотрицательности и нормировки по строкам для переходных вероятностей. Пусть H — блочно-диагональная $(rT \times rT)$-матрица вида

$$H = \begin{bmatrix} a_1 d_1 & & & \\ & a_2 d_2 & & \\ & & \ddots & \\ & & & a_r d_r \end{bmatrix}, \qquad (5.2.3)$$

где a_i, $i = 1, 2, \ldots, r$, — скаляры, а d_i, $i = 1, 2, \ldots, r$, — диагональные $(T \times T)$-матрицы. В этом случае весовая матрица

$$H'H = \begin{bmatrix} a_1^2 d_1' d_1 & & & \\ & a_2^2 d_2' d_2 & & \\ & & \ddots & \\ & & & a_r^2 d_r' d_r \end{bmatrix}. \qquad (5.2.4)$$

Хотя при этом вектор p_j можно оценивать независимо от p_i $(i \neq j)$, применение обычного метода наименьших квадратов может привести к нарушению условий нормировки. Взяв для примера случай (2×2), получаем, что сумма оценок переходных вероятностей по строкам равна:

$$\tilde{\tilde{\mathrm{p}}}_1 + \tilde{\tilde{\mathrm{p}}}_2 = (X' d_1' d_1 X)^{-1} X' d_1' d_1 \mathrm{y}_1 + (X' d_2' d_2 X)^{-1} X' d_2' d_2 \mathrm{y}_2, \quad (5.2.5)$$

что может не совпадать с η_2, если d_i не имеет специального вида. Таким образом, имеет смысл пользоваться только оценками с ограничениями, т. е. оценками, удовлетворяющими как (1.1.5), так и (1.1.6).

5.3. ВАРИАНТЫ ВЫБОРА ВЕСОВ

Ослабление влияния гетероскедастичности между уравнениями и (или) внутри уравнений возможно при различных выборах весов, порождающих соответствующие оценки по методу наименьших взвешенных квадратов.

(1) *Различные веса для разных уравнений и одинаковые веса внутри каждого уравнения*, т. е. $d_i = I$ и a_i одного из приводимых видов:

(а) обратно пропорциональные оценкам дисперсий σ^2 ошибок i-го уравнения, т. е.

$$a_i = (T-r)/[(y_i - X\tilde{p}_i)'(y_i - X\tilde{p}_i)], \tag{5.3.1}$$

где $\tilde{p}_i$ — обычная оценка без ограничений;

(б) обратно пропорциональные средним относительным частотам состояний y_i, т. е.

$$a_i = T/(\sum_i y_i(t)); \tag{5.3.2}$$

(в) обратно пропорциональные произведениям средних частот состояния i и остальных состояний, т. е.

$$a_i = \left[\frac{\sum_i y_i(t)}{T}\,\frac{1-\sum_i y_i(t)}{T}\right]^{-1}. \tag{5.3.3}$$

(2) *Различные веса внутри и между уравнениями.*

Дисперсии ошибок приближенно принимаются пропорциональными истинным относительным частотам заполнения:

$$\sigma_j^2(t) \propto \begin{cases} q_i(t)/N(t) & \text{при } q_j(t) \leqslant 0{,}5; \\ (1-q_j(t))/N(t) & \text{при } q_j(t) \geqslant 0{,}5. \end{cases} \tag{5.3.4}$$

Это значит, что дисперсия ошибки максимальна при $q_j(t) = 1/2$ и обращается в нуль в граничных точках 0 и 1. Поскольку большинство наблюдаемых частот заполнения в рассматриваемых экспериментальных моделях меньше 0,5, а также для простоты построим весовую матрицу следующим образом:

$$a_j^2 d_j' d_j = \begin{bmatrix} N(1)/q_j(1) & & & \\ & N(2)/q_j(2) & & \\ & & \ddots & \\ & & & N(T)/q_j(T) \end{bmatrix} \tag{5.3.5}$$

при всех j от 1 до r, причем вместо $q_j(t)$ можно подставить наблюдаемые относительные частоты $y_j(t)$.

При таком подходе выбор весов оказывается в некоторой степени произвольным. Можно пользоваться и другими весами; соответствующие примеры есть в работах [72] и [114]. Теперь наша задача состоит в том, чтобы по выборочным данным научиться определять, дают ли указанные веса выигрыш в смысле уменьшения дисперсии оценки по сравнению с классическими оценками по методу наименьших квадратов с ограничениями.

В следующих главах будут рассмотрены различные формы оценок с ограничениями и весами вида (5.2.4) в записи (5.3.5). Результаты статистических экспериментов с весами (5.3.1), (5.3.2) и (5.3.3) даны в следующем параграфе.

5.4. РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Аналогично изложенному в главе 4 рассмотрим 50 оценок по методу наименьших взвешенных квадратов с ограничениями для выборок объема 25, 50 и 100. Вычисляются средние значения и среднеквадратические ошибки оценок для различных объемов выборок.

а) Взвешивание по средним частотам состояний. Результаты использования весов (5.3.2), обратных к средним частотам состояний за период наблюдений, а именно средние значения и среднеквадратические ошибки для 50 оценок по выборкам объема 25, 50 и 100, даны в табл. 5.1.

Таблица 5.1

Оценки по методу наименьших взвешенных квадратов при взвешивании по средним частотам состояний

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}5193 & 0{,}4181 & 0{,}0613 & 0{,}0013 \\ 0{,}1213 & 0{,}4632 & 0{,}3998 & 0{,}0157 \\ 0{,}0144 & 0{,}1003 & 0{,}6371 & 0{,}2482 \\ 0{,}0066 & 0{,}0267 & 0{,}1325 & 0{,}8342 \end{bmatrix}$	$\begin{bmatrix} 0{,}1372 & 0{,}1388 & 0{,}1176 & 0{,}0057 \\ 0{,}0954 & 0{,}1565 & 0{,}1294 & 0{,}0297 \\ 0{,}0286 & 0{,}0658 & 0{,}1291 & 0{,}0894 \\ 0{,}0162 & 0{,}0452 & 0{,}1090 & 0{,}1116 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5801 & 0{,}3851 & 0{,}0348 & 0{,}0000 \\ 0{,}1019 & 0{,}4880 & 0{,}4008 & 0{,}0093 \\ 0{,}0061 & 0{,}1021 & 0{,}6588 & 0{,}2330 \\ 0{,}0052 & 0{,}0177 & 0{,}1224 & 0{,}8547 \end{bmatrix}$	$\begin{bmatrix} 0{,}0833 & 0{,}1120 & 0{,}0775 & 0{,}0000 \\ 0{,}0694 & 0{,}1115 & 0{,}0838 & 0{,}0192 \\ 0{,}0157 & 0{,}0566 & 0{,}0950 & 0{,}0674 \\ 0{,}0106 & 0{,}0369 & 0{,}0731 & 0{,}0795 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}5757 & 0{,}3923 & 0{,}0320 & 0{,}0000 \\ 0{,}1161 & 0{,}4773 & 0{,}4005 & 0{,}0061 \\ 0{,}0036 & 0{,}1063 & 0{,}6881 & 0{,}2020 \\ 0{,}0025 & 0{,}0116 & 0{,}0913 & 0{,}8946 \end{bmatrix}$	$\begin{bmatrix} 0{,}0808 & 0{,}1054 & 0{,}0709 & 0{,}0000 \\ 0{,}0672 & 0{,}1122 & 0{,}0779 & 0{,}0141 \\ 0{,}0088 & 0{,}0398 & 0{,}0599 & 0{,}0397 \\ 0{,}0057 & 0{,}0237 & 0{,}0448 & 0{,}0422 \end{bmatrix}$

Средние значения оказываются ближе к истинным параметрам популяции, а среднеквадратические ошибки меньше, чем при применении метода наименьших (невзвешенных) квадратов с ограничениями. Суммы средних абсолютных отклонений оценок от параметров популяции для выборок объема 25, 50 и 100 составляют соответственно 0,5025, 0,2660 и 0,1611. Таким образом, новые оценки имеют меньшие среднеквадратическую и абсолютную ошибки по сравнению с оценками по методу наименьших квадратов без взвешивания. Их отличает также свойство, которое заключается в том, что с увеличением выборки оценки лучше приближаются к истинным значениям.

б) Взвешивание по оценкам дисперсий ошибок. Экспериментальные результаты для этого случая приводятся в табл. 5.2.

Таблица 5.2

Оценки по методу наименьших взвешенных квадратов при взвешивании по дисперсиям ошибок

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}5238 & 0{,}4129 & 0{,}0618 & 0{,}0015 \\ 0{,}1189 & 0{,}4668 & 0{,}4005 & 0{,}0138 \\ 0{,}0139 & 0{,}1010 & 0{,}6375 & 0{,}2476 \\ 0{,}0064 & 0{,}0251 & 0{,}1304 & 0{,}8381 \end{bmatrix}$	$\begin{bmatrix} 0{,}1343 & 0{,}1356 & 0{,}1187 & 0{,}0060 \\ 0{,}0946 & 0{,}1549 & 0{,}1316 & 0{,}0284 \\ 0{,}0279 & 0{,}0655 & 0{,}1268 & 0{,}0880 \\ 0{,}0157 & 0{,}0434 & 0{,}1032 & 0{,}1066 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5828 & 0{,}3811 & 0{,}0360 & 0{,}0001 \\ 0{,}1001 & 0{,}4926 & 0{,}3990 & 0{,}0083 \\ 0{,}0060 & 0{,}1008 & 0{,}6624 & 0{,}2308 \\ 0{,}0051 & 0{,}0164 & 0{,}1193 & 0{,}8592 \end{bmatrix}$	$\begin{bmatrix} 0{,}0819 & 0{,}1130 & 0{,}0797 & 0{,}0000 \\ 0{,}0674 & 0{,}1119 & 0{,}0845 & 0{,}0177 \\ 0{,}0158 & 0{,}0534 & 0{,}0910 & 0{,}0642 \\ 0{,}0104 & 0{,}0352 & 0{,}0703 & 0{,}0751 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}5799 & 0{,}3844 & 0{,}0357 & 0{,}0000 \\ 0{,}1126 & 0{,}4879 & 0{,}3945 & 0{,}0050 \\ 0{,}0037 & 0{,}1026 & 0{,}6922 & 0{,}2015 \\ 0{,}0028 & 0{,}0113 & 0{,}0892 & 0{,}8967 \end{bmatrix}$	$\begin{bmatrix} 0{,}0806 & 0{,}1119 & 0{,}0748 & 0{,}0000 \\ 0{,}0667 & 0{,}1160 & 0{,}0833 & 0{,}0126 \\ 0{,}0087 & 0{,}0377 & 0{,}0573 & 0{,}0378 \\ 0{,}0063 & 0{,}0229 & 0{,}0430 & 0{,}0406 \end{bmatrix}$

Эти оценки демонстрируют явное превосходство над оценками без взвешивания (глава 4), поскольку среднеквадратические ошибки при всех объемах выборок меньше. Суммы средних абсолютных отклонений для выборок объема 25, 50 и 100 составляют соответственно 0,4755, 0,2472 и 0,1519.

в) Взвешивание по произведениям средних частот состояний. Оценки с ограничениями и взвешиванием по произведениям средних частот состояния i и остальных состояний приведены в табл. 5.3. Оценки по методу наименьших взвешенных квадратов вновь оказываются лучше невзвешенных оценок. По сравнению с другими взвешенными оценками среднеквадратические ошибки в данном случае несколько больше. Если, однако, проводить сравнение трех способов взвешивания по агрегированным среднеквадратическим ошибкам

(см. табл. 5.4), то становится трудно отдать предпочтение одному из них.

Взвешенные оценки во всех случаях превосходят невзвешенные. Среди разных способов взвешивания наименее удачно, по-видимому, взвешивание по произведениям средних частот. При выборках объема 25 и 50 веса, обратно пропорциональные оценкам дисперсий ошибок, дают лучшие показатели, чем веса, обратно пропорциональные средним частотам. При больших выборках, напротив, лучшие результаты обеспечивает взвешивание по средним частотам состояний.

Для исследования вопроса о распределении оценок с помощью D-критерия Колмогорова — Смирнова проверялась гипотеза о соответствии 50 оценок по выборке объема 50 при взвешивании по оценкам

Таблица 5.3

Оценки по методу наименьших взвешенных квадратов при взвешивании по произведениям средних частот состояний

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,5161 & 0,4180 & 0,0646 & 0,0013 \\ 0,1229 & 0,4638 & 0,3980 & 0,0153 \\ 0,0146 & 0,0999 & 0,6362 & 0,2493 \\ 0,0063 & 0,0264 & 0,1339 & 0,8334 \end{bmatrix}$	$\begin{bmatrix} 0,1392 & 0,1416 & 0,1236 & 0,0058 \\ 0,0962 & 0,1569 & 0,1330 & 0,0293 \\ 0,0290 & 0,0653 & 0,1301 & 0,0905 \\ 0,0159 & 0,0446 & 0,1103 & 0,1130 \end{bmatrix}$
50	$\begin{bmatrix} 0,5780 & 0,3859 & 0,0361 & 0,0000 \\ 0,1028 & 0,4869 & 0,4010 & 0,0093 \\ 0,0063 & 0,1022 & 0,6582 & 0,2333 \\ 0,0049 & 0,0176 & 0,1232 & 0,8543 \end{bmatrix}$	$\begin{bmatrix} 0,0838 & 0,1130 & 0,0795 & 0,0000 \\ 0,0692 & 0,1123 & 0,0852 & 0,0191 \\ 0,0161 & 0,0563 & 0,0954 & 0,0677 \\ 0,0102 & 0,0367 & 0,0736 & 0,0800 \end{bmatrix}$
100	$\begin{bmatrix} 0,5738 & 0,3918 & 0,0344 & 0,0000 \\ 0,1172 & 0,4779 & 0,3988 & 0,0061 \\ 0,0039 & 0,1060 & 0,6880 & 0,2021 \\ 0,0022 & 0,0115 & 0,0920 & 0,8943 \end{bmatrix}$	$\begin{bmatrix} 0,0818 & 0,1083 & 0,0755 & 0,0000 \\ 0,0680 & 0,1144 & 0,0824 & 0,0142 \\ 0,0091 & 0,0396 & 0,0609 & 0,0398 \\ 0,0055 & 0,0235 & 0,0453 & 0,0425 \end{bmatrix}$

Таблица 5.4

Агрегированные среднеквадратические ошибки для оценок по методу наименьших квадратов

Объем выборки	Без взвешивания	Со взвешиванием		
		по средним частотам состояний	по произведениям частот состояний	по дисперсиям ошибок
25	1,4795	1,4052	1,3812	1,4243
50	1,0191	0,9915	0,9715	0,9981
100	0,8520	0,7931	0,8002	0,8108

Таблица 5.5

Характеристики оценок параметров с нулевыми истинными значениями, полученных по методу наименьших квадратов с ограничениями при взвешивании по дисперсиям ошибок

Параметр	Число нулевых оценок	Число положительных оценок	Максимальное значение	Среднее значение	Стандартное отклонение
p_{13}	30	20	0,3807	0,0360	0,0711
p_{14}	49	1	0,0037	0,0001	0,0005
p_{24}	35	15	0,0547	0,0083	0,0156
p_{31}	38	12	0,0590	0,0060	0,0146
p_{41}	33	17	0,0317	0,0051	0,0089
p_{42}	33	17	0,1415	0,0164	0,0311

дисперсий ошибок для выборки из нормальной совокупности. Полученные для ненулевых параметров значения D-статистики таковы:

$$\begin{bmatrix} 0{,}1004 & 0{,}0806 & \text{——} & \text{——} \\ 0{,}0720 & 0{,}0705 & 0{,}0951 & \text{——} \\ \text{——} & 0{,}0683 & 0{,}1301 & 0{,}0763 \\ \text{——} & \text{——} & 0{,}0663 & 0{,}0894 \end{bmatrix}. \qquad (5.4.1)$$

Все числа в (5.4.1) меньше допустимого предела 0,17, отвечающего для выборок объема 50 10%-ному уровню значимости. Таким образом, можно принять гипотезу о нормальности распределения ошибок.

Результаты для оценок нулевых параметров сведены в табл. 5.5. Сравнивая ее с табл. 4.7 в главе 4, легко увидеть, что за исключением p_{41} распределения этих оценок характеризуются меньшими дисперсиями и более близкими к нулю средними значениями.

В целом взвешенные оценки имеют меньшие среднеквадратические ошибки, сохраняя при этом состоятельность. В следующей главе приводятся экспериментальные результаты по оценкам с различными весами как для самих уравнений, так и внутри уравнений.

5.5. РЕЗУЛЬТАТЫ ДЛЯ ЗАДАЧИ О ВЫБОРЕ МАРКИ СИГАРЕТ

В сформулированной в параграфе 4.9 «задаче курильщика» оценки по методу наименьших взвешенных квадратов с ограничениями при взвешивании по средним частотам состояний следующие:

$$\tilde{\tilde{P}} = [\tilde{\tilde{p}}_{ij}] = \begin{bmatrix} 0{,}6790 & 0{,}1416 & 0{,}1794 \\ 0{,}0000 & 0{,}8691 & 0{,}1309 \\ 0{,}3881 & 0{,}0000 & 0{,}6119 \end{bmatrix}. \qquad (5.5.1)$$

Этот результат весьма близок к оценке, полученной методом наименьших квадратов с ограничениями без взвешивания.

Глава 6 ● ОЦЕНКИ ПО ОБОБЩЕННОМУ МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ

Вырожденность ковариационной матрицы ошибок в марковских моделях уже была отмечена ранее. Из-за вырожденности этой матрицы обобщенный метод наименьших квадратов Эйткена оказывается неприменимым. Внимательное исследование обнаруживает, однако, что вырожденность возникает из-за избыточных переменных, так что проблема становится разрешимой после исключения избыточных переменных. Этот подход развивается в данной главе.

6.1. НЕСФЕРИЧЕСКИЕ ОШИБКИ

Переформулируем статистическую модель, описанную в главе 3, записав множество r уравнений марковской цепи в виде одного соотношения:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_r \end{bmatrix} = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_r \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_r \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_r \end{bmatrix} \tag{6.1.1}$$

или в более компактной форме:

$$\mathbf{y} = X\mathbf{p} + \mathbf{u}, \tag{6.1.2}$$

где $X_1 = X_2 = \ldots = X_r$,

$$E(\mathbf{u}) = 0, \tag{6.1.3}$$

$$E(\mathbf{u}\mathbf{u}') = \Sigma. \tag{6.1.4}$$

Здесь Σ—ковариационная матрица размера $(Tr \times Tr)$. Предположим, что значения $\mathbf{u}$ не зависят от X, по крайней мере в одни и те же моменты времени, так что ковариационные матрицы $\mathbf{y}$ и $\mathbf{u}$ одинаковы. Далее предполагается также, что $\mathbf{y}$, множество относительных частот состояний, порождается полиномиальным распределением со средними значениями $q_j(t)$, дисперсиями $q_j(t)\ [1 - q_j(t)]/N(t)$ и ковариациями

$-q_i(t)\,q_j(t)/N(t)$. Поэтому $\mathbf{u}$ — вектор с нулевым средним и ковариационной $(r\,T \times rT)$-матрицей вида

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1r} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{r1} & \Sigma_{r2} & \cdots & \Sigma_{rr} \end{bmatrix} \tag{6.1.5}$$

с подматрицами размера $(T \times T)$:

$$\Sigma_{ij} = \begin{bmatrix} -\dfrac{q_i(1)\,q_j(1)}{N(1)} & & & \\ & -\dfrac{q_i(2)\,q_j(2)}{N(2)} & & \\ & & \ddots & \\ & & & -\dfrac{q_i(T)\,q_j(T)}{N(T)} \end{bmatrix}, \quad i \neq j; \tag{6.1.6}$$

$$\Sigma_{ii} = \begin{bmatrix} \dfrac{q_i(1)\,[1-q_i(1)]}{N(1)} & & & \\ & \dfrac{q_i(2)\,[1-q_i(2)]}{N(2)} & & \\ & & \ddots & \\ & & & \dfrac{q_i(T)\,[1-q_i(T)]}{N(T)} \end{bmatrix}. \tag{6.1.7}$$

Отметим некоторые свойства этой ковариационной структуры.

(1) В каждом уравнении модели (6.1.1) ошибки гетероскедастичны.

(2) Ошибки в разных уравнениях зависимы в пределах одного и того же выборочного интервала.

(3) Автокорреляция ошибок отсутствует.

Свойство (1) объясняется особенностями полиномиального распределения в схеме Лексиса[1]. Свойство (2) следует из зависимости наблюдаемых относительных частот. Свойство (3) определяется тем, что, хотя наблюдения в каждый период времени порождаются од-

[1] О полиномиальном распределении в схеме Лексиса см. [93], [40*].

ной и той же популяцией, они могут быть некоррелированы с предыдущими наблюдениями из-за различного числа наблюдений $N(t)$ в каждом периоде. Во всяком случае для системы (6.1.1) в целом характерны несферические ошибки наблюдений.

6.2. ИЗБЫТОЧНЫЕ ПАРАМЕТРЫ И РЕДУЦИРОВАННАЯ МОДЕЛЬ

Ковариационная матрица (6.1.5) симметрична и вырождена. Вырожденность определяется тем, что сумма элементов в каждой строке равна нулю, поскольку

$$\sum_{l=1}^{r} q_l(t) = 1 \quad \text{при всех } t. \tag{6.2.1}$$

Поэтому непосредственное применение обобщенного метода наименьших квадратов Эйткена для получения наилучшей линейной несмещенной оценки невозможно.

Исследуя модель (6.1.1), отметим, что элементы векторов $\mathbf{y}_l$ и матриц X_l представляют собой частоты состояний, элементы векторов $\mathbf{p}_l$ — вероятности, а векторы $\mathbf{u}_l$ зависимы для разных уравнений в соответствии с (6.1.5). Таким образом, уравнения в (6.1.1) линейно-зависимы, и, если $r-1$ уравнение в (6.1.1) известно, остающееся определяется соотношением (6.2.1). Другими словами, тождество (6.2.1) дублирует информацию о модели, содержащуюся в (6.1.1), так что существует множество избыточных параметров.

Поскольку имеются только $r-1$ независимый вектор наблюдений $\mathbf{y}_j$ и столько же независимых векторов параметров $\mathbf{p}_j$ и поскольку тождеством (6.2.1) пренебрегать нельзя, одно из уравнений модели (6.1.1)[1] можно опустить. Не ограничивая общности, можно опустить последнее уравнение, так как указанные уравнения инвариантны относительно перестановок. Таким образом, редуцированная модель записывается в виде

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{r-1} \end{bmatrix} = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_{r-1} \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_{r-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_{r-1} \end{bmatrix} \tag{6.2.2}$$

или короче

$$\mathbf{y}_* = X_* \mathbf{p}_* + \mathbf{u}_*, \tag{6.2.3}$$

где звездочками отмечены подмножества наблюдений, параметров и ошибок. Ошибки $\mathbf{u}_*$ при этом имеют следующие характеристики:

$$E\mathbf{u}_* = 0; \tag{6.2.4}$$

$$E\mathbf{u}_*\mathbf{u}_*' = \Sigma_*, \tag{6.2.5}$$

где Σ_* — невырожденная подматрица Σ размера $((r-1)\,T \times (r-1)\,T)$.

[1] Такое исключение одного уравнения аналогично процедуре определения стационарного вектора для регулярной марковской цепи [60].

6.3. ОБРАТИМОСТЬ КОВАРИАЦИОННОЙ МАТРИЦЫ ОШИБОК

Ковариационная матрица ошибок Σ_* представляет собой подматрицу матрицы (6.1.5), в которой опущены последняя строка (Σ_{rj}, $j = 1, 2, \ldots, r$) и последний столбец блоков (Σ_{ir}, $i = 1, 2, \ldots, r$). Обратная к ней матрица Σ_*^{-1} может служить весовой матрицей при построении оценки по методу наименьших взвешенных квадратов. Чтобы упростить обращение Σ_*, введем несколько новых обозначений.

Пусть $V_*(t)$ — сечение матрицы Σ_* в момент t, т. е. матрица размера $((r-1) \times (r-1))$, имеющая вид

$$V_*(t) = \frac{1}{N(t)} \begin{bmatrix} q_1(t)[1-q_1(t)] & -q_1(t)q_2(t) & \ldots & -q_1(t)q_{r-1}(t) \\ -q_2(t)q_1(t) & q_2(t)[1-q_2(t)] & \ldots & -q_2(t)q_{r-1}(t) \\ \vdots & \vdots & & \vdots \\ -q_{r-1}(t)q_1(t) & -q_{r-1}(t)q_2(t) & \ldots & q_{r-1}(t)[1-q_{r-1}(t)] \end{bmatrix}. \tag{6.3.1}$$

С помощью прямого (кронекеровского) произведения матриц можно записать:

$$V_*(t) \otimes I_t = \Sigma_*. \tag{6.3.2}$$

Другими словами, каждый элемент $V_*(t)$ преобразуется в диагональный блок с элементами, индексы t которых меняются в естественном порядке от 1 до T. Простым примером, иллюстрирующим эти обозначения в случае гетероскедастичной ковариационной матрицы, может служить

$$\begin{bmatrix} \sigma^2(1) I_1 & & & \\ & \sigma^2(2) I_2 & & \\ & & \ddots & \\ & & & \sigma^2(T) I_T \end{bmatrix} = \sigma^2(t) \otimes I_t. \tag{6.3.3}$$

Отметим следующие свойства кронекеровского умножения. Если $a(t)$ и $b(t)$ — скалярные сечения, то

$$a(t) \otimes I_t \cdot b(t) \otimes I_t = a(t)\, b(t) \otimes I_t. \tag{6.3.4}$$

Если $A(t)$ и $B(t)$ — матричные сечения одинаковых размеров, то

$$A(t) \otimes I_t \cdot B(t) \otimes I_t = A(t)\, B(t) \otimes I_t. \tag{6.3.5}$$

Доказательство этих равенств проводится непосредственным перемножением матриц. Наиболее важное и интересное соотношение — равенство

$$\Sigma_*^{-1} = V_*^{-1}(t) \otimes I_t, \tag{6.3.6}$$

доказываемое перемножением (6.3.6) и (6.3.2) с учетом (6.3.5). Итак, для обращения Σ_* достаточно обратить матрицу сечения $V_*(t)$ и применить затем кронекеровское умножение.

Определитель $V_*(t)$ равен

$$\det V_*(t) = \frac{\left[\prod_{k=1}^{r-1} q_k(t)\right]\left[1-\sum_{k=1}^{r-1} q_k(t)\right]}{N(t)^{r-1}}, \tag{6.3.7}$$

что легко получается после элементарных преобразований (6.3.1). Аналогично выводу (6.3.7) алгебраическое дополнение (i, j)-го элемента матрицы $V_*(t)$, обозначаемого V_{*ij}, равно

$$V_{*ij} = \frac{\prod_{k=1}^{r-1} q_k(t)}{N(t)^{r-2}} \quad \text{при } i \neq j \tag{6.3.8}$$

и

$$V_{*jj} = \frac{\prod_{k \neq j}^{r-1} q_k(t)\left[1-\sum_{k \neq j}^{r-1} q_k(t)\right]}{N(t)^{r-2}}. \tag{6.3.9}$$

Следовательно, элементы V_*^{ij} обратной матрицы $V_*^{-1}(t)$ равны:

$$V_*^{ij} = \frac{N(t)}{q_r(t)} + \frac{N(t)}{q_j(t)} \quad \text{при } i = j; \tag{6.3.10}$$

$$V_*^{ij} = \frac{N(t)}{q_r(t)} \quad \text{при } i \neq j, \tag{6.3.11}$$

а значит,

$$V_*^{-1}(t) = \begin{bmatrix} \frac{N(t)}{q_r(t)} + \frac{N(t)}{q_1(t)} & \frac{N(t)}{q_r(t)} & \cdots & \frac{N(t)}{q_r(t)} \\ \frac{N(t)}{q_r(t)} & \frac{N(t)}{q_r(t)} + \frac{N(t)}{q_2(t)} & \cdots & \frac{N(t)}{q_r(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{N(t)}{q_r(t)} & \frac{N(t)}{q_r(t)} & \cdots & \frac{N(t)}{q_r(t)} + \frac{N(t)}{q_{r-1}(t)} \end{bmatrix}. \tag{6.3.12}$$

Поэтому матрица, обратная к Σ_*, будет $((r-1)\,T \times (r-1)\,T)$-матрицей с $(r-1)^2$ диагональными подматрицами размера $(T \times T)$, так что (см. также [61], [123])

$$\Sigma_*^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} & \cdots & \Sigma^{1,r-1} \\ \Sigma^{21} & \Sigma^{22} & \cdots & \Sigma^{2,r-1} \\ \vdots & \vdots & & \vdots \\ \Sigma^{r-1,1} & \Sigma^{r-2,2} & \cdots & \Sigma^{r-1,r-1} \end{bmatrix}, \tag{6.3.13}$$

где

$$\Sigma^{ij} = \begin{bmatrix} \frac{N(1)}{q_r(1)} & & & \\ & \frac{N(2)}{q_r(2)} & & \\ & & \ddots & \\ & & & \frac{N(T)}{q_r(T)} \end{bmatrix} \quad \text{при} \quad i \neq j \qquad (6.3.14)$$

и

$$\Sigma^{ii} = \begin{bmatrix} \frac{N(1)}{q_r(t)} + \frac{N(1)}{q_1(1)} & & & \\ & \frac{N(2)}{q_r(2)} + \frac{N(2)}{q_2(2)} & & \\ & & \ddots & \\ & & & \frac{N(T)}{q_r(T)} + \frac{N(T)}{q_{r-1}(T)} \end{bmatrix}. \qquad (6.3.15)$$

6.4. ОЦЕНКИ ПО ОБОБЩЕННОМУ МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ

При заданной модели (6.2.2) или (6.2.3) с допущениями (6.2.4) и (6.2.5) в рамках обобщенного метода наименьших квадратов Эйткена задача состоит в минимизации

$$(\mathbf{y}_* - X_* \mathbf{p}_*)' \; \Sigma_*^{-1} \; (\mathbf{y}_* - X_* \mathbf{p}_*) \qquad (6.4.1)$$

при ограничениях

$$R\mathbf{p}_* \geqslant \boldsymbol{\eta}_r, \qquad (6.4.2)$$

$$\mathbf{p}_* \geqslant 0. \qquad (6.4.3)$$

Здесь $\boldsymbol{\eta}_r$ — $(r \times 1)$-вектор единиц, а R (подматрица G) — матрица вида $[I_1, I_2, \ldots, I_{r-1}]$, где каждая I — единичная $(r \times r)$-матрица. Если бы ограничения (6.4.2) и (6.4.3) отсутствовали, то можно было бы, продифференцировав (6.4.1) по $\mathbf{p}_*$ и приравняв производную нулю, из уравнения

$$-2X_*' \Sigma_*^{-1}(\mathbf{y}_* - X_* \hat{\mathbf{p}}_*) = 0 \qquad (6.4.4)$$

найти оценку

$$\hat{\mathbf{p}}_* = (X_*' \Sigma_*^{-1} X_*)^{-1} X_*' \Sigma_*^{-1} \mathbf{y}_*. \qquad (6.4.5)$$

Ковариационная матрица оценки $\mathbf{p}_*$ согласно обобщению теоремы Гаусса — Маркова, принадлежащему Эйткену [2], равна

$$V(\mathbf{p}_*) = (X_*'\Sigma_*^{-1} X_*)^{-1}. \tag{6.4.6}$$

Исключенный параметр $\mathbf{p}_r$ оценивается с помощью равенства

$$\hat{\mathbf{p}}_r = \eta_r - R\hat{\mathbf{p}}_*. \tag{6.4.7}$$

Отметим, что оценку (6.4.5) можно получить, если известна ковариационная матрица ошибок Σ_*. В действительности матрица Σ_*, как правило, неизвестна. Поскольку Σ_* — функция относительных частот $q_j(t)$, для получения состоятельной оценки Σ_* нужно заменить в (6.3.13) неизвестные истинные значения $q_j(t)$ известными наблюденными значениями $y_j(t)$. В параграфе 3.2., а) было показано, что $y_j(t)$ — состоятельная оценка $q_j(t)$.

Заметим также, что оценка (6.4.5) может не удовлетворять ограничениям (6.4.2) и (6.4.3), и, в частности, оценка по методу наименьших взвешенных квадратов без ограничений $\hat{\mathbf{p}}_*$ может оказаться отрицательной или превосходящей 1. Для учета ограничений (6.4.2) и (6.4.3) при минимизации целевой функции (6.4.1) можно воспользоваться алгоритмом квадратичного программирования. Так же как в главе 3, с помощью теоремы приводимости[1] нетрудно получить следующую эквивалентную задачу линейного программирования: максимизировать

$$(X_*'\Sigma_*^{-1}\mathbf{y}_* - X_*'\Sigma_*^{-1}X_*\hat{\mathbf{p}}_*^c)'\mathbf{p}_* \tag{6.4.8}$$

при ограничениях

$$R\mathbf{p}_* \leqslant \eta_r \tag{6.4.2}$$

и

$$\mathbf{p}_* \geqslant 0, \tag{6.4.3}$$

где $\hat{\mathbf{p}}_*^c$ — искомое оптимальное решение. Двойственная задача имеет вид: минимизировать

$$\lambda'\eta_r \tag{6.4.9}$$

при ограничениях

$$R'\lambda \geqslant X_*'\Sigma_*^{-1}\mathbf{y}_* - X_*'\Sigma_*^{-1}X_*\hat{\mathbf{p}}_*^c \tag{6.4.10}$$

и

$$\lambda \geqslant 0, \tag{6.4.11}$$

где λ — вектор-столбец двойственных переменных.

Основываясь на этих линейных формулировках задачи, обозначим теперь $\hat{\mathbf{p}}^c$ через $\mathbf{p}_*$ и определим следующую задачу: максимизировать

$$(X_*\Sigma_*^{-1}\mathbf{y}_* - X_*'\Sigma_*^{-1}X_*\mathbf{p}_*)'\mathbf{p}_* - \lambda'\eta_r = -\lambda'\mathbf{p}_r - \beta'\mathbf{p}_* \leqslant 0 \tag{6.4.12}$$

[1] См., например, [17], [66].

при ограничениях

$$R\mathbf{p}_* + \mathbf{p}_r = \boldsymbol{\eta}_r, \tag{6.4.13}$$

$$R'\boldsymbol{\lambda} + (X_*'\Sigma_*^{-1}X_*)\,\mathbf{p}_* - \boldsymbol{\beta} = X_*'\Sigma_*^{-1}\mathbf{y}_* \tag{6.4.14}$$

и

$$\mathbf{p}_*, \mathbf{p}_r, \boldsymbol{\lambda}, \boldsymbol{\beta} \geqslant 0, \tag{6.4.15}$$

где $\mathbf{p}_r$ и $\boldsymbol{\beta}$ — векторы дополнительных переменных. В этой задаче $\mathbf{p}_r$ также представляет собой подвектор $\mathbf{p}$, опущенный при формировании модели (6.2.2). Для приведенной задачи оптимальное значение целевой функции равно нулю ввиду эквивалентности прямой и двойственной задач. Поэтому из правой части (6.4.12) следует, что если оптимальное решение $\hat{\mathbf{p}}_*^c$ существует, то

$$\boldsymbol{\lambda}'\,\hat{\mathbf{p}}_r^c + \boldsymbol{\beta}'\,\hat{\mathbf{p}}_*^c = 0. \tag{6.4.16}$$

Для выполнения (6.4.16) нужно, чтобы по крайней мере по одному вектору из двух пар $\boldsymbol{\lambda}'$ и $\hat{\mathbf{p}}_r^c$ или $\boldsymbol{\beta}'$ и $\hat{\mathbf{p}}_*^c$ обращалось в нуль, поскольку все они неотрицательны. При этом условии сформулированная задача разрешима с помощью стандартного симплекс-метода; табл. 6.1 представляет собой соответствующую симплекс-таблицу.

Таблица 6.1

Симплекс-таблица для получения оценок по обобщенному методу наименьших квадратов с ограничениями

$\boldsymbol{\beta}_0$	$\boldsymbol{\lambda} \geqslant 0$	$\mathbf{p}_* \geqslant 0$	$\mathbf{p}_r \geqslant 0$	$\boldsymbol{\beta} \geqslant 0$
$\boldsymbol{\eta}_r$	0	R	I	0
$X_*'\Sigma_*^{-1}\mathbf{y}_*$	R'	$X_*'\Sigma_*^{-1}X_*$	0	$-I$

6.5. РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКИХ ЭКСПЕРИМЕНТОВ

При практических вычислениях оценок по методу наименьших квадратов как с ограничениями, так и без них приходится сталкиваться с тем, что для существования Σ_*^{-1} нужно, чтобы величины $q_j(t)$ или $y_j(t)$ не были равны нулю. Во избежание нулевых частот прибегают к агрегированию данных для периодов от 3-го до 14-го, а некоторые из данных меняют (например, нулевые $y_j(t)$ заменены на 0,0001). Средние значения и среднеквадратические ошибки для оценок без ограничений, полученных в результате 50 экспериментов с выборками объема 25, 50 и 100, сведены в табл. 6.2.

Табл. 6.2 иллюстрирует уменьшение среднеквадратической ошибки с ростом объема выборки. Поскольку агрегированные данные, приведенные здесь, несколько отличаются от данных из предыдущих глав, прямое сравнение невозможно. В дальнейшем для рассмотренных

Таблица 6.2

Оценки по обобщенному методу наименьших квадратов без ограничений

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}4151 & 0{,}5778 & 0{,}1033 & -0{,}0962 \\ 0{,}1349 & 0{,}3788 & 0{,}4520 & 0{,}0343 \\ 0{,}0312 & 0{,}1220 & 0{,}5747 & 0{,}2675 \\ 0{,}0120 & -0{,}0388 & 0{,}1978 & 0{,}8342 \end{bmatrix}$	$\begin{bmatrix} 0{,}2743 & 0{,}3364 & 0{,}4506 & 0{,}3646 \\ 0{,}1380 & 0{,}2459 & 0{,}3144 & 0{,}2361 \\ 0{,}1094 & 0{,}1445 & 0{,}2235 & 0{,}1500 \\ 0{,}1323 & 0{,}0230 & 0{,}1993 & 0{,}1819 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}4754 & 0{,}5287 & 0{,}0057 & -0{,}0087 \\ 0{,}1390 & 0{,}3846 & 0{,}5188 & -0{,}0424 \\ 0{,}0149 & 0{,}1518 & 0{,}5540 & 0{,}2793 \\ -0{,}0088 & -0{,}0457 & 0{,}2153 & 0{,}8932 \end{bmatrix}$	$\begin{bmatrix} 0{,}1858 & 0{,}2635 & 0{,}2941 & 0{,}1981 \\ 0{,}1359 & 0{,}2213 & 0{,}2774 & 0{,}2007 \\ 0{,}1174 & 0{,}1846 & 0{,}2160 & 0{,}1576 \\ 0{,}0925 & 0{,}1418 & 0{,}1697 & 0{,}1238 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}5059 & 0{,}5511 & -0{,}0347 & -0{,}0221 \\ 0{,}1727 & 0{,}3032 & 0{,}5310 & -0{,}0068 \\ -0{,}0162 & 0{,}1982 & 0{,}5988 & 0{,}2193 \\ -0{,}0002 & -0{,}0495 & 0{,}1598 & 0{,}8899 \end{bmatrix}$	$\begin{bmatrix} 0{,}1737 & 0{,}2762 & 0{,}3139 & 0{,}2519 \\ 0{,}1490 & 0{,}2634 & 0{,}3066 & 0{,}2604 \\ 0{,}0755 & 0{,}1387 & 0{,}1531 & 0{,}1064 \\ 0{,}0496 & 0{,}0943 & 0{,}1083 & 0{,}0715 \end{bmatrix}$

ранее методов оценивания будут предложены некоторые новые оценки, основанные на измененных данных, их сравнение проводится в главе 12.

Статистики среднего и среднеквадратической ошибки для оценок с ограничениями приведены в табл. 6.3.

Таблица 6.3

Оценки по обобщенному методу наименьших квадратов с ограничениями

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}4003 & 0{,}4875 & 0{,}1035 & 0{,}0088 \\ 0{,}1370 & 0{,}4456 & 0{,}4061 & 0{,}0113 \\ 0{,}0260 & 0{,}0743 & 0{,}6777 & 0{,}2219 \\ 0{,}0125 & 0{,}0064 & 0{,}1083 & 0{,}8728 \end{bmatrix}$	$\begin{bmatrix} 0{,}2601 & 0{,}2113 & 0{,}1871 & 0{,}0224 \\ 0{,}1007 & 0{,}1876 & 0{,}1588 & 0{,}0238 \\ 0{,}0449 & 0{,}0801 & 0{,}1344 & 0{,}0937 \\ 0{,}0257 & 0{,}0158 & 0{,}0991 & 0{,}0983 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}4787 & 0{,}4583 & 0{,}0592 & 0{,}0038 \\ 0{,}1194 & 0{,}4781 & 0{,}3948 & 0{,}0077 \\ 0{,}0112 & 0{,}0775 & 0{,}7106 & 0{,}2007 \\ 0{,}0046 & 0{,}0096 & 0{,}0955 & 0{,}8722 \end{bmatrix}$	$\begin{bmatrix} 0{,}1736 & 0{,}1740 & 0{,}1131 & 0{,}0111 \\ 0{,}0978 & 0{,}1555 & 0{,}1113 & 0{,}0209 \\ 0{,}0269 & 0{,}0670 & 0{,}0932 & 0{,}0648 \\ 0{,}0100 & 0{,}0242 & 0{,}0639 & 0{,}1299 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}5231 & 0{,}4197 & 0{,}0565 & 0{,}0007 \\ 0{,}1261 & 0{,}4769 & 0{,}3914 & 0{,}0056 \\ 0{,}0048 & 0{,}0962 & 0{,}7144 & 0{,}1846 \\ 0{,}0012 & 0{,}0043 & 0{,}0849 & 0{,}9096 \end{bmatrix}$	$\begin{bmatrix} 0{,}1362 & 0{,}1483 & 0{,}1053 & 0{,}0034 \\ 0{,}0824 & 0{,}1237 & 0{,}0951 & 0{,}0140 \\ 0{,}0116 & 0{,}0411 & 0{,}0553 & 0{,}0399 \\ 0{,}0038 & 0{,}0142 & 0{,}0514 & 0{,}0493 \end{bmatrix}$

В этой таблице вновь заметно уменьшение среднеквадратической ошибки с ростом объема выборки. Сравнивая оценки с ограничениями и без них, видим, что первые дают результаты, более близкие к истинным значениям параметров (4.1.1). Суммы среднеквадратических ошибок для обоих методов оценивания сведены в табл. 6.4.

Таблица 6.4

Суммы среднеквадратических ошибок оценок по обобщенному методу наименьших квадратов с ограничениями и без ограничений

Объем выборки	Без ограничений	С ограничениями
25	3,6036	1,7440
50	2,9800	1,3371
100	2,7390	0,9750

При таком сравнении оценки с ограничениями значительно превосходят оценки без ограничений, особенно для больших выборок. С ростом объема выборки отношение среднеквадратических ошибок оценок без ограничений и с ними растет примерно от 2 до 3.

С помощью критерия согласия Колмогорова — Смирнова проверялась нормальность распределения каждого элемента матрицы оценок. Оценки без ограничений все нормально распределены, поскольку вычисленные значения D-статистики меньше табличного предела 0,17, допустимого для выборки объема 50 при 10%-ном уровне значимости.

Оценки с ограничениями также нормально распределены, за исключением тех элементов, истинные значения которых равны нулю, а именно p_{13}, p_{14}, p_{24}, p_{31}, p_{41} и p_{42}. Для этих параметров более половины оценок отличаются от нуля. Процент таких оценок можно найти в в табл. 6.5. Вообще говоря, с ростом объема выборки доля оценок, отличающихся от нуля, растет.

Таблица 6.5

Доля оценок нулевых параметров, отличающихся от нуля, в случае обобщенного метода наименьших квадратов, %

Параметры	Объем выборки		
	25	50	100
p_{13}	54	56	54
p_{14}	76	82	92
p_{24}	64	80	76
p_{31}	42	70	72
p_{41}	54	72	88
p_{42}	80	70	88

Глава 7 ● ОЦЕНКИ ПО КРИТЕРИЮ χ^2

7.1. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

Пусть в каждом из T периодов времени имеются наблюдения, относящиеся к $N(t)$ микрообъектам. Производится классификация наблюдений, при которой каждое из наблюдений в момент t относится к одному из r непересекающихся и в совокупности охватывающих все возможные результаты наблюдений классов (марковских состояний). Если $q_j(t)$ — вероятность того, что наблюдение (микрообъект) в период t попадает в состояние j, то

$$\sum_j q_j(t) = 1. \tag{7.1.1}$$

Таким образом, задача сведена к исследованию системы из T полиномиальных распределений. Если через $n_j(t)$ обозначить число микрообъектов, попадающих в класс j в момент t, то величины

$$y_j(t) = n_j(t)/N(t), \quad j = 1, 2, \ldots, r \tag{7.1.2}$$

представляют собой соответствующие наблюдаемые относительные частоты состояний. Предположим, что вероятности $q_j(t)$ порождаются марковской цепью первого порядка и являются функциями множества неизвестных постоянных параметров p_{ij} ($i, j = 1, 2, \ldots, r$), так что

$$q_j(t) = \sum_{i=1}^{r} y_i(t-1)\, p_{ij}. \tag{7.1.3}$$

Если в последнем периоде наблюдаются $y_i(t-1)$, то истинные относительные частоты состояний в текущем периоде равны $q_j(t)$, $j = 1, 2, \ldots, r$, в то время как текущие наблюдаемые относительные частоты заполнения — $y_j(t)$, $j = 1, 2, \ldots, r$. Различие между $q_j(t)$ и $y_j(t)$ объясняется выборочными ошибками. Для проверки соответствия наблюдаемых и теорети-

ческих частот, $N(t)\,y_j(t)$ и $N(t)\,q_j(t)$, можно воспользоваться следующей статистикой χ^2:

$$\chi^2 = \sum_{t=1}^{T} \sum_{j=1}^{r} \frac{[N(t)\,y_j(t) - N(t)q_j(t)]^2}{[N(t)\,q_j(t)]}. \qquad (7.1.4)$$

Статистика (7.1.4) распределена по закону χ^2 с $T(r-1)$ степенями свободы, и оценки, полученные в результате минимизации выражения (7.1.4), называются оценками по критерию χ^2.

7.2. ОЦЕНКИ ПО КРИТЕРИЮ χ^2 С ОГРАНИЧЕНИЯМИ

Для получения оценок по критерию χ^2 можно дифференцировать (7.1.4) по p_{ij}[1]. Прежде чем дифференцировать (7.1.4) по p_{ij} с целью отыскания минимума, необходимо выразить одну из $q_j(t)$, скажем, $q_r(t)$, через остальные с помощью отношений (7.1.1) и условия

$$\sum_{j=1}^{r} y_j(t) = 1. \qquad (7.2.1)$$

После этого (7.1.4) примет вид

$$\chi^2 = \sum_{t=1}^{T} \sum_{j=1}^{r-1} N(t)\,[(y_j(t) - q_j(t))^2/q_j(t)] + \sum_{t=1}^{T} N(t)\,[(y_r(t) -$$

$$- q_r(t))^2/q_r(t)] = \sum_{t=1}^{T} \sum_{j=1}^{r-1} [N(t)/q_j(t)]\,(y_j(t) - q_j(t))^2 +$$

$$+ \sum_{t=1}^{T} \left([N(t)/q_r(t)] \times \sum_{j=1}^{r-1} (y_j(t) - q_j(t))^2 \right) = \sum_{t=1}^{T} \sum_{j=1}^{r-1} [N(t)/q_j(t)] \times$$

$$\times (y_j(t) - q_j(t))^2 + \sum_{t=1}^{T} ([N(t)/q_r(t)] \sum_{l=1}^{r-1} \sum_{j=1}^{r-1} (y_l(t) - q_l(t)) \times$$

$$\times (y_j(t) - q_j(t))). \qquad (7.2.2)$$

Для упрощения записи (7.2.2) введем обозначения $\mathbf{y}'_* = [\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_{r-1}]$, где $\mathbf{y}'_j = [y_j(1), y_j(2), \ldots, y_j(T)]$, $j = 1, 2, \ldots, r-1$, $t = 1, 2, \ldots, T$, $X_* \mathbf{p}_* = \{q_j(t)\}$, $j = 1, 2, \ldots, r-1$, $t = 1, 2, \ldots, T$, D_* — диагональная $(T(r-1) \times T(r-1))$-матрица с элементами $N(t)/q_j(t)$, $j = 1, 2, \ldots, r-1$, $t = 1, 2, \ldots, T$, и S_* — $(T(r-1) \times$

[1] Правая часть (7.1.4) — функция $q_j(t)$, которые в свою очередь зависят от p_{ij} в соответствии с (7.1.3).

$\times T(r-1))$-матрица, составленная из $(r-1)^2$ диагональных матриц размера $(T \times T)$, диагональные элементы которых равны $N(t)/q_r(t)$, $t = 1, 2, \ldots, T$. При этом выражение (7.2.2) для критерия χ^2 запишется в виде

$$\chi^2 = (\mathbf{y}_* - X_* \mathbf{p}_*)' D_* (\mathbf{y}_* - X_* \mathbf{p}_*) + (\mathbf{y}_* - X_* \mathbf{p}_*)' S_* (\mathbf{y}_* - X_* \mathbf{p}_*), \tag{7.2.3}$$

или, что еще проще, в виде

$$\chi^2 = (\mathbf{y}_* - X_* \mathbf{p}_*)' \Sigma_*^{-1} (\mathbf{y}_* - X_* \mathbf{p}_*), \tag{7.2.4}$$

где матрица Σ_*^{-1} есть сумма D_* и S_*, которая была определена ранее в главе 6 формулой (6.3.12). Таким образом, выражение для χ^2 — не что иное, как та же квадратичная форма, которая уже рассматривалась при анализе модели из главы 6. Поэтому, минимизируя (7.2.4) по $\mathbf{p}_*$, получим оценку по критерию χ^2 без ограничений[1]:

$$\hat{\mathbf{p}}^* = (X_*' \Sigma_*^{-1} X_*)^{-1} X_*' \Sigma_*^{-1} \mathbf{y}_*, \tag{7.2.5}$$

совпадающую с оценкой по обобщенному методу наименьших квадратов без ограничений. Оценка с ограничениями получается при минимизации (7.2.4) по $\mathbf{p}_*$ при условиях

$$R\mathbf{p}_* < \boldsymbol{\eta}_r \tag{7.2.6}$$

и

$$\mathbf{p}_* \geqslant \mathbf{0}_r. \tag{7.2.7}$$

В результате находим уже известную нам оценку по обобщенному методу наименьших квадратов с ограничениями. Обе оценки можно построить при условии, что Σ_*^{-1} известна. На практике обычно пользуются оценкой $\hat{\Sigma}_*^{-1}$.

7.3. ОЦЕНКИ ПО МОДИФИЦИРОВАННОМУ КРИТЕРИЮ χ^2 С ОГРАНИЧЕНИЯМИ

Для оценки весовой матрицы Σ_*^{-1} истинные относительные частоты состояний заменяются наблюдаемыми $y_j(t)$, поскольку последние представляют собой состоятельные оценки относительных частот. В результате выражение для χ^2 примет вид

$$(\chi')^2 = (\mathbf{y}_* - X_* \mathbf{p}_*)' \hat{\Sigma}_*^{-1} (\mathbf{y}_* - X_* \mathbf{p}_*), \tag{7.3.1}$$

---

[1] Формула (7.2.5) выводится в предположении, что элементы Σ_* известны, и поэтому представляет собой точку условного минимума. Если элементы Σ_* содержат неизвестные параметры, то для получения точки безусловного минимума необходимо применять итеративные численные методы.

откуда, проводя в обратном порядке выкладки (7.2.2), имеем:

$$(\chi')^2 = \sum_{t=1}^{T} \sum_{j=1}^{r-1} \frac{[N(t)\, y_j(t) - N(t)\, q_j(t)]^2}{[N(t)\, y_j(t)]}, \qquad (7.3.2)$$

где предполагается, что $y_j(t) \neq 0$. Выражение (7.3.2) называют модифицированным критерием χ^2, а соответствующую оценку — оценкой по модифицированному критерию χ^2. Можно показать [82], что обе оценки — по обычному и модифицированному критериям χ^2 — наилучшие асимптотически нормальные оценки.

7.4. ЭКВИВАЛЕНТНАЯ МОДЕЛЬ

Возможен еще один подход к задаче минимизации критерия χ^2 (7.1.4). Прежде всего отметим, что в соответствии с условием (7.1.1) имеется только $r-1$ независимых величин $q_j(t)$ при каждом t. Альтернатива состоит в том, чтобы временно пренебречь соотношением (7.1.1) и записать (7.1.4) как

$$\chi^2 = \sum_{t=1}^{T} \sum_{j=1}^{r} [y_j(t) - q_j(t)]\,[N(t)/q_j(t)]\,[y_j(t) - q_j(t)]. \qquad (7.4.1)$$

Теперь, положив $y' = (y_1', y_2', \ldots, y_r')$, где $y_j' = [y_j(1), y_j(2), \ldots, y_j(T)]$, и обозначив через $X\mathbf{p}$ вектор $q_j(t)$ ($j = 1, 2, \ldots, r$; $t = 1, 2, \ldots, T$), а через D — диагональную ($Tr \times Tr$)-матрицу с элементами $N(t)/q_j(t)$ ($j = 1, 2, \ldots, r$; $t = 1, 2, \ldots, T$), можно записать (7.4.1) в виде

$$\chi^2 = (\mathbf{y} - X\mathbf{p})'\, D\, (\mathbf{y} - X\mathbf{p}). \qquad (7.4.2)$$

Дифференцируя (7.4.2) по $\mathbf{p}$ и предполагая все p_{ij} независимыми[1], получим систему нормальных уравнений:

$$X'DX\mathbf{p} - X'D\mathbf{y} = 0. \qquad (7.4.3)$$

Оценка без ограничений получается в результате решения уравнений (7.4.3) в виде

$$\tilde{\tilde{\mathbf{p}}} = (X'\,DX)^{-1} X'\,D\mathbf{y}. \qquad (7.4.4)$$

Эта оценка может и не совпадать с оценкой по обобщенному методу наименьших квадратов без ограничений:

$$\begin{bmatrix} \hat{\mathbf{p}}_* \\ \hat{\mathbf{p}}_r \end{bmatrix} = \begin{bmatrix} (X_*'\,\Sigma_*^{-1} X_*)^{-1} X_*'\,\Sigma_*^{-1} \mathbf{y}_* \\ \mathbf{\eta}_r - R\,\hat{\mathbf{p}}_* \end{bmatrix}. \qquad (7.4.5)$$

[1] Это допущение, очевидно, расходится со сделанными ранее при описании модели предположениями о p_{ij}, оно используется для получения оценки без ограничений.

Оценка с ограничениями по альтернативной модели (7.4.1) получается в результате минимизации (7.4.1) при условиях

$$G\mathbf{p} = \boldsymbol{\eta}_r \tag{7.4.6}$$

и

$$\mathbf{p} \geqslant 0, \tag{7.4.7}$$

где G — $(r \times r^2)$-матрица из r единичных блоков, расположенных в строку, т. е. $[I_1 I_2 \ldots I_r]$; каждая I — единичная $(r \times r)$-матрица. Как и в предыдущих главах, применяя теорему о двойственности и комбинированную формулировку задачи, получаем задачу квадратичного программирования, которую можно представить симплекс-таблицей.

Таблица 7.1

Симплекс-таблица для оценок по критерию χ^2 в предположении независимости параметров

$\mathbf{p}_0$	$\boldsymbol{\lambda}_1$	$\mathbf{p}$	$\boldsymbol{\lambda}_2$	$\boldsymbol{\beta}$
$\boldsymbol{\eta}_r$	0	G	0	0
$X'D\mathbf{y}$	G'	$X'DX$	$-G'$	$-I$

В этой таблице $\boldsymbol{\lambda}_1$ и $\boldsymbol{\lambda}_2$ — векторы множителей Лагранжа и $\boldsymbol{\beta}$ — вектор дополнительных переменных.

Оценка с ограничениями, получающаяся из табл. 7.1, та же, что и оценка, порождаемая процедурой, описанной в параграфе 7.3, или оценка по обобщенному методу наименьших квадратов с ограничениями из главы 6, так как в конечном итоге во всех трех случаях минимизируется одна и та же целевая функция при тех же самых ограничениях.

7.5. ЧИСЛОВОЙ ПРИМЕР

Для того чтобы проиллюстрировать применимость результатов параграфов 7.3 и 7.4 и подчеркнуть важность учета ограничений, рассмотрим задачу оценки для поглощающей цепи со следующими наблюдениями:

$$[X_1, X_2] = \begin{bmatrix} 0{,}5000 & 0{,}5000 \\ 0{,}75 & 0{,}25 \\ 0{,}88 & 0{,}12 \\ 0{,}94 & 0{,}06 \\ 0{,}97 & 0{,}03 \end{bmatrix}, \tag{7.5.1}$$

причем $N(t) = N(t+1) = 100$. Следуя процедуре, описанной в параграфе 7.3, получим оценку Σ_*^{-1}, в данном случае равную

$$\Sigma_*^{-1} = \begin{bmatrix} 533{,}33 & & & \\ & 946{,}37 & & \\ & & 1773{,}00 & \\ & & & 3436{,}40 \end{bmatrix}. \tag{7.5.2}$$

Тогда

$$X_*' \hat{\Sigma}_*^{-1} X_* = \begin{bmatrix} 5075{,}4790 & 691{,}9386 \\ 691{,}9386 & 230{,}4220 \end{bmatrix}, \tag{7.5.3}$$

$$X_* \hat{\Sigma}_*^{-1} \mathbf{y}_* = \begin{bmatrix} 5425{,}000 \\ 808{,}333 \end{bmatrix}. \tag{7.5.4}$$

Обратная к $X_*' \hat{\Sigma}_*^{-1} X_*$ матрица есть

$$(X_*' \hat{\Sigma}_*^{-1} X_*)^{-1} = \begin{bmatrix} 0{,}0003 & -0{,}0010 \\ -0{,}0010 & 0{,}0073 \end{bmatrix}. \tag{7.5.5}$$

Следовательно, оценка по критерию χ^2, или оценка по обобщенному методу наименьших квадратов вектора $\mathbf{p}_1$, есть

$$\begin{bmatrix} \hat{p}_{11} \\ \hat{p}_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 0{,}5051 \end{bmatrix}. \tag{7.5.6}$$

Оценкой исключенного вектора $\mathbf{p}_2$ будет

$$\begin{bmatrix} \hat{p}_{12} \\ \hat{p}_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0{,}4949 \end{bmatrix}. \tag{7.5.7}$$

Таким образом, оценка переходной матрицы имеет вид

$$\hat{P} = \begin{bmatrix} 1 & 0 \\ 0{,}5051 & 0{,}4949 \end{bmatrix}. \tag{7.5.8}$$

В данном случае оценка без ограничений удовлетворяет условию неотрицательности. Поэтому $\hat{P} = \hat{P}_c$.

Далее, следуя процедуре из параграфа 7.4, получим весовую матрицу

$$D = \begin{bmatrix} 133{,}33 & & & & & & & \\ & 113{,}64 & & & & & & \\ & & 106{,}38 & & & & & \\ & & & 103{,}09 & & & & \\ & & & & 400{,}00 & & & \\ & & & & & 833{,}33 & & \\ & & & & & & 1666{,}67 & \\ & & & & & & & 333{,}33 \end{bmatrix}. \tag{7.5.9}$$

Перекрестные произведения, взвешенные с помощью D, равны:

$$X'DX = \begin{bmatrix} 270{,}7295 & 71{,}6886 & & \\ 71{,}6886 & 42{,}3387 & & \\ & & 4804{,}7500 & 620{,}2500 \\ & & 620{,}2500 & 188{,}0833 \end{bmatrix}; \quad (7.5.10)$$

$$X'D\mathbf{y} = \begin{bmatrix} 307 \\ 93 \\ 307 \\ 93 \end{bmatrix}. \quad (7.5.11)$$

Обратная матрица к $X'DX$ равна:

$$(X'DX)^{-1} = \begin{bmatrix} 0{,}0067 & -0{,}0013 & & \\ -0{,}0013 & 0{,}0428 & & \\ & & 0{,}0004 & -0{,}0012 \\ & & -0{,}0012 & 0{,}0093 \end{bmatrix}. \quad (7.5.12)$$

Оценкой переходной матрицы без ограничений в этом случае будет

$$\tilde{\tilde{P}} = [\tilde{\tilde{\mathbf{p}}}_1 \; \tilde{\tilde{\mathbf{p}}}_2] = \begin{bmatrix} 1{,}0012 & 0{,}0001 \\ 0{,}5013 & 0{,}4941 \end{bmatrix}. \quad (7.5.13)$$

Оценка $\tilde{\tilde{p}}_{11}$ противоречит определению вероятности, так как превосходит 1. Оценка с ограничениями, вычисленная по табл. 7.1, при 6 итерациях равна:

$$\hat{P}^c = \begin{bmatrix} 1 & 0 \\ 0{,}5051 & 0{,}4949 \end{bmatrix}, \quad (7.5.14)$$

т. е. она совпадает с оценкой по обобщенному методу наименьших квадратов. Для упрощения вычислений рекомендуется процедура с применением обобщенного метода наименьших квадратов, так как при этом в процессе вычислений исключается ряд параметров.

Глава 8 ● МАКРООЦЕНКИ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Как уже указывалось в предыдущей главе, предполагается, что наблюдаемые относительные частоты порождаются системой полиномиальных распределений. Поскольку распределение выборки известно, можно применить метод максимального правдоподобия для отыскания оценок неизвестных параметров. Хорошо известно, что, в то время как оценки по критерию χ^2 получаются путем минимизации критерия, оценки максимального правдоподобия определяются в результате максимизации функции правдоподобия. Обе оценки обладают одинаковыми свойствами при больших выборках [62]. Эти два подхода можно рассматривать как прямую и двойственную задачи поиска седловой точки.

8.1. ПОЛИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ В СХЕМЕ ЛЕКСИСА

Допустим, что наблюдения $y_j(t)$ $(j = 1, 2, \ldots, r)$, относительные частоты событий, попадающих в j-ю категорию в период t, порождаются множеством из T серий независимых испытаний с $N(t)$ опытами в серии t, причем вероятность попасть в каждое состояние для каждого отдельного индивидуума (микрообъекта) постоянна в данной последовательности испытаний, но она меняется от одной серии к другой. Такую вероятностную модель можно рассматривать как полиномиальное распределение в предположениях схемы Лексиса [93]. Другими словами, в t-й последовательности любое из событий $s_1, s_2, \ldots, s_r$ может произойти в каждом отдельном испытании с вероятностями $q_1(t), q_2(t), \ldots, q_r(t)$ соответственно, причем

$$\sum_{j=1}^{r} q_j(t) = 1, \tag{8.1.1}$$

$$q_j(t) = \sum_{i=1}^{r} y_i(t-1)\, p_{ij}. \tag{8.1.2}$$

Таким образом, если за период t произведено $N(t)$ опытов, вероятность того, что s_1 осуществится $n_1(t)$ раз, s_2 — $n_2(t)$ раз и т. д., где

$$\sum_{i=1}^{r} n_i(t) = N(t), \tag{8.1.3}$$

определяется формулой

$$f(\mathbf{n}(t) \mid \mathbf{q}(t), N(t)) = \frac{N(t)!}{\prod\limits_{i=1}^{r} n_i(t)!} \prod_{j=1}^{r} q_j(t)^{n_j(t)} \tag{8.1.4}$$

с ограничениями (8.1.1). Здесь $\mathbf{n}(t)$ — вектор с элементами $n_i(t)$, $\mathbf{q}(t)' = [q_1(t), q_2(t), \ldots, q_r(t)]$. Совместное распределение вероятностей для результатов серий независимых испытаний в моменты $t = 1, 2, \ldots, T$ есть

$$f(\mathbf{n} \mid \mathbf{q}, \mathbf{N}) = \prod_{t=1}^{T} \frac{N(t)!}{\prod\limits_{i=1}^{r} n_i(t)!} \prod_{j=1}^{r} q_j(t)^{n_j(t)}, \tag{8.1.5}$$

где $\mathbf{q}' = [\mathbf{q}(1)', \mathbf{q}(2)', \ldots, \mathbf{q}(T)']$, $\mathbf{N}' = [N(1), N(2), \ldots, N(T)]$, $\mathbf{n}' = [\mathbf{n}(1)', \ldots, \mathbf{n}(T)']$, $q_j(t)$ определены в (8.1.2)[1]. Поскольку справедливы условия нормировки (8.1.1) и (8.1.3), имеется только $r-1$ независимых переменных $n_j(t)$ и $r-1$ переменных $q_j(t)$ при каждом t. Выразим $n_r(t)$ и $q_r(t)$ через остальные переменные из условий (8.1.1) и (8.1.3). Тогда (8.1.5) перепишется в виде

$$f(\mathbf{n} \mid \mathbf{q}, \mathbf{N}) = \prod_{t=1}^{T} \left[\frac{N(t)}{\left[\prod\limits_{i} n_i(t)!\right]\left[N(t) - \sum\limits_{k} n_k(t)\right]!} \prod_{j} q_j(t)^{n_j(t)} \times \right.$$
$$\left. \times \left(1 - \sum_{k} q_k(t)\right)^{N(t) - \sum\limits_{k} n_k(t)} \right] \tag{8.1.6}$$

при $i, j, k = 1, 2, \ldots, r-1$. Учитывая (8.1.2) и определив $n_i(t)$ по формулам

$$n_i(t) = N(t)\, y_i(t),\ i = 1, 2, \ldots, r, \tag{8.1.7}$$

можно записать выражение для функции правдоподобия, т. е. совместного распределения (8.1.6), рассматриваемого как функция переход-

[1] Аналогичная задача рассмотрена в [62].

ных вероятностей p_{ij} при заданных характеристиках выборки и наблюденных относительных частотах состояний:

$$L(\mathbf{p}\,|\,\mathbf{y}) = \prod_{t=1}^{T} \frac{N(t)}{\prod\limits_{m} (N(t)\, y_m(t))! \left(N(t) - \sum\limits_{k} N(t)\, y_k(t-1)\right)!} \times$$

$$\times \prod_{j} \left(\sum_{i} y_i(t-1)\, p_{ij}\right)^{N(t)\, y_j(t)} \left(1 - \sum_{k} \sum_{i} y_i(t-1)\, p_{ik}\right)^{N(t) - \sum\limits_{k} N(t) y_k(t)}. \tag{8.1.8}$$

Здесь $m,\ j,\ k = 1, 2, \ldots, r-1$; $i = 1, 2, \ldots, r$; $\mathbf{p}$ — вектор-столбец параметров p_{ij} $(i, j = 1, 2, \ldots, r)$ и $\mathbf{y}$ — вектор с элементами $y_i(t)$ и $y_i(t-1)$ $(t = 1, 2, \ldots, T)$. Поскольку переходные вероятности должны удовлетворять некоторым ограничениям, оценки максимального правдоподобия для p_{ij} определяются путем максимизации (8.1.8) при условиях:

$$\sum_{j} p_{ij} = 1,\ i = 1, 2, \ldots, r; \tag{8.1.9}$$

$$0 \leqslant p_{ij} \leqslant 1 \text{ при всех } i,\ j. \tag{8.1.10}$$

Максимизируя функцию правдоподобия (8.1.8) без учета (8.1.9) и (8.1.10), можно и не получить оценок максимального правдоподобия.

8.2. МОДА ФУНКЦИИ ПРАВДОПОДОБИЯ

Как указано в предыдущем разделе, мода функции правдоподобия (8.1.8) может оказаться за пределами области допустимых значений (8.1.10). Поэтому мы вынуждены искать локальный максимум внутри области (8.1.10), а не абсолютный максимум, который может и не попасть в область (8.1.10). Первый определяет оценку с ограничениями, второй — без ограничений. Иногда абсолютный максимум может оказаться внутри допустимой области, и в этом случае локальный максимум является и глобальным, а оценка без ограничений совпадает с оценкой с ограничениями.

Вначале не будем учитывать ограничения (8.1.8) и (8.1.10), наложенные на p_{ij}. Следуя обычной процедуре вывода оценок максимального правдоподобия для параметров функции правдоподобия, приравниваем нулю первые производные логарифмической функции правдоподобия по p_{ij}. Таким образом получаем уравнение

$$\frac{\partial \log L}{\partial p_{ij}} = \sum_{t} \left[\frac{\partial \log L}{\partial q_j(t)} \cdot \frac{\partial q_j(t)}{\partial p_{ij}} \right] = 0 \tag{8.2.1}$$

$$(j = 1, 2, \ldots, r-1),$$

поскольку

$$\frac{\partial \log L}{\partial q_j(t)} = \frac{n_j(t)}{q_j(t)} - \frac{N(t) - \sum\limits_k n_k(t)}{1 - \sum\limits_k q_k(t)} \tag{8.2.2}$$

$(k = 1, 2, \ldots, r-1)$,

$$\frac{\partial q_j(t)}{\partial p_{ij}} = \frac{\partial}{\partial p_{ij}} \sum_l y_l(t-1)\, p_{lj} = y_i(t-1) \tag{8.2.3}$$

$$(i = 1, 2, \ldots, r).$$

Итак, необходимые условия максимума следующие:

$$\sum_t \left[\frac{n_j(t)}{q_j(t)} - \frac{N(t) - \sum\limits_k n_k(t)}{1 - \sum\limits_k q_k(t)} \right] y_i(t-1) = 0 \tag{8.2.4}$$

$$(i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, r-1).$$

Уравнения (8.2.4) можно преобразовать к виду

$$\sum_{t=1}^{T} y_i(t-1) \left(\frac{N(t)}{q_r(t)} + \frac{N(t)}{q_j(t)} \right) y_i(t) + \sum_{k \neq j}^{r-1} \sum_{t=1}^{T} y_i(t-1) \frac{N(t)}{q_r(t)} y_k(t) -$$

$$- \left[\sum_{t=1}^{T} y_i(t-1) \left(\frac{N(t)}{q_r(t)} + \frac{N(t)}{q_j(t)} \right) \sum_{l=1}^{r} y_l(t-1)\, p_{lj} + \right.$$

$$\left. + \sum_{k \neq j}^{r-1} \sum_{t=1}^{T} y_i(t-1) \frac{N(t)}{q_r(t)} \sum_{l=1}^{r} y_l(t-1) p_{lk} \right] = 0 \tag{8.2.5}$$

$$(i = 1, 2, \ldots, r;\ j = 1, 2, \ldots, r-1).$$

В матричных обозначениях эти $r(r-1)$ уравнений можно записать так:

$$X_*' \Sigma_*^{-1} \mathbf{y}_* - X_*' \Sigma_*^{-1} X_* \mathbf{p}_* = \mathbf{0}, \tag{8.2.6}$$

где $\mathbf{p}_*$ — вектор-столбец параметров из $(r-1)$ подвекторов; X_* — блочно-диагональная матрица размера $(T(r-1) \times r(r-1))$ с блоками

$$X_j = \begin{bmatrix} y_1(0) & y_2(0) & \cdots & y_r(0) \\ y_1(1) & y_2(1) & \cdots & y_r(1) \\ \vdots & \vdots & \ddots & \\ y_1(T-1) & y_2(T-1) & \cdots & y_r(T-1) \end{bmatrix}, \quad j = 1, 2, \ldots, r-1, \tag{8.2.7}$$

на главной диагонали; $\mathbf{y}_*$ есть $(T\,(r-1)\times 1)$-вектор из подвекторов

$$\mathbf{y}_j = \begin{bmatrix} y_j(1) \\ y_j(2) \\ \vdots \\ y_j(T) \end{bmatrix} \quad (j = 1, 2, \ldots, r-1) \tag{8.2.8}$$

и Σ_*^{-1} представляет собой $(T\,(r-1)\times T\,(r-1))$-матрицу из $(r-1)^2$ диагональных подматриц размера $(T\times T)$. Другими словами,

$$\Sigma_*^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} & \ldots & \Sigma^{1,r-1} \\ \vdots & & \vdots & \vdots \\ \Sigma^{r-1,1} & \ldots & & \Sigma^{r-1,r-1} \end{bmatrix}, \tag{8.2.9}$$

где

$$\Sigma^{ij} = \begin{bmatrix} \frac{N(1)}{q_r(1)} & & & \\ & \frac{N(2)}{q_r(2)} & & \\ & & \ddots & \\ & & & \frac{N(T)}{q_r(T)} \end{bmatrix} (i \neq j), \tag{8.2.10}$$

$$\Sigma^{jj} = \begin{bmatrix} \frac{N(1)}{q_j(1)} + \frac{N(1)}{q_r(1)} & & & \\ & \frac{N(2)}{q_j(2)} + \frac{N(2)}{q_r(2)} & & \\ & & \ddots & \\ & & & \frac{N(T)}{q_j(T)} + \frac{N(T)}{q_r(T)} \end{bmatrix}. \tag{8.2.11}$$

Матрица Σ_* — ковариационная матрица величин $y_j(t)$, $j = 1, 2, \ldots, r-1$ [61]. Это значит, что Σ_* — матрица из $(r-1)^2$ диагональных блоков Σ_{ij}, таких, что Σ_{ij} есть диагональная $(T\times T)$-матрица с элементами $-[q_i(t)\,q_j(t)]/N(t)$ при $t = 1, 2, \ldots, T$; $i \neq j$, а Σ_{ii} представляет собой диагональную $(T\times T)$-матрицу с элементами $[q_i(t)\times\times(1-q_i(t))]/N(t)$ при $t = 1, 2, \ldots, T$. Итак, если ковариационная матрица Σ_* известна, то оценка максимального правдоподобия без ограничений есть

$$\hat{\mathbf{p}}_* = (X_*'\,\Sigma_*^{-1}\,X_*)^{-1}\,X_*'\,\Sigma_*^{-1}\,\mathbf{y}_*, \tag{8.2.12а}$$

причем исключенный вектор параметров $\mathbf{p}_r$ оценивается из соотношения

$$\hat{\mathbf{p}}_r = \boldsymbol{\eta}_r - R\hat{\mathbf{p}}_*. \tag{8.2.12б}$$

Здесь R есть $(r \times (r-1)\, r)$-матрица вида $[I_1, I_2, \ldots, I_{r-1}]$, где каждая I_l есть единичная матрица размера $(r \times r)$, а η_r — $(r \times 1)$-вектор с единичными компонентами. Решение $(\hat{p}_*, \hat{p}_r)$ единственно и не зависит от того, какой именно столбец P обозначен через p_r, т. е. решение инвариантно относительно подстановок номеров состояний. Оценка (8.2.12) совпадает с оценкой по обобщенному методу наименьших квадратов, описанной в главе 6.

Ковариационную матрицу оценок (8.2.12) можно найти, обратив информационную матрицу[1], определяемую по Фишеру [37] формулой

$$I p_* = -E\,[\partial^2 \log L / \partial p_{lj} \partial p_{kh}]. \tag{8.2.13}$$

Для вычисления (8.2.13) продифференцируем левую часть (8.2.5), после чего, применяя операцию математического ожидания, получим

$$-E \cdot \frac{\partial^2 \log L}{\partial p_{lj}\, \partial p_{kh}} =$$

$$= \begin{cases} \sum\limits_{t=1}^{T} q_l(t-1) \left(\dfrac{N(t)}{q_j(t)} + \dfrac{N(t)}{q_r(t)} \right) q_k(t-1) \left(\dfrac{N(t-1)-1}{N(t-1)} \right), & j = h; \\ \sum\limits_{t=1}^{T} q_l(t-1) \left(\dfrac{N(t)}{q_r(t)} \right) q_k(t-1) \left(\dfrac{N(t-1)-1}{N(t-1)} \right), & j \neq h. \end{cases} \tag{8.2.14}$$

Вывод (8.2.14) основан на следующих соотношениях:

$$E\,(y_l\,(t)) = q_l\,(t);$$
$$E\,(y_l\,(t)\, y_j\,(t)) = [(N\,(t) - 1)/N\,(t)]\, q_l\,(t)\, q_j\,(t);$$
$$E\,(y_l\,(t)\, y_j\,(t-1)) = q_l\,(t)\, q_j\,(t-1);$$
$$E\,(y_l\,(t)\, y_j\,(t-1)\, y_k\,(t-1)) = [(N\,(t-1) - 1)/N\,(t-1)] \times$$
$$\times q_l\,(t)\, q_j(t-1)\, q_k\,(t-1) \text{ и т. д.}$$

При известных $N\,(t)$ вычисления проводятся непосредственно; если объем выборки при каждом t достаточно велик, то поправочный коэффициент $[(N\,(t-1) - 1)/N\,(t-1)]$ можно принять равным 1, а величины $q_l\,(t-1)$, $q_k\,(t-1)$ оценить с помощью соответствующих наблюдаемых относительных частот $y_l\,(t-1)$. Итак, оцениваемая ковариационная матрица $\hat{p}_*$ приближенно равна:

$$V\,(\hat{p}_*) = (X'_* \Sigma_*^{-1} X_*)^{-1}. \tag{8.2.15}$$

Ковариационную матрицу всех параметров можно получить, если перенумеровать состояния системы в новом порядке и повторить описанную процедуру. Отметим, что расширенная матрица $V\,(\hat{p})$ вырождена.

[1] Примеры можно найти в [124].

Благодаря специальной структуре X_* и Σ_* матрица $V(\hat{\mathbf{p}})$ симметрична; сумма ее элементов в любой строке или столбце равна нулю. В качестве примера в табл. 8.1 приведена ковариационная матрица для задачи Телсера [110].

8.3. МАКРООЦЕНКИ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Формулы (8.2.12) определяют оценку без ограничений, так как при ее получении ограничения (8.1.9) и (8.1.10) не принимались во внимание. Для соблюдения всех условий, предъявляемых к переходным вероятностям, требуется оценка максимального правдоподобия с ограничениями. Таким образом, переформулируя исходную задачу, необходимо максимизировать логарифм функции правдоподобия (8.1.8) при ограничениях (8.1.9) и (8.1.10), которые можно записать в виде

$$R\mathbf{p}_* \leqslant \boldsymbol{\eta}_r, \tag{8.3.1}$$

$$\mathbf{p}_* \geqslant 0. \tag{8.3.2}$$

Поскольку нелинейная функция правдоподобия (8.1.8) максимизируется при линейных ограничениях (8.3.1) и (8.3.2), можно, как и в главе 3, применить теорему приводимости нелинейного программирования для отыскания решения задачи с ограничениями. Такой подход приводит к следующей задаче линейного программирования: максимизировать

$$\left(X_* \Sigma_*^{-1} \mathbf{y}_* - X_*' \Sigma_*^{-1} X_* \hat{\mathbf{p}}^c\right)' \mathbf{p}_* \tag{8.3.3}$$

при ограничениях (8.3.1) и (8.3.2). Постоянная часть (8.3.3) получается из формулы (8.2.6), она представляет значение ее левой части в неизвестной оптимальной точке $\hat{\mathbf{p}}_*^c$, т. е. уравнение касательной плоскости к функции правдоподобия в точке локального максимума. Применяя теорему двойственности Дорна [30], приходим к следующей формулировке двойственной задачи: минимизировать

$$\boldsymbol{\lambda}' \boldsymbol{\eta}_r \tag{8.3.4}$$

при условиях

$$R' \boldsymbol{\lambda} \geqslant \left(X_* \Sigma_*^{-1} \mathbf{y}_* - X_*' \Sigma_*^{-1} X_* \hat{\mathbf{p}}_*^c\right) \tag{8.3.5}$$

и

$$\boldsymbol{\lambda} \geqslant 0, \tag{8.3.6}$$

где $\boldsymbol{\lambda}$ — вектор-столбец двойственных переменных. Поскольку как прямая, так и двойственная задачи содержат неизвестное оптимальное решение $\hat{\mathbf{p}}_*^c$, можно сформулировать следующую комбинированную задачу: максимизировать

$$(X_*' \Sigma_*^{-1} \mathbf{y}_* - X_*' \Sigma_*^{-1} X_* \mathbf{p}_*)' \mathbf{p}_* - \boldsymbol{\lambda}' \boldsymbol{\eta}_r = -\boldsymbol{\alpha}' \mathbf{p}_* - \boldsymbol{\lambda}' \mathbf{p}_r \leqslant 0 \tag{8.3.7}$$

Таблица 8.1

Ковариации оценок максимального правдоподобия без ограничений в задаче Телсера

	p_{11}	p_{21}	p_{31}	p_{12}	p_{22}	p_{32}	p_{13}	p_{23}	p_{33}
p_{11}	0,0495								
p_{21}	0,0086	0,0147							
p_{31}	—0,0746	—0,0280	0,1308						
p_{12}	—0,0239	—0,0056	0,0375	0,0433					
p_{22}	—0,0056	—0,0082	0,0164	0,0103	0,0151				
p_{32}	0,0375	0,0164	—0,0682	—0,0692	—0,0300	0,1256			
p_{13}	—0,0257	—0,0031	0,0370	—0,0799	—0,0047	0,0317	0,0456		
p_{23}	—0,0030	—0,0065	0,0116	—0,0048	0,0068	0,0137	0,0078	0,0133	
p_{33}	0,0370	0,0116	—0,0626	0,0316	0,0316	—0,0574	—0,0687	—0,0242	0,1200

при ограничениях

$$R\mathbf{p}_* + \mathbf{p}_r = \eta_r, \tag{8.3.8}$$

$$R'\lambda + (X'_* \Sigma_*^{-1} X_*)\mathbf{p}_* - \alpha = X'_* \Sigma_*^{-1} \mathbf{y}_*, \tag{8.3.9}$$

$$\mathbf{p}_*, \mathbf{p}_r, \lambda, \alpha \geqslant 0, \tag{8.3.10}$$

где $\mathbf{p}_r$ и α — векторы дополнительных переменных. Это та же задача, которая получалась в случае применения обобщенного метода наименьших квадратов с ограничениями, и соответствующая симплекс-таблица совпадает с табл. 6.1 из главы 6.

Ранее при построении оценок p_{ij} предполагалась известной матрица Σ_*^{-1}. На практике Σ_*^{-1} неизвестна и она представляет собой функцию неизвестных p_{ij}. Эту трудность удается обойти с помощью итеративной процедуры с обратной связью для корректировки оценок p_{ij}[1]. Таким образом, первоначально неизвестные элементы $q_j(t)$ в матрице Σ_*^{-1} заменяются своими состоятельными оценками $y_j(t)$. Решая задачу, получаем набор оценок первого приближения $\hat{P}^c(1) = \{\hat{p}_{ij}^c(1)\}$. Тогда, полагая

$$\hat{q}_j^c(t) = \sum_{l=1}^{r} y_l(t-1)\,\hat{p}_{lj}^c(1)$$

и оценивая Σ_*^{-1} с помощью $\hat{q}_j^c(t)$, находим новое решение $\hat{p}_{ij}^c(2)$. Итерации продолжаются до тех пор, пока

$$\hat{p}_{ij}^c(n+1) = \hat{p}_{ij}^c(n).$$

В результате такой рекуррентной процедуры квадратичного программирования получаются искомые оценки[2].

8.4. РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Ввиду совпадения оценок максимального правдоподобия с оценками по обобщенному методу наименьших квадратов и оценками по критерию χ^2 результаты статистических экспериментов те же самые, что и в главе 6.

8.5. НЕКОТОРЫЕ ПРИЛОЖЕНИЯ

а) Результаты для «задачи курильщика». В отличие от первоначального допущения, что данные были получены из фиксированной выборки курильщиков, предположим теперь, что число курильщиков пропорционально числу проданных сигарет. Исходя из среднедушевого потребления — одна пачка в день — было подсчитано число курильщиков в каждом году, на основе чего вычислялась весовая матрица для

[1] О последовательных приближениях для оценок максимального правдоподобия см. [62].

[2] Условия $0 \leqslant p_{ij} < 1$ обеспечивают сходимость итерационного процесса.

оценок максимального правдоподобия. При таком предположении начальным решением для оценок максимального правдоподобия будет

$$\hat{p}^e = [\hat{p}^e_{ij}] = \begin{bmatrix} 0{,}6774 & 0{,}1294 & 0{,}1933 \\ 0{,}0000 & 0{,}8828 & 0{,}1172 \\ 0{,}3935 & 0{,}0000 & 0{,}6065 \end{bmatrix}, \qquad (8.5.1)$$

что весьма близко к результатам по методу наименьших взвешенных квадратов с ограничениями (5.5.1).

При изложении процедуры оценивания по методу максимального правдоподобия указывалось, что более точные оценки получаются, если применять итеративный алгоритм вычислений. Результаты применения его к задаче Телсера представлены в табл. 8.2.

Таблица 8.2

Оценки максимального правдоподобия в задаче о покупателях сигарет трех основных сортов («Camel», «Lucky Strike» и «Chesterfield»), полученные с применением итеративного алгоритма квадратичного программирования

Этапы	Оценки	Итеративная ошибка
Начальное решение	$\begin{bmatrix} 0{,}6776 & 0{,}1290 & 0{,}1934 \\ 0 & 0{,}8828 & 0{,}1172 \\ 0{,}3935 & 0 & 0{,}6065 \end{bmatrix}$	—
1-я итерация	$\begin{bmatrix} 0{,}6699 & 0{,}1421 & 0{,}1880 \\ 0 & 0{,}8710 & 0{,}1290 \\ 0{,}4000 & 0 & 0{,}6000 \end{bmatrix}$	0,0628
2-я итерация	$\begin{bmatrix} 0{,}6689 & 0{,}1424 & 0{,}1887 \\ 0 & 0{,}8707 & 0{,}1293 \\ 0{,}4012 & 0 & 0{,}5988 \end{bmatrix}$	0,0049
3-я итерация	$\begin{bmatrix} 0{,}6689 & 0{,}1425 & 0{,}1886 \\ 0 & 0{,}8707 & 0{,}1293 \\ 0{,}4011 & 0 & 0{,}5989 \end{bmatrix}$	0,0004

Из этой таблицы видно, что решение на четвертом шаге совпадает с предыдущим решением с точностью до четвертого знака. Поэтому последнее решение можно практически считать точным значением оценки максимального правдоподобия. В рассматриваемом случае отличие от начального решения незначительное. Сходимость рекуррентной процедуры хорошо заметна по величине итеративной ошибки, вычисляемой как сумма модулей разностей соответствующих оценок на двух последовательных шагах. В нашем случае итеративная ошибка быстро стремится к нулю.

б) Результаты для задачи об имущественном положении фермеров. В дополнение к задаче Телсера о поведении курильщиков рассмотрим теперь задачу о поведении арендаторов до и после земельной реформы

на Тайване. Другими словами, в рамках марковской модели рассматриваются бихевиористские системы, отражающие упорядоченные во времени изменения в имущественном положении фермеров до и после проведения новой аграрной политики. Программа земельной реформы, принятая в 1949 г., начала эффективно реализовываться в 1953 г.

Хотя упорядоченные во времени данные об имущественном положении отсутствуют, можно воспользоваться данными о процентных долях арендаторов, частичных владельцев и владельцев среди фермеров на острове с 1900 по 1966 г. Для нашего примера предположим, что эти агрегированные данные порождены марковским процессом первого порядка и что система переходных вероятностей оставалась стационарной в течение 1941—1952 гг. Предположим также, что после 1952 г. благодаря земельной реформе переходные вероятности изменились по сравнению с предыдущим периодом, но в дальнейшем оставались постоянными в период 1953—1966 гг. Данные об относительных частотах состояний за это время сведены в табл. 8.3.

Таблица 8.3

Число семей фермеров и их классификация (Тайвань, 1941—1966 гг.)

Год	Общее число семей фермеров	Доли, %		
		владельцы	частичные владельцы	арендаторы
1941	440 105	31	31	38
42	452 462	31	31	38
43	470 374	31	30	39
44	482 776	31	30	39
45	500 533	30	29	41
46	527 016	33	28	39
47	553 308	32	27	41
48	597 333	35	26	39
49	620 875	36	25	39
50	638 062	36	26	38
51	661 125	38	25	37
52	676 750	39	26	35
1953	702 325	55	24	21
54	716 582	57	24	19
55	732 555	59	24	17
56	746 318	60	23	17
57	759 234	60	23	17
58	769 925	61	23	16
59	780 402	62	23	15
60	785 592	64	21	15
61	800 835	65	21	14
62	809 917	65	21	14
63	824 560	66	21	13
64	834 827	66	21	13
65	847 242	67	20	13
66	854 203	67	21	12

Источник. Taiwan Agricultural Yearbook, Department of Agriculture and Forestry, Provincial Government of Taiwan, 1941—1967.

Построенные на их основе оценки максимального правдоподобия переходных вероятностей для обоих периодов приведены в табл. 8.4.

Таблица 8.4

Оценки максимального правдоподобия матрицы переходных вероятностей для периодов до и после проведения аграрной реформы на Тайване

Состояние	1941—1952 гг.			1953—1966 гг.		
	Арендаторы	Частичные владельцы	Владельцы	Арендаторы	Частичные владельцы	Владельцы
Арендаторы	0,86	0,06	0,08	0,66	0,34	0
Частичные владельцы	0,18	0,82	0	0,21	0,54	0,25
Владельцы	0	0,07	0,93	0	0,07	0,93

Полученные результаты, по-видимому, согласуются с ожидаемым экономическим эффектом и позволяют сделать вывод, что проводимая земельная реформа предоставила арендаторам хорошую возможность для изменения их имущественного положения.

Глава 9 ● БАЙЕСОВСКИЙ АНАЛИЗ МАКРОМОДЕЛИ

В главе 2 при выводе выражения для апостериорного распределения переходных вероятностей в соответствии с теоремой Байеса использовалось сочетание априорного распределения (в данном случае многомерного бета-распределения) и функции правдоподобия, построенной по микроданным. В этой главе проводятся аналогичные вычисления апостериорного распределения, но уже по макроданным, с помощью которого будут получены оценки переходных вероятностей, в частности байесовские оценки.

9.1. БАЙЕСОВСКИЙ АНАЛИЗ: АПРИОРНОЕ РАСПРЕДЕЛЕНИЕ

В этом разделе будет следующая общепринятая запись плотности вероятности многомерного бета-распределения [75], [77] для элементов i-й строки матрицы переходных вероятностей:

$$f(\mathbf{p}_i \mid \mathbf{a}_i) = \frac{\Gamma(\boldsymbol{\eta}' \mathbf{a}_i)}{\prod_{j=1}^{r} \Gamma(a_{ij})} \prod_{j=1}^{r} p_{ij}^{a_{ij}-1}, \tag{9.1.1}$$

где $\sum_{j=1}^{r} p_{ij} = 1$, $0 \leqslant p_{ij}$, $j = 1, 2, \ldots, r$, $\mathbf{p}_i' = (p_{i1}, p_{i2}, \ldots, p_{ij})$ — вектор-столбец переходных вероятностей, $\boldsymbol{\eta}' = (1, 1, \ldots, 1)$ — $(1 \times r)$-вектор-столбец единиц, $\mathbf{a}_i' = (a_{i1}, a_{i2}, \ldots, a_{ir})$ — вектор априорных значений параметров $(a_{ij} > 0)$, через $\Gamma(\cdot)$ обозначена гамма-функция.

Из (9.1.1) можно получить маргинальную плотность вероятности для отдельной переходной вероятности. Например, для вероятности p_{ij} такой плотностью вероятности будет плотность вероятности бета-распределения вида[1]

[1] При выводе (9.1.2) из (9.1.1) неоднократно применялись формулы

$$B(m, n) = [1/(a^{m+n-1})] \int_0^a y^{m-1}(a-y)^{n-1}\, dy$$

и

$$B(m, n) = B(n, m).$$

$$f(p_{ij} \mid \mathrm{a}_i) = \frac{1}{B\left(a_{ij}, \sum\limits_{k \neq j}^{r} a_{ik}\right)} p_{ij}^{a_{ij}-1} (1-p_{ij})^{\sum\limits_{k \neq j}^{r} a_{ik}-1}, \qquad (9.1.2)$$

где $0 \leqslant p_{ij} \leqslant 1$ и через $B(\cdot)$ обозначена бета-функция. Известно, что среднее и дисперсия априорного распределения, вычисленные по плотности вероятности (9.1.2), определяются формулами

$$E(p_{ij}) = a_{ij} \Big/ \sum_{j=1}^{r} a_{ij} = a_{ij}/a_i \qquad (9.1.3)$$

и

$$\mathrm{Var}(p_{ij}) = \frac{a_{ij}(a_i - a_{ij})}{a_i^2 (a_i + 1)} = \frac{E(p_{ij})(1 - E(p_{ij}))}{a_i + 1}, \qquad (9.1.4)$$

где $a_i = \sum\limits_{j=1}^{r} a_{ij}$ и a_{ij} — положительные параметры априорной плотности (9.1.2). Далее, априорная ковариация между любыми двумя элементами p_{ik} и p_{ih} i-й строки матрицы переходных вероятностей может быть вычислена по плотности маргинального двумерного распределения вероятностей для p_{ik} и p_{ih}. Она равна:

$$\mathrm{Cov}(p_{ik}, p_{ih}) = -\frac{a_{ik} a_{ih}}{a_i^2 (a_i + 1)} = \frac{E p_{ik} E p_{ih}}{a_i + 1}. \qquad (9.1.5)$$

Если априори предполагается, что все строки матрицы переходных вероятностей распределены независимо в соответствии с плотностью вероятности вида (9.1.1), то совместная априорная плотность вероятности элементов матрицы $P = \{p_{ij}\}$ определяется призведением функций вида (9.1.1). Другими словами, совместная априорная плотность вероятности равна:

$$f(P \mid A) = \prod_{i=1}^{r} f(\mathrm{p}_i \mid \mathrm{a}_i) = \prod_{i=1}^{r} \frac{\Gamma(a_i)}{\prod\limits_{j=1}^{r} \Gamma(a_{ij})} \prod_{j=1}^{r} p_{ij}^{a_{ij}-1}, \qquad (9.1.6)$$

где $A = \{a_{ij}\}$ — $(r \times r)$-матрица априорных значений параметров, $\eta' \mathrm{p}_i = 1$ для $i = 1, 2, \ldots, r$ и $0 \leqslant p_{ij} \leqslant 1$. Совместная плотность (9.1.6) представлена в матричной бета-форме [75].

Априорная плотность вероятности (9.1.6) далее будет сочетаться с функцией правдоподобия для получения апостериорной плотности вероятности элементов матрицы P. По-видимому, (9.1.6) в ряде приложений будет достаточно удобным представлением априорной информации.

9.2. ФУНКЦИЯ ПРАВДОПОДОБИЯ И АПОСТЕРИОРНАЯ ПЛОТНОСТЬ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Предположим, что наши данные представлены в форме агрегированных относительных частот, являющихся относительным числом микрообъектов в каждом состоянии и в каждом из T временных интервалов. Как уже объяснялось в главе 8, можно полагать, что эти данные получены с помощью мультиномиального процесса по схеме Лексиса. При этом предположении функция правдоподобия имеет вид[1]

$$L(\mathbf{p}_* \mid \mathbf{y}) = \prod_{t=1}^{T} \frac{N(t)!}{\prod_m [N(t)\, y_m(t)]!\, [N(t) - \sum_k N(t)\, y_k(t)]!} \times$$

$$\times \prod_j \left[\sum_i y_i(t-1)\, p_{ij}\right]^{N(t)\, y_j(t)} \left[1 - \sum_k \sum_i y_i(t-1)\, p_{ik}\right]^{N(t)\,[1 - \Sigma_k y_k(t)]}, \tag{9.2.1}$$

где $j, k, m = 1, 2, \ldots, r-1$, $i = 1, 2, \ldots, r$. Далее, в (9.2.1) $\mathbf{p}'_* = (\mathbf{p}'_1, \mathbf{p}'_2, \ldots, \mathbf{p}'_{r-1})$ — вектор переходных вероятностей, каждая из компонент p_j которого представляет собой $(r \times 1)$-вектор, $\mathbf{y}' = [\mathbf{y}(0)', \mathbf{y}(1)', \ldots, \mathbf{y}(T)']$ — вектор наблюдаемых относительных частот. Отметим, что вектор начальных наблюдений $\mathbf{y}(0)$ предполагается заданным.

При совместном использовании (9.1.6) и функции правдоподобия (9.2.1) можно получить апостериорную плотность вероятности для переходных вероятностей вида

$$f(\mathbf{p}_* \mid \mathbf{n}) = K \left[\prod_{t=1}^{T} \prod_{j=1}^{r-1} q_j(t)^{n_j(t)} \left[1 - \sum_{m=1}^{r-1} q_m(t)\right]^{N(t) - \sum_{m=1}^{r-1} n_m(t)} \times\right.$$

$$\left.\times \prod_{i=1}^{r} \prod_{j=1}^{r-1} p_{ij}^{a_{ij}-1} \left[1 - \sum_{k=1}^{r-1} p_{ik}\right]^{a_i - \left(\sum_{k=1}^{r-1} a_{ik} - 1\right)}\right], \tag{9.2.2}$$

где K — константа нормировки, не зависящая от переходных вероятностей, $n_j(t) = N(t)\, y_j(t)$, $\mathbf{n}' = [\mathbf{n}(0)', \mathbf{n}(1)', \ldots, \mathbf{n}(T)']$ и $q_j(t) = \sum_{i=1}^{r} y_i(t-1)\, p_{ij}$. В дальнейшем будут проанализированы несколько свойств апостериорной плотности вероятности (9.2.2).

---

[1] Эта функция правдоподобия подробно рассмотрена в главе 8.

9.3. МОДА АПОСТЕРИОРНОГО РАСПРЕДЕЛЕНИЯ

В этом разделе введем в рассмотрение величину p_*, равную моде апостериорного распределения (9.2.2). Мода, как известно, служит мерой локализации максимума апостериорной плотности вероятности. Если считать, что апостериорное распределение представляет собой характеристику степени нашего незнания, то вероятность нахождения истинного значения оцениваемого параметра в малой окрестности p_* больше вероятности нахождения истинного значения в малой окрестности любой другой точки.

Очевидно, что мода апостериорного распределения должна принадлежать следующей допустимой области пространства параметров:

$$\sum_{j=1}^{r} p_{ij} = 1, \qquad i = 1, 2, \ldots, r \tag{9.3.1}$$

и

$$0 \leqslant p_{ij} \leqslant 1 \text{ для всех } i, j. \tag{9.3.2}$$

Вначале при отыскании моды распределения (9.2.2) не будем учитывать условие (9.3.2). Если оказывается, что решение этой задачи удовлетворяет (9.3.2), тогда в допустимой области пространства параметров получаем точечную оценку. Если же (9.3.2) не удовлетворяется, тогда придется воспользоваться аппаратом математического программирования, который рассматривается в последующих параграфах этой главы.

Логарифмируя (9.2.2), дифференцируя полученное выражение по p_{ij} и приравнивая производные нулю, получим

$$\sum_{t=1}^{T} \left[\frac{n_j(t)}{q_j(t)} - \frac{N(t) - \sum\limits_{m=1}^{r-1} n_m(t)}{1 - \sum\limits_{m=1}^{r-1} q_m(t)} \right] y_i(t-1) + \frac{a_{ij} - 1}{p_{ij}} -$$

$$- \frac{a_i - \left(\sum\limits_{k=1}^{r-1} a_{ik} - 1 \right)}{1 - \sum\limits_{k=1}^{r-1} p_{ik}} = 0, \tag{9.3.3}$$

где $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, r-1$. Решение этих $r(r-1)$ уравнений относительно $r(r-1)$ величин p_{ij}, удовлетворяющих (9.3.1), оказывается трудной задачей. Однако заметим, что $r(r-1)$ соотношений в (9.3.3) могут быть представлены в следующем виде:

$$\sum_{t=1}^{T} y_i(t-1)\left[\frac{N(t)}{q_r(t)}+\frac{N(t)}{q_j(t)}\right]y_i(t)+\sum_{t=1}^{T}\sum_{k\neq j}^{r-1} y_i(t-1)\times$$

$$\times\left[\frac{N(t)}{q_r(t)}\right]y_k(t)-\sum_{t=1}^{T} y_i(t-1)\left[\frac{N(t)}{q_r(t)}+\frac{N(t)}{q_j(t)}\right]\sum_{l=1}^{r} y_i(t-1)\,p_{lj}-$$

$$-\sum_{t=1}^{T}\sum_{k\neq 1}^{r-1} y_i(t-1)\left[\frac{N(t)}{q_r(t)}\right]\sum_{l=1}^{r-1} y_i(t-1)\,p_{lk}+$$

$$+\left[\frac{a_i+1}{p_{ir}}+\frac{a_i+1}{p_{ij}}\right]\left[\frac{a_{ij}-1}{a_i+1}\right]+\sum_{k\neq j}^{r-1}\left[\frac{a_i+1}{p_{ir}}\right]\left[\frac{a_{ik}-1}{a_i+1}\right]-$$

$$-\left[\frac{a_i+1}{p_{ir}}+\frac{a_i+1}{p_{ij}}\right]\left[\frac{a_i-r}{a_i+1}\right]p_{ij}-\sum_{k\neq j}^{r-1}\left[\frac{a_i+1}{p_{ir}}\right]\left[\frac{a_i-r}{a_i+1}\right]p_{ik}=0, \qquad (9.3.4)$$

где $i=1, 2, \ldots, r$, $j=1, 2, \ldots, r-1$. В матричных обозначениях (9.3.4) может быть переписано:

$$X'_*\Sigma_*^{-1}\mathbf{y}_*-X'_*\Sigma_*^{-1}X_*\mathbf{p}_*+\Sigma_0^{-1}\mathbf{f}-\Sigma_0^{-1}J\mathbf{p}_*=0, \qquad (9.3.5)$$

где Σ_0^{-1} — блочная $(r(r-1)\times r(r-1))$-матрица, состоящая из $(r-1)^2$ диагональных подматриц Σ_0^{kj}, $k, j=1, 2, \ldots, r-1$. Каждая из подматриц Σ_0^{jj}, $j=1, 2, \ldots, r-1$, расположенных на диагонали матрицы Σ_0^{-1}, представляет собой $(r\times r)$-матрицу с диагональными элементами вида $(a_i+1)(1/p_{ir}+1/p_{ij})$, $i=1, 2, \ldots, r$. Каждая из подматриц Σ_0^{kj}, $k\neq j$, расположенных вне диагонали матрицы Σ_0^{-1}, также будет $(r\times r)$-матрицей, но с диагональными элементами, равными $(a_i+1)/p_{ir}$, $i=1, 2, \ldots, r$[1]. В соотношении (9.3.5) матрица J есть диагональная $(r(r-1)\times r(r-1))$-матрица, состоящая из $(r-1)$ одинаковых диагональных блоков размера $(r\times r)$ с элементами вида $(a_i-r)/(a_i+1)$, $i=1, 2, \ldots, r$, а $\mathbf{f}$ — $(r(r-1)\times 1)$-вектор-стол-

[1] В качестве более сильного предположения об априорном распределении можно принять, вообще говоря, ниоткуда не следующее предположение о том, что все p_{ij} независимы и каждая переходная вероятность p_{ij} распределена в соответствии с одномерным бета-распределением. Тогда плотность вероятности совместного распределения $r(r-1)$ независимых параметров $(p_{ij}, i=1, 2, \ldots, r; j=1, 2, \ldots, r-1)$ имеет вид

$$f(\mathbf{p}_*)=\prod_{ij}[1/B(a_{ij}, b_{ij})]\times p_{ij}^{a_{ij}-1}(1-p_{ij})^{b_{ij}-1}.$$

Эта априорная плотность — частный случай (9.1.6), когда ковариации (9.1.5) равны нулю. Окончательная байесовская оценка аналогична (9.3.5), она отличается от нее лишь тем, что матрица Σ_0^{-1} становится диагональной, так как соответствующие ковариации обращаются в нуль.

бец, компоненты которого равны $(a_{ij} - 1)/(a_i + 1)$, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, r - 1$.

Поскольку матрицы Σ_*^{-1} и Σ_0^{-1} содержат элементы с неизвестными параметрами p_{ij}, определим моду апостериорного распределения (9.2.2) следующим образом. Подставим априорные средние значения для p_{ij} в элементы матриц Σ_*^{-1} и Σ_0^{-1} и обозначим полученные матрицы соответственно через $\hat{\Sigma}_*^{-1}$ и $\hat{\Sigma}_0^{-1}$. Тогда решение уравнения (9.3.5) дает первое приближение моды апостериорного распределения:

$$\tilde{\mathbf{p}}_*(1) = (X_*' \hat{\Sigma}_*^{-1} X_* + \hat{\Sigma}_0^{-1} J)^{-1} (X_*' \hat{\Sigma}_*^{-1} \mathbf{y}_* + \hat{\Sigma}_0^{-1} \mathbf{f}). \qquad (9.3.6а)$$

Это приближение может быть улучшено путем подстановки элементов вектора $\tilde{\mathbf{p}}_*(1)$ в выражения для Σ_*^{-1} и Σ_0^{-1} и повторного решения уравнения (9.3.5) относительно нового приближения моды $\tilde{\mathbf{p}}_*(2)$. Подобная итеративная процедура может быть продолжена до тех пор, пока результаты не установятся на некотором значении $\tilde{\mathbf{p}}_*$, которое и будет модой апостериорного распределения. Мода распределения исключенного вектора параметров $\mathbf{p}_r$ дается формулой

$$\tilde{\mathbf{p}}_r = \eta_r - R\tilde{\mathbf{p}}_*. \qquad (9.3.6б)$$

Как будет показано далее, приближенная оценка ковариационной матрицы апостериорного распределения (9.2.2) записывается в виде

$$V(\mathbf{p}_*) = (X_*' \hat{\Sigma}_*^{-1} X_* + \hat{\Sigma}_0^{-1} J)^{-1}, \qquad (9.3.7а)$$

где $\hat{\Sigma}_*^{-1}$ и $\hat{\Sigma}_0^{-1}$ — определенные ранее матрицы, в элементах которых p_{ij} заменено соответствующими компонентами вектора моды апостериорного распределения, т. е. компонентами вектора $\tilde{\mathbf{p}}_*$. Наконец, приближенная оценка апостериорной ковариационной матрицы исключенного вектора параметров дается формулой

$$\mathrm{Var}(\mathbf{p}_r) = RV(\mathbf{p}_*)R' = R(X_*' \hat{\Sigma}_*^{-1} X_* + \hat{\Sigma}_0^{-1} J)^{-1} R'. \qquad (9.3.7б)$$

9.4. СОПОСТАВЛЕНИЕ С НЕКОТОРЫМИ РЕЗУЛЬТАТАМИ ТЕОРИИ ВЫБОРОК

Хорошо известно[1], что при некоторых не очень жестких общих условиях апостериорное распределение с увеличением объема выборки сходится к нормальному распределению со средним значением (модой), равным оценке максимального правдоподобия. Далее, при больших выборках ковариационная матрица апостериорного распределения может быть достаточно хорошо приближена матрицей, обратной к оценке информационной матрицы. Имеются следующие эвристические

[1] См., например, [52], [71], [122].

соображения: в тех случаях, когда объем выборки достаточно велик, выборочная информация доминирует над априорной информацией (если априорная информация неполная) и апостериорное распределение хорошо аппроксимируется функцией правдоподобия. Эта функция при некоторых общих условиях приобретает форму нормального распределения с математическим ожиданием, равным оценке максимального правдоподобия. Поэтому не слишком удивительно, что для больших выборок результаты при байесовском подходе и методе максимального правдоподобия совпадают. Однако следует понимать, что результаты, получаемые с помощью этих методов, интерпретируются совершенно по-разному.

Интересно также отметить, что в тех случаях, когда параметры a_{ij} априорного распределения велики, матрица J в (9.3.5) с элементами $(a_i - 1)/(a_i + 1)$ будет близка к единичной матрице. Далее, при больших a_{ij} координаты вектора $\mathbf{f}$ в (9.3.5) вида $(a_{ij} - r)/(a_i + 1)$ близки к величинам a_{ij}/a_i, т. е. априорным средним. При этих условиях из (9.3.5) можно получить соотношение

$$\mathbf{p}_* = (X_*' \Sigma_*^{-1} X_* + \Sigma_0^{-1})^{-1} (X_*' \Sigma_*^{-1} \mathbf{y}_* + \Sigma_0^{-1} \mathbf{p}_0), \tag{9.4.1}$$

где $\mathbf{p}_0$ — априорное среднее значение вектора $\mathbf{p}_*$ с координатами a_{ij}/a_i, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, r - 1$. Если с помощью выборочных данных получить оценки матриц Σ_* и Σ_0^{-1} и эти оценки подставить в (9.4.1), то в условиях сделанных предположений о величинах a_{ij} получим моду апостериорного распределения.

Выражение (9.4.1) может быть также получено с помощью модификации предложенного в [113] выборочного метода для включения априорной информации в анализ обычной линейной регрессионной модели[1]. В данном случае мы поступим следующим образом. Рассмотрим модель наблюдений

$$\mathbf{y}_* = X_* \mathbf{p}_* + \mathbf{u}_*, \tag{9.4.2}$$

где

$$E\mathbf{u}_* = \mathbf{0}, \tag{9.4.3}$$

$$E\mathbf{u}_* \mathbf{u}_*' = \Sigma_*, \tag{9.4.4}$$

а матрица Σ_* предполагается невырожденной. В методе из [113] априорная информация учитывается в виде

$$\mathbf{p}_0 = \mathbf{p}_* + \mathbf{d}. \tag{9.4.5}$$

---

[1] В работе [117] в связи с анализом многомерной регрессионной модели было показано, что метод, описанный в [113], дает оценку вектора коэффициентов уравнения регрессии, которая совпадает со средним значением главного нормального члена асимптотического разложения, аппроксимирующего апостериорное распределение вектора коэффициентов регрессии.

В этом выражении $\mathbf{p}_0$ — вектор наблюдений, $\mathbf{p}_*$ — вектор фиксированных неизвестных параметров, $\mathbf{d}$ — вектор ошибок, обладающий следующими свойствами:

$$E\mathbf{d} = 0 \tag{9.4.6}$$

и

$$E\mathbf{d}\mathbf{d}' = \Sigma_0. \tag{9.4.7}$$

В (9.4.7) Σ_0 — невырожденная матрица с заданными заранее элементами. Далее, предполагается также, что

$$E\mathbf{d}\mathbf{u}' = 0, \tag{9.4.8}$$

т. е. ошибки в уравнениях (9.4.2) некоррелированы с ошибками в соотношениях (9.4.5), которые характеризуют точность априорной информации.

Теперь (9.4.2) и (9.4.5) перепишем в виде

$$\begin{pmatrix} \mathbf{y}_* \\ \mathbf{p}_0 \end{pmatrix} = \begin{pmatrix} X_* \\ I \end{pmatrix} \mathbf{p}_* + \begin{pmatrix} \mathbf{u}_* \\ \mathbf{d} \end{pmatrix}. \tag{9.4.9}$$

Из анализа (9.4.9) становится ясно, что рассматриваемый метод состоит в дописывании дополнительных наблюдений (элементов вектора $\mathbf{p}_0$) к вектору наблюдений $\mathbf{y}_*$[1].

При ранее введенных предположениях ковариационная матрица вектора ошибок в (9.4.9) дается формулой

$$\operatorname{Var}\begin{pmatrix} \mathbf{u}_* \\ \mathbf{d} \end{pmatrix} = \begin{pmatrix} \Sigma_* & 0 \\ 0 & \Sigma_0 \end{pmatrix}. \tag{9.4.10}$$

Применение обобщенного метода наименьших квадратов для решения системы (9.4.9) дает

$$\bar{\mathbf{p}}_* = \left[\begin{pmatrix} X_* \\ I \end{pmatrix}' \begin{pmatrix} \Sigma_* & 0 \\ 0 & \Sigma_0 \end{pmatrix}^{-1} \begin{pmatrix} X_* \\ I \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} X_* \\ I \end{pmatrix}' \begin{pmatrix} \Sigma_* & 0 \\ 0 & \Sigma_0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_* \\ \mathbf{p}_0 \end{pmatrix}\right] =$$

$$= (X_*' \Sigma_*^{-1} X_* + \Sigma_0^{-1})^{-1} (X_*' \Sigma_*^{-1} \mathbf{y}_* + \Sigma_0^{-1} \mathbf{p}_0). \tag{9.4.11}$$

Заметим, что по виду (9.4.11) в точности совпадает с (9.4.1).

Несмотря на то что $\bar{\mathbf{p}}_*$ в (9.4.11) — оптимальная оценка по методу наименьших квадратов, следует иметь в виду, что $\bar{\mathbf{p}}_*$ зависит от матрицы Σ_*, элементы которой неизвестны и зависят, как было показано,

[1] Из (9.4.9), (9.4.6) следует, что $E\mathbf{p}_0 = \mathbf{p}_*$, т. е. предполагается, что среднее значение вектора дополнительных наблюдений равно истинному значению вектора $\mathbf{p}_*$. Этому очень сильному условию должна удовлетворять априорная информация. В случае байесовского подхода это условие не накладывается.

от подлежащих оценке параметров p_{ij}. Если подобно тому, как это делается в обычной регрессионной задаче, подставить выборочную оценку матрицы Σ_* в (9.4.11), то оценка $\bar{\mathbf{p}}_*$ оказывается достаточно хорошей в случае больших выборок. Однако в этом случае выборочная информация доминирует над априорной информацией. Заметим, что элементы матрицы $X_*'\Sigma_*^{-1}X_*$ растут по величине с увеличением объема выборки, в то время как элементы матрицы Σ_0^{-1} не изменяются. Априорная информация, представленная выражениями (9.4.5) — (9.4.7), не включает в себя ограничения вида $0 \leqslant p_{ij} \leqslant 1$ и $\sum_{j=1}^{r} p_{ij} = 1$, $i = 1, 2, \ldots, r$. Таким образом, $\bar{\mathbf{p}}_*$ или приближение вектора $\bar{\mathbf{p}}_*$, построенное по оценке матрицы Σ_*, может привести к таким оценкам переходных вероятностей, которые не попадают в область допустимых значений пространства параметров.

9.5. МАКРОБАЙЕСОВСКАЯ ОЦЕНКА ПЕРЕХОДНОЙ ВЕРОЯТНОСТИ

Вектор $\tilde{\mathbf{p}}_*$ в (9.3.6а) будет приемлемой оценкой моды апостериорного распределения (9.2.2) при условии, что $\tilde{\mathbf{p}}_*$ попадает в допустимую область пространства параметров, т. е. удовлетворяет ограничению $0 \leqslant \tilde{p}_{ij} \leqslant 1$ и условию нормировки. Если в действительности эти условия выполняются, то найденные $\tilde{p}_{ij}$ могут служить точечными оценками. С другой стороны, может оказаться, что (9.3.6) дает оценки, лежащие вне допустимой области пространства параметров. Для этого случая далее рассматривается процедура максимизации апостериорной плотности вероятности при условии ограничений (9.3.1), (9.3.2).

Действуя, как и ранее, сведем исходную задачу квадратичного программирования к эквивалентной задаче линейного программирования: максимизировать

$$\left(X_*'\Sigma_*^{-1}\mathbf{y}_* - X_*'\Sigma_*^{-1}X_*\tilde{\mathbf{p}}_*^c + \Sigma_0^{-1}\mathbf{f} - \Sigma_0^{-1}J\tilde{\mathbf{p}}_*^c\right)'\mathbf{p}_* \qquad (9.5.1)$$

при условиях

$$R\mathbf{p}_* \leqslant \boldsymbol{\eta}_r \qquad (9.5.2)$$

и

$$\mathbf{p}_* \geqslant 0, \qquad (9.5.3)$$

где $R = (I_1, I_2, \ldots, I_{r-1})$, I_i — единичная $(r \times r)$-матрица, $\boldsymbol{\eta}_r$ есть $(r \times 1)$-вектор-столбец единиц. Согласно теореме двойственности [27], если поставленная прямая задача имеет оптимальное решение, то существует оптимальное решение следующей двойственной задачи: минимизировать

$$\boldsymbol{\lambda}'\boldsymbol{\eta}_r \qquad (9.5.4)$$

при условиях

$$R'\boldsymbol{\lambda} \geqslant X_*'\Sigma_*^{-1}\mathbf{y}_* + \Sigma_0^{-1}\mathbf{f} - X_*'\Sigma_*^{-1}X_*\tilde{\mathbf{p}}_*^c - \Sigma_0^{-1}J\tilde{\mathbf{p}}_*^c \qquad (9.5.5)$$

$$\lambda \geqslant 0, \tag{9.5.6}$$

где λ — вектор-столбец двойственных переменных размера $(r \times 1)$.

Решения прямой и двойственной задач можно получить сразу, решая следующую задачу математического программирования: минимизировать

$$(X_*' \Sigma_*^{-1} y_* + \Sigma_0^{-1} f - X_*' \Sigma_*^{-1} X_* p_* - \Sigma_0^{-1} J p)' p_* - \lambda' \eta_r =$$

$$= -\alpha p_* - \lambda' p_r \leqslant 0 \tag{9.5.7}$$

при условиях

$$R p_* + p_r = \eta_r, \tag{9.5.8}$$

$$R' \lambda + (X_*' \Sigma_*^{-1} X_* + \Sigma_0^{-1} J) p_* - \alpha = X_*' \Sigma_*^{-1} y_* + \Sigma_0^{-1} f \tag{9.5.9}$$

и

$$p_*, \lambda, p_r, \alpha \geqslant 0, \tag{9.5.10}$$

где p_r и α — $(r \times 1)$-векторы дополнительных переменных. Вектор p_r исчезает в случае, когда матрица Σ_* невырожденная. Симплекс-таблицу для этой задачи представляет собой табл. 9.1.

Таблица 9.1

Симплекс-таблица для байесовской оценки в случае априорного многомерного бета-распределения

B_0	$\lambda \geqslant 0$	$p_* \geqslant 0$	$p_r \geqslant 0$	$\alpha \geqslant 0$
η_r	0	R	I	0
$X_*' \Sigma_*^{-1} y_* + \Sigma_0^{-1} f$	R'	$X_*' \Sigma_*^{-1} X_* + \Sigma_0^{-1} J$	0	$-I$

В случае, когда параметры плотности вероятности многомерного априорного бета-распределения велики, априорная плотность вероятности стремится к нормальной при стремлении J к I и f к p_0. В этом случае симплекс-таблица преобразуется к виду табл. 9.2.

Таблица 9.2

Симплекс-таблица для байесовской оценки в случае априорного многомерного нормального распределения

B_0	$\lambda \geqslant 0$	$p_* \geqslant 0$	$p_r \geqslant 0$	$\alpha \geqslant 0$
η_r	0	R	I	0
$X_*' \Sigma_*^{-1} y_* + \Sigma_0^{-1} p_0$	R'	$X_*' \Sigma_*^{-1} X_* + \Sigma_0^{-1}$	0	$-I$

В параграфе 9.2 рассматривалось апостериорное распределение переходных вероятностей, которое объединяло в себе априорную и выборочную информацию. В случае (2×2)-матрицы переходных вероятностей (т. е. в случае двух состояний) для получения нормированного двумерного апостериорного распределения параметров p_{11} и p_{21} и частных апостериорных плотностей может применяться численная техника двумерного интегрирования. Эти распределения могут оказаться необходимыми для вынесения вероятностных суждений относительно параметров p_{11} и p_{21}. Далее, в силу того что $p_{11} + p_{12} = 1$ и $p_{21} + p_{22} = 1$, относительно p_{12} и p_{22} также могут быть сделаны апостериорные вероятностные выводы. Кроме того, если задана функция потерь, то оптимальные точечные оценки могут быть найдены численно. Например, если функция потерь — квадратичная функция, то, как известно, среднее значение апостериорного распределения будет оптимальной оценкой. В случае двух состояний среднее значение апостериорного распределения может быть вычислено с помощью техники численного интегрирования. Таким образом, в случае двух состояний может быть проведен полный анализ апостериорного распределения.

Для систем с более чем двумя состояниями число параметров матрицы переходных вероятностей сравнительно велико. В этом случае известные авторам обычные методы численного интегрирования неудобны для анализа апостериорного распределения. Учитывая сказанное, а также то, что апостериорное распределение достаточно сложное, мы привели выражение для моды апостериорного распределения и приближенное выражение для ковариационной матрицы. Для дальнейшего исследования апостериорного распределения необходимо воспользоваться аналитическими и численными методами.

Анализ, проведенный в первых параграфах этой главы, показывает, что моду апостериорного распределения переходных вероятностей можно рассматривать как взвешенное среднее выборочной и априорной информации (см., например, (9.3.6а)). Матрицы Σ_{θ}^{-1} и Σ_{0}^{-1}, обратные соответственно к матрице ошибок и априорной ковариационной матрице, играют важную роль при взвешивании выборочной и априорной информации о значениях переходных вероятностей. Согласно байесовскому подходу, если априорная плотность вероятности характеризуется малой дисперсией, то область, в которой апостериорная плотность отлична от нуля, будет существенно зависеть от априорных средних значений переходных вероятностей. Однако это характерно в случае малых и средних выборок. С другой стороны, если априорная плотность вероятности сильно «размыта», то выборочная информация будет играть более значительную роль при определении области, соответствующей отличным от нуля значениям апостериорной плотности. Если, например, в качестве априорного распределения переходных вероятностей выбрано равномерное распределение, то апостериорное распределение пропорционально функции правдоподобия. Естественно, что при этом мода апостериорного распределения совпадает с оцен-

кой максимального правдоподобия[1]. В случае больших выборок выборочная информация оказывает наибольшее влияние на апостериорное распределение, и, таким образом, вид априорного распределения, если только он точно не известен, мало сказывается на свойствах апостериорного распределения.

Во многих случаях интересно выяснить, как выборочная информация видоизменяет априорные предположения о переходных вероятностях. Если априорная информация представима в виде многомерной плотности вероятности бета-распределения (9.1.1), то можно объединить это распределение и функцию правдоподобия и сравнить свойства результирующего апостериорного распределения с соответствующими свойствами априорного распределения. Кроме того, если априорная информация учитывается при анализе выборочных данных, то окончательные выводы, как уже говорилось, будут основываться как на выборочной, так и на априорной информации.

9.7. ЧИСЛОВОЙ ПРИМЕР

Для иллюстрации техники вычислений рассмотрим простую поглощающую цепь, определенную в главе 7. Агрегированные данные (7.5.1) необходимы для получения матричных произведений $X_*' \Sigma_*^{-1} X_*$ и $X_*' \Sigma_*^{-1} y_*$, определенных согласно (7.5.3) и (7.5.4).

Плотность вероятности многомерного бета-распределения в этом случае, т. е. в случае марковской цепи с двумя состояниями, превращается в одномерную плотность вероятности бета-распределения. Перечень априорных значений параметров приводен в табл. 9.3.

Таблица 9.3

Априорные значения параметров бета-распределений

p_{ij}	a_{ij}	$a_i - a_{ij}$	Среднее значение	Дисперсия
p_{11}	99	1	0,9900	0,0001
p_{12}	1	99	0,0100	0,0001
p_{21}	50	50	0,5000	0,0025
p_{22}	50	50	0,5000	0,0025

Необходимые для вычисления оценок матричные произведения для выборочной и априорной информации имеют вид:

$$X_*' \Sigma_*^{-1} X_* + \Sigma_0^{-1} = \begin{bmatrix} 5075{,}4790 & 691{,}9386 \\ 691{,}9386 & 230{,}4220 \end{bmatrix} + \begin{bmatrix} 9898{,}9900 & 0 \\ 0 & 392{,}0000 \end{bmatrix} =$$

$$= \begin{bmatrix} 14974{,}4690 & 691{,}9386 \\ 691{,}9386 & 622{,}4220 \end{bmatrix} \qquad (9.7.1)$$

[1] Среднее значение апостериорного распределения, т. е. оптимальная точечная оценка при квадратичной функции потерь, может заметно отличаться от оценки максимального правдоподобия в случае малых выборок.

и

$$X_*' \Sigma_*^{-1} X_* + \Sigma_0^{-1} \mathbf{p}_0 = \begin{bmatrix} 5425{,}00 \\ 808{,}33 \end{bmatrix} +$$

$$+ \begin{bmatrix} 9898{,}99 & 0 \\ 0 & 392{,}00 \end{bmatrix} \begin{bmatrix} 0{,}99 \\ 0{,}50 \end{bmatrix} = \begin{bmatrix} 15323{,}99 \\ 1004{,}33 \end{bmatrix}. \qquad (9.7.2)$$

Оценки без учета ограничений равны:

$$\begin{bmatrix} \tilde{p}_{11} \\ \tilde{p}_{21} \end{bmatrix} = \begin{bmatrix} 0{,}00007 & -0{,}00008 \\ -0{,}00008 & 0{,}00169 \end{bmatrix} \begin{bmatrix} 15323{,}99 \\ 1004{,}33 \end{bmatrix} = \begin{bmatrix} 1{,}0002 \\ 0{,}5017 \end{bmatrix} \qquad (9.7.3)$$

и

$$\begin{bmatrix} \tilde{p}_{12} \\ \tilde{p}_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1{,}0002 \\ 0{,}5017 \end{bmatrix} = \begin{bmatrix} -0{,}0002 \\ 0{,}4933 \end{bmatrix}. \qquad (9.7.4)$$

Оценки (9.7.3) и (9.7.4) не удовлетворяют условию неотрицательности, и, следовательно, в этом случае необходимо воспользоваться итеративной процедурой. В случае квадратичных итераций с самого начала вводятся в базис p_{12} и p_{22}. Точно после трех итераций оценка p_{11} вытесняет оценку p_{12}, а p_{21} и α_1 вытесняют искусственные переменные. Получаемое решение будет оптимальным. Окончательные байесовские оценки равны

$$\tilde{p}^e = \begin{bmatrix} 1 & 0 \\ 0{,}5019 & 0{,}4981 \end{bmatrix}. \qquad (9.7.5)$$

9.8. РЕЗУЛЬТАТЫ ВЫБОРОЧНОГО ЭКСПЕРИМЕНТА

В описываемом далее выборочном эксперименте взяты дополнительные данные, относящиеся к интервалам наблюдений с 3-го по 14-й. Выборочная информация $X_*' \Sigma_*^{-1} X_*$ и $X_*' \Sigma_*^{-1} \mathbf{y}_*$ точно та же, что и в случае оценивания по методу максимального правдоподобия, она эквивалентна также выборочной информации при оценивании по обобщенному методу наименьших квадратов (см. главу 6). Априорные данные о многомерном бета-распределении приведены в табл. 9.4. Данные этой таблицы получены при $a_j = 1000$, поскольку в этом эксперименте наблюдалось 1000 объектов. Объекты генерировались с помощью имитационной модели, описанной в главе. 4. Дисперсии априорных распределений малы, так что следует ожидать более сильного влияния априорной, а не выборочной информации.

а) Случай совокупности независимых одномерных априорных бета-распределений с положительным эксцессом. В качестве начального эксперимента рассматривается случай, когда дисперсии априорных распределений не учитываются и предполагается, что каждая переходная вероятность p_{ij} характеризуется независимой априорной плотностью

вероятности бета-распределения. Средние значения и среднеквадратические ошибки для 50 байесовских оценок без ограничений, подсчитанные по выборкам объема 25, 50 и 100, даны в табл. 9.5.

Таблица 9.4

Значения параметров априорного многомерного бета-распределения при условии наблюдения микрообъектов

p_{ij}	a_{ij}	$a_i - a_{ij}$	Среднее значение	Дисперсия
p_{11}	599	401	0,5990	0,00023995
p_{12}	399	601	0,3990	0,00023956
p_{13}	1	999	0,0010	0,00000099
p_{14}	1	999	0,0010	0,00000099
p_{21}	100	900	0,1000	0,00008991
p_{22}	499	501	0,4990	0,00024974
p_{23}	400	600	0,4000	0,00023976
p_{24}	1	999	0,0010	0,00000099
p_{31}	1	999	0,0010	0,00000099
p_{32}	100	900	0,1000	0,00008991
p_{33}	699	301	0,6990	0,00021019
p_{34}	200	800	0,2000	0,00014984
p_{41}	1	999	0,0010	0,00000099
p_{42}	1	999	0,0010	0,00000099
p_{43}	100	900	0,1000	0,00008991
p_{44}	898	102	0,8980	0,00009150

Таблица 9.5

Средние значения и среднеквадратические ошибки байесовских оценок без учета ограничений в случае совокупности независимых априорных бета-распределений с положительным эксцессом (интервалы наблюдений с 3-го по 14-й)

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,6047 & 0,3986 & -0,0000 & 0,0027 \\ 0,0990 & 0,4983 & 0,4002 & 0,0025 \\ 0,0000 & 0,0982 & 0,7001 & 0,2018 \\ 0,0000 & 0,0000 & 0,0991 & 0,9009 \end{bmatrix}$	$\begin{bmatrix} 0,0428 & 0,0024 & 0,0000 & 0,0045 \\ 0,0017 & 0,0040 & 0,0030 & 0,0069 \\ 0,0000 & 0,0026 & 0,0032 & 0,0044 \\ 0,0000 & 0,0000 & 0,0015 & 0,0015 \end{bmatrix}$
50	$\begin{bmatrix} 0,5995 & 0,4004 & 0,0000 & 0,0000 \\ 0,0987 & 0,5000 & 0,4024 & -0,0011 \\ 0,0000 & 0,0979 & 0,7010 & 0,2012 \\ 0,0000 & 0,0000 & 0,0994 & 0,9006 \end{bmatrix}$	$\begin{bmatrix} 0,0035 & 0,0032 & 0,0000 & 0,0060 \\ 0,0024 & 0,0047 & 0,0050 & 0,0087 \\ 0,0000 & 0,0043 & 0,0040 & 0,0054 \\ 0,0000 & 0,0000 & 0,0013 & 0,0013 \end{bmatrix}$
100	$\begin{bmatrix} 0,6007 & 0,4008 & 0,0000 & -0,0015 \\ 0,0991 & 0,5007 & 0,4035 & -0,0033 \\ 0,0000 & 0,0989 & 0,7012 & 0,1999 \\ 0,0000 & 0,0000 & 0,0991 & 0,9009 \end{bmatrix}$	$\begin{bmatrix} 0,0039 & 0,0035 & 0,0000 & 0,0061 \\ 0,0030 & 0,0047 & 0,0056 & 0,0092 \\ 0,0000 & 0,0026 & 0,0036 & 0,0044 \\ 0,0000 & 0,0000 & 0,0017 & 0,0018 \end{bmatrix}$

Байесовские оценки, полученные без учета ограничений, лучше, чем оценки максимального правдоподобия или оценки по обобщенному методу наименьших квадратов в том смысле, что среди оценок из табл. 9.5 имеется лишь несколько отрицательных. Хотя для выборки объема 25 все средние значения оценок положительны, это вовсе не означает, что все эти оценки допустимы. Агрегированные ошибки, найденные путем суммирования всех элементов матрицы среднеквадратических ошибок, равны 0,0383, 0,0500 и 0,0503 для выборок объема 25, 50 и 100 соответственно. Величины этих ошибок относительно малы по сравнению с агрегированными среднеквадратическими ошибками оценок максимального правдоподобия и оценок по обобщенному методу наименьших квадратов, которые соответственно равны 3,6036, 2,9800 и 2,7936. Интересно отметить, что с увеличением объема выборки оценки существенно не улучшаются. Естественно, что это происходит вследствие преобладающего влияния выбранного априорного распределения, которое характеризуется положительным эксцессом. Другими словами, в этом примере оценки переходных вероятностей определяются в основном априорной информацией и почти не зависят от функции правдоподобия. Следовательно, мы имеем дело со случаем, прямо противоположным тому, когда получаемые оценки устойчивы к выбору априорного распределения. Поэтому опасно задавать малую дисперсию в априорном распределении, если априорные знания не надежны.

В силу того что оценки без ограничений не являются, вообще говоря, допустимыми, были также вычислены байесовские оценки с учетом ограничений по трем выборкам объема 25, 50 и 100. Средние значения и среднеквадратические ошибки приведены в табл. 9.6.

Таблица 9.6

Средние значения и среднеквадратические ошибки байесовских оценок с учетом ограничений в случае совокупности независимых априорных бета-распределений с положительным эксцессом (интервалы наблюдений с 3-го по 14-й)

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}5984 & 0{,}3983 & 0{,}0000 & 0{,}0032 \\ 0{,}0987 & 0{,}4997 & 0{,}3995 & 0{,}0041 \\ 0{,}0000 & 0{,}0982 & 0{,}7002 & 0{,}2016 \\ 0{,}0000 & 0{,}0000 & 0{,}0989 & 0{,}9009 \end{bmatrix}$	$\begin{bmatrix} 0{,}0025 & 0{,}0024 & 0{,}0000 & 0{,}0043 \\ 0{,}0016 & 0{,}0037 & 0{,}0024 & 0{,}0058 \\ 0{,}0000 & 0{,}0026 & 0{,}0034 & 0{,}0046 \\ 0{,}0000 & 0{,}0000 & 0{,}0019 & 0{,}0015 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5984 & 0{,}3993 & 0{,}0000 & 0{,}0023 \\ 0{,}0981 & 0{,}4982 & 0{,}4006 & 0{,}0031 \\ 0{,}0000 & 0{,}0981 & 0{,}7016 & 0{,}2003 \\ 0{,}0000 & 0{,}0000 & 0{,}0994 & 0{,}9006 \end{bmatrix}$	$\begin{bmatrix} 0{,}0028 & 0{,}0020 & 0{,}0000 & 0{,}0039 \\ 0{,}0027 & 0{,}0039 & 0{,}0028 & 0{,}0054 \\ 0{,}0000 & 0{,}0043 & 0{,}0048 & 0{,}0063 \\ 0{,}0000 & 0{,}0000 & 0{,}0012 & 0{,}0013 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}5991 & 0{,}3992 & 0{,}0000 & 0{,}0016 \\ 0{,}0984 & 0{,}4985 & 0{,}4012 & 0{,}0019 \\ 0{,}0000 & 0{,}0995 & 0{,}7025 & 0{,}1980 \\ 0{,}0000 & 0{,}0000 & 0{,}0092 & 0{,}9008 \end{bmatrix}$	$\begin{bmatrix} 0{,}0029 & 0{,}0022 & 0{,}0000 & 0{,}0031 \\ 0{,}0032 & 0{,}0037 & 0{,}0027 & 0{,}0036 \\ 0{,}0000 & 0{,}0027 & 0{,}0051 & 0{,}0065 \\ 0{,}0000 & 0{,}0000 & 0{,}0017 & 0{,}0017 \end{bmatrix}$

При проверке по различным критериям оценки, полученные с учетом ограничений, оказываются лучше оценок без ограничений. Агрегированные среднеквадратические ошибки оценок с ограничениями для выборок объема 25, 50 и 100 соответственно равны 0,0365, 0,0415 и 0,0391.

Применение критерия проверки нормальности, основанного на D-статистике Колмогорова — Смирнова, показало, что ни один из элементов матрицы оценок без ограничений не характеризуется распределением, которое значимо отклоняется от нормального при 10%-ном уровне значимости. Распределения оценок с ограничениями также не имеют значимых отклонений от нормального распределения, за исключением распределений тех оценок, которые соответствуют равным нулю истинным значениям. Доли оценок, принявших нулевое значение, совпадающее с истинным нулевым значением переходной вероятности, представлены в табл. 9.7.

Таблица 9.7

Доли (%) байесовских оценок с учетом ограничений, значения которых совпали с истинными значениями переходных вероятностей

Переходные вероятности	Объем выборки		
	25	50	100
p_{13}	100	98	100
p_{14}	64	52	64
p_{24}	36	52	64
p_{31}	100	100	94
p_{41}	100	100	100
p_{42}	100	100	100

Доли оценок, значения которых совпали с соответствующим истинным значением, более равномерно распределены в случае выборки объема 100 по сравнению с выборками объема 50 и 25.

б) Случай многомерного априорного бета-распределения с положительным эксцессом. Во втором эксперименте учитывались дисперсии априорных распределений. В нем было взято многомерное бета-распределение. В табл. 9.8 даны средние значения и среднеквадратические ошибки для 50 байесовских оценок без ограничений, вычисленных по выборкам объема 25, 50 и 100.

Агрегированные ошибки, найденные путем суммирования всех элементов матрицы среднеквадратических ошибок, равны 0,0233, 0,0326 и 0,0335 для выборок объема 25, 50 и 100 соответственно. Эти результаты несколько лучше результатов, полученных при рассмотрении совокупности одномерных бета-распределений. Однако, по существу, результаты первого и второго экспериментов идентичны, что объясняется преобладающим влиянием априорного распределения.

В силу того что для нескольких оценок без ограничений нарушено условие неотрицательности, в табл. 9.9 содержатся оценки с учетом ограничений. Агрегированные среднеквадратические ошибки в этом эксперименте равны 0,0232, 0,0325 и 0,0334 соответственно для выборок

объема 25, 50 и 100. Эти результаты отличаются незначительно от случая оценок без ограничений и, естественно, очень близки к результатам, полученным в эксперименте с совокупностью бета-распределений. Все описанные в предыдущем разделе результаты проверки нормальности применимы к данным табл. 9.8 и 9.9.

Таблица 9.8

Средние значения и среднеквадратические ошибки байесовских оценок без учета ограничений в случае априорного многомерного бета-распределения с положительным эксцессом

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}6001 & 0{,}3999 & 0{,}6000 & 0{,}0000 \\ 0{,}0993 & 0{,}4992 & 0{,}4015 & 0{,}0000 \\ 0{,}0000 & 0{,}0981 & 0{,}7022 & 0{,}1997 \\ 0{,}0000 & -0{,}0000 & 0{,}0993 & 0{,}9007 \end{bmatrix}$	$\begin{bmatrix} 0{,}0021 & 0{,}0021 & 0{,}0000 & 0{,}0000 \\ 0{,}0013 & 0{,}0035 & 0{,}0034 & 0{,}0000 \\ 0{,}0000 & 0{,}0026 & 0{,}0038 & 0{,}0022 \\ 0{,}0000 & 0{,}0000 & 0{,}0011 & 0{,}0011 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5995 & 0{,}4005 & 0{,}0000 & -0{,}0000 \\ 0{,}0986 & 0{,}4991 & 0{,}7033 & 0{,}1990 \\ -0{,}0000 & 0{,}0976 & 0{,}7033 & 0{,}1990 \\ -0{,}0000 & -0{,}0000 & 0{,}0995 & 0{,}9005 \end{bmatrix}$	$\begin{bmatrix} 0{,}0026 & 0{,}0026 & 0{,}0000 & 0{,}0000 \\ 0{,}0024 & 0{,}0039 & 0{,}0042 & 0{,}0000 \\ 0{,}0000 & 0{,}0045 & 0{,}0060 & 0{,}0035 \\ 0{,}0000 & 0{,}0001 & 0{,}0011 & 0{,}0012 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}6000 & 0{,}3999 & 0{,}0000 & -0{,}0000 \\ 0{,}0988 & 0{,}4989 & 0{,}4023 & -0{,}0000 \\ -0{,}0000 & 0{,}0987 & 0{,}7029 & 0{,}1983 \\ -0{,}0000 & -0{,}0000 & 0{,}0993 & 0{,}9007 \end{bmatrix}$	$\begin{bmatrix} 0{,}0037 & 0{,}0037 & 0{,}0000 & 0{,}0000 \\ 0{,}0030 & 0{,}0043 & 0{,}0039 & 0{,}0000 \\ 0{,}0000 & 0{,}0024 & 0{,}0048 & 0{,}0039 \\ 0{,}0000 & 0{,}0000 & 0{,}0015 & 0{,}0016 \end{bmatrix}$

Таблица 9.9

Средние значения и среднеквадратические ошибки байесовских оценок с учетом ограничений в случае априорного многомерного бета-распределения с положительным эксцессом

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}6001 & 0{,}3999 & 0{,}0000 & 0{,}0000 \\ 0{,}0993 & 0{,}4992 & 0{,}4015 & 0{,}0000 \\ 0{,}0000 & 0{,}0981 & 0{,}7022 & 0{,}1997 \\ 0{,}0000 & 0{,}0000 & 0{,}0993 & 0{,}9006 \end{bmatrix}$	$\begin{bmatrix} 0{,}0021 & 0{,}0021 & 0{,}0000 & 0{,}0000 \\ 0{,}0013 & 0{,}0035 & 0{,}0034 & 0{,}0000 \\ 0{,}0000 & 0{,}0026 & 0{,}0038 & 0{,}0022 \\ 0{,}0000 & 0{,}0000 & 0{,}0011 & 0{,}0011 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5995 & 0{,}4005 & 0{,}0000 & 0{,}0000 \\ 0{,}0986 & 0{,}4990 & 0{,}4023 & 0{,}0000 \\ 0{,}0000 & 0{,}0977 & 0{,}7033 & 0{,}1990 \\ 0{,}0000 & 0{,}0000 & 0{,}0995 & 0{,}9005 \end{bmatrix}$	$\begin{bmatrix} 0{,}0026 & 0{,}0026 & 0{,}0000 & 0{,}0000 \\ 0{,}0024 & 0{,}0039 & 0{,}0042 & 0{,}0000 \\ 0{,}0000 & 0{,}0045 & 0{,}0061 & 0{,}0036 \\ 0{,}0000 & 0{,}0000 & 0{,}0011 & 0{,}0012 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}6000 & 0{,}3999 & 0{,}0000 & 0{,}0000 \\ 0{,}0988 & 0{,}4989 & 0{,}4023 & 0{,}0000 \\ 0{,}0000 & 0{,}0987 & 0{,}7029 & 0{,}1933 \\ 0{,}0000 & 0{,}0000 & 0{,}0993 & 0{,}9007 \end{bmatrix}$	$\begin{bmatrix} 0{,}0037 & 0{,}0037 & 0{,}0000 & 0{,}0000 \\ 0{,}0030 & 0{,}0043 & 0{,}0039 & 0{,}0000 \\ 0{,}0000 & 0{,}0025 & 0{,}0048 & 0{,}0040 \\ 0{,}0000 & 0{,}0000 & 0{,}0016 & 0{,}0016 \end{bmatrix}$

в) Случай априорного распределения с отрицательным эксцессом. В случае, когда априорных сведений очень мало, априорное распределение должно иметь большую дисперсию. Для того чтобы это осмыслить, выполним эксперимент в предположении, что априорная информация мала. Предположим известным лишь только то, что оцениваемые параметры всегда принимают значения между нулем и единицей. Можно определить все параметры совокупности одномерных бета-распределений так, что средние значения всех переходных вероятностей будут равны, т. е. микрообъекты будут иметь равные возможности перейти в любое другое марковское состояние. Можно также задать слишком большую величину дисперсии априорного распределения, так что три стандартных отклонения будут покрывать почти всю область, в которой плотность вероятности отлична от нуля. До проведения крупномасштабного выборочного эксперимента случайным образом был отобран эксперимент на шестом интервале наблюдения с выборкой объема 50. Исследовались случаи отличных друг от друга параметров. Полученные результаты представлены в табл. 9.10.

Таблица 9.10

Исследование влияния различных априорных параметров на качество байесовской оценки при фиксированном априорном среднем значении (в качестве выборочной информации взята шестая выборка объема 50)

Параметры априорного бета-распределения					Значение критерия χ^2		Среднеквадратические ошибки	
a_{ij}	$a_i - a_{ij}$	среднее значение	дисперсия	σ	без ограничений	с ограничениями	без ограничений	с ограничениями
1,5	4,5	0,25	0,0269	0,16	13,5565	16,6946	0,0012	0,0014
1,4	4,2	0,25	0,0284	0,17	13,5084	16,5379	0,0012	0,0014
1,3	3,9	0,25	0,0302	0,17	13,4034	16,3857	0,0011	0,0014
1,2	3,6	0,25	0,0323	0,18	13,4223	16,2413	0,0011	0,0014
1,1	3,3	0,25	0,0347	0,18	13,3863	16,1094	0,0011	0,0013
1,0	3,0	0,25	0,0375	0,19	13,3575	15,9963	0,0011	0,0013
0,9	2,7	0,25	0,0408	0,20	13,3395	15,9105	0,0011	0,0013
0,8	2,4	0,25	0,0446	0,21	13,3388	15,8638	0,0011	0,0013
0,7	2,1	0,25	0,0493	0,22	13,3692	15,8721	0,0011	0,0013
0,6	1,8	0,25	0,0551	0,23	13,4617	15,9664	0,0010	0,0013
0,5	1,5	0,25	0,0625	0,25	13,7078	15,2290	0,0010	0,0013
0,4	1,2	0,25	0,0721	0,27	23,0324	16,6977	0,0018	0,0013
0,3	0,9	0,25	0,0952	0,29	17,2145	19,0623	0,0013	0,0014
0,2	0,6	0,25	0,1042	0,32	506,3499	19,8018	0,0032	0,0014
0,1	0,3	0,25	0,1339	0,37	153,3647	21,9741	0,1087	0,0017

В этом эксперименте при значениях параметров $a_{ij} = 0{,}8$ и $a_i - a_{ij} = 2{,}4$ величина критерия χ^2 минимальна. При значениях именно этих параметров осуществлялся крупномасштабный эксперимент. Результаты этого эксперимента приведены в табл. 9.11. Они показывают, что при выборе априорного распределения с отрицательным эксцессом априорным данным присваивается малый вес. При этом полученная

Таблица 9.11

Средние значения и среднеквадратические ошибки байесовских оценок без учета ограничений в случае априорного многомерного бета-распределения с отрицательным эксцессом (интервалы наблюдений с 3-го по 14-й)

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,3467 & 0,4291 & 0,1864 & 0,0376 \\ 0,1701 & 0,4280 & 0,4205 & -0,0186 \\ 0,0473 & 0,1461 & 0,5474 & 0,2592 \\ -0,0189 & 0,0512 & 0,1939 & 0,8761 \end{bmatrix}$	$\begin{bmatrix} 0,2872 & 0,1477 & 0,2657 & 0,2055 \\ 0,1158 & 0,1530 & 0,1398 & 0,1363 \\ 0,0927 & 0,1049 & 0,2028 & 0,1272 \\ 0,0736 & 0,0811 & 0,1447 & 0,1013 \end{bmatrix}$
50	$\begin{bmatrix} 0,4087 & 0,4453 & 0,1407 & 0,0053 \\ 0,1788 & 0,4259 & 0,4345 & -0,0392 \\ 0,0214 & 0,1625 & 0,5453 & 0,2709 \\ -0,0208 & -0,0556 & 0,2232 & 0,8532 \end{bmatrix}$	$\begin{bmatrix} 0,2194 & 0,1421 & 0,2111 & 0,1419 \\ 0,1266 & 0,1463 & 0,1363 & 0,1487 \\ 0,0904 & 0,1344 & 0,1935 & 0,1310 \\ 0,0769 & 0,1069 & 0,1562 & 0,1061 \end{bmatrix}$
100	$\begin{bmatrix} 0,4342 & 0,4447 & 0,1216 & -0,0005 \\ 0,2166 & 0,3824 & 0,4208 & -0,0198 \\ -0,0080 & 0,1935 & 0,5930 & 0,2215 \\ -0,0120 & 0,0561 & 0,1752 & 0,8928 \end{bmatrix}$	$\begin{bmatrix} 0,1906 & 0,1286 & 0,1980 & 0,1541 \\ 0,1455 & 0,1510 & 0,1242 & 0,1652 \\ 0,0651 & 0,1259 & 0,1319 & 0,0832 \\ 0,0490 & 0,0850 & 0,1049 & 0,0627 \end{bmatrix}$

Таблица 9.12

Средние значения и среднеквадратические ошибки байесовских оценок с учетом ограничений в случае априорного многомерного бета-распределения с отрицательным эксцессом (интервалы наблюдений с 3-го по 14-й)

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,3631 & 0,4261 & 0,1841 & 0,0268 \\ 0,1578 & 0,4403 & 0,3952 & 0,0067 \\ 0,0305 & 0,0995 & 0,6244 & 0,2456 \\ 0,0089 & 0,0021 & 0,1288 & 0,8602 \end{bmatrix}$	$\begin{bmatrix} 0,2820 & 0,1169 & 0,2176 & 0,0524 \\ 0,0967 & 0,1366 & 0,1097 & 0,0178 \\ 0,0464 & 0,0709 & 0,1417 & 0,1066 \\ 0,0208 & 0,0079 & 0,0960 & 0,0972 \end{bmatrix}$
50	$\begin{bmatrix} 0,4403 & 0,4436 & 0,1126 & 0,0036 \\ 0,1418 & 0,4607 & 0,3918 & 0,0057 \\ 0,0115 & 0,0976 & 0,6696 & 0,2213 \\ 0,0035 & 0,0035 & 0,1185 & 0,8745 \end{bmatrix}$	$\begin{bmatrix} 0,1865 & 0,1277 & 0,1448 & 0,0108 \\ 0,0958 & 0,1223 & 0,0882 & 0,0166 \\ 0,0259 & 0,0575 & 0,0923 & 0,0711 \\ 0,0081 & 0,0129 & 0,0683 & 0,0715 \end{bmatrix}$
100	$\begin{bmatrix} 0,4808 & 0,4266 & 0,0913 & 0,0013 \\ 0,1518 & 0,4531 & 0,3911 & 0,0041 \\ 0,0047 & 0,1107 & 0,6882 & 0,1963 \\ 0,0004 & 0,0023 & 0,0986 & 0,8988 \end{bmatrix}$	$\begin{bmatrix} 0,1494 & 0,1116 & 0,1226 & 0,0050 \\ 0,0859 & 0,0999 & 0,0761 & 0,0113 \\ 0,0119 & 0,0392 & 0,0526 & 0,0393 \\ 0,0015 & 0,0082 & 0,0513 & 0,0510 \end{bmatrix}$

матрица оценок удовлетворяет условию нормировки по строке, несмотря на то, что априорное среднее значение каждой переходной вероятности равно 0,25. Поскольку агрегированные среднеквадратические ошибки для выборок объема 25, 50 и 100 соответственно принимают

значения 2,3793, 2,2678 и 1,9659, свойство состоятельности оценок по-прежнему сохраняется.

В рассматриваемом эксперименте были вычислены оценки с учетом ограничений. Эти оценки представлены в табл. 9.12. Их асимптотическое поведение таково, что с увеличением объема выборки они приближаются к истинным значениям переходных вероятностей. Агрегированные ошибки для выборок объема 25, 50 и 100 соответственно равны 1,5973, 1,2003 и 0,9167.

На основании результатов описанных экспериментов можно сделать вывод о том, что если имеется априорная информация, то ее следует учитывать при получении апостериорного распределения по теореме Байеса. Если априорная информация достаточно полная, то для придания значительного веса априорным данным следует выбрать априорное распределение с малой дисперсией. Если же априорная информация не очень надежна, то для того, чтобы обеспечить малый вес априорных данных, нужно взять априорное распределение с большой дисперсией. В этом заключается принцип устойчивого оценивания.

9.9. РЕЗУЛЬТАТЫ РЕШЕНИЯ ЗАДАЧИ ТЕЛСЕРА

При получении результатов в задаче Телсера ранее использовалась только выборочная информация. Предположим теперь, что имеются следующие априорные сведения:

(1) у курильщика равные возможности сменить сорт сигарет и продолжать курить сигареты ранее выбранного сорта;

(2) если курильщик захочет сменить сорт сигарет, то у него окажутся равными возможности выбрать любой другой сорт.

При этих условиях априорные средние значения переходных вероятностей равны:

$$\begin{array}{l} \\ \text{«Camel»} \\ \text{«Lucky Strike»} \\ \text{«Chesterfield»} \end{array} \begin{array}{c} \begin{array}{ccc} \text{«Camel»} & \text{«Lucky Strike»} & \text{«Chesterfield»} \end{array} \\ \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} \end{array}. \qquad (9.9.1)$$

Если степень достоверности матрицы переходных вероятностей не слишком высока, то матрица ковариаций априорного многомерного бета-распределения должна быть выбрана с бо́льшими значениями элементов по сравнению со средними значениями. Предположим, что ковариационная матрица задана табл. 9.13.

Параметры априорного многомерного бета-распределения a_{ij}, вычисленные по априорным средним и ковариациям, равны:

$$[a_{ij}] = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}. \qquad (9.9.2)$$

Соединение приведенной априорной информации с выборочными данными дает байесовские оценки, представленные в табл. 9.14.

Таблица 9.13

Априорная матрица ковариаций переходных вероятностей в задаче Телсера

	p_{11}	p_{21}	p_{31}	p_{12}	p_{22}	p_{32}
p_{11}	0,0500			—0,0250		
p_{21}		0,0375			—0,0250	
p_{31}			0,0375			—0,0250
p_{12}	—0,0250			0,0500		
p_{22}		—0,0250			0,0375	
p_{32}			—0,0250			0,0375

Таблица 9.14

Байесовские оценки переходных вероятностей в задаче Телсера, вычисленные с помощью итеративного метода квадратического программирования для трех сортов сигарет: «Camel», «Lucky Strike», «Chesterfield»

Этапы	Оценки	Невязка итерации
Начальное приближение	$\begin{bmatrix} 0,7802 & 0,1251 & 0,0947 \\ 0 & 0,8910 & 0,1090 \\ 0,2545 & 0 & 0,7455 \end{bmatrix}$	—
Решение после 1-й итерации	$\begin{bmatrix} 0,7830 & 0,1220 & 0,0950 \\ 0 & 0,8944 & 0,1056 \\ 0,2517 & 0 & 0,7485 \end{bmatrix}$	0,0112
Решение после 2-й итерации	$\begin{bmatrix} 0,7835 & 0,1216 & 0,0949 \\ 0 & 0,8948 & 0,1052 \\ 0,2509 & 0 & 0,7491 \end{bmatrix}$	0,0022
Решение после 3-й итерации	$\begin{bmatrix} 0,7836 & 0,1216 & 0,0948 \\ 0 & 0,8948 & 0,1052 \\ 0,2508 & 0 & 0,7492 \end{bmatrix}$	0,0004

Глава 10 ● ОЦЕНИВАНИЕ ПО КРИТЕРИЮ НАИМЕНЬШИХ АБСОЛЮТНЫХ ОТКЛОНЕНИЙ

10.1. ОПИСАНИЕ СТАТИСТИЧЕСКОЙ МОДЕЛИ. ПОСТАНОВКА ЗАДАЧИ

Для оценки переходных вероятностей марковской цепи рассмотрим совокупность непересекающихся регрессионных гиперплоскостей, которые можно записать в векторной форме:

$$\mathbf{y} = X\mathbf{p} + \mathbf{u}. \tag{10.1.1}$$

Помимо критерия наименьших квадратов для вычисления оценок параметров при наличии ограничений типа равенств или неравенств применяется критерий наименьших абсолютных отклонений (в общем случае взвешенных). С помощью этого критерия получают оценку минимума абсолютных отклонений [37], [118]. В случае минимизации суммы абсолютных отклонений наша задача может быть сформулирована следующим образом: найти вектор $\overline{\mathbf{p}}$, такой, чтобы сумма абсолютных отклонений

$$\mathbf{a}'\boldsymbol{\eta}_{rT} \tag{10.1.2}$$

была минимальной при наличии ограничений

$$\mathbf{y} = X\mathbf{p} + \mathbf{u}, \tag{10.1.3}$$

$$G\mathbf{p} = \boldsymbol{\eta}_r \tag{10.1.4}$$

и

$$\mathbf{p} \geqslant 0. \tag{10.1.5}$$

Здесь $\mathbf{a}' = (u_1(t),\ u_2(t),\ \ldots,\ u_r(t))$ есть $(1 \times rT)$-вектор-строка с элементами $u_l(t) = \|u_l(1)|, |u_l(2)|, \ldots, |u_l(T)|\|$, $\mathbf{u}$ — вектор ошибок с элементами, принимающими либо положительные, либо отрицательные значения, $\boldsymbol{\eta}_{rT}$ и $\boldsymbol{\eta}_r$ — векторы-столбцы единиц размера rT и r соответственно. Если имеются наблюдения $\mathbf{y}$ и X, то задача (10.1.2) — (10.1.5) может быть решена с помощью методов линейного программирования.

10.2. СВЕДЕНИЕ К ЗАДАЧЕ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Для того чтобы сформулировать нашу задачу как задачу линейного программирования, необходимо преобразовать вектор отклонений $\mathbf{u}$ так, чтобы он в одном и том же виде появился и в выражении для целевой функции (10.1.2) и в ограничениях (10.1.3). Представим абсолютное значение $|u_j(t)|$ в виде

$$|u_j(t)| = \max(0, u_j(t)) + \max(-u_j(t), 0). \qquad (10.2.1)$$

Тогда сама величина $u_j(t)$ может быть записана как

$$u_j(t) = \max(0, u_j(t)) - \max(-u_j(t), 0), \qquad (10.2.2)$$

так что $u_j(t)$ может принимать как положительное, так и отрицательное значение. Обозначим через $\mathbf{u}_1$ вектор величин $\max(0, u_j(t))$, а через $\mathbf{u}_2$ — вектор $\max(-u_j(t), 0)$. Векторы $\mathbf{u}_1$ и $\mathbf{u}_2$ обладают свойством неотрицательности и, кроме того, их соответствующие элементы являются взаимодополнениями в том смысле, что эти элементы одновременно не могут принимать положительные значения. Иначе говоря, при условии, что компоненты вектора $\mathbf{u}$ принимают как положительные, так и отрицательные значения, формируется пара дополнений $\mathbf{u}_1$ и $\mathbf{u}_2$ так, что $\mathbf{u}_1$ и $\mathbf{u}_2$ имеют неотрицательные элементы, $\mathbf{u}_1$ соответствует положительным, а $\mathbf{u}_2$ — отрицательным отклонениям.

а) Минимизация суммы абсолютных отклонений без учета весов. Если задача оптимизации заключается в том, что должна быть минимизирована сумма абсолютных отклонений без учета весов, то задача линейного программирования состоит в следующем: минимизировать

$$(\mathbf{u}_1 + \mathbf{u}_2)' \boldsymbol{\eta}_r T \qquad (10.2.3)$$

при условиях

$$\mathbf{y} = X\mathbf{p} + \mathbf{u}_1 - \mathbf{u}_2, \qquad (10.2.4)$$

$$G\mathbf{p} = \boldsymbol{\eta}_r \qquad (10.2.5)$$

и

$$\mathbf{p}, \mathbf{u}_1, \mathbf{u}_2 \geqslant \mathbf{0}. \qquad (10.2.6)$$

Эта задача может быть решена с помощью симплекс-метода. Симплекс-таблица для рассматриваемой задачи минимизации представлена табл. 10.1.

В этой таблице вектор ошибок $\mathbf{u}_1$ может быть первоначально включен в базис для того, чтобы уменьшить ошибки округления, возникающие вследствие итеративного процесса вычислений. Последнее ограничение $G\mathbf{p} = \boldsymbol{\eta}_r$ должно иметь искусственные переменные в базисе. Число m вводится для придания высокой цены искусственным переменным.

б) Минимизация взвешенной суммы абсолютных отклонений. Целевую функцию в предыдущей задаче представляет собой сумма абсолютных отклонений в отсутствие весов. Если измерения неравноточные, то каждому наблюдению может быть приписан различный вес. Точно так же, как и в случае взвешенных наименьших квадратов, когда

матрица $H'H$ выбирается в качестве весовой матрицы при минимизации квадратичной функции потерь вида

$$\Phi(\mathbf{p}) = (\mathbf{y} - X\mathbf{p})\, H'H\, (\mathbf{y} - X\mathbf{p})', \tag{10.2.7}$$

можно воспользоваться матрицей H как весовой матрицей при минимизации функции потерь в виде суммы абсолютных отклонений:

$$\psi(\mathbf{p}) = (H\mathbf{a})'\, \eta_{rT}. \tag{10.2.8}$$

Как один из возможных способов определения матрицы H можно предложить следующий:

$$H'H = \Sigma^{-1}, \tag{10.2.9}$$

где Σ — ковариационная матрица вектора $\mathbf{u}$.

Таблица 10.1

Симплекс-таблица линейного программирования для оценки по методу наименьших абсолютных отклонений

C_j		0	η_{rT}	η_{rT}
	B_0	$\mathbf{p}$	$\mathbf{u}_1$	$\mathbf{u}_2$
η_{rT}	$\mathbf{y}$	X	I	$-I$
$\mathbf{m}$	η_r	G	0	0

Если компоненты вектора $\mathbf{u}$ — некоррелированные гетероскедастические случайные величины, то H представляет собой диагональную матрицу элементов, обратных по отношению к стандартным отклонениям. Таким образом, для придания меньшего веса наблюдениям с большими стандартными отклонениями и приписывания большего веса наблюдениям с малыми стандартными отклонениями все наблюдения делятся на соответствующие стандартные отклонения, и после этого применяется процедура, описанная в предыдущем параграфе. Другими словами, уравнение (10.1.1) домножается слева на матрицу H:

$$H\mathbf{y} = HX\mathbf{p} + H\mathbf{u}, \tag{10.2.10}$$

и затем минимизируется выражение

$$H\mathbf{a}'\eta_{rT} \tag{10.2.11}$$

или

$$(\mathbf{u}_1 + \mathbf{u}_2)'\, H'\eta_{rT} \tag{10.2.12}$$

при наличии ограничений (10.2.4) — (10.2.6). Это означает, что так называемым «вектором цен» в задаче вместо η_{rT} становится $H'\eta_{rT}$. Симплекс-таблица для этой задачи представлена табл. 10.2. Следует иметь в виду, что векторы $H'\eta_{rT}$, $\mathbf{p}$, $\mathbf{u}_1$ и $\mathbf{u}_2$ в столбцах B_0 и C_j симплекс-таблицы представлены в транспонированном виде.

Таблица 10.2

Симплекс-таблица линейного программирования для оценки по методу наименьших абсолютных отклонений с учетом взвешивания

C_j		0	$H'\eta_{rT}$	$H'\eta_{rT}$
	B_0	$\mathbf{p}$	$\mathbf{u}_1$	$\mathbf{u}_2$
$H'\eta_{rT}$	$\mathbf{y}$	X	I	$-I$
$\mathbf{m}$	η_r	G	0	0

10.3. РЕЗУЛЬТАТЫ ВЫБОРОЧНОГО ЭКСПЕРИМЕНТА

В экспериментах участвовали исходные агрегированные данные, полученные на интервалах наблюдения с 1-го по 16-й. Для каждой задачи размер симплекс-таблицы (64 × 136), т. е. имеются 136 переменных (включая искусственные переменные) и 64 ограничения. Для каждой из выборок объема 25, 50 и 100 решалось 50 задач. Для решения отдельной задачи требовалось в среднем около 90 итераций, что занимало около 60с машинного времени на вычислительной системе IBM 7094.

Величины средних значений и среднеквадратических ошибок для оценок наименьших абсолютных отклонений по каждой из выборок представлены в табл. 10.3.

Таблица 10.3

Средние значения и среднеквадратические ошибки оценок по методу наименьших абсолютных отклонений

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0{,}5096 & 0{,}4059 & 0{,}0784 & 0{,}1481 \\ 0{,}1155 & 0{,}4962 & 0{,}3701 & 0{,}0126 \\ 0{,}0222 & 0{,}0864 & 0{,}6357 & 0{,}2357 \\ 0{,}0044 & 0{,}0244 & 0{,}1195 & 0{,}8517 \end{bmatrix}$	$\begin{bmatrix} 0{,}1481 & 0{,}1661 & 0{,}1500 & 0{,}0154 \\ 0{,}0957 & 0{,}1554 & 0{,}1578 & 0{,}0324 \\ 0{,}0408 & 0{,}0642 & 0{,}1147 & 0{,}0948 \\ 0{,}0136 & 0{,}0431 & 0{,}1057 & 0{,}1066 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}5633 & 0{,}3905 & 0{,}0558 & 0{,}0004 \\ 0{,}1059 & 0{,}5110 & 0{,}3788 & 0{,}0043 \\ 0{,}0094 & 0{,}0840 & 0{,}6652 & 0{,}2354 \\ 0{,}0052 & 0{,}0191 & 0{,}1186 & 0{,}8571 \end{bmatrix}$	$\begin{bmatrix} 0{,}1028 & 0{,}1416 & 0{,}1078 & 0{,}0030 \\ 0{,}0743 & 0{,}1455 & 0{,}1254 & 0{,}0125 \\ 0{,}0220 & 0{,}0578 & 0{,}1049 & 0{,}0811 \\ 0{,}0108 & 0{,}0388 & 0{,}0824 & 0{,}0870 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}5810 & 0{,}3770 & 0{,}0426 & 0{,}0000 \\ 0{,}1063 & 0{,}5050 & 0{,}3876 & 0{,}0012 \\ 0{,}0035 & 0{,}0947 & 0{,}6906 & 0{,}2111 \\ 0{,}0037 & 0{,}0110 & 0{,}0971 & 0{,}8882 \end{bmatrix}$	$\begin{bmatrix} 0{,}0847 & 0{,}1227 & 0{,}0910 & 0{,}0000 \\ 0{,}0716 & 0{,}1343 & 0{,}1103 & 0{,}0062 \\ 0{,}0094 & 0{,}0515 & 0{,}0726 & 0{,}0487 \\ 0{,}0081 & 0{,}0242 & 0{,}0552 & 0{,}0547 \end{bmatrix}$

С увеличением объема выборки наблюдается значительное улучшение как средних значений точечных оценок, так и среднеквадратических ошибок, причем суммирование элементов матриц среднеквадратических ошибок дает величины 1,5044, 1,1978 и 0,0453 соответственно для выборок объема 25, 50 и 100. Таким образом, оказывается, что результаты выборочного эксперимента отражают свойство состоятельности оценок, полученных по критерию наименьших отклонений. Однако эти оценки хуже по сравнению с любыми другими, за исключением оценок по методу наименьших квадратов без учета ограничений, что следует из сопоставления агрегированных среднеквадратических ошибок. Этот вывод согласуется с результатами модельного исследования метода оценивания по критерию наименьших абсолютных отклонений, описанного в [6]. Заметим, однако, что в методе, рассматриваемом в этом параграфе, учитываются ограничения на переходные вероятности.

Оценки по критерию наименьших взвешенных абсолютных отклонений не вычислялись, так как предполагалось, что оценки, получаемые по этому критерию, менее эффективны по сравнению с обсуждавшимися ранее. Кроме того, вычисление их требует значительных затрат машинного времени[1].

Гипотезы относительно того, что оценки по методу наименьших абсолютных отклонений распределены нормально, проверялись с помощью D-статистики Колмогорова—Смирнова. Эта статистика вычислялась для оценок ненулевых элементов матрицы переходных вероятностей по выборкам объема 25, 50 и 100. Подсчитанные величины D-статистики перечислены в табл. 10.4.

Значения D-статистики, приведенные в табл. 10.4, согласуются с гипотезой о том, что оценки по методу наименьших абсолютных отклонений распределены нормально (для 5%-ного уровня значимости критическое значение D-статистики составляет 0,16), за исключением оценки p_{31} в случае выборки объема 25. Для этой оценки величина D равна 0,1725 и близка к 10%-ному критическому значению ($D = 0{,}17$). Для оценок переходных вероятностей p_{13}, p_{14}, p_{24}, p_{41} и p_{42} проценты совпадения величины оценки с истинным значением (нулем) приведены в табл. 10.5. Согласно данным этой таблицы доля оценок, совпавших по величине с истинным значением, увеличивается с возрастанием объема выборки, что отражает свойство состоятельности оценок. Величина оценки переходной вероятности p_{14} при объеме выборки 100 имеет 100%-ную вероятность совпадения с истинным значением. Выборочные стандартные отклонения, вычисленные по найденным в эксперименте оценкам, представлены в табл. 10.6. Эти стандартные отклонения уменьшаются с увеличением объема выборки. Кроме того, стандартные отклонения, приведенные в табл. 10.6,

[1] На вычислительной системе IBM 7094 требуется около 60 с для вычисления серии оценок по критерию наименьших абсолютных отклонений в одной задаче, в то время как на вычисление оценок максимального правдоподобия с учетом ограничений и байесовских оценок с ограничениями уходит лишь 12 с машинного времени.

Таблица 10.4

Значения D-статистики Колмогорова—Смирнова для оценок по методу наименьших абсолютных отклонений

Объем выборки	Значения D-статистики
25	$\begin{bmatrix} 0{,}0992 & 0{,}0818 & — & — \\ 0{,}1725 & 0{,}0636 & 0{,}0832 & — \\ — & 0{,}0843 & 0{,}0807 & 0{,}1354 \\ — & — & 0{,}1575 & 0{,}1022 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}0887 & 0{,}0776 & — & — \\ 0{,}1103 & 0{,}0959 & 0{,}0723 & — \\ — & 0{,}0738 & 0{,}0691 & 0{,}1208 \\ — & — & 0{,}0823 & 0{,}1019 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}1036 & 0{,}0856 & — & — \\ 0{,}1077 & 0{,}0682 & 0{,}1452 & — \\ — & 0{,}0624 & 0{,}0910 & 0{,}0707 \\ — & — & 0{,}0603 & 0{,}0761 \end{bmatrix}$

Таблица 10.5

Процент совпадения с истинными значениями для некоторых из оценок, вычисленных по методу наименьших абсолютных отклонений

Переходные вероятности	Объем выборки		
	25	50	100
p_{13}	52	56	58
p_{14}	98	98	100
p_{24}	78	84	94
p_{31}	58	68	78
p_{41}	82	64	68
p_{42}	50	58	70

несколько больше стандартных отклонений для других оценок с ограничениями.

В заключение следует отметить, что в тех случаях, когда для установления качества оценок в повторных испытаниях применяются обычные критерии, оценки по методу наименьших абсолютных отклонений оказываются хуже других, ранее обсуждавшихся оценок. Однако этот метод учитывает ограничения на переходные вероятности, и на основании полученных результатов представляется, что оценки по критерию наименьших абсолютных отклонений, по-видимому, обеспечивают приемлемые результаты при оценке переходных вероятностей.

Таблица 10.6

Стандартные отклонения оценок по методу наименьших абсолютных отклонений

Объем выборки	Стандартные отклонения
25	$\begin{bmatrix} 0{,}1173 & 0{,}1160 & 0{,}1279 & 0{,}0152 \\ 0{,}0944 & 0{,}1554 & 0{,}1549 & 0{,}0299 \\ 0{,}0342 & 0{,}0627 & 0{,}1058 & 0{,}0879 \\ 0{,}0129 & 0{,}0365 & 0{,}1039 & 0{,}0951 \end{bmatrix}$
50	$\begin{bmatrix} 0{,}0961 & 0{,}1403 & 0{,}0923 & 0{,}0030 \\ 0{,}0740 & 0{,}1450 & 0{,}1236 & 0{,}0118 \\ 0{,}0199 & 0{,}0555 & 0{,}0990 & 0{,}0729 \\ 0{,}0094 & 0{,}0337 & 0{,}0802 & 0{,}0757 \end{bmatrix}$
100	$\begin{bmatrix} 0{,}0825 & 0{,}1206 & 0{,}0804 & 0{,}0000 \\ 0{,}0713 & 0{,}1342 & 0{,}1096 & 0{,}0061 \\ 0{,}0087 & 0{,}0512 & 0{,}0720 & 0{,}0474 \\ 0{,}0072 & 0{,}0216 & 0{,}0552 & 0{,}0534 \end{bmatrix}$

Глава 11 ● ПРОГНОЗИРОВАНИЕ И КРИТЕРИЙ СОГЛАСИЯ χ^2

Одна из целей оценивания параметров экономической модели заключается в том, чтобы обеспечить основу для прогнозирования исходов экономических преобразований в будущем. В этой главе будут рассмотрены задача прогноза величины относительных частот признаков на будущее и критерий соответствия модели выборочным данным. Этот критерий позволяет оценивать качество модели по бо́льшему объему данных.

11.1. ПРЕДСКАЗАННЫЕ ОТНОСИТЕЛЬНЫЕ ЧАСТОТЫ

Согласно изложенному в главе 1 в марковском случае можно записать следующее соотношение:

$$q_j(t) = \sum_{i=1}^{r} y_i(t-1)\, p_{ij}, \qquad (11.1.1)$$

где $y_i\ (t-1)$ — наблюдаемая относительная частота, $q_j(t)$ — истинная относительная частота. При включении в модель ошибок измерений $u_j\ (t)$ по формуле

$$y_j\ (t) = q_j\ (t) + u_j\ (t) \qquad (11.1.2)$$

имеет место следующее соотношение для наблюдаемых относительных частот:

$$y_j(t) = \sum_{i=1}^{r} y_i(t-1)\, p_{ij} + u_j(t). \qquad (11.1.3)$$

Если обозначить через $\hat{p}_{ij}$ оценку вероятности p_{ij}, то предсказанные частоты определяются по формуле[1]

---

[1] Здесь следует иметь в виду различие между прогнозом в пределах выборочного интервала наблюдений и прогнозом за пределы этого интервала. В последнем случае $\hat{y}_j\ (T_\alpha) = \sum_{i=1}^{r} \hat{y}_i\ (T_\alpha)\, p_{ij}$, $\alpha = T+1,\ T+2, \ldots$ и $\hat{y}_i\ (T_\alpha)$, $\hat{p}_{ij}$ подвержены ошибкам. По-видимому, это приводит к тому, что с увеличением α дисперсия ошибки прогноза возрастает.

$$\hat{y}_j(t) = \sum_{i=1}^{r} y_i(t-1)\,\hat{p}_{ij}, \quad j = 1, 2, \ldots, r. \qquad (11.1.4)$$

В качестве примера применения формулы (11.1.4) рассмотрим случай, когда по наблюдениям 1000 объектов (практически здесь мы имеем дело с бесконечной совокупностью) в моменты времени $t = 3, 4, \ldots, 14$ были получены оценки максимального правдоподобия для переходных вероятностей:

$$[\hat{p}_{ij}] = \begin{bmatrix} 0{,}5613 & 0{,}3540 & 0{,}0847 & 0{,}0000 \\ 0{,}1212 & 0{,}5358 & 0{,}3430 & 0{,}0000 \\ 0{,}0000 & 0{,}0879 & 0{,}7209 & 0{,}1912 \\ 0{,}0000 & 0{,}0000 & 0{,}0861 & 0{,}9139 \end{bmatrix}. \qquad (11.1.5)$$

Наблюдаемые и предсказанные с учетом (11.1.5) относительные частоты приведены в табл. 11.1.

В качестве суммарной ошибки всех прогнозов может быть рассмотрена среднеквадратическая ошибка, которая в данной задаче составляет лишь 0,00066. Более тонкой мерой ошибки будет критерий χ^2, который обсуждается в следующем параграфе.

Таблица 11.1

Наблюдаемые и предсказанные относительные частоты в модельной задаче с матрицей переходных вероятностей (4.1.1) при прогнозе с помощью оценок максимального правдоподобия (11.1.5)

Интервал наблюдений	Наблюдаемые относительные частоты				Предсказанные относительные частоты			
	s_1	s_2	s_3	s_4	s_1	s_2	s_3	s_4
2	0,4060	0,4280	0,1660	0,0000	—	—	—	—
3	0,2890	0,3850	0,2960	0,0300	0,2798	0,3877	0,3008	0,0317
4	0,2060	0,3250	0,3930	0,0760	0,2089	0,3346	0,3725	0,0840
5	0,1520	0,2900	0,4130	0,1450	0,1550	0,2816	0,4188	0,1446
6	0,1150	0,2620	0,4030	0,2200	0,1205	0,2455	0,4255	0,2115
7	0,0980	0,2170	0,3990	0,2860	0,0963	0,2165	0,4091	0,2781
8	0,0080	0,1700	0,3990	0,3430	0,0813	0,1861	0,3950	0,3377
9	0,0630	0,1750	0,3750	0,3870	0,0700	0,1573	0,3829	0,3898
10	0,0590	0,1430	0,3660	0,4320	0,0566	0,1491	0,3690	0,4254
11	0,0480	0,1180	0,3660	0,4680	0,0504	0,1297	0,3551	0,4648
12	0,0400	0,1130	0,3550	0,4920	0,0412	0,1124	0,3487	0,4977
13	0,0420	0,1040	0,3360	0,5180	0,0361	0,1059	0,3404	0,5175
14	0,0330	0,1050	0,3330	0,5290	0,0362	0,1001	0,3260	0,5377
15	—	—	—	—	0,0313	0,0972	0,3244	0,5471

11.2. КРИТЕРИЙ ПРОВЕРКИ ГИПОТЕЗ χ^2

Известно [62], что в асимптотике полиномиальное распределение микрообъектов $n_j(t)$, $j = 1, 2, \ldots, r$ представляет собой r-мерное нормальное распределение и что квадратичная форма в показателе экспоненты этого распределения подчиняется распределению χ^2 с числом степеней свободы, равным рангу $r - 1$. На этих фактах основывается критерий Пирсона для проверки простых гипотез о том, что наблюдаемые и предсказанные относительные частоты одинаково распределены. Статистика Пирсона представляет собой квадратичную форму и имеет вид

$$\chi^2_{(r-1)} = \sum_l (n_l(t) - N(t)\,\hat{y}_l(t))^2 / N(t)\,\hat{y}_l(t). \qquad (11.2.1)$$

Это выражение может быть переписано как

$$\chi^2_{(r-1)} = \sum_l N(t)\,(y_l(t) - \hat{y}_l(t))^2 / \hat{y}_l(t). \qquad (11.2.2)$$

Величины, определенные в (11.2.1) и (11.2.2), имеют распределение χ^2 с $(r - 1)$-й степенью свободы. Это обстоятельство лежит в основе критерия проверки гипотез χ^2. Предположим, что осуществляется прогноз относительных частот на T шагов вперед. Тогда в силу свойства аддитивности распределения χ^2 получается, что выражение

$$\chi^2_{(r-1)T} = \sum_t^T \sum_l^r N(t)\,(y_l(t) - \hat{y}_l(t))^2 / \hat{y}_l(t) \qquad (11.2.3)$$

подчиняется распределению χ^2 с $(r - 1)\,T$ степенями свободы. Предполагается, что предсказанные частоты отличны от нуля.

Статистика в форме (11.2.3) может быть применена для того, чтобы проверить, являются ли наблюдаемые частоты (или агрегированные данные) «обычными» или «необычными» исходами в предположении, что наблюдения генерировались соотношением (11.1.1). Напомним, что это соотношение получено из марковского процесса, относительно которого предполагается зависимость событий в данный момент времени лишь от событий в непосредственно предшествующий момент времени. Если исходы окажутся «необычными» при проверке по критерию χ^2, то ориентировочно можно сделать вывод о том, что, по-видимому, данные не подчиняются (11.1.1). Мы говорим «по-видимому» потому, что появление «необычного» (маловероятного) события возможно в случае, когда предположения о процессе, порождающем экспериментальные данные, верны. Кроме того, критерий χ^2 рассмотренного типа не позволяет установить степень достоверности принятой модели даже в простейшем случае.

Величина статистики χ^2 для предсказанных относительных частот, приведенных в табл. 11.1, составляет 15,1998, что существенно меньше табличного значения, равного 47,1420 при 36 степенях свободы и 10%-ном уровне значимости. Таким образом, на основе традиционного метода проверки гипотез по выборочным данным можно заключить, что полученный результат согласуется с принятой гипотезой о том, что агрегированные данные являются исходами марковского процесса, заданного соотношением (11.1.1).

В предыдущих главах было показано, что оценки максимального правдоподобия, а также оценки по обобщенному методу наименьших квадратов и методу минимума χ^2 обладают одинаковыми свойствами при большом объеме выборки. Вспоминая, что оценка по минимуму χ^2 получается при минимизации выражения для величины χ^2 вида (7.2.2), приведенного в главе 7, можно предположить, что этой оценке соответствует минимальное значение критерия χ^2. В действительности это характерно для случая больших выборок, потому что тогда оцененные веса, содержащие оценки истинных относительных частот, близки к истинным весам.

Иначе говоря, при малом объеме выборки оценкам по критерию минимума χ^2 может и не соответствовать минимальное значение величины χ^2. Однако оценка по минимуму χ^2 обеспечивает наименьшее значение (среди других оценок) модифицированной величине χ^2, определенной (7.3.2). В случае больших выборок выражения (7.2.2) и (7.3.2) асимптотически эквивалентны.

11.3. РЕЗУЛЬТАТЫ ВЫБОРОЧНОГО ЭКСПЕРИМЕНТА

Было просчитано более чем 1500 задач для случая, когда матрица переходных вероятностей имела размер (4×4). Переходные вероятности оценивались по методу квадратичного программирования. В каждой задаче на IBM 7094 вычислялись оценки с учетом и без учета ограничений. Были найдены величины предсказанных относительных частот и, наконец, была вычислена статистика χ^2 с целью применения критерия χ^2. Ни один из результатов проверки по этому критерию не отвергает гипотезы о том, что данные порождены марковским процессом. Средние значения и выборочные χ^2-статистики для выборок объема 25, 50 и 100 представлены в табл. 11.2.

Результаты эксперимента с 1000 объектов также приведены в этой таблице для иллюстрации основной тенденции поведения величины χ^2. Для этого эксперимента вычислены значения модифицированной величины χ^2, минимум которой соответствует оценке максимального правдоподобия. В терминах этой меры классическая оценка оказывается несколько хуже. Все значения величины χ^2 в табл. 11.2 меньше табличного значения $\chi^2 = 47{,}14$ при 36 степенях свободы и 10%-ном уровне значимости.

Таблица 11.2

Средние значения статистики χ^2 для оценок, найденных в выборочном эксперименте

Оценка	Объем выборки			Совокупность, включающая 1000 объектов	
	25	50	100	χ^2	модифицированная величина
По обычному методу наименьших квадратов с учетом ограничений	14,2943	17,6977	17,6358	15,9293	16,1164
По методу наименьших квадратов с учетом ограничений и взвешивания (веса равны величинам, обратным к средним относительным частотам)	14,2899	17,6459	17,5826	16,0009	15,1996
Оценка максимального правдоподобия (по обобщенному методу наименьших квадратов, по методу минимума χ^2)	16,1373	19,5361	18,4240	15,1998	15,1526
Байесовская оценка при априорных средних, приблизительно равных истинным значениям параметров	19,0119	22,1620	21,2785	—	—

Глава 12 ● СРАВНЕНИЕ РАЗЛИЧНЫХ ОЦЕНОК

В предыдущих главах были введены различные оценки и описаны выборочные эксперименты, иллюстрирующие их практическое применение. В этой главе будет дано краткое изложение экспериментальных результатов и сравнение качества различных оценок.

12.1. ОСНОВА ДЛЯ СРАВНЕНИЯ

Для проведения корректного сравнения оценок их вычисление должно осуществляться по одной и той же выборке данных. В предыдущих экспериментах было две выборки. Одну выборку составили данные, полученные в моменты времени с 1-го по 16-й, другую — данные, соответствующие моментам времени с 3-го по 14-й. Для сравнения оценок использовалась вторая выборка. Поэтому были пересчитаны классические оценки без весов и оценки с весами, равными обратным величинам к средним относительным частотам. Средние значения и среднеквадратические ошибки относительно истинных частот для классических оценок без ограничений, классических оценок с учетом ограничений и взвешенных оценок с учетом ограничений представлены в табл. 12.1, 12.2 и 12.3.

12.2. АГРЕГИРОВАННАЯ СРЕДНЕКВАДРАТИЧЕСКАЯ ОШИБКА И МЕРА УКЛОНЕНИЯ

Прежде всего возникает вопрос о том, как сравнивать матрицы среднеквадратических ошибок для разных оценок в случае, когда элементы одной матрицы отличаются от соответствующих элементов другой, причем невязки этих элементов принимают как положительные, так и отрицательные значения.
Один из возможных критериев — сравнение агрегированных среднеквадратических ошибок, которые представляют собой сумму элементов матрицы среднеквадратических ошибок. При использовании этой меры требуется, чтобы параметры имели одинаковые веса, кроме того, не учитывается зависимость между ошибками оценок вероятностей p_{ij}.

Таблица 12.1

Классические оценки без учета ограничений, вычисленные по исправленным данным (интервалы наблюдений с 3-го по 14-й)

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,4416 & 0,5808 & 0,1231 & -0,1455 \\ 0,1605 & 0,3631 & 0,4439 & 0,0326 \\ 0,0117 & 0,1388 & 0,5398 & 0,3097 \\ 0,0064 & -0,0201 & 0,2151 & 0,7986 \end{bmatrix}$	$\begin{bmatrix} 0,2611 & 0,3534 & 0,4454 & 0,4007 \\ 0,1522 & 0,2422 & 0,3022 & 0,2533 \\ 0,1360 & 0,1515 & 0,2407 & 0,1747 \\ 0,0938 & 0,1068 & 0,2054 & 0,1635 \end{bmatrix}$
50	$\begin{bmatrix} 0,5041 & 0,5134 & 0,0167 & -0,0342 \\ 0,1544 & 0,3715 & 0,5155 & -0,0414 \\ -0,0127 & 0,1795 & 0,5284 & 0,3047 \\ 0,0147 & -0,0514 & 0,2294 & 0,8073 \end{bmatrix}$	$\begin{bmatrix} 0,1668 & 0,2680 & 0,2816 & 0,2027 \\ 0,1323 & 0,2208 & 0,2652 & 0,1989 \\ 0,1220 & 0,1861 & 0,2343 & 0,1664 \\ 0,0888 & 0,1299 & 0,1808 & 0,1376 \end{bmatrix}$
100	$\begin{bmatrix} 0,5161 & 0,5494 & -0,0063 & -0,0591 \\ 0,8151 & 0,2913 & 0,5166 & 0,0071 \\ -0,0300 & 0,2107 & 0,5787 & 0,2406 \\ 0,0109 & -0,0513 & 0,1748 & 0,8656 \end{bmatrix}$	$\begin{bmatrix} 0,1605 & 0,2670 & 0,3039 & 0,2135 \\ 0,1600 & 0,2599 & 0,2960 & 0,2160 \\ 0,0885 & 0,1492 & 0,1689 & 0,1067 \\ 0,0569 & 0,1015 & 0,1165 & 0,0766 \end{bmatrix}$

Таблица 12.2

Классические оценки с учетом ограничений, вычисленные по исправленным данным (интервалы наблюдений с 3-го по 14-й)

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,3756 & 0,4788 & 0,1400 & 0,0056 \\ 0,1628 & 0,4242 & 0,3957 & 0,0172 \\ 0,0377 & 0,1071 & 0,5949 & 0,2604 \\ 0,0082 & 0,0130 & 0,1687 & 0,8100 \end{bmatrix}$	$\begin{bmatrix} 0,2914 & 0,2408 & 0,2445 & 0,0203 \\ 0,1295 & 0,1906 & 0,1682 & 0,0385 \\ 0,0613 & 0,0796 & 0,1679 & 0,1096 \\ 0,0195 & 0,0288 & 0,1244 & 0,1392 \end{bmatrix}$
50	$\begin{bmatrix} 0,4721 & 0,4587 & 0,0677 & 0,0015 \\ 0,1389 & 0,4409 & 0,4112 & 0,0091 \\ 0,0168 & 0,1077 & 0,6242 & 0,2513 \\ 0,0034 & 0,0147 & 0,1599 & 0,8220 \end{bmatrix}$	$\begin{bmatrix} 0,1766 & 0,1815 & 0,1255 & 0,0061 \\ 0,1017 & 0,1577 & 0,1186 & 0,0241 \\ 0,0329 & 0,0705 & 0,1286 & 0,0827 \\ 0,0084 & 0,0319 & 0,0985 & 0,1081 \end{bmatrix}$
100	$\begin{bmatrix} 0,4947 & 0,4268 & 0,0785 & 0,0000 \\ 0,1553 & 0,4452 & 0,3932 & 0,0063 \\ 0,0069 & 0,1151 & 0,6608 & 0,2172 \\ 0,0009 & 0,0063 & 0,1222 & 0,8706 \end{bmatrix}$	$\begin{bmatrix} 0,1520 & 0,1685 & 0,1336 & 0,0000 \\ 0,0964 & 0,1481 & 0,1064 & 0,0149 \\ 0,0135 & 0,0453 & 0,0730 & 0,0412 \\ 0,0028 & 0,0177 & 0,0518 & 0,0539 \end{bmatrix}$

Таблица 12.3

Оценки, вычисленные по исправленным данным, с учетом ограничений и взвешивания на величины, обратные к средним относительным частотам (интервалы наблюдений с 3-го по 14-й)

Объем выборки	Средние значения	Среднеквадратические ошибки
25	$\begin{bmatrix} 0,4123 & 0,4660 & 0,1157 & 0,0052 \\ 0,1509 & 0,4338 & 0,3989 & 0,0184 \\ 0,0304 & 0,1061 & 0,6058 & 0,2577 \\ 0,0120 & 0,0144 & 0,1607 & 0,8132 \end{bmatrix}$	$\begin{bmatrix} 0,2598 & 0,2206 & 0,1997 & 0,0196 \\ 0,1163 & 0,1862 & 0,1482 & 0,0403 \\ 0,0546 & 0,0811 & 0,1589 & 0,1081 \\ 0,0249 & 0,0313 & 0,1265 & 0,1335 \end{bmatrix}$
50	$\begin{bmatrix} 0,4985 & 0,4442 & 0,0559 & 0,0016 \\ 0,1295 & 0,4546 & 0,4065 & 0,0093 \\ 0,0137 & 0,1038 & 0,6349 & 0,2476 \\ 0,0055 & 0,0153 & 0,1525 & 0,8268 \end{bmatrix}$	$\begin{bmatrix} 0,1557 & 0,1667 & 0,1090 & 0,0066 \\ 0,0969 & 0,1484 & 0,1102 & 0,0244 \\ 0,0301 & 0,0698 & 0,1207 & 0,0801 \\ 0,0122 & 0,0325 & 0,0925 & 0,1039 \end{bmatrix}$
100	$\begin{bmatrix} 0,5216 & 0,4162 & 0,0622 & 0,0000 \\ 0,1417 & 0,4537 & 0,3578 & 0,0068 \\ 0,0056 & 0,1132 & 0,6672 & 0,2142 \\ 0,0020 & 0,0073 & 0,1165 & 0,8742 \end{bmatrix}$	$\begin{bmatrix} 0,1335 & 0,1504 & 0,1098 & 0,0000 \\ 0,0887 & 0,1326 & 0,0917 & 0,0162 \\ 0,0115 & 0,0453 & 0,0669 & 0,0400 \\ 0,0052 & 0,0195 & 0,0486 & 0,0509 \end{bmatrix}$

Агрегированные среднеквадратические ошибки результатов, сведенных в табл. 4.2, 4.5 и 4.6 из главы 4, табл. 5.1, 5.2 и 5.3 из главы 5 и табл. 10.3 из главы 10, представлены в табл. 12.4. Эта группа результатов получена по выборочным данным, относящимся к моментам времени с 1-го по 16-й.

На основании данных табл. 12.4 представляется, что, несмотря на произвол в выборе весов, оценка по методу наименьших квадратов с учетом ограничений характеризуется меньшим значением агрегированной среднеквадратической ошибки по сравнению с другими оценками по методу наименьших квадратов и методу наименьших абсолют-

Таблица 12.4

Сумма значений элементов матрицы среднеквадратических ошибок для нескольких типов оценок (интервалы наблюдений с 1-го по 16-й)

Объем выборки	Оценка по методу максимального правдоподобия, вычисленная по микроданным	Оценка по методу наименьших квадратов без учета ограничений	Оценка по методу наименьших абсолютных отклонений с учетом ограничений	Оценки по методу наименьших квадратов с учетом ограничений			
				без взвешивания	при взвешивании на средние значения относительных частот	Оценка матрицы корреляций ошибок	произведение выборочных относительных частот
25	0,4025	2,4944	1,5014	1,4795	1,4052	1,3812	1,4243
50	0,3155	2,0537	1,1978	1,0191	0,9915	0,9715	0,9981
100	0,2190	1,9526	1,0453	0,8520	0,7931	0,8002	0,8108

Таблица 12.5

Агрегированные среднеквадратические ошибки для различных оценок (интервалы наблюдений с 3-го по 14-й)

Оценка	Объем выборки					
	25		50		100	
	без ограничений	с ограничениями	без ограничений	с ограничениями	без ограничений	с ограничениями
По обычному методу наименьших квадратов с учетом ограничений	3,6829	2,0641	2,9920	1,4536	2,7418	1,1191
По методу наименьших квадратов с учетом ограничений и взвешивания	3,6829	1,9095	2,9920	1,3597	2,7418	1,0109
Оценка максимального правдоподобия (по обобщенному методу наименьших квадратов, по методу минимума χ^2)	3,5181	1,7440	2,9800	1,2726	2,7936	0,9750
Байесовская оценка в случае априорного распределения с отрицательным эксцессом и равных априорных средних	2,3793	1,5973	2,2678	1,2003	1,9659	0,9167
Байесовская оценка в случае априорного распределения с положительным эксцессом и априорных средних, приблизительно равных истинным значениям параметров	0,0383	0,0265	0,0500	0,0415	0,0503	0,0391

ных отклонений. Кроме того, оказывается, что имеется значительное улучшение эффективности (если определять ее по величине агрегированной среднеквадратической ошибки) при переходе от оценок без учета ограничений к оценкам по методу наименьших квадратов с ограничениями. Однако если имеются не только относительные частоты, но и микроданные, то оценка максимального правдоподобия, полученная по микроданным, значительно лучше (в смысле величины агрегированной среднеквадратической ошибки) по сравнению с любой из макрооценок, вычисленных в описываемом эксперименте. Как уже отмечалось в главе 2, можно вычислить также байесовскую оценку, основанную на микроданных.

В табл. 12.5 перечислены агрегированные среднеквадратические ошибки оценок, приведенных в табл. 6.2, 6.3 из главы 6, табл. 9.5, 9.6, 9.11 и 9.12 из главы 9 и табл. 12.1, 12.2 и 12.3 настоящей главы. Эти результаты вычислены по выборочным данным, соответствующим моментам времени с 3-го по 14-й.

Вообще говоря, в смысле среднеквадратического критерия оценки максимального правдоподобия, оценки по обобщенному методу наименьших квадратов и по критерию минимума χ^2 лучше оценок

по методу наименьших квадратов без учета ограничений, а также оценок с весами, которые равны величинам, обратным к наблюдаемым относительным частотам. Байесовская оценка характеризуется меньшим по сравнению с оценкой максимального правдоподобия значением агрегированной среднеквадратической ошибки при любом объеме выборки. Степень улучшения зависит от параметров выбранной совокупности бета-распределений. Полученные результаты показывают, что если имеется априорная информация о значениях параметров распределений, то учет этой информации приводит к значительному снижению среднеквадратических ошибок.

Для сравнения дисперсий различных оценок были вычислены агрегированные выборочные дисперсии. Эти данные представлены в табл. 12.6.

Таблица 12.6

Агрегированные дисперсии различных оценок (интервалы наблюдений с 3-го по 14-й)

Оценка	Объем выборки					
	25		50		100	
	без ограничений	с ограничениями	без ограничений	с ограничениями	без ограничений	с ограничениями
По обычному методу наименьших квадратов с учетом ограничений	0,8255	0,2597	0,4740	0,1334	0,4290	0,0997
По методу наименьших квадратов с учетом ограничений и взвешивания	0,8255	0,2252	0,4740	0,1217	0,4290	0,0836
Оценка максимального правдоподобия (по обобщенному методу наименьших квадратов, по методу минимума χ^2)	0,8047	0,2085	0,4938	0,1237	0,4288	0,0832
Байесовская оценка в случае априорного распределения с отрицательным эксцессом и равных априорных средних	0,2379	0,1211	0,2178	0,0859	0,1694	0,0557
Байесовская оценка в случае априорного распределения с положительным эксцессом и априорных средних, приблизительно равных истинным значениям параметров	0,0001	0,0000	0,0001	0,0000	0,0001	0,0000

Среди небайесовских оценок оценка максимального правдоподобия (оценка по обобщенному методу наименьших квадратов, оценка по критерию минимума χ^2) лучшая в том смысле, что она имеет наименьшую дисперсию. Учет априорной информации при вычислении байесовских оценок приводит к значительному уменьшению дисперсии по сравнению с другими случаями, когда эта информация принималась

во внимание при подсчете оценок. В случае, когда в описываемом эксперименте выбиралось априорное распределение с положительным эксцессом, наблюдалось значительное уменьшение дисперсии оценок. С другой стороны, при априорном распределении с отрицательным эксцессом дисперсия байесовской оценки оказывалась меньше дисперсии оценки максимального правдоподобия, но это отличие не столь велико, сколь в случае байесовской оценки, основанной на априорном распределении с положительным эксцессом.

12.3. КРИТЕРИЙ УИЛКОКСОНА И КОЭФФИЦИЕНТ СОГЛАСИЯ КЕНДЭЛА

В дополнение к рассмотренному критерию для измерения качества оценок применялись также знако-ранговый критерий попарного сравнения Уилкоксона [100] и критерий ранговой связи, основанный на коэффициенте согласия Кендэла [100]. Первый критерий предназначен для проверки гипотезы о том, что абсолютные отклонения двух оценок от истинного значения параметра одинаковые. Второй критерий обеспечивает способ проверки гипотезы об отсутствии ранговой связи между различными оценками.

В описываемом здесь эксперименте проводилось вычисление 50 оценок по выборкам объема 50 и осуществлялось две группы сравнений. В первой группе все оценки были получены из исходных экспериментальных данных и сравнивались четыре различные оценки: оценка по методу наименьших квадратов с учетом ограничений и взвешивания по среднему (МНКОВ), оценка по обычному методу наименьших квадратов с учетом ограничений (МНКО), оценка по методу наименьших абсолютных отклонений (МНАО) и оценка по обычному методу наименьших квадратов без учета ограничений (МНК). Соответствующие результаты представлены в табл. 12.7 и 12.8. Для проведения второй группы сравнений все оценки были вычислены по измененным данным, в которых не содержалось нулевых относительных частот. Здесь сравнивались оценки МНК, МНКОВ, оценки максимального правдоподобия с учетом (МПО) и без учета (МП) ограничений, а также еще два типа байесовских оценок с ограничениями: байесовская оценка при отрицательном эксцессе априорного распределения и равных априорных средних (в таблице она названа «байесовская оценка II»), байесовская оценка при положительном эксцессе априорного распределения и априорных средних, совпадающих с истинными значениями оцениваемых параметров (в таблице она названа «байесовская оценка I»). Сравниваемые данные представлены в табл. 12.10. Результаты по обеим группам сравнений оказались непротиворечивыми.

Применение критерия Уилкоксона показывает, что имеется значимое отличие между оценками с учетом и без учета ограничений при 5%-ном уровне значимости. Все результаты статистических испытаний оказались значимыми, за исключением проверок p_{11} и p_{44}. Сравнение оценок МНКОВ и МНКО в двенадцати из шестнадцати случаев оказалось в пользу оценки МНКОВ, хотя большинство оценок МНКОВ

Таблица 12.7

Знако-ранговый критерий попарного сравнения (критерий Уилкоксона)

Тип оценки	p_{11}	p_{12}	p_{13}	p_{14}	p_{21}	p_{22}	p_{23}	p_{24}	p_{31}	p_{32}	p_{33}	p_{34}	p_{41}	p_{42}	p_{43}	p_{44}
МНКОВ	−1,61	−1,08	−2,28	−1,00		−1,62	−2,93	−0,50	−3,06		−1,70	−2,95			−0,98	−1,77
МНКО					0,191					−1,14			−3,38	−1,14		
МНКОВ	−0,95	−0,61	−0,53		0,03	−1,20	−1,05									
МНАО				−1,00				−2,86	−0,82	−0,78	−0,99	−2,01	0,79	−1,14	−1,26	−2,99
МНКОВ	−2,01	−3,20	−5,49	−6,15	−3,73	−3,91	−4,84	−6,15	−5,94	−5,10	−4,11	−6,14	−5,76	−3,65	−3,65	−1,35
МНК																
МНКО	−0,80	−0,58	−0,19		−0,32	−1,98	−0,77									
МНАО				−1,00				−2,86	−1,23	−0,69	−1,23	−2,16	0,10	−1,06	−1,31	−2,20
МНКО	−1,46	−3,13	−5,53	−6,15	−3,62	−3,82	−4,78	−6,15	−5,79	−5,20	−4,28	−3,69	−6,15	−5,75	−3,77	−1,19
МНК																
МНАО МНК	−1,11	−1,75	−4,62	−6,15	−3,10	−1,63	−3,47	−6,15	−6,15	−4,49	−3,63	−4,60	−5,89	−5,21	−3,15	−2,84

Таблица 12.8

Ранжированные ряды оценок и коэффициент согласия Кендэла

Параметры		Ранжированные ряды оценок				Сумма квадратов отклонений	Коэффициент согласия
истинное значение	оценка	МНКОВ	МНКО	МНАО	МНК		
0,6	p_{11}	113,5	119,5	131,0	116,0	319,5	0,0256
0,4	p_{12}	110,5	117,5	123,0	149,0	846,5	0,0679
0	p_{13}	99,0	108,5	113,5	179,0	3996,5	0,3824
0	p_{14}	100,0	101,0	99,0	200,0	7502,0	0,9871
0,1	p_{21}	113,0	113,0	122,0	152,0	1026,0	0,0827
0,5	p_{22}	111,0	116,0	124,0	149,0	854,0	0,0683
0,4	p_{23}	97,0	118,0	122,0	163,0	2286,0	0,1829
0	p_{24}	105,5	107,5	87,0	200,0	7755,5	0,8499
0	p_{31}	99,5	113,5	97,0	190,0	5791,5	0,6178
0,1	p_{32}	114,5	109,0	110,0	166,5	2313,5	0,1869
0,7	p_{33}	107,0	118,5	109,0	165,5	2262,5	0,1817
0,2	p_{34}	106,5	130,5	103,0	160,0	2081,0	0,1675
0	p_{41}	107,0	92,5	104,5	196,0	6841,5	0,6893
0	p_{42}	108,5	100,5	105,0	186,0	4993,5	0,5135
0,1	p_{43}	115,5	118,5	114,0	152,0	982,5	0,0788
0,9	p_{44}	120,5	135,5	108,0	136,0	540,5	0,0435

не дает значительного улучшения даже при 10%-ном уровне значимости (см. табл. 12.7). При сравнении оценок МНКОВ и МНАО оказывается, что для большинства элементов матрицы переходных вероятностей разность оценок незначима, за исключением элемента p_{24}, для которого результат сравнения оценок МНКОВ и МНАО в пользу последней. Вообще говоря, метод наименьших абсолютных отклонений позволяет оценивать нулевые параметры (такие, как p_{14}, p_{41} и т. д.) очень хорошо, в то время как метод наименьших квадратов с учетом ограничений и взвешивания дает более хорошие оценки для ненулевых истинных параметров. При сравнении оценок МНКО и МНАО наблюдается ситуация, аналогичная взаимосвязи между оценками МНКОВ и МНАО: большинство из оценок МНКО и МНАО значимо не отличаются.

В табл. 12.9 приведены данные о числе сравнений, в которых наблюдалось значимое улучшение оценок. В табл. 12.8 представлены результаты применения коэффициента согласия Кендэла. Этот коэффициент, характеризующий степень связи между ранжированными рядами различных оценок, вычислялся по следующей схеме. По одним и тем же данным определялись четыре типа оценок (МНКОВ, МНКО, МНАО и МНК) для каждой вероятности перехода. Сравнивались абсолютные отклонения этих четырех оценок от соответствующего истинного значения, и оценкам приписывались ранги 1, 2, 3 и 4. Первый ранг присваивался оценке, наиболее близкой к истинному значению, т. е. оценке с минимальным абсолютным отклонением. Второй ранг присваивался оценке с минимальным абсолютным отклонением среди оставшихся непроранжированных трех оценок и т. д. Если две оценки оказывались одинаковыми, то для того, чтобы их не упорядочивать, каж-

Таблица 12.9

Число параметров p_{ij}, для которых оценка r оказалась лучше оценки s по критерию Уилкоксона при 5%-ном уровне значимости (интервалы наблюдений с 1-го по 10-й)

r	s			
	МНКОВ	МНКО	МНАО	МНК
МНКОВ	—	6	0	15
МНКО	2	—	1	14
МНАО	2	3	—	14
МНК	0	0	0	—

дой из оценок присваивался средний ранг (например, 1,5 или 2,5 и т. д.). Каждый параметр матрицы переходных вероятностей оценивался отдельно, причем оценка каждого из четырех типов вычислялась 50 раз по 50 различным выборкам. Таким образом, для каждого параметра было получено 50 наборов рангов. Затем для каждой оценки была вычислена сумма 50 рангов. Эти суммы для всех четырех оценок приведены в табл. 12.8. Если все 50 наборов рангов находятся в полном согласии с наилучшей оценкой, то одна из оценок должна иметь 50 рангов, равных 1, и, следовательно, сумму рангов, равную 50. Ясно, что степень согласия между 50 наборами рангов отражается в степени разброса между четырьмя суммами рангов. Коэффициент согласия будет функцией от этой степени разброса и вычисляется как отношение дисперсии рангов к максимально возможной дисперсии, которая должна была бы получиться при полном согласии между 50 наборами рангов[1].

Таким образом, коэффициент согласия принимает значения от 0 до 1. Если бы все четыре оценки имели одинаковую сумму рангов и находились бы в полном согласии, то коэффициент Кендэла был бы равен единице. Как следует из табл. 12.8, коэффициент согласия не принимает больших значений, за исключением случаев оценок параметров p_{14}, p_{24}, p_{31}, p_{41} и p_{42}, истинные значения которых равны нулю.

[1] Если в каждом наборе рангов отсутствуют одинаковые ранги, то коэффициент согласия Кендэла W определяется по формуле

$$W = \frac{s}{k^2 (N^3 - N)/12},$$

где s — сумма квадратов отклонений от среднего значения сумм рангов; k — 50 наборов рангов; N — 4 оценки.

Если в отдельных наборах рангов имеются одинаковые ранги, то должен быть подсчитан поправочный коэффициент $T = \Sigma (t^3 - t)/12$, где t — число наблюдений в группе, совпавших для фиксированного ранга. Суммирование осуществляется по всем совпадениям в любом из 50 наборов рангов. В этом случае коэффициент согласия Кендэла есть

$$W = \frac{s}{k^2 (N^3 - N)/12 - k\Sigma_T T},$$

где $\Sigma_T T$ означает суммирование величин T для всех k наборов рангов [100].

При анализе суммарных рангов можно обнаружить, что различия между оценками с учетом ограничений незначительны, однако все оценки с учетом ограничений лучше, чем обычные оценки по классическому методу наименьших квадратов. Например, для параметра p_{14} суммарные ранги для оценок МНКОВ, МНКО и МНАО соответственно равны 100, 101 и 99, в то время как оценка МНК имеет ранг 200. Это означает, что оценка МНК самая плохая. Проверка значимости суммы квадратов отклонений рангов от их средних по критерию χ^2 показывает также, что большинство дисперсий не являются достаточно большими для того, чтобы можно было сделать вывод о значимом различии рангов.

По результатам второй группы сравнений, приведенным в табл. 12.10, оценки без учета ограничений оказываются неприемлемыми. Эти оценки упорядочиваются следующим образом: байесовские оценки без ограничений, оценки максимального правдоподобия без ограничений (МП) и оценки по методу наименьших квадратов без ограничений. Байесовские оценки без ограничений намного лучше оценок МП и МНК. Однако улучшение оценок МП по сравнению с оценками МНК не очень значительно. Среди всех оценок с учетом ограничений байесовская оценка наилучшая. Оценки МП лучше, чем оценки МНКОВ, в том смысле, что для одиннадцати из шестнадцати переходных вероятностей оценки МП оказались ближе к истинным значениям. Результаты применения критерия Уилкоксона показывают, что только три переходные вероятности имеют оценки, значимо раз-

Таблица 12.10

Число параметров p_{ij} из шестнадцати, для которых оценка r оказалась лучше оценки s по критерию Уилкоксона при 5%-ном уровне значимости (интервалы наблюдений с 3-го по 14-й)

r	s							
	Байесовская I	Байесовская II	по методу максимального правдоподобия с ограничениями	МНКОВ	МНКО	Байесовская I без ограничений	по методу максимального правдоподобия без ограничений	МНК
Байесовская I	—	14	14	14	14	9	16	16
Байесовская II	0	—	7	9	10	—	—	—
По методу максимального правдоподобия с ограничениями	0	3	—	3	6	2	12	15
МНКОВ	1	3	2	—	10	2	15	15
МНКО	1	2	1	1	—	2	14	15
Байесовская I без ограничений	2	—	14	14	14	—	16	16
По методу максимального правдоподобия без ограничений	0	—	3	0	0	0	—	2
МНК	0	—	0	0	0	0	1	—

личающиеся при 5%-ном уровне значимости. Оценка МНКОВ также лучше по сравнению с оценкой МНКО, и не более половины элементов имеют значимое различие. Байесовские оценки без ограничений неприемлемы (в том смысле, что некоторые из них оказываются отрицательными), но эти оценки вполне близки по абсолютному отклонению к истинным значениям параметров. В действительности байесовские оценки без ограничений ближе к истинным значениям по сравнению с другими небайесовскими оценками. Конечно, байесовские оценки с равными априорными средними и большими априорными дисперсиями (байесовские оценки II) не такие хорошие, как байесовские оценки с истинными априорными средними и малыми априорными дисперсиями (байесовские оценки I). Но байесовские оценки II лучше, чем оценки МП, МНКОВ, а также МНКО в том смысле, что большинство из байесовских оценок II оказываются близкими к истинным значениям параметров.

Коэффициент согласия Кендэла, вычисленный для всех рассмотренных оценок, принимал значения в диапазоне от 0,5 до 0,8. Это свидетельствует о высокой степени связи в упорядоченном ряде оценок. Полученные результаты говорят и о том, что значимым является следующее упорядочение оценок: байесовские оценки с учетом и без учета ограничений имеют ранг 1, оценки МП с ограничениями, МНКОВ и МНКО имеют ранг 2 и оценки МП, МНК имеют ранг 3. В каждой группе оценок одного и того же ранга оценки могут быть проранжированы заново, однако такие оценки не различаются значимо.

В заключение следует отметить, что ранжированные ряды оценок, получаемые при попарном сравнении оценок с ограничениями, вообще говоря, аналогичны ранжированным рядам, выстраиваемым по критерию среднеквадратической ошибки. Однако имеются отдельные результаты статистических испытаний, которые не позволяют окончательно сформулировать подобный вывод. Единственное заключение, которое можно сделать, состоит в том, что оценки с ограничениями лучше оценок без учета ограничений и что оценки максимального правдоподобия и байесовские оценки дают выигрыш в эффективности.

12.4. КРАТКАЯ СВОДКА РЕЗУЛЬТАТОВ

В заключение перечислим основные результаты эксперимента, которые состоят в следующем.

1. Оценка максимального правдоподобия, найденная по микроданным, лучшая в смысле критерия минимума среднеквадратической ошибки по сравнению с любой другой оценкой, вычисленной по агрегированным данным (см. табл. 12.4). Среди оценок, получаемых по агрегированным данным, оценка с учетом ограничений лучше, чем оценка без учета ограничений, оценка по методу наименьших квадратов с ограничениями лучше оценки по методу наименьших абсолютных отклонений с ограничениями, оценка с учетом ограничений и взвешивания лучше, чем оценка с ограничениями, но без учета весов. Среди оценок, получаемых с учетом взвешивания, оценка максимального правдоподобия (или оценка по обобщенному методу наименьших квад-

ратов, по методу минимума χ^2) наилучшая. При наличии априорной информации байесовская оценка, имеющая априорное распределение с положительным эксцессом и априорные средние, приближенно равные истинным значениям параметров, характеризуется наименьшей среднеквадратической ошибкой по сравнению с любой другой оценкой.

2. При анализе дисперсий сравниваемых оценок оказывается, что эти оценки могут быть упорядочены точно так же, как и в пункте 1. При переходе от классической оценки максимального правдоподобия с учетом взвешивания к байесовской оценке дисперсия их уменьшается.

Правильный учет априорной информации в байесовском случае существенно снижает дисперсию оценки (см. табл. 12.6).

3. При анализе смещения рассмотренных оценок было обнаружено, что с увеличением объема выборки смещение оценок стремится к нулю.

4. Распределение оценок приближается к нормальному достаточно быстро. Например, критерий проверки нормальности (критерий Колмогорова — Смирнова, см. главу 6) показал, что ни одна из оценок ненулевых параметров не отвергает гипотезы о нормальности распределения при 10%-ном уровне значимости.

3. Результаты применения критерия согласия показали, что ни одна из проверок не дала оснований отвергнуть гипотезу о том, что данные порождены определенным марковским процессом.

Поскольку при оценивании матрицы весов вместо истинных относительных частот были взяты наблюдаемые относительные частоты, оценка по модифицированному методу минимума χ^2 характеризуется минимальным значением модифицированного критерия χ^2.

Глава 13 ● ЗАКЛЮЧЕНИЕ

В вводной главе был поставлен вопрос, возможно ли использовать данные об агрегированных относительных частотах для оценивания вероятностей перехода в предположении о том, что поведение микрообъектов описывается марковской моделью первого порядка. На основании анализа и результатов, представленных в предыдущих главах, на этот вопрос сейчас может быть дан утвердительный ответ.

При поиске приемлемых оценок вначале были рассмотрены оценки по методу наименьших квадратов с учетом и без учета ограничений и в серии экспериментов методом Монте-Карло были исследованы выборочные свойства этих оценок. Результаты проведенных экспериментов подтвердили теоретические результаты, и в случае большой выборки были продемонстрированы сходимость оценок к истинным значениям параметров и сходимость распределений оценок к нормальному распределению. Кроме того, было обнаружено, что при сравнении по критерию минимума среднеквадратической ошибки оценка с учетом ограничений оказывается лучше оценки без ограничений. Наконец, оценивание с учетом ограничений обеспечивает попадание оценок в допустимую область пространства параметров.

На следующем этапе исследования рассматривались несколько оценок по методу наименьших квадратов (МНК) с учетом ограничений и взвешивания. Веса были введены с целью уменьшить влияние гетероскедастичности. Результаты экспериментов, проведенных по методу Монте-Карло, показали, что оценки по методу наименьших квадратов с учетом взвешивания характеризуются меньшей среднеквадратической ошибкой по сравнению с обычными оценками МНК.

Кроме того, оценки МНК с учетом взвешивания сохраняют свойство состоятельности и в случае большой выборки имеют распределение, близкое к нормальному. Таким образом, полученные экспериментальные результаты подтверждают теоретические выводы.

При попытке избежать трудностей, связанных с вырожденностью ковариационной матрицы ошибок наблюдений, было обнаружено, что имеется возможность исключить из рассматриваемой системы одно стохастическое уравнение без какой-либо потери информации и получить таким образом редуцированную систему или систему меньшего порядка. Для параметров редуцированной системы с помощью обобщенного метода наименьших квадратов были получены оценки с ограничениями и без ограничений, которые учитывали влияние гетероскедастичности. Затем с учетом условия нормировки были оценены параметры исключенного уравнения. Было установлено, что полученные таким образом оценки не зависят от того, какое из уравнений исходной системы было исключено для получения редуцированной системы. Результаты экспериментов по методу Монте-Карло показали, что оценка, которую дает обобщенный метод наименьших квадратов, оказывается лучше, чем любая из ранее упомянутых оценок. Результаты экспериментов подтвердили свойство состоятельности этих оценок и тот факт, что при большой выборке данных распределения оценок близки к нормальному распределению.

Была исследована еще одна оценка, получаемая по критерию минимума χ^2. Для случая известной ковариационной матрицы ошибок наблюдений было показано, что эта оценка совпадает с той, которую дает обобщенный метод наименьших квадратов. В случае, когда неизвестная ковариационная матрица ошибок заменяется на ее состоятельную оценку, получается приближенная оценка, названная оценкой по модифицированному методу минимума χ^2. Эта оценка и оценка, получаемая по методу минимума χ^2, в случае известной ковариационной матрицы ошибок являются наилучшими асимптотически нормальными оценками.

При применении метода максимального правдоподобия (МП) для получения оценки по агрегированным относительным частотам оказалось, что оценка максимального правдоподобия по форме в точности совпадает с оценками по методу минимума χ^2 и обобщенному МНК. В действительности задачи оценивания, решаемые с помощью метода максимального правдоподобия и метода минимума χ^2, представляют собой двойственные задачи отыскания седловой точки. Была предложена итеративная процедура с обратной связью, при которой по последовательным оценкам переходных вероятностей оценивают параметры ковариационной матрицы ошибок. Поскольку эта процедура требует условий ограничений на параметры типа неравенства, ее можно рассматривать как рекуррентный алгоритм квадратичного программирования. Проведенные эксперименты с этим алгоритмом показывают, что он сходится за три или четыре итерации. Решения не расходятся, так как оценки принадлежат замкнутому интервалу [0, 1].

При байесовском подходе в исследовании марковской модели первого порядка по агрегированным данным в качестве априорного распределения параметров применялось многомерное бета-распре-

деление. В этом априорном распределении учитывается информация о том, что значения переходных вероятностей принадлежат замкнутому интервалу от нуля до единицы и что переходные вероятности также удовлетворяют условию нормировки. Далее, оказывается возможным учитывать априорную информацию о значениях оцениваемых параметров, если такая информация имеется. Действительно, если при построении функции правдоподобия используется большой объем априорной информации, то эта функция в случае малых и средних выборок будет оказывать определенное влияние на положение области, в которой плотность вероятности апостериорного распределения отлична от нуля, на разброс и форму апостериорного распределения. С другой стороны, если априорное распределение относительно плоское в той области, где функция правдоподобия принимает заметно отличные от нуля значения, то характеристики апостериорного распределения будут в значительной степени определяться формой функции правдоподобия. Хорошо известно, что в любом случае с ростом объема выборки свойства апостериорного распределения будут определяться видом функции правдоподобия. Таким образом, при больших выборках функция правдоподобия хорошо аппроксимирует апостериорное распределение, которое имеет в этом случае форму многомерного нормального распределения со средним значением, равным оценке максимального правдоподобия.

Представляло большой интерес исследовать выборочные свойства оценок. Поэтому были выполнены выборочные эксперименты, в которых определялись свойства двух байесовских оценок. Одна из них вычислялась на основе априорного распределения с отрицательным эксцессом, в то время как для отыскания другой оценки было взято априорное распределение с положительным эксцессом.

Первое априорное распределение по сравнению со вторым менее информативно в задаче оценивания параметров. Результаты выборочных экспериментов показали, что рассмотренные байесовские оценки лучше всех других оценок, в особенности в случае априорного распределения с положительным эксцессом. Однако даже при априорном распределении с отрицательным эксцессом для байесовских оценок результаты оказывались лучше, чем для оценок максимального правдоподобия (оценок по обобщенному методу наименьших квадратов, оценок по методу минимума χ^2). На основе этих экспериментальных результатов и некоторых общих соображений было установлено преимущество байесовского подхода к анализу модели и данных, которые она порождает.

Совершенно другим подходом к оценке параметров является процедура оценивания, которая минимизирует сумму абсолютных отклонений. В этой процедуре с помощью обычного алгоритма линейного программирования удается учесть ограничение неотрицательности параметров и условие нормировки. Результаты выборочных экспериментов показали, что получаемые таким образом оценки сходятся к истинным значениям параметров и распределены нормально при больших выборках. Однако оценка по методу наименьших абсолютных

отклонений неприемлема в том смысле, что она характеризуется бо́льшей среднеквадратической ошибкой по сравнению с другими оценками.

Прогнозирование величин относительных частот и применение критерия согласия χ^2 показало, что хотя оценки без учета ограничений могут обеспечить хорошие прогнозы, однако в некоторых случаях предсказанные относительные частоты оказываются отрицательными и, следовательно, лишенными смысла. Показатель качества прогноза, измеренный по величине критерия χ^2, может применяться для проверки гипотезы о том, что экспериментальные данные порождены марковской моделью. При подобном тесте результаты оказывались в согласии с гипотезой о том, что наблюдаемые данные представляют собой реализации марковского процесса.

При подведении итогов выборочных экспериментов было проведено сравнение различных оценок по критерию агрегированной среднеквадратической ошибки. Результаты экспериментов показывают, что оценки максимального правдоподобия (оценки по обобщенному методу наименьших квадратов, оценки по методу минимума χ^2) наилучшие среди всех небайесовских оценок. Рассмотренные байесовские оценки оказываются лучше всех других отчасти из-за того, что в них учитывается как выборочная, так и априорная информация. Кроме того, сравнивались дисперсии различных оценок. Вывод, вытекающий из этого сравнения, в точности совпадает с выводом, сделанным при сравнении оценок по величине среднеквадратической ошибки.

Применение знако-рангового критерия попарного сравнения (критерия Уилкоксона) также дало результаты, аналогичные тем, которые были получены при сравнении оценок по критерию среднеквадратической ошибки. Хотя результаты испытаний оказывались не всегда логичными, в результате последовательного применения нескольких критериев для установления относительного качества различных оценок можно предложить следующее упорядочение оценок: байесовские оценки (наилучшие), оценки максимального правдоподобия, оценки по методу наименьших квадратов с учетом ограничений и взвешивания, оценки по методу наименьших квадратов с учетом ограничений, оценки по методу наименьших абсолютных отклонений и обычные оценки по методу наименьших квадратов.

До сих пор при применении марковской вероятностной модели одна из основных трудностей была связана с необходимостью собирать подробные экспериментальные данные, упорядоченные во времени. В силу того что исследованные в настоящей книге методы являются алгоритмическими, они должны пройти большой путь развития в направлении уменьшения или ограничения проблемы, связанной с накоплением исходных данных. Тогда эти методы найдут многочисленные приложения в тех задачах, в которых экспериментальные данные согласуются с марковской моделью первого порядка.

Следует отметить, что все представленные в книге результаты получены в предположении о постоянстве переходных вероятностей в течение периода наблюдений (предположение стационарности). Это

предположение справедливо не всегда, поскольку развитие экономики может и не быть стационарным процессом, в особенности тогда, когда экономика подвержена действию таких внешних возмущающих факторов, как войны, переход к новой экономической политике и т. д. Поэтому переходные вероятности могут меняться во времени. Это обстоятельство находит отражение в так называемой нестационарной марковской цепи. В нестационарном случае остается нерешенной задача о том, как проследить изменения переходных вероятностей при отсутствии наблюдений микрообъектов, располагая лишь агрегированными по своей природе данными. Один из возможных подходов к подобной задаче оценивания рассматривается в приложении *В*.

Приложение А

МЕТОД ОЦЕНИВАНИЯ С ПОМОЩЬЮ ПСЕВДООБРАТНОЙ МАТРИЦЫ

В главе 5 отмечалось, что обобщенный метод наименьших квадратов нельзя применить в случае регрессионной модели для переходных вероятностей в силу вырожденности соответствующей ковариационной матрицы ошибок наблюдений. В этой ситуации в качестве одного из возможных подходов к решению задачи был предложен метод оценивания с помощью псевдообратной матрицы. В связи с этим цель предпринятого в приложении A исследования состоит в выяснении возможностей такого регрессионного метода при оценке переходных вероятностей.

А.1. Обобщенный метод наименьших квадратов в случае вырожденной ковариационной матрицы ошибок

В этом разделе вначале будет поставлена задача в достаточно общем виде. Рассмотрим линейную модель

$$\mathbf{y} = X\mathbf{p} + \mathbf{u}, \tag{А.1.1}$$

где $\mathbf{y}$ — это $(T \times 1)$-вектор, X — $(T \times k)$-матрица наблюдений, $\mathbf{p}$ — $(k \times 1)$-вектор оцениваемых параметров и $\mathbf{u}$ — $(T \times 1)$-вектор ошибок, удовлетворяющий условиям

$$E\mathbf{u} = \mathbf{0} \tag{А.1.2}$$

и

$$E\mathbf{u}\mathbf{u}' = \Sigma. \tag{А.1.3}$$

В соотношении (А.1.3) матрица Σ — вырожденная матрица ранга g, который меньше общего числа наблюдений T[1].

[1] Когда задача оценки параметров марковской цепи сформулирована как регрессионная задача, $\mathbf{y}$ есть $(rT \times 1)$-вектор, X — $(rT \times r^2)$-матрица, $\mathbf{p}$ — $(r^2 \times 1)$-вектор, $\mathbf{u}$ — $(rT \times 1)$-вектор и матрица Σ имеет ранг $(r-1)T$, который меньше полного ранга rT.

Если бы Σ была невырожденной матрицей, то наиболее эффективный способ оценки $\mathbf{p}$ заключался бы в преобразовании наблюдений, заданных (А.1.1), к виду, при котором приведенные ошибки были бы гомоскедастическими, с последующим вычислением оценки по методу наименьших квадратов. Подобное преобразование можно было бы осуществить путем перемножения невырожденной $(T \times T)$-матрицы H' и соотношения (А.1.1):

$$H'\mathbf{y} = H'X\mathbf{p} + H'\mathbf{u}, \tag{А.1.4}$$

выбрав матрицу H' так, чтобы

$$EH'\mathbf{u} = 0 \tag{А.1.5}$$

и

$$EH'\mathbf{u}\mathbf{u}'H = H'\Sigma H = I_T, \tag{А.1.6}$$

где I_T — единичная $(T \times T)$-матрица. В результате применения метода наименьших квадратов для оценки параметров в (А.1.4) получим, что

$$\overline{\mathbf{p}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathbf{y}, \tag{А.1.7}$$

где

$$\Sigma^{-1} = HH'. \tag{А.1.8}$$

Поскольку в нашем случае ковариационная матрица ошибок Σ вырожденная, то матрицы Σ^{-1} не существует и, следовательно, не существует подобной матрицы H.

Если Σ — вырожденная матрица ранга g, то предлагается отыскать $(T \times g)$-матрицу K ранга g, которая преобразует (А.1.1) к виду

$$K'\mathbf{y} = K'X\mathbf{p} + K'\mathbf{u} \tag{А.1.9}$$

так, что

$$EK'\mathbf{u} = 0 \tag{А.1.10}$$

и

$$EK'\mathbf{u}\mathbf{u}'K = K'\Sigma K = I_g, \tag{А.1.11}$$

где I_g — единичная $(g \times g)$-матрица[1]. Таким образом, K представляет собой обобщение матрицы H, так как если Σ имеет полный ранг, то K совпадает с H. Применение метода наименьших квадратов для оценки вектора $\mathbf{p}$ в соотношении (А.1.9) дает следующее выражение для оценки:

$$\overline{\mathbf{p}} = (X'\Sigma^{+}X)^{-1}X'\Sigma^{+}\mathbf{y}, \tag{А.1.12}$$

---

[1] Следует отметить, что в этом случае T наблюдений преобразуются в g наблюдений. Таким образом, возникает вопрос, эффективна ли оценка, основанная на соотношении (А.1.9) (см. также сноску 1 на с. 140).

где

$$\Sigma^{+} = KK'. \tag{A.1.13}$$

Матрица Σ^{+} называется псевдообратной матрицей для Σ и обладает следующими свойствами[1]:

$$\Sigma\Sigma^{+}\Sigma = \Sigma; \tag{A.1.14}$$

$$\Sigma^{+}\Sigma\Sigma^{+} = \Sigma^{+}; \tag{A.1.15}$$

$$(\Sigma\Sigma^{+})' = \Sigma\Sigma^{+}; \tag{A.1.16}$$

$$(\Sigma^{+}\Sigma)' = \Sigma^{+}\Sigma. \tag{A.1.17}$$

В дальнейшем будет доказано существование матрицы K и справедливость соотношений (A.1.12) и (A.1.13).

Вначале рассмотрим случай вырожденной матрицы Σ. Поскольку Σ — вещественная симметричная матрица, существует ортогональная матрица U — матрица собственных векторов — такая, что $U'\Sigma U$ является диагональной матрицей, элементы которой равны характеристическим числам матрицы Σ, т. е.

$$U'\Sigma U = \Lambda. \tag{A.1.18}$$

Если Σ — вырожденная матрица ранга g, то $(T-g)$ характеристических чисел должны быть равны нулю. Таким образом,

$$\Lambda = \begin{bmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_g & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} = \left[\begin{array}{c|c} \lambda & \\ \hline & 0 \end{array}\right], \tag{A.1.19}$$

где собственные числа λ_i могут быть расположены в порядке убывания: $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \ldots \geqslant \lambda_g$. В силу этого можно показать, что

$$\Lambda^{-\frac{1}{2}} U' \Sigma U \Lambda^{-\frac{1}{2}} = \begin{bmatrix} I_g & \\ & 0 \end{bmatrix}, \tag{A.1.20}$$

где $\Lambda^{-\frac{1}{2}}$ — диагональная матрица, в которой g элементов равны $\lambda_i^{-\frac{1}{2}}$ и $(T-g)$ элементов равны нулю. Предположим, что ортогональная матрица U разбита на два блока:

$$U = (U_1 \vdots U_2), \tag{A.1.21}$$

[1] Более подробно псевдообратные матрицы рассматриваются в работах [14], [87], [90], [91], [92].

где блок U_1 содержит g столбцов, а блок U_2 содержит $(T-g)$ столбцов. Тогда можно записать, что

$$U\Lambda^{-\frac{1}{2}} = (U_1 \mid U_2)\left(\begin{array}{c|c} \lambda^{-\frac{1}{2}} & \\ \hline & 0 \end{array}\right) = (U_1\lambda^{-\frac{1}{2}} \mid 0), \qquad (A.1.22)$$

где $U_1\lambda^{-\frac{1}{2}}$ есть $(T \times g)$-матрица, которую мы обозначим через K, т. е.

$$K = U_1\lambda^{-\frac{1}{2}}. \qquad (A.1.23)$$

Таким образом, имеет место следующее соотношение:

$$K'\Sigma K = \lambda^{-\frac{1}{2}}\, U_1'\Sigma U_1\lambda^{-\frac{1}{2}} = I_g, \qquad (A.1.24)$$

что и доказывает существование матрицы K.

Далее из (A.1.23) следует, что

$$U_1 = K\lambda^{\frac{1}{2}}, \qquad (A.1.25)$$

где $\lambda^{\frac{1}{2}}$ — диагональная $(g \times g)$-матрица, элементы которой равны λ_i. Поскольку матрица U имеет T ортогональных векторов, то $U^{-1} = U'$. Поэтому для ковариационной матрицы Σ можно записать соотношение

$$\Sigma = U\Lambda U' = (U_1 \mid U_2)\left(\begin{array}{c|c} \lambda & \\ \hline & 0 \end{array}\right)\left(\begin{array}{c} U_1' \\ \hline U_2' \end{array}\right) = U_1\Lambda U_1', \qquad (A.1.26)$$

которое с учетом (A.1.25) принимает вид

$$\Sigma = U_1\lambda U_1' = K\lambda^{\frac{1}{2}}\lambda\lambda^{\frac{1}{2}}K' = K\lambda^2K' = (K\lambda)\,(K\lambda)'. \qquad (A.1.27)$$

Согласно [29], если вырожденная матрица Σ может быть представлена в виде произведения прямоугольных матриц $(K\lambda)$ полного ранга, то, принимая во внимание (A.1.23), можно получить, что

$$K'K\lambda = \lambda^{-\frac{1}{2}}\, U_1'U_1\lambda^{-\frac{1}{2}}\lambda = I_g \qquad (A.1.23a)$$

и псевдообратная матрица определяется соотношением

$$\Sigma^+ = (K\lambda)\,(\lambda K'K\lambda)^{-1}\,(\lambda K'K\lambda)^{-1}\,(K\lambda)' = KK'. \qquad (A.1.28)$$

Это соотношение завершает доказательство (А.1.13).

Полученный результат позволяет при выполнении некоторых условий[1] найти наилучшую линейную несмещенную оценку вектора параметров модели (А.1.9) по критерию наименьших квадратов. Эта оценка имеет вид

$$\bar{\mathbf{p}} = (X' KK' X)^{-1} X' KK' \mathbf{y} = (X' \Sigma^{+} X)^{-1} X' \Sigma^{+} \mathbf{y}, \qquad (А.1.29)$$

если матрица $(X'\Sigma^{+}X)^{-1}$ существует. Дисперсия оценки $\bar{\mathbf{p}}$ выражается следующим образом:

$$V(\bar{\mathbf{p}}) = (X' KK' X')^{-1} = (X' \Sigma^{+} X)^{-1}. \qquad (А.1.30)$$

В случае, когда Σ — невырожденная матрица, матрицы K и H равны между собой, а оценки (А.1.7) и (А.1.12) совпадают. Поэтому оценка (А.1.12) или (А.1.29) в определенном смысле будет более общей.

А.2. Вычисление псевдообратной матрицы для ковариационной матрицы ошибок в случае марковской вероятностной модели

Ковариационная матрица ошибок наблюдений, представляющих собой относительные частоты, имеет вид[2]

---

[1] Поскольку исходные T наблюдений были преобразованы в g наблюдений ($g < T$), то $(T - g)$ наблюдений были неявно исключены из рассмотрения. Именно эти исключенные наблюдения связаны с $(T - g)$ собственными числами, равными нулю, и соответствующими собственными векторами. Пусть U_2 есть $(T \times (T - g))$-матрица, которая преобразует соотношение

$$U_2' \mathbf{y} = U_2' X \mathbf{p} + U_2' \mathbf{u},$$

так что

$$EU_2' \mathbf{u} = 0 \qquad (а)$$

и

$$EU_2' \mathbf{u}\mathbf{u}' U_2 = 0. \qquad (б)$$

Из (А.1.8) следует, что ковариация ошибок этого преобразования равна нулю, так как Σ — вырожденная матрица. Соотношения (а) и (б) означают, что $U_2' \mathbf{u} = 0$ и что полученная таким образом точная связь

$$U_2' \mathbf{y} = U_2' X \mathbf{p} \qquad (в)$$

является линейным ограничением на параметры типа равенств. Поэтому оценка (А.1.29) будет эффективной, если $U_2' X = 0$, т. е. когда это ограничение исчезает [112]. Если $U_2' X \neq 0$, то это ограничение должно быть учтено. В нашем случае матрица $U_2' = (I, I, \ldots, I)$ представляет собой $(T \times rT)$-матрицу, состоящую из r единичных матриц. Ограничение (в), когда оно не исчезает, эквивалентно равенству $\eta_r = G\mathbf{p}$, если обе части (в) домножены на матрицу $(X_1' X_1)^{-1} X_1'$. Это ограничение уже рассматривалось в случае применения метода наименьших квадратов с учетом ограничений. Даже если ограничения не учитывать, то оценка (А.1.29) автоматически будет удовлетворять ограничению (в) в силу того, что наблюдениями являются относительные частоты. Поэтому оценка (А.1.29) в нашем случае эффективна.

[2] Эта матрица рассматривалась в главах 5—9.

$$\Sigma = \begin{bmatrix} \frac{q_1(t)(1-q_1(t))}{N(t)} & \frac{-q_1(t)q_2(t)}{N(t)} & \cdots & \frac{-q_1(t)q_r(t)}{N(t)} \\ \frac{-q_2(t)q_1(t)}{N(t)} & \frac{q_2(t)(1-q_2(t))}{N(t)} & \cdots & \frac{-q_2(t)q_r(t)}{N(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-q_r(t)q_1(t)}{N(t)} & \frac{-q_r(t)q_2(t)}{N(t)} & \cdots & \frac{q_r(t)(1-q_r(t))}{N(t)} \end{bmatrix} \otimes I^T_{t=1} =$$

$$= V(t) \otimes I^t_{t=1}, \qquad (A.2.1)$$

где через $V(t)$ обозначена матрица, диагональные и недиагональные элементы которой соответственно равны $q_i(t)(1-q_i(t))/N(t)$ и $-q_i(t)q_j(t)/N(t)$, $I^T_{t=1}$ определена в главе 6. Поскольку имеет место соотношение

$$\Sigma^+ = V(t)^+ \otimes I^T_{t=1}, \qquad (A.2.2)$$

для определения Σ^+ необходимо вычислить $V(t)^+$. Для отыскания $V(t)^+$, Σ^+ можно воспользоваться методом, описанным в предыдущем параграфе, т. е. сначала найти характеристические векторы, соответствующие положительным собственным числам, и определить матрицы U_1 и λ (имеющие одинаковый вид), затем следует вычислить матрицу K путем деления i-го столбца матрицы U_1 на величину квадратного корня из i-го собственного числа. Матрица KK' тогда будет искомой псевдообратной матрицей[1].

[1] В качестве примера рассмотрим (2×2)-матрицу

$$\Sigma_2 = \begin{bmatrix} \frac{p_1 p_2}{N} & \frac{-p_1 p_2}{N} \\ \frac{-p_1 p_2}{N} & \frac{p_1 p_2}{N} \end{bmatrix},$$

где $p_1 + p_2 = 1$. Характеристическими числами этой матрицы являются $p_1 p_2/N$ и 0. Соответствующие характеристические векторы после нормировки равны:

$$U_1 = \sqrt{1/2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ для } \lambda_1 = 2p_1 p_2/N \text{ и } U_2 = \sqrt{-1/2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ для } \lambda_2 = 0.$$

Таким образом, матрица K равна:

$$K = \frac{\sqrt{N}}{\sqrt{p_1 p_2}} \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix},$$

и искомая псевдообратная матрица имеет вид

$$\Sigma_2^+ = KK' = \frac{\sqrt{N}}{\sqrt{p_1 p_2}} \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} (1/2 \quad -1/2) \left(\frac{\sqrt{N}}{\sqrt{p_1 p_2}} \right) =$$

$$= \begin{bmatrix} N/4p_1 p_2 & -N/4p_1 p_2 \\ -N/4p_1 p_2 & N/4p_1 p_2 \end{bmatrix}.$$

Если размер матрицы Σ велик, то аналитический расчет псевдообратной матрицы Σ^+ достаточно сложен. В этом случае для вычисления можно воспользоваться методом Бута [14], который вкратце заключается в следующем. Если матрица A вырожденная, то вначале A может быть записана как

$$A = \begin{bmatrix} D & E \\ F & G \end{bmatrix}, \tag{A.2.3}$$

где подматрица D имеет тот же ранг, что и A, и существует матрица D^{-1}. Блочные строки (столбцы) матрицы A линейно-зависимы. Иначе говоря, блочная строка (FG) будет линейной комбинацией строки (DE), а блочный столбец (EG) — линейной комбинацией столбца (DF). Пусть $F = HD$, $G = HE$, $E = DK$, $G = FK$, тогда матрица A может быть записана в виде

$$A = \begin{bmatrix} D & E \\ F & G \end{bmatrix} = \begin{bmatrix} D & E \\ HD & HE \end{bmatrix} = \begin{bmatrix} D & DK \\ HD & HDK \end{bmatrix} = \begin{bmatrix} I \\ H \end{bmatrix} D [I \quad K] = RDQ', \tag{A.2.4}$$

где $R' = (IH)$ и $Q' = (IK)$. Тогда псевдообратная матрица A^+ определяется выражением

$$A^+ = Q\,(Q'Q)^{-1}\,D^{-1}\,\,(R'R)^{-1}\,R'. \tag{A.2.5}$$

Применяя формулы (A.2.4) и (A.2.5) для решения нашей задачи, получим следующие соотношения для R и Q:

$$Q = R = \begin{bmatrix} I_{r-1} \\ -1 -1 \ldots -1 \end{bmatrix} = \begin{bmatrix} I_{r-1} \\ -\eta'_{r-1} \end{bmatrix}, \tag{A.2.6}$$

где η_{r-1} есть $((r-1) \times 1)$-вектор-столбец единиц, а I_{r-1} — единичная $((r-1) \times (r-1))$-матрица. Матрицы Q и R совпадают, так как $H = K = -\eta_r$. Таким образом, получаем, что

$$Q'\,Q = R'\,R = [I_{r-1} - \eta_r] \begin{bmatrix} I_{r-1} \\ -\eta'_r \end{bmatrix} = [I_{r-1} + \eta_r\,\eta'_r] = [I_{r-1} + E_{r-1}], \tag{A.2.7}$$

где E_{r-1} — $((r-1) \times (r-1))$-матрица, все элементы которой равны единице. Следовательно, матрица $Q'Q$ (или $R'R$) есть $((r-1) \times \times (r-1))$-матрица, диагональные элементы которой равны 2, а все остальные элементы равны 1. Определитель матрицы $Q'Q$ равен

$$\det Q'Q = r. \tag{A.2.8}$$

Алгебраические дополнения $(Q'Q)_{ii}$ и $(Q'Q)_{ij}$ $(i \neq j)$ соответственно равны:

$$(Q'Q)_{ii} = r - 1 \tag{A.2.9}$$

и

$$(Q'Q)_{ij} = -1. \tag{A.2.10}$$

Тогда матрица, обратная к $Q'Q$ (или $R'R$), определяется по формуле

$$(Q'Q)^{-1} = \begin{bmatrix} \frac{r-1}{r} & -\frac{1}{r} & -\frac{1}{r} & \dots & -\frac{1}{r} \\ -\frac{1}{r} & \frac{r-1}{r} & -\frac{1}{r} & \dots & -\frac{1}{r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{r} & -\frac{1}{r} & -\frac{1}{r} & \dots & \frac{r-1}{r} \end{bmatrix} = \left[I_{r-1} - \frac{1}{r} E_{r-1} \right]. \tag{A.2.11}$$

Выражения для $Q(Q'Q)^{-1}$ и $(R'R)^{-1} R'$ совпадают и равны:

$$Q(Q'Q)^{-1} = \begin{bmatrix} I_{r-1} \\ \hline -\eta'_{r-1} \end{bmatrix} \left[I_{r-1} - \frac{1}{r} E_{r-1} \right] =$$

$$= \begin{bmatrix} \left(I_{r-1} - \frac{1}{r} E_{r-1} \right) \\ \hline \left(-\eta'_{r-1} + \frac{1}{r} \eta'_{r-1} E_{r-1} \right) \end{bmatrix} = \begin{bmatrix} \left(I_{r-1} - \frac{1}{r} E_{r-1} \right) \\ \hline \left(-\frac{1}{r} \eta'_{r-1} \right) \end{bmatrix}. \tag{A.2.12}$$

С другой стороны, ранг матрицы D равен $r - 1$ и матрица D^{-1} равна[1]:

$$D^{-1} = \begin{bmatrix} \frac{N(t)}{q_1(t)} + \frac{N(t)}{q_r(t)} & \frac{N(t)}{q_r(t)} & \dots & \frac{N(t)}{q_r(t)} \\ \frac{N(t)}{q_r(t)} & \frac{N(t)}{q_2(t)} + \frac{N(t)}{q_r(t)} & \dots & \frac{N(t)}{q_r(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{N(t)}{q_r(t)} & \frac{N(t)}{q_r(t)} & \dots & \frac{N(t)}{q_{r-1}(t)} + \frac{N(t)}{q_r(t)} \end{bmatrix} =$$

$$= \left[\frac{N(t)}{q_r(t)} E_{r-1} + A \right], \tag{A.2.13}$$

[1] По поводу матрицы D^{-1} см. главу 6.

где A — диагональная $((r-1)\times(r-1))$-матрица, элементы которой равны $N(t)/q_i(t)$ $(i=1, 2, \ldots, r-1)$. Обозначим временно $q_i(t)$ через q_i, а $N(t)$ через N. Применяя формулу (А.2.5), получим, что псевдообратная матрица для матрицы V равна:

$$V^+(t)=\begin{bmatrix} \left(I_{r-1}-\frac{1}{r}E_{r-1}\right) \\ \hline \left(-\frac{1}{r}\eta'_{r-1}\right) \end{bmatrix}\left[\frac{N}{q_r}E_{r-1}+A\right]\left[\left(I_{r-1}-\frac{1}{r}E_{r-1}\right)\times\right.$$

$$\left.\times\left(-\frac{1}{r}\eta_{r-1}\right)\right]. \qquad (А.2.14)$$

После операций умножения в (А.2.14) можно представить $V^+(t)$ в виде блочной матрицы с четырьмя блоками:

$$V^+=\begin{bmatrix} V_1 & \mathbf{V}_2 \\ \mathbf{V}_3 & V_4 \end{bmatrix}, \qquad (А.2.15)$$

где V_1 есть $((r-1)\times(r-1))$-матрица, $\mathbf{V}_2$ — $((r-1)\times 1)$-вектор-столбец, $\mathbf{V}_3$ — $(1\times(r-1))$-вектор-строка и V_4 — скалярная величина, соответственно равные:

$$V_1=(N/q_r)\,E_{r-1}+A-((r-1)/r)\,E_{r-1}-(1/r)\,\eta_{r-1}\,\mathbf{B}-((r-1)/r)\times$$
$$\times(N/q_r)\,E_{r-1}-(1/r)\,\mathbf{B}'\eta_{r-1}+((r-1)^2/r^2)\,(N/q_r)\,\times$$
$$\times E_{r-1}+(1/r^2)\,\eta_{r-1}\,\mathbf{B}E_{r-1}; \qquad (А.2.16)$$

$$\mathbf{V}_2=\mathbf{V}'_3=-((r-1)/r)\,(N/q_r)\,\eta_{r-1}-(1/r)\,\mathbf{B}'+((r-1)^2/r^2)\,(N/q_r)\,\eta_{r-1}+$$
$$+(1/r^2)\sum_{i=1}^{r-1}(N/q_i)\,\eta_{r-1} \qquad (А.2.17)$$

и

$$V_4=((r-1)^2/r^2)\,(N/q_r)\,\eta_{r-1}+(1/r^2)\,\mathbf{B}\,\eta_{r-1}. \qquad (А.2.18)$$

Здесь A — матрица, определение которой было дано в тексте сразу же после формулы (А.2.13), $\mathbf{B}$ — вектор-строка $[N/q_1, N/q_2, \ldots, N/q_{r-1}]$. В качестве примера рассмотрим случай, когда $r=2$. При этом матрица V^+ принимает вид

$$V^+(t)=\begin{bmatrix} \frac{N}{4q_1q_2} & -\frac{N}{4q_1q_2} \\ -\frac{N}{4q_1q_2} & \frac{N}{4q_1q_2} \end{bmatrix}. \qquad (А.2.19)$$

При $r = 3$ матрица $V^+(t)$ равна:

$$V^+(t) =$$

$$= \begin{bmatrix} \frac{4N}{9q_1} + \frac{N}{9q_2} + \frac{N}{9q_3} & -\frac{2N}{9q_1} - \frac{2N}{9q_2} + \frac{N}{9q_3} & -\frac{2N}{9q_1} + \frac{N}{9q_2} - \frac{2N}{9q_3} \\ -\frac{2N}{9q_1} - \frac{2N}{9q_2} + \frac{N}{9q_3} & \frac{N}{9q_1} + \frac{4N}{9q_2} + \frac{N}{9q_3} & \frac{N}{9q_1} - \frac{2N}{9q_2} - \frac{2N}{9q_3} \\ -\frac{2N}{9q_1} + \frac{N}{9q_2} - \frac{2N}{9q_3} & \frac{N}{9q_1} - \frac{2N}{9q_2} - \frac{2N}{9q_3} & \frac{N}{9q_1} + \frac{N}{9q_2} + \frac{4N}{9q_3} \end{bmatrix}. \tag{A.2.20}$$

A.3. Мультиколлинеарный случай

Определив псевдообратную матрицу Σ^+, можно перейти к вычислению оценки (A.1.12). Для того чтобы применить формулу (A.1.12), необходимо выяснить, существует ли $(r^2 \times r^2)$-матрица $(X'\Sigma^+X)^{-1}$. Пусть матрица Σ^+ разбита на блоки диагональных подматриц s_{ij} и представлена в виде

$$\Sigma^+ = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1r} \\ s_{21} & s_{22} & \dots & s_{2r} \\ \vdots & \vdots & & \vdots \\ s_{r1} & s_{r2} & \dots & s_{rr} \end{bmatrix}, \tag{A.3.1}$$

где каждая матрица s_{ij} имеет размер $(T \times T)$. Тогда матричное произведение $(X'\Sigma^+X)$ равно:

$$X'\Sigma^+X = \begin{bmatrix} X_1' s_{11} X_1 & X_1' s_{12} X_2 & \dots & X_1' s_{1r} X_r \\ X_2' s_{21} X_1 & X_2' s_{22} X_2 & \dots & X_2' s_{2r} X_r \\ \vdots & \vdots & \ddots & \vdots \\ X_r' s_{r1} X_1 & X_r' s_{r2} X_2 & \dots & X_r' s_{rr} X_r \end{bmatrix}. \tag{A.3.2}$$

Поскольку в выражении (A.3.2) матрица Σ вырожденная, матрица Σ^+ также вырожденная. Суммы блочных строк матрицы Σ^+ равны нулю. Более того, так как $X_1 = X_2 = \dots = X_r = x$, суммы блочных строк матрицы (A.3.2) также равны нулю, т. е.

$$\sum_{j=1}^{r} x' s_{ij} x = x' \sum_{j=1}^{r} s_{ij} x = 0. \tag{A.3.3}$$

Таким образом, можно заключить, что матрица $(X'\Sigma^+X)$ является вырожденной и матрицы $(X'\Sigma^+X)^{-1}$ не существует. Следовательно, оценка $\bar{p}$ не может быть вычислена по формуле (A.1.12). Мы рассмотрели так называемый мультиколлинеарный случай.

А.4. Условие нормировки и редуцированная весовая матрица

Как было показано, применение метода вычисления псевдообратной матрицы при расчете марковской вероятностной модели связано с трудностями, обусловленными проблемой мультиколлинеарности. Однако даже этот случай не полностью безнадежный. Так, в работе [53] показано, что если удается получить независимые оценки соответствующего числа параметров модели, то по тем же выборочным данным могут быть вычислены оценки остальных параметров. Для параметров модели имеет место соотношение

$$\sum_{T}^{T} \mathrm{p}_j = \eta_r, \tag{А.4.1}$$

поэтому, если известны векторы $\mathrm{p}_1, \mathrm{p}_2, \ldots, \mathrm{p}_{r-1}$, вектор p_r также известен в силу (А.4.1). Таким образом, можно считать, что T последних строк матрицы $(X'\Sigma^+X)$ линейно-зависимы от первых $(r-1)\,T$ строк, и рассмотреть вместо полной системы нормальных уравнений

$$(X'\Sigma^+X)\,\overline{\mathrm{p}} = (X'\Sigma^+\mathrm{y}) \tag{А.4.2}$$

редуцированную систему вида

$$\begin{bmatrix} x' s_{11} x & x' s_{12} x & \ldots & x' s_{1r} x \\ x' s_{21} x & x' s_{22} x & \ldots & x' s_{2r} x \\ \vdots & \vdots & \ddots & \vdots \\ x' s_{r-1,1} x & x' s_{r-1,2} x & & x' s_{r-1,r} x \end{bmatrix} \times$$

$$\times \begin{bmatrix} \overline{\mathrm{p}}_1 \\ \overline{\mathrm{p}}_2 \\ \vdots \\ \overline{\mathrm{p}}_r \end{bmatrix} = \begin{bmatrix} \sum_T x' s_{1l}\, \mathrm{y}_l \\ \sum_T x' s_{2l}\, \mathrm{y}_l \\ \vdots \\ \sum_T x' s_{r-1,l}\, \mathrm{y}_l \end{bmatrix}. \tag{А.4.3}$$

С помощью (А.4.1) и соотношения

$$\mathrm{y}_r = \eta_r - \sum_{l=1}^{r-1} \mathrm{y}_l \tag{А.4.4}$$

последние r столбцов матрицы $(X'\Sigma^+X)$ редуцированной системы могут быть выражены через другие столбцы. Например, рассмотрим совокупность уравнений, соответствующих первым T строкам в системе уравнений (А.4.3):

$$x' s_{11} x\, \mathrm{p}_1 + x' s_{12} x\, \mathrm{p}_2 + \ldots + x' s_{1r} x\, \mathrm{p}_r = \Sigma_l\, x' s_{1l}\, \mathrm{y}_l. \tag{А.4.5}$$

Уравнения (А.4.5) эквивалентны уравнениям:

$$x' s_{11} x \mathbf{p}_1 + x' s_{12} x \mathbf{p}_2 + \ldots + x' s_{1r} x (\boldsymbol{\eta}_r - \mathbf{p}_1 - \mathbf{p}_2 - \ldots - \mathbf{p}_{r-1}) =$$

$$= \sum_{l=1}^{r-1} x' s_{1l} \mathbf{y}_l + x' s_{1r} (\boldsymbol{\eta}_r - \mathbf{y}_1 - \mathbf{y}_2 - \ldots - \mathbf{y}_{r-1}). \qquad (A.4.6)$$

В результате простых преобразований (А.4.6) принимает вид

$$x' (s_{11} - s_{1r}) x \mathbf{p}_1 + x' (s_{12} - s_{1r}) x \mathbf{p}_2 + \ldots + x' (s_{1,r-1} - s_{1r}) x \mathbf{p}_{r-1} =$$

$$= \sum_{l=1}^{r-1} x' (s_{1l} - s_{1r}) \mathbf{y}_l. \qquad (A.4.7)$$

Очевидно, что теперь система (А.4.3) может быть переписана следующим образом:

$$\begin{bmatrix} x' (s_{11} - s_{1r}) x & \ldots & x' (s_{1,r-1} - s_{1r}) x \\ x' (s_{21} - s_{2r}) x & \ldots & x' (s_{2,r-1} - s_{2r}) x \\ \vdots & \ddots & \vdots \\ x' (s_{r-1,1} - s_{r-1,r}) x & \ldots & x' (s_{r-1,r-1} - s_{r-1,r}) x \end{bmatrix} \begin{bmatrix} \overline{\mathbf{p}}_1 \\ \overline{\mathbf{p}}_2 \\ \vdots \\ \overline{\mathbf{p}}_{r-1} \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_{l}^{r-1} x' (s_{1l} - s_{1r}) \mathbf{y}_l \\ \vdots \\ \sum_{l}^{r-1} x' (s_{r-1,l} - s_{r-1,r}) \mathbf{y}_l \end{bmatrix}. \qquad (A.4.8)$$

Следовательно, редуцированная весовая матрица имеет вид

$$\Sigma^* = \begin{bmatrix} s_{11} - s_{1r} & & s_{1,r-1} - s_{1r} \\ s_{21} - s_{2r} & & s_{2,r-1} - s_{2r} \\ \vdots & & \vdots \\ s_{r-1} - s_{r-1,r} & \ldots & s_{r-1,r-1} - s_{r-1,r} \end{bmatrix}, \qquad (A.4.9)$$

где Σ^* есть $((r-1)\,T \times (r-1)\,T)$-матрица. Редуцированная матрица V^* может быть выражена через величины $q_l\,(t)$, $N\,(t)$ (сокращенно q_l, N) и представлена в виде

$$V^* = [V_1 - \mathbf{V}_2 \;\; \boldsymbol{\eta}'_{r-1}] = [(1/r) \;\; (N/q_r) \;\; E_{r-1} + (A - (1/r)\, \boldsymbol{\eta}_{r-1}\, \mathbf{B})], \qquad (A.4.10)$$

где V_1, $\mathbf{V}_2$, A и $\mathbf{B}$ были определены в параграфе А.2[1]. Например, для случая, когда $r = 2$, редуцированная весовая матрица V^* равна:

$$V^* = (1/2)(N/q_2) + (N/q_1) - (1/2)(N/q_1) = N/2\,q_1 q_2, \tag{А.4.11}$$

а при $r = 3$ матрица V^* имеет вид

$$V^* = (1/3)(N/q_3)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \frac{N}{q_1} & \\ & \frac{N}{q_2} \end{bmatrix} - (1/3)\begin{bmatrix} \frac{N}{q_1} & \frac{N}{q_2} \\ \frac{N}{q_1} & \frac{N}{q_2} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{2N}{3q_1} + \frac{N}{3q_3} & -\frac{N}{3q_2} + \frac{N}{3q_3} \\ -\frac{N}{3q_1} + \frac{N}{3q_3} & \frac{2N}{3q_2} + \frac{N}{3q_3} \end{bmatrix}. \tag{А.4.12}$$

Таким образом, искомые оценки векторов $\mathbf{p}_1, \ldots, \mathbf{p}_{r-1}$ могут быть вычислены по формуле

$$\overline{\mathbf{p}}_* = (X_*' \Sigma^* X_*)^{-1} X_*' \Sigma^* \mathbf{y}_*, \tag{А.4.13}$$

где через X_* и $\mathbf{y}_*$ обозначены соответственно матрица, имеющая $(r-1)$ диагональных блоков X_i, и вектор с координатами $\mathbf{y}_1, \ldots, \mathbf{y}_{r-1}$. Последний вектор $\mathbf{p}_r$ может быть найден из соотношения

$$\overline{\mathbf{p}}_r = \eta_r - \sum_{j=1}^{r-1} \overline{\mathbf{p}}_j. \tag{А.4.14}$$

Заметим, что формула (А.4.13) дает оценку без учета ограничений. Оценка с учетом ограничений может быть получена при минимизации выражения

$$(\mathbf{y}_* - X_* \mathbf{p}_*)\, \Sigma^* (\mathbf{y}_* - X_* \mathbf{p}_*) \tag{А.4.15}$$

при наличии ограничений

$$\sum_{j=1}^{r} \mathbf{p}_j = \eta_r \tag{А.4.16}$$

и

$$\mathbf{p}_j \geqslant 0,\ j = 1, 2, \ldots, r-1. \tag{А.4.17}$$

---

[1] Редуцированная матрица V^* входит в следующее выражение для Σ^*:

$$\Sigma^* = V^* \otimes I_{l=1}^T.$$

Симплекс-таблица для этой задачи квадратичного программирования аналогична симплекс-таблице для соответствующей задачи, рассмотренной в главе 6. Однако в отличие от оценок по методу максимального правдоподобия и по обобщенному методу наименьших квадратов оценка с учетом ограничений и взвешивания в соответствии с Σ^* не является единственной. Другими словами, если различные составляющие векторы $\mathbf{p}_j$ исключаются из рассмотрения и заменяются невязками соотношения (А.4.16), то получаемые решения будут различаться. Лишь в случае (2×2)-матрицы переходных вероятностей, когда редуцированная весовая матрица представляет собой скаляр, получаемая оценка совпадает с оценкой по обобщенному методу наименьших квадратов или по методу максимального правдоподобия.

В качестве иллюстрации существования нескольких решений задачи (А.4.15) с учетом ограничений (А.4.16) и (А.4.17) рассмотрим пример Телсера [110]. Оценки с учетом и без учета ограничений в случае редуцированной матрицы Σ^*, полученной из псевдообратной матрицы Σ^+, приведены в табл. А.1. Эти оценки вычислены для трех задач, отличающихся первоначально исключенным вектором $\mathbf{p}_j$.

Таблица А.1

Оценки с учетом и без учета ограничений, вычисленные при взвешивании на редуцированную матрицу Σ^* для задачи Телсера

Исключенный вектор $\mathbf{p}_j$	Оценка без учета ограничений	Оценки с учетом ограничений
$\mathbf{p}_3$	$\begin{bmatrix} 0{,}5748 & 0{,}2975 & 0{,}1277 \\ -0{,}0782 & 0{,}9791 & 0{,}0991 \\ 0{,}6195 & -0{,}3363 & 0{,}7168 \end{bmatrix}$	$\begin{bmatrix} 0{,}6140 & 0{,}1145 & 0{,}2144 \\ 0 & 0{,}8963 & 0{,}1037 \\ 0{,}4794 & 0 & 0{,}5206 \end{bmatrix}$
$\mathbf{p}_2$	$\begin{bmatrix} 0{,}5748 & 0{,}2975 & 0{,}1277 \\ -0{,}0782 & 0{,}9791 & 0{,}0991 \\ 0{,}6195 & -0{,}3363 & 0{,}7168 \end{bmatrix}$	$\begin{bmatrix} 0{,}6759 & 0{,}1320 & 0{,}1921 \\ 0 & 0{,}8783 & 0{,}1262 \\ 0{,}3979 & 0 & 0{,}6021 \end{bmatrix}$
$\mathbf{p}_1$	$\begin{bmatrix} 0{,}5748 & 0{,}2975 & 0{,}1277 \\ -0{,}0782 & 0{,}9791 & 0{,}0991 \\ 0{,}6195 & -0{,}3363 & 0{,}7168 \end{bmatrix}$	$\begin{bmatrix} 0{,}7692 & 0{,}1127 & 0{,}1181 \\ 0 & 0{,}8980 & 0{,}1020 \\ 0{,}2729 & 0 & 0{,}7271 \end{bmatrix}$

Интересно отметить, что оценки без учета ограничений оказываются одними и теми же, независимо от исключенного $\mathbf{p}_j$ (см. табл. А.1). В то же время соответствующие оценки с учетом ограничений различаются, поскольку имеющиеся условия не обеспечивают полного набора ограничений в задаче минимизации.

А.5. Существование и свойства решения задачи оценивания с учетом ограничений в случае редуцированной весовой матрицы

Естественна попытка выявить случаи, когда метод наименьших квадратов с учетом ограничений и взвешивания в соответствии с Σ^* дает единственное решение задачи оценивания. Проведенное исследование показало, что для получения единственного решения необходимо,

чтобы весовая матрица была симметричной, независимо от того, какой набор параметров первоначально исключается. В нашей задаче симметричная редуцированная весовая матрица может быть получена следующим образом. Прежде чем осуществить переход от системы нормальных уравнений (А.4.2) к редуцированной системе (А.4.3), вычтем последнее матричное уравнение системы (А.4.2) из каждого матричного уравнения этой системы и получим совокупность $(r-1)$ систем нормальных уравнений, которая в матричной записи имеет следующий вид:

$$\begin{bmatrix} x'(s_{11}-s_{r1})x & x'(s_{12}-s_{r2})x & \dots & x'(s_{1r}-s_{rr})x \\ x'(s_{21}-s_{r1})x & x'(s_{22}-s_{r2})x & \dots & x'(s_{2r}-s_{rr})x \\ \vdots & \vdots & & \vdots \\ x'(s_{r-1,1}-s_{r1})x & x'(s_{r-1,2}-s_{r2})x & \dots & x'(s_{r-1,r}-s_{rr})x \end{bmatrix} \begin{bmatrix} \bar{\mathbf{p}}_1 \\ \bar{\mathbf{p}}_2 \\ \vdots \\ \bar{\mathbf{p}}_r \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_l x'(s_{1l}-s_{rl})\,\mathbf{y}_l \\ \sum_l x'(s_{2l}-s_{rl})\,\mathbf{y}_l \\ \vdots \\ \sum_l x'(s_{r-1,l}-s_{rl})\,\mathbf{y}_l \end{bmatrix}. \qquad (A.5.1)$$

После выполнения этих преобразований для получения единственного решения можно решать все уравнения системы (А.5.1) одновременно. В таком случае, учитывая (А.4.1) и (А.4.4), можно выразить $\mathbf{y}_r$ через остальные $\mathbf{y}_i$ $(i = 1, 2, \dots, r-1)$, а $\mathbf{p}_r$ — через остальные $\mathbf{p}_i$ $(i = 1, 2, \dots, r-1)$. Тогда система уравнений (А.5.1) принимает вид

$$\begin{bmatrix} x'b_{11}x & x'b_{12}x & \dots & x'b_{1,r-1}x \\ x'b_{21}x & x'b_{22}x & \dots & x'b_{2,r-1}x \\ \vdots & \vdots & & \vdots \\ x'b_{r-1,1}x & x'b_{r-1,2}x & \dots & x'b_{r-1,r-1}x \end{bmatrix} \begin{bmatrix} \bar{\mathbf{p}}_1 \\ \bar{\mathbf{p}}_2 \\ \vdots \\ \bar{\mathbf{p}}_{r-1} \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_l x'b_{1l}\,\mathbf{y}_l \\ \sum_l x'b_{2l}\,\mathbf{y}_l \\ \sum_l x'b_{r-1,l}\,\mathbf{y}_l \end{bmatrix}, \qquad (A.5.2)$$

где

$$b_{lj} = s_{lj} - s_{rj} - s_{lr} + s_{rr}. \qquad (A.5.3)$$

Таким образом, весовая матрица системы преобразована к известной симметричной редуцированной матрице:

$$\Sigma_*^{-1} = [b_{ij}] = \begin{bmatrix} \frac{N(t)}{q_r(t)} + \frac{N(t)}{q_1(t)} & \frac{N(t)}{q_r(t)} & \dots & \frac{N(t)}{q_r(t)} \\ \frac{N(t)}{q_r(t)} & \frac{N(t)}{q_r(t)} + \frac{N(t)}{q_2(t)} & \dots & \frac{N(t)}{q_r(t)} \\ \vdots & \vdots & & \vdots \\ \frac{N(t)}{q_r(t)} & \frac{N(t)}{q_r(t)} & \dots & \frac{N(t)}{q_r(t)} + \frac{N(t)}{q_{r-1}(t)} \end{bmatrix},$$

и сама система нормальных уравнений приобретает вид

$$(X_*' \Sigma_*^{-1} X_*) \bar{\mathbf{p}}_* = X_*' \Sigma_*^{-1} \mathbf{y}_*. \tag{A.5.4}$$

Заметим, что система уравнений (A.5.4) совпадает с системой нормальных уравнений, получаемой по обобщенному методу наименьших квадратов для модели, в которой опущено последнее уравнение. Как показано в главе 6, система (A.5.4) с учетом ограничений имеет единственное решение.

A.6. Заключение

Подводя итоги проведенного в приложении *A* исследования, хотелось бы более подробно обсудить взаимосвязь между исходной моделью и моделью, в которой исключены несколько параметров. Излагаемые далее результаты позволяют предложить метод вычисления матрицы Σ_*^{-1} по матрице Σ^* или по диагональной весовой матрице, которая используется при оценивании по критерию минимума χ^2.

Рассмотрим сначала задачу максимизации функции

$$\Phi(\mathbf{p}) = -(\mathbf{y} - X\mathbf{p})' \, S \, (\mathbf{y} - X\mathbf{p}) \tag{A.6.1}$$

при наличии ограничений

$$G\mathbf{p} = \boldsymbol{\eta}_r \tag{A.6.2}$$

и

$$\mathbf{p} \geqslant 0, \tag{A.6.3}$$

где $\mathbf{y}$ — $(r \times 1)$-вектор, X — блочно-диагональная $(rT \times r^2)$-матрица, блоки которой имеют одинаковый размер $(T \times r)$, $\mathbf{p}$ — $(r^2 \times 1)$-вектор, составленный из r векторов $\mathbf{p}_j$ $(j = 1, 2, \dots, r)$ одинакового размера $(r \times 1)$, S — весовая $(rT \times rT)$-матрица, G — $(r \times r^2)$-матрица из r единичных $(r \times r)$-подматриц и $\boldsymbol{\eta}_r$ — $(r \times 1)$-вектор-столбец единиц[1].

[1] Все эти матрицы и векторы были рассмотрены в главе 3.

Целевая функция (А.6.1) может быть представлена в виде суммы матричных произведений:

$$\Phi(\mathrm{p}) = -\sum_{l=1}^{r} (\mathrm{y}_l - x\mathrm{p}_l)' S_{ll} (\mathrm{y}_l - x\mathrm{p}_l) -$$

$$- \sum_{l=1}^{r} \sum_{j \neq l}^{r} (\mathrm{y}_l - x\mathrm{p}_l)' S_{lj} (\mathrm{y}_j - x\mathrm{p}_j), \qquad (\text{А.6.4})$$

где S_{lj} $(l, j = 1, 2, \ldots, r)$ — подматрицы размера $(T \times T)$ матрицы S, которая может быть записана в блочной форме:

$$S = \begin{bmatrix} S_{11} & S_{12} & \ldots & S_{1r} \\ S_{12} & S_{22} & \ldots & S_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ S_{r1} & S_{r2} & \ldots & S_{rr} \end{bmatrix}. \qquad (\text{А.6.5})$$

Если в (А.6.4) выделить слагаемые, соответствующие r-му матричному произведению, то (А.6.4) приобретает вид

$$\Phi(\mathrm{p}) = -\sum_{l=1}^{r-1} (\mathrm{y}_l - x\mathrm{p}_l)' S_{ll} (\mathrm{y}_l - x\mathrm{p}_l) - (\mathrm{y}_r - x\mathrm{p}_r)' S_{rr} (\mathrm{y}_r - x\mathrm{p}_r) -$$

$$- \sum_{l=1}^{r-1} \sum_{j \neq l}^{r-1} (\mathrm{y}_l - x\mathrm{p}_l)' S_{lj} (\mathrm{y}_j - x\mathrm{p}_j) -$$

$$- \sum_{j=1}^{r-1} (\mathrm{y}_r - x\mathrm{p}_r)' S_{rj} (\mathrm{y}_j - x\mathrm{p}_j) -$$

$$- \sum_{l=1}^{r-1} (\mathrm{y}_l - x\mathrm{p}_l)' S_{lr} (\mathrm{y}_r - x\mathrm{p}_r). \qquad (\text{А.6.6})$$

Учитывая ограничение (А.6.2), которое эквивалентно соотношению

$$\sum_{j=1}^{r} \mathrm{p}_j = \eta_r, \qquad (\text{А.6.7})$$

можно уменьшить число параметров p_l целевой функции (А.6.6) от r^2 до $(r-1)\,r$, заменив без потери общности r последних параметров на линейную комбинацию первых $(r-1)\,r$ параметров. Обозначим

редуцированный вектор параметров через p_*. Тогда целевая функция выражается как

$$\Phi(p_*) = -\sum_{l=1}^{r-1}(y_l - xp_l)' S_{ll}(y_l - xp_l) -$$

$$-\sum_{l,j}^{r-1}(y_l - xp_l)' S_{rr}(y_j - xp_j) -$$

$$-\sum_{l}^{r-1}\sum_{j\neq l}^{r-1}(y_l - xp_l)' S_{lj}(y_j - xp_j) +$$

$$+\sum_{l,j}^{r-1}(y_l - xp_l)' S_{rj}(y_j - xp_j) +$$

$$+\sum_{l,j}^{r-1}(y_l - xp_l)' S_{lr}(y_j - xp_j). \qquad (A.6.8)$$

Наконец, функция (A.6.8) может быть записана как

$$\Phi(p_*) = -\sum_{l=1}^{r-1}(y_l - xp_l)'(S_{ll} + S_{rr} - S_{rl} - S_{lr})(y_l - xp_l) -$$

$$-\sum_{l}^{r-1}\sum_{j\neq l}^{r-1}(y_l - xp_l)'(S_{rr} + S_{lj} - S_{rj} - S_{lr})(y_j - xp_j) =$$

$$= -\sum_{l,j}^{r-1}(y_l - xp_l)'(S_{rr} + S_{lj} - S_{rj} - S_{lr})(y_j - xp_j). \qquad (A.6.9)$$

Прибегнув к компактной матричной записи, перепишем (A.6.9) в виде

$$\Phi(p_*) = -(y_* - X_* p_*)' M (y_* - X_* p_*), \qquad (A.6.10)$$

где M — $((r-1)T \times (r-1)T)$-матрица, блоки M_{lj} которой имеют одинаковый размер $(T \times r)$ и равны

$$M_{lj} = S_{rr} + S_{lj} - S_{rj} - S_{lr}. \qquad (A.6.11)$$

Таким образом, матрицы $S_{rr} + S_{ll} - S_{rl} - S_{lr}$ представляют собой диагональные блоки, а матрицы $S_{rr} + S_{lj} - S_{rj} - S_{lr}$ — недиагональные блоки. Это означает, что матрица M может быть получена из матрицы S по следующей схеме. Рассматривая блоки S_{lj} как обычные элементы матрицы, можно преобразовать матрицу S к матрице M, если вычесть последний столбец матрицы S из каждого столбца этой матрицы, а затем из всех строк полученной матрицы вычесть r-ю строку.

На основании изложенного можно сформулировать следующую теорему.

Максимизация функции $\Phi(\mathbf{p}) = -(\mathbf{y} - X\mathbf{p})' \; S(\mathbf{y} - X\mathbf{p})$ при ограничениях $G\mathbf{p} = \boldsymbol{\eta}_r$ и $\mathbf{p} \geqslant \mathbf{0}$ эквивалентна максимизации функции $\Phi(\mathbf{p}_*) = -(\mathbf{y}_* - X_*\mathbf{p}_*)' \; M(\mathbf{y}_* - X_*\mathbf{p}_*)$ при ограничениях $R\mathbf{p}_* \leqslant \boldsymbol{\eta}_r$ и $\mathbf{p}_* \geqslant \mathbf{0}$ тогда и только тогда, когда матрица M получена в результате преобразования элементов матрицы S, такого, что $M_{ij} = S_{ij} + S_{rr} - S_{rj} - S_{ir}$, где M_{ij} и S_{ij} — $(T \times T)$-подматрицы матриц M и S соответственно, R — матрица, которая в отличие от G состоит из $(r - 1)$ единичных $(r \times r)$-матриц, и, наконец, X, $\mathbf{y}$ и X_*, $\mathbf{y}_*$ — соответственно исходные и редуцированные матрицы наблюдаемых относительных частот, которые появляются в задаче оценивания переходных вероятностей марковского процесса при постановке этой задачи как задачи оценки параметров многомерной регрессии.

Примерами матриц S, приводимых с помощью описанных преобразований к матрице $M = \Sigma_*^{-1}$, служат матрица Σ^+ и матрица

$$\begin{bmatrix} N(t)/q_1(t) & & & \\ & N(t)/q_2(t) & & \\ & & \ddots & \\ & & & N(t)/q_r(t) \end{bmatrix} \otimes I^T_{t-1}, \tag{A.6.12}$$

которая представляет собой весовую матрицу. Ею пользуются при оценивании по критерию минимума χ^2. Преобразование матрицы (A.6.12) в матрицу Σ_*^{-1} осуществляется очень просто. Для того чтобы показать, как матрица Σ^+ преобразуется в матрицу Σ_*^{-1}, воспользуемся формулой (A.2.5) и запишем для Σ^+ матрицу сечения $V(t)^+$:

$$V(t)^+ = Q(Q'Q)^{-1} \; V(t)_*^{-1} (Q'Q)^{-1} Q'. \tag{A.6.13}$$

Если обе части матричного соотношения (A.6.13) домножить слева на матрицу Q', а справа — на матрицу Q, то получим

$$Q'V(t)^+ \; Q = V(t)_*^{-1} \tag{A.6.14}$$

или

$$Q'V(t)^+ \; Q \otimes I^T_{t-1} = \Sigma_*^{-1}. \tag{A.6.15}$$

Формула (A.6.15) устанавливает точный алгоритм преобразования матрицы Σ^+ к матрице Σ_*^{-1}.

Во многих задачах, возникающих в экономике, социологии, биологии и других областях, модели применяются для описания измерений относительных частот, в качестве которых выступают дискретные случайные величины. Например, в экономике встречаются задачи, связанные с математическим описанием изменений бинарных случайных величин, таких, как «покупают, не покупают», «используют, не используют», или многозначных дискретных случайных величин типа «число потребителей в каждой группе», «число фирм в разных группах». Некоторые статистические модели, предложенные для анализа изменений бинарных случайных величин и прошедшие соответствующую экспериментальную проверку, описаны в [123].

Б.1. Описание модели

Предположим, что в момент времени t наблюдаемое значение относительной частоты j-го состояния (или j-й группы) $y_j(t)$, вычисленное по $n_j(t)$ реализациям, определяется формулой

$$y_j(t) = q_j(t) + u_j(t),\ t = 1, 2, \ldots, T,\quad i, j = 1, 2, \ldots, r, \tag{Б.1.1}$$

где $q_j(t)$ — истинная величина относительной частоты j-го состояния в момент времени t и $u_j(t)$ — ошибки наблюдений. Относительно $u_j(t)$ предполагается, что эти ошибки распределены независимо в соответствии с полиномиальным распределением, вектор средних значений которого равен нулю, а дисперсии и ковариации соответственно равны $q_j(t)\,[1 - q_j(t)]/N(t)$ и $-q_i(t)\,q_j(t)/N(t)$. Через $N(t) = \sum_j n_j(t)$ обозначено общее число реализаций.

Далее, предположим, что истинные относительные частоты по меньшей мере в некотором диапазоне своих значений связаны с определяющими или базовыми переменными линейной по параметрам зависимостью вида

$$q_i = X_i \gamma_i,\ i = 1, 2, \ldots, r, \tag{Б.1.2}$$

где $\mathbf{q}_i$ — $(T \times 1)$-вектор истинных относительных частот i-го состояния, X_i — $(T \times K_i)$-матрица (ранга K_i) наблюдений над K_i неслучайными переменными и $\boldsymbol{\gamma}_i$ — $(K_i \times 1)$-вектор неизвестных коэффициентов. С учетом (Б.1.1) уравнение (Б.1.2) может быть переписано в терминах наблюдаемых относительных частот:

$$\mathbf{y}_i = X_i \boldsymbol{\gamma}_i + \mathbf{u}_i, \; i = 1, 2, \ldots, r, \tag{Б.1.3}$$

где $\mathbf{y}_i$, $\mathbf{u}_i$ — $(T \times 1)$-векторы с ранее определенными свойствами. Уравнения (Б.1.3) могут быть представлены в виде следующей системы уравнений:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_r \end{bmatrix} = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_r \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \vdots \\ \boldsymbol{\gamma}_r \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_r \end{bmatrix} \tag{Б.1.4}$$

или в компактной матричной записи:

$$\mathbf{y} = X\boldsymbol{\gamma} + \mathbf{u}. \tag{Б.1.5}$$

Предполагается, что $\mathbf{y}_i$ — несмещенные оценки истинных относительных частот $\mathbf{q}_i$, т. е.

$$E(\mathbf{y}_i) = X_i \boldsymbol{\gamma}_i = \mathbf{q}_i \text{ для всех } i. \tag{Б.1.6}$$

Поэтому вектор ошибок $\mathbf{u}$ характеризуется следующими свойствами:

$$E(\mathbf{u}) = E(\mathbf{y} - X\boldsymbol{\gamma}) = \mathbf{q} - X\boldsymbol{\gamma} = 0 \tag{Б.1.7}$$

и

$$E(\mathbf{u}\mathbf{u}') = V(\mathbf{u}) = V(\mathbf{y} - X\boldsymbol{\gamma}) = V(\mathbf{y}). \tag{Б.1.8}$$

Ковариационная матрица вектора $\mathbf{y}$ в данный момент времени t имеет вид

$$V(t) = \frac{1}{N(t)} \begin{bmatrix} q_1(t)[1-q_1(t)] & -q_1(t)q_2(t) & \ldots & -q_1(t)q_r(t) \\ -q_2(t)q_1(t) & q_2(t)[1-q_2(t)] & \ldots & -q_2(t)q_r(t) \\ \vdots & \vdots & & \vdots \\ -q_r(t)q_1(t) & -q_r(t)q_2(t) & \ldots & q_r(t)[1-q_r(t)] \end{bmatrix}. \tag{Б.1.9}$$

В силу того, что

$$q_1(t) + q_2(t) + \ldots + q_r(t) = 1, \tag{Б.1.10}$$

при анализе выборочных данных генеральной совокупности, распределенной в соответствии с полиномиальным распределением, число микрообъектов в любом одном состоянии может не рассматриваться, поскольку это число определяется разностью между объемом выборки и суммарным числом объектов в других состояниях. Таким образом, для анализа остается $(r-1)$ независимых полиномиально распределенных совокупностей, характеризуемых вектором средних $\mathbf{q}_*$ и ковариационной матрицей $V_*(t)$, которая является $((r-1)\times(r-1))$-подматрицей матрицы $V(t)$, определенной в (Б.1.9). Если предположить, что $V_*(t)$ — матрица полного ранга, то существует обратная к ней матрица вида

$$V_*(t)^{-1} = N(t)\begin{bmatrix} \frac{1}{q_1(t)}+\frac{1}{q_r(t)} & \frac{1}{q_r(t)} & \cdots & \frac{1}{q_r(t)} \\ \frac{1}{q_r(t)} & \frac{1}{q_2(t)}+\frac{1}{q_r(t)} & \cdots & \frac{1}{q_r(t)} \\ \vdots & \vdots & & \vdots \\ \frac{1}{q_r(t)} & \frac{1}{q_r(t)} & \cdots & \frac{1}{q_{r-1}(t)}+\frac{1}{q_r(t)} \end{bmatrix}, \qquad (Б.1.11)$$

где звездочка снизу означает случай $(r-1)$ независимых состояний.

Ковариационная матрица вектора $\mathbf{u}_*$ в случае T интервалов наблюдений определяется по формуле

$$E(\mathbf{u}_*\mathbf{u}_*') = V_*(t)\otimes I, \qquad (Б.1.12)$$

если построение выборки осуществляется по пуассоновской схеме. Если же выборка получена по схеме Лексиса, то

$$E(\mathbf{u}_*\mathbf{u}_*') = V_*(t)\otimes I_{t=1}^{T}. \qquad (Б.1.13)$$

Б.2. Оценка без учета ограничений

Предположим, что

$$E(\mathbf{u}_*\mathbf{u}_*') = \Sigma_*, \qquad (Б.2.1)$$

и что Σ_* — матрица полного ранга, тогда для оценки параметров модели (Б.1.4), (Б.1.7) и (Б.1.8) может быть применен обобщенный метод наименьших квадратов. При этом наилучшей линейной несмещенной оценкой вектора параметров γ_* будет

$$\mathbf{g}_* = (X_*'\Sigma_*^{-1}X_*)^{-1}X_*'\Sigma_*^{-1}\mathbf{y}_*, \qquad (Б.2.2)$$

где γ_* — подмножество множества γ и Σ_* — известная матрица. В практических задачах матрица Σ_* неизвестна, поэтому ее обычно заменяют на матрицу $\hat{\Sigma}_*$, которая представляет собой состоятельную

оценку ковариационной матрицы вектора $\mathbf{y}$ [123]. В предположении о том, что матрица $\hat{\Sigma}_*$ имеет полный ранг, матрица $\hat{\Sigma}_*^{-1}$ может быть получена, если в (Б.1.11) заменить $q_l(t)$ на $y_l(t)$. В случае T периодов наблюдений в зависимости от схемы эксперимента для матрицы $\hat{\Sigma}_*^{-1}$ имеют место равенства, аналогичные (Б.1.12) и (Б.1.13).

Следует отметить, что оценка

$$\mathbf{g}_* = (X_*' \hat{\Sigma}_*^{-1} X_*)^{-1} X_*' \hat{\Sigma}_*^{-1} \mathbf{y}_* \qquad (Б.2.3)$$

приближенная по отношению к наилучшей линейной несмещенной оценке. Оценка (Б.2.3) может быть улучшена, если воспользоваться процедурой последовательных приближений, на каждом шаге которой $q_l(t)$ в формуле (Б.1.11) заменяются на $y_l(t)$, вычисленные на предыдущем шаге, определяются оценка $\mathbf{g}_*$ согласно (Б.2.3) и новые значения $\hat{y}_l(t)$ по формуле

$$\hat{\mathbf{y}}_*(t) = X_* \mathbf{g}_*. \qquad (Б.2.4)$$

Предполагая, что на каждом шаге матрица Σ_* имеет полный ранг, можно продолжать эту процедуру до тех пор, пока оценки $\mathbf{g}_*$, полученные в двух последовательных итерациях, не будут совпадать с достаточной степенью точности. К сожалению, выборочные свойства оценки, получаемой таким образом, не исследованы.

Ковариационная матрица вектора $\mathbf{g}_*$ имеет вид

$$(X_*' \Sigma_*^{-1} X_*)^{-1} \qquad (Б.2.5)$$

и может быть оценена выражением

$$(X_*' \hat{\Sigma}_*^{-1} X_*)^{-1}. \qquad (Б.2.6)$$

Б.3. Оценка с учетом ограничений

Предсказываемое значение вектора $\hat{\mathbf{y}}_*$ определяется по формуле (Б.2.4). Поскольку на X_* и $\mathbf{g}_*$ не накладывалось никаких ограничений, то не исключено, что значения отдельных компонент вектора $\hat{\mathbf{y}}_*$ окажутся при прогнозе вне интервала [0, 1]. Для устранения подобных случаев могут потребоваться некоторые ограничения.

Для обеспечения неотрицательности компонент вектора $\hat{\mathbf{y}}_*$ подобным ограничением будет неравенство

$$X_* \mathbf{g}_* = \hat{\mathbf{y}}_* \geqslant \mathbf{0}. \qquad (Б.3.1)$$

Если известно, что матрица X_* состоит из неотрицательных элементов, то один из приемов, обеспечивающих выполнение (Б.3.1), заключается в введении ограничений неотрицательности на компоненты вектора γ_*. Если же про элементы матрицы X_* известно, что их значения лежат

в интервале $[a, b]$, то эта информация также может оказаться полезной для конструирования соответствующих ограничений на γ_{*l}. Вообще говоря, ограничения на компоненты вектора γ_* не так уж необходимы.

Ограничение, которое исключает появление значений $\hat{y}_{*l}$, бо́льших единицы, будет аналогом (Б.1.10) и может быть записано в виде неравенства

$$G\hat{y}_*^c \leqslant \eta_T \tag{Б.3.2}$$

или эквивалентного условия

$$GX_* g_*^c \leqslant \eta_T, \tag{Б.3.3}$$

где G — $(T \times (r-1)\ T)$-матрица вида $[I_1 I_2 \ldots\ I_{r-1}]$ (I_j — единичная $(T \times T)$-матрица, $j = 1, 2, \ldots, r-1$) и η_T — вектор-столбец размера T, все компоненты которого равны единице.

Для того чтобы исключить избыточность в системе ограничений, на знаки величин γ_{*l} ограничения не накладываются. При постановке задачи оценивания с учетом ограничений как задачи математического программирования вектор неизвестных параметров γ_* представляется следующим образом:

$$\gamma_* = [\gamma_{*1} \vdots -\gamma_{*2}], \tag{Б.3.4}$$

где векторы γ_{*1} и γ_{*2} удовлетворяют условиям

$$\gamma'_{*1}\gamma_{*2} = 0 \tag{Б.3.5}$$

и

$$\gamma_{*1},\ \gamma_{*2} \geqslant 0. \tag{Б.3.6}$$

При этом исходное регрессионное уравнение для $(r-1)$ независимых состояний приобретает вид

$$y_* = X_*\gamma_{*1} - X_*\gamma_{*2} + u_* \tag{Б.3.7}$$

или в компактной записи

$$y_* = Z_*\gamma_* + u_*, \tag{Б.3.8}$$

где

$$Z_* = [X_* \vdots -X_*] \tag{Б.3.9}$$

и

$$\gamma_* = \begin{pmatrix} \gamma_{*1} \\ \gamma_{*2} \end{pmatrix}. \tag{Б.3.10}$$

Задача оценивания заключается в минимизации

$$\Phi = \mathbf{u}_*'\mathbf{u}_* = (\mathbf{y}_* - Z_*\boldsymbol{\gamma}_*)' \ \Sigma_*^{-1} \ (\mathbf{y}_* - Z_*\boldsymbol{\gamma}_*) \qquad (Б.3.11)$$

при наличии ограничений

$$\begin{cases} GX_*\,\boldsymbol{\gamma}_{*1} - GX_*\,\boldsymbol{\gamma}_{*2} \leqslant \boldsymbol{\eta}_T; & (Б.3.12) \\ \boldsymbol{\gamma}_{*1},\ \boldsymbol{\gamma}_{*2} \geqslant \mathbf{0} & (Б.3.13) \end{cases}$$

или

$$\begin{cases} GZ_*\,\boldsymbol{\gamma}_* \leqslant \boldsymbol{\eta}_T; & (Б.3.14) \\ \boldsymbol{\gamma}_* \geqslant \mathbf{0} & (Б.3.15) \end{cases}$$

и ограничения (Б.3.1) или

$$Z_*\boldsymbol{\gamma}_* \geqslant 0. \qquad (Б.3.16)$$

Поступая аналогично тому, как это делалось в главе 3 и последующих главах, сведем нашу задачу квадратичного программирования к задаче максимизации линейной формы:

$$(Z_*'\,\Sigma_*^{-1}\,\mathbf{y}_* - Z_*'\,\Sigma_*^{-1}\,Z_*\,\mathbf{g}_*^c)'\,\boldsymbol{\gamma}_* \qquad (Б.3.17)$$

при наличии ограничений

$$\begin{pmatrix} GZ_* \\ -Z_* \end{pmatrix} \boldsymbol{\gamma}_* \leqslant \begin{pmatrix} \boldsymbol{\eta}_T \\ 0 \end{pmatrix} \qquad (Б.3.18)$$

и

$$\boldsymbol{\gamma}_* \geqslant 0. \qquad (Б.3.19)$$

Двойственная задача заключается в минимизации

$$\boldsymbol{\lambda}' \begin{pmatrix} \boldsymbol{\eta}_T \\ 0 \end{pmatrix} \qquad (Б.3.20)$$

при наличии ограничений

$$\begin{pmatrix} GZ_* \\ -Z_* \end{pmatrix}' \boldsymbol{\lambda} \geqslant Z_*'\,\Sigma_*^{-1}\,\mathbf{y}_* - Z_*'\,\Sigma_*^{-1}\,Z_*\,\boldsymbol{\gamma}_* \qquad (Б.3.21)$$

и

$$\boldsymbol{\lambda} \geqslant 0. \qquad (Б.3.22)$$

Решения прямой и двойственной задач получаются одновременно в результате максимизации выражения

$$(Z_*'\,Z_*^{-1}\,\mathbf{y}_* - Z_*'\,\Sigma_*^{-1}\,Z_*\,\boldsymbol{\gamma}_*)'\,\boldsymbol{\gamma}_* - \boldsymbol{\lambda}'\begin{pmatrix} \boldsymbol{\eta}_T \\ 0 \end{pmatrix} \qquad (Б.3.23)$$

при наличии ограничений

$$\begin{pmatrix} GZ_* \\ -Z_* \end{pmatrix} \gamma_* + \mathbf{w} = \begin{pmatrix} \eta_T \\ 0 \end{pmatrix}, \tag{Б.3.24}$$

$$\begin{pmatrix} GZ_* \\ -Z_* \end{pmatrix}' \lambda + Z_*' \Sigma_*^{-1} Z_* \gamma_* - \mathbf{v} = Z_*' \Sigma_*' \mathbf{y}_* \tag{Б.3.25}$$

и

$$\lambda, \gamma_* \geqslant 0. \tag{Б.3.26}$$

Переписывая (Б.3.23) с учетом (Б.3.24) и (Б.3.25), получаем уравнение

$$-\lambda^{\tau\prime} \mathbf{w}^{\tau} - \mathbf{g}_*^{c\prime} \mathbf{v}^{\tau} = 0, \tag{Б.3.27}$$

где верхними индексами помечены оптимальные величины. Это уравнение означает, что, поскольку λ и $\mathbf{w}$, γ_* и $\mathbf{v}$ — взаимодополняющие векторы, в случае оптимального решения по меньшей мере одна из соответствующих компонент каждой пары векторов должна быть равна нулю.

В приведенной постановке задача может быть решена с помощью симплекс-метода. Симплекс-таблица этой задачи представлена табл.Б.1. Табл. Б.2 составлена по исходным X_*, γ_* и т. д. Последние два ограничения, записанные в этой таблице, — противоположные неравенства и

Таблица Б.1

Симплекс-таблица для обобщенной линейной модели

p_0	λ_1	λ_2	γ_*	$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{v}$
η_T 0 $Z_*' \Sigma_*^{-1} \mathbf{y}_*$	$Z_*' G_*'$	$-Z_*'$	GZ_* $-Z_*$ $Z_*' \Sigma_*^{-1} Z_*$	I_T	I_T	$-I_{2(r-1)}$

Таблица Б.2

Симплекс-таблица для обобщенной линейной модели

p_0	λ_1	λ_2	γ_{*1}	γ_{*2}	$\mathbf{w}_1$	$\mathbf{w}_2$	$\mathbf{v}_1$	$\mathbf{v}_2$
η_T 0 $X_*' \Sigma_*^{-1} \mathbf{y}_*$ $-X_*' \Sigma_*^{-1} \mathbf{y}_*$	$X_*' G'$ $-X_*' G'$	$-X_*'$ X_*'	GX_* $-X_*$ $X_*' \Sigma_*^{-1} X_*$ $-X_*' \Sigma_*^{-1} X_*$	$-GX_*$ X_* $-X_*' \Sigma_*^{-1} X_*$ $X' \Sigma_*^{-1} X_*$	I_T	I_T	$-I_{r-1}$	$-I_{r-1}$

в совокупности эквивалентны ограничению типа равенства. В рассматриваемом случае нет ограничений на знаки компонент вектора γ_*, а ограничение на соответствующие двойственные переменные приведено в форме равенства. Поэтому симплекс-таблицу Б.2 можно упростить, учитывая, что v_1 и v_2 всегда будут равны нулю. Введение переменной w_1 в действительности означает, что относительная частота $y_r(t)$ исключена из рассмотрения. Упрощенная симплекс-таблица представлена табл. Б.3.

Таблица Б.3

Упрощенная симплекс-таблица для обобщенной линейной модели

p_0	λ_1	λ_2	γ_{*1}	γ_{*2}	y_{rt}	w_2
η_T			GX_*	$-GX_*$	I_T	
0			$-X_*$	X_*		I_T
$X'_*\Sigma_*^{-1}y_*$	X'_*G'	$-X'_*$	$X'_*\Sigma_*^{-1}X_*$	$-X'_*\Sigma_*^{-1}X_*$		

В табл. Б.3 переменные (λ_1, y_{rt}), (λ_2, w_2) и (γ_{*1}, γ_{*2}) представляют собой пары взаимодополняющих переменных. Размер этой таблицы — $\left(2T+\sum_{l=1}^{r-1}K_l\right)\times\left(4T+2\sum_{l=1}^{r-1}K_l\right)$.

Если матрица Σ_* неизвестна, то, действуя как и ранее, заменяем Σ_* на матрицу $\hat{\Sigma}_*$, которая будет состоятельной оценкой ковариационной матрицы. Относительно этой матрицы предполагается, что она имеет полный ранг. Решение соответствующей задачи математического программирования дает искомые оценки, удовлетворяющие ограничениям (Б.3.1) и (Б.3.2).

Б.4. Метод совместного оценивания совокупности базовых переменных

Предложенные в параграфах Б.1 — Б.3 процедуры могут применяться при оценке базовых переменных системы уравнений, каждое из которых содержит дискретные случайные переменные [123]. Вообще говоря, встречаются задачи, в которых имеется несколько систем уравнений, подобных рассмотренным в параграфе Б.1. Если зависимые дискретные переменные различных систем уравнений коррелированы между собой, то процедура совместного оценивания базовых переменных дает оценки, которые асимптотически более эффективны по сравнению с оценками, полученными в предыдущих параграфах [121]. Метод вычисления оценок без учета ограничений в случае корреляции зависимых переменных описан в [123]. Для того чтобы оцениваемые относительные частоты всей совокупности систем уравнений удовлетворяли системе ограничений типа (Б.3.1) и (Б.3.2), рекомендуется воспользоваться обобщением постановки задачи, сделанным в параграфе Б.3.

Приложение В

ОЦЕНИВАНИЕ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ НЕОДНОРОДНОЙ МАРКОВСКОЙ ЦЕПИ

Статистическая модель, рассматриваемая в этой книге, приводит к задаче оценки переходных вероятностей однородной марковской цепи, поскольку предполагается, что переходные вероятности постоянны в течение всего интервала наблюдений. Однако во многих случаях более естественно предположить, что переходные вероятности меняются во времени (или в течение интервала наблюдений) и что эти вероятности — функции некоторых определяющих или базовых переменных. Иначе говоря, изменение базовых переменных приводит к изменению переходных вероятностей. Например, при исследовании задачи потребления сигарет разных сортов Телсер [110] предполагает, что вероятность перехода от одного сорта сигарет к другому зависит от относительных цен различных сортов и от относительных расходов фирм-производителей на рекламу.

В этом разделе будет рассмотрен один из возможных методов оценки изменяющихся во времени (нестационарных) переходных вероятностей по экспериментальным данным, представляющим собой упорядоченные во времени наблюдаемые относительные частоты и наблюдаемые значения соответствующих базовых переменных.

В.1. Описание модели

Для того чтобы получить обобщение задачи оценки переходных вероятностей на нестационарный случай, запишем с помощью введенных ранее понятий неоднородную марковскую цепь первого порядка:

$$y_j(t) = \sum_{i=1}^{r} y_i(t-1)\, p_{ij}(t) + u_j(t), \quad j = 1, 2, \ldots, r, \qquad (B.1.1)$$

где параметры $p_{ij}(t)$ могут меняться во времени. Если помимо (В.1.1) не существует никакой другой информации, то невозможно оценить переменные параметры $p_{ij}(t)$, так как в каждый момент времени имеется r^2 неизвестных параметров. В рассмотренных ранее задачах

предположение о постоянстве переходных вероятностей приводит к тому, что

$$p_{ij}(t) = p_{ij} \text{ для всех } t,\ i,\ j \tag{B.1.2}$$

и число неизвестных уменьшается от r^2T до r^2.

Как показано в главе 3, оценки этих неизвестных параметров могут быть найдены при решении системы из rT уравнений при условии, что $T \geqslant r$.

Если в случае неоднородной марковской цепи предположить существование линейной зависимости между $p_{ij}(t)$ и некоторыми внешними по отношению к модели базовыми переменными $z_k(t)$, то можно получить дополнительную информацию в виде

$$p_{ij}(t) = \sum_{k=1}^{M} \delta_{ijk} z_k(t) + v_{ij}(t) \text{ для всех } i,\ j,\ t, \tag{B.1.3}$$

где δ_{ijk} — параметры, $v_{ij}(t)$ — случайные переменные, а $z_k(t)$ — внешние базовые переменные. Заметим, что (B.1.3) представляет собой обобщение соотношений (B.1.2), из которых (B.1.3) получается, если учесть изменение во времени базовых и случайных переменных. Теперь число неизвестных параметров модели уменьшается от r^2T до r^2M и эти параметры могут быть оценены из rT уравнений наблюдений при условии, что $T \geqslant rM$. Введение в модель базовых переменных в нестационарном случае уменьшает число степеней свободы. Однако это уменьшение может быть компенсировано увеличением числа наблюдений. Следует отметить, что при применении стационарных моделей в практических задачах увеличение числа наблюдений (t) может увеличить вероятность отвергнуть гипотезу стационарности.

Запишем rT уравнений типа (B.1.1) и r^2T уравнений типа (B.1.3) в компактной матричной форме:

$$\mathbf{y} = X\mathbf{p} + \mathbf{u} \tag{B.1.4}$$

и

$$\mathbf{p} = Z\boldsymbol{\delta} + \mathbf{v}. \tag{B.1.5}$$

В (B.1.4) $\mathbf{y}$ есть $(rT \times 1)$-вектор относительных частот $y_j(t)$, $j = 1, 2, \ldots, r$, $t = 1, 2, \ldots, T$, $\mathbf{u}$ и $\mathbf{v}$ — соответственно $(rT \times 1)$- и $(r^2T \times 1)$-векторы ошибок наблюдений, X — $(rT \times r^2T)$-матрица относительных частот $y_i(t-1)$ из диагональных блоков:

$$X = \begin{bmatrix} [y(t-1) \otimes I_T] & & & \\ & [y(t-1) \otimes I_T] & & \\ & & \ddots & \\ & & & [y(t-1) \otimes I_T] \end{bmatrix},$$

где каждый блок представляет собой $(T \times rT)$-матрицу, а $\mathbf{y}(t) = (y_1(t), y_2(t), \ldots, y_r(t))$ для $t = 0, 1, \ldots, T-1$. В уравнении (В.1.5) $\mathbf{p}$ есть $(r^2T \times 1)$-вектор нестационарных переходных вероятностей $p_{lj}(t)$, составленный следующим образом:

$$p = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_r \end{bmatrix}, \quad \text{где } \mathbf{p}_j = \begin{bmatrix} \mathbf{p}_{1j} \\ \mathbf{p}_{2j} \\ \vdots \\ \mathbf{p}_{rj} \end{bmatrix} \quad \text{и} \quad p_{lj} = \begin{bmatrix} p_{lj}(1) \\ p_{lj}(2) \\ \vdots \\ p_{lj}(T) \end{bmatrix},$$

Z — блочно-диагональная $(r^2T \times r^2M)$-матрица вида

$$Z = \begin{bmatrix} Z_1 & & & \\ & Z_2 & & \\ & & \ddots & \\ & & & Z_{r^2} \end{bmatrix},$$

в которой $Z_1 = Z_2 = \ldots = Z_{r^2}$, а каждый блок Z_l является $(T \times M)$-матрицей, δ — $(r^2M \times 1)$-вектор параметров, устроенный так, что

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_1 \\ \vdots \\ \delta_r \end{bmatrix}, \quad \text{где } \delta_j = \begin{bmatrix} \delta_{1j} \\ \delta_{2j} \\ \vdots \\ \delta_{rj} \end{bmatrix} \quad \text{и} \quad \delta_{lj} = \begin{bmatrix} \delta_{lj1} \\ \delta_{lj2} \\ \vdots \\ \delta_{ljm} \end{bmatrix}.$$

Относительно случайных переменных $\mathbf{u}$ и $\mathbf{v}$ предполагается, что

$$E\mathbf{u} = 0, \tag{В.1.6}$$

$$E\ \mathbf{u}\mathbf{u}' = \Sigma, \tag{В.1.7}$$

$$E\mathbf{v} = 0, \tag{В.1.8}$$

$$E\ \mathbf{v}\mathbf{v}' = \Omega, \tag{В.1.9}$$

$$E\ \mathbf{u}\mathbf{v} = 0. \tag{В.1.10}$$

В условии (В.1.7) Σ — $(rT \times rT)$-матрица из блоков диагональных подматриц. Диагональные элементы блоков, расположенных на диагонали и вне диагонали матрицы Σ, соответственно равны $q_l(t)(1 - q_l(t))/N(t)$ и $-q_l(t)\,q_j(t)/N(t)$, где $q_l(t)$ — истинные относительные частоты или математические ожидания величин $y_l(t)$. В условии (В.1.9) Ω — $(r^2T \times r^2T)$-матрица из диагональных блоков. В этой матрице элементы блоков, расположенных на диагонали и вне диагонали, соответственно равны $p_{lj}(t)(1 - p_{lj}(t))/n_l(t-1)$ и $-p_{lj}(t) \times p_{lk}(t)/n_l(t-1)$, где $n_l(t-1) = N(t)\,y_l(t-1)$. Матрицы Σ и Ω вырожденные. Наша задача заключается в определении вектора па-

раметров δ в предположении о том, что модели (В.1.4), (В.1.5) верны, т. е. нестационарные переходные вероятности $\mathbf{p}$ определяются значениями базовых переменных, а относительные частоты изменяются в соответствии с уравнением неоднородной марковской цепи первого порядка.

При подстановке (В.1.5) в (В.1.4) получается, что

$$\mathbf{y} = XZ\delta + \mathbf{w}, \tag{В.1.11}$$

где

$$\mathbf{w} = X\mathbf{v} + \mathbf{u}, \tag{В.1.12}$$

$$E\mathbf{w} = XE\mathbf{v} + E\mathbf{u} = 0, \tag{В.1.13}$$

и

$$E\mathbf{w}\mathbf{w}' = \omega = (X\Omega X') + \Sigma. \tag{В.1.14}$$

Ковариационная матрица ω вырожденная, так как матрицы Σ и $X\Omega X'$ имеют строки, в которых сумма элементов равна нулю. Это свойство матрицы Σ обсуждалось в главах 5,6. Аналогичным свойством характеризуется матрица $X\Omega X'$, поскольку она является матрицей из диагональных блоков, причем диагональные элементы блоков, расположенных на диагонали матрицы $X\Omega X'$, равны $p_{lj}(t)(1-p_{lj}(t)) \times \times y_l(t-1)/N(t)$, а элементы блоков, расположенных вне диагонали матрицы $X\Omega X'$, равны $-p_{lj}(t)\,p_{lk}(t)\,y_l(t-1)/N(t)$. Следовательно, матрица, равная сумме $X\Omega X' + \Sigma$, характеризуется тем, что сумма элементов в любой строке этой матрицы равна нулю. Для того чтобы обойти трудность, связанную с вырожденностью ковариационной матрицы ошибок наблюдений, воспользуемся стандартным приемом построения редуцированной модели. Пометив звездочкой соответствующие матрицы, которые получаются при исключении из рассмотрения уравнений для r-х относительных частот y_r, можно записать редуцированную модель:

$$\mathbf{y}_* = X_* Z_* \delta_* + \mathbf{w}_*, \tag{В.1.15}$$

где

$$E\mathbf{w}_* = 0, \tag{В.1.16}$$

$$E\,\mathbf{w}_*\mathbf{w}_*' = \omega_* \tag{В.1.17}$$

и ω_* — невырожденная матрица, элементы которой представляют собой функции от $q_l(t)$ и $p_{lj}(t)$. Хотя для обратной матрицы ω_*^{-1} нет аналитической записи, эта матрица может быть определена численно, если $q_j(t)$ заменить на $y_j(t)$, а $p_{lj}(t)$ — на приближенную оценку и воспользоваться процедурой последовательных приближений, которая применялась для вычисления оценки максимального правдоподобия и байесовской оценки в главах 8, 9. Иные трудности связаны с отысканием приближенной оценки $p_{lj}(t)$ и обеспечением сходимости процедуры последовательных приближений. Эти вопросы требуют дальнейшего исследования.

В том случае, когда можно ослабить предположение о случайном характере связи между нестационарными переходными вероятностями и базовыми переменными и считать, что эта связь чисто детерминированная, то случайная величина $\mathbf{v}$ в уравнении (В.1.5) исчезает и ковариационная матрица ω_* вычисляется проще. Действительно, в этом случае ω_* совпадает с Σ_*, для которой обратная матрица Σ_*^{-1} состоит из диагональных блоков. Диагональные элементы диагональных и недиагональных блоков матрицы Σ_*^{-1} соответственно равны $q_j(t)/N(t) + q_r(t)/N(t)$ и $q_r(t)/N(t)$. При решении практической задачи вместо истинных относительных частот $q_j(t)$ можно подставить $y_j(t)$ и для вычисления Σ_*^{-1} применить процедуру последовательных приближений, рассмотренную в главах 6, 8 и 9. В дальнейшем вместо Σ_* будем брать ω_*, так как матрица ω_* фактически более общая по отношению к Σ_*.

В.2. Оценка без учета ограничений

Если не вводить никаких ограничений, то оценка $\tilde{\delta}_*$ по обобщенному методу наименьших квадратов равна

$$\tilde{\delta}_* = (Z_*' X_*' \omega_*^{-1} X_* Z_*)^{-1} X_*' X_*' \omega_*^{-1} \mathbf{y}_* \qquad (\text{В.2.1})$$

при условии, что $X_* Z_*$ имеет ранг, равный числу столбцов (т. е. необходимым, но не достаточным условием существования (В.2.1) будет неравенство $T \geqslant rM$). Тогда векторы переходных вероятностей определяются следующим образом:

$$\hat{\mathbf{p}}_*^e = Z_* \tilde{\delta}_* \qquad (\text{В.2.2})$$

и

$$\hat{\mathbf{p}}_r^e = [\eta_{rT} - R\hat{\mathbf{p}}_*^e, \qquad (\text{В.2.3})$$

где η_{rT} — $(rT \times 1)$-вектор-столбец единиц, R — $(rT \times r(r-1)T)$-матрица из $(r-1)$ единичных подматриц размера $(rT \times rT)$.

В.3. Оценка с учетом ограничений

Поскольку на оценку (В.2.1) не наложено никаких ограничений, нет никаких гарантий относительно того, что компоненты векторов переходных вероятностей $\mathbf{p}_*$ и $\mathbf{p}_r$ попадут в интервал [0, 1]. Поэтому следует рассмотреть процедуру вычисления оценки по методу наименьших квадратов с учетом необходимых ограничений, которыми в данном случае будут

$$R\mathbf{p}_* \leqslant \eta_{rT} \qquad (\text{В.3.1})$$

и

$$\mathbf{p}_* \geqslant 0. \qquad (\text{В.3.2})$$

При оценивании вектора параметров δ_* ограничения (В.3.1) и (В.3.2) должны быть приведены к виду

$$RZ_*\delta_* \leqslant \eta_{rT} \tag{В.3.3}$$

и

$$Z_*\delta_* \geqslant 0, \tag{В.3.4}$$

где ограничения на знаки компонент вектора δ_* не накладываются. Таким образом, мы приходим к следующей задаче квадратичного программирования: требуется найти вектор $\tilde{\delta}_*^c$, максимизирующий выражение

$$-(\mathbf{y}_* - X_*Z_*\delta_*)' \ \omega_*^{-1} \cdot (\mathbf{y}_* - X_*Z_*\delta_*) \tag{В.3.5}$$

при ограничениях (В.3.3) и (В.3.4). Поскольку знаки компонент вектора δ_* могут быть любыми, введем взаимодополняющие векторы δ_Δ и $\delta_{\Delta\Delta}$ с неотрицательными компонентами, такие, что

$$\delta_* = \delta_\Delta + \delta_{\Delta\Delta} \tag{В.3.6}$$

и

$$\delta_\Delta' \delta_{\Delta\Delta} = 0. \tag{В.3.7}$$

Тогда исходную систему уравнений можно преобразовать в систему вида

$$\mathbf{y}_* = X_* Z_* \delta_\Delta - X_* Z_* \delta_{\Delta\Delta} + \mathbf{w}_* \tag{В.3.8}$$

и сформулировать обычную задачу квадратичного программирования: максимизировать

$$(\mathbf{y}_* - X_* Z_* \delta_\Delta + X_* Z_* \delta_{\Delta\Delta})' \ \omega_*^{-1} (\mathbf{y}_* - X_* Z_* \delta_\Delta + X_* Z_* \delta_{\Delta\Delta}) \tag{В.3.9}$$

при наличии ограничений

$$RZ_* \delta_\Delta - RZ_* \delta_{\Delta\Delta} \leqslant \eta_{rT}, \tag{В.3.10}$$

$$Z_* \delta_\Delta - Z_* \delta_{\Delta\Delta} \geqslant 0 \tag{В.3.11}$$

и

$$\delta_\Delta, \ \delta_{\Delta\Delta} \geqslant 0. \tag{В.3.12}$$

Табл. В.1 для рассмотренного в приложении Б и в главах 4, 6 метода совместного решения прямой и двойственной задач представляет собой симплекс-таблицу. В ней λ_1, λ_2 — двойственные переменные, α_1, α_2 — начальные искусственные переменные, λ_1 и α_1, λ_2 и α_2, δ_Δ и $\delta_{\Delta\Delta}$ — пары взаимодополняющих переменных. Заметим, что только один элемент каждой такой пары может входить в базис.

Таблица В.1

Симплекс-таблица для вычисления нестационарных переходных вероятностей

B_0	λ_1	λ_2	δ_Δ	$\delta_{\Delta\Delta}$	α_1	α_2
η_{rT}			RZ_*	$-RZ_*$	I	
0			$-Z_*$	Z_*		I
$Z'_* X'_* \omega_*^{-1} y_*$	$Z'_* R'$	$-Z'_*$	$Z'_* X'_* \omega_*^{-1} X_* Z_*$	$-Z'_* X'_* \omega_*^{-1} X_* Z_*$		

В.4. Заключительные замечания

В приведенной ранее задаче переходные вероятности определяются через $Z_*\tilde{\delta}_*^c$, причем ограничения оказываются эффективными лишь в том случае, если базовые переменные Z меняются в заданном диапазоне, который может быть определен из анализа прошлых наблюдений. Однако если нельзя утверждать, что значения базовых переменных в будущем не покинут этого диапазона, то нет совершенно никаких гарантий в том, что предсказанные переходные вероятности p_{ij} будут принадлежать интервалу [0, 1]. Если же будущие значения базовых переменных известны и составляют некоторую матрицу $\overline{Z}_*$, то на основе $\overline{Z}_*$ можно сформировать дополнительную систему ограничений типа (В.3.10) — (В.3.12). С помощью неоднородной марковской цепи первого порядка могут быть правильно предсказаны не только переходные вероятности, но и будущие значения относительных частот, если будущие переходные вероятности точно известны.

Приложение Г ● ПРОГРАММА НА ФОРТРАНЕ ДЛЯ ВЫЧИСЛЕНИЯ КЛАССИЧЕСКИХ И БАЙЕСОВСКИХ ОЦЕНОК ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ

Приведенная в этом приложении программа позволяет вычислять различные оценки переходных вероятностей. Перечень оценок, которые можно получить с помощью данной программы, включает в себя оценки с учетом и без учета ограничений, классические оценки, оценки с учетом взвешивания наблюдений, оценки, вычисляемые по псевдообратной матрице (смещенные на первом шаге), двухшаговые оценки, оценки по критерию наименьших абсолютных отклонений, определяемые методом линейного программирования, оценки максимального правдоподобия и байесовские апостериорные оценки.

С помощью соответствующей управляющей карты можно вычислить любую оценку из тех, которые рассмотрены в этой книге, с выводом на печать промежуточных результатов и перфорацией матрицы финальных вероятностей на перфокарты. Можно также осуществить прогноз относительных частот и проверки по критерию согласия χ^2. Для решения каждой отдельной задачи необходимо набить лишь несколько перфокарт со значениями наблюдаемых относительных частот (или агрегированных данных). За один выход на машину может быть решена последовательность нескольких задач при неоднократном обращении к процедуре ввода.

По результатам решения этих задач могут быть также вычислены сводные результаты: средние, дисперсии, стандартные отклонения, среднеквадратические ошибки оценок и т. д. Программа позволяет решать задачи с числом состояний, не большим 6, и весовой матрицей, размер которой не превышает (96 × 96).

Г.1. Колода перфокарт с исходной информацией

Рассмотрим в качестве примера этап подготовки колоды перфокарт с исходной информацией для задачи отыскания оценки максимального правдоподобия. Колода перфокарт состоит из двух управляющих карт и колоды с числовыми данными. Первая управляющая карта содержит слово SAMPLE, перфорируемое в ко-

лонках 1—6. Это слово означает, что информация, отперфорированная на последующих картах, является выборочной информацией. Колонки с 9-й по 72-ю карты SAMPLE могут быть заняты под заголовок или идентификатор задачи. Вторая управляющая карта предназначена для выбора конкретного варианта счета. Состав и расположение информации на этой карте будут рассмотрены в параграфе Г.8. Непосредственно вслед за управляющими картами помещаются карты с числовыми данными, отперфорированными в соответствии с выбранным форматом. Этот формат описывается во второй управляющей карте, которая задает конкретный вариант счета. Перфокарты с экспериментальными данными должны быть отперфорированы так, чтобы на одной перфокарте размещалась одна строка расширенной матрицы относительных частот. Другими словами, вначале на карту перфорируются данные о числе микрообъектов в каждом состоянии, а затем — общее число наблюдений в соответствующий момент времени. В случае, когда число наблюдений в каждый момент времени одно и то же, может быть применена другая процедура ввода. Вызов этой процедуры осуществляется, если в колонках с 10-й по 15-ю управляющей карты выбора варианта счета набить общее число наблюдений, одинаковое для всех экспериментов, и не перфорировать это число на картах с данными об относительных частотах.

Г.2. Объединение априорной и выборочной информации

При вычислении байесовской оценки можно учитывать априорную информацию. Для этого требуется другая управляющая карта, в первых шести колонках которой отперфорировано слово PRIORI. Карта PRIORI располагается вслед за колодой с выборочными данными. Непосредственно за этой картой помещается карта, первые шесть колонок которой отведены под число последующих перфокарт с априорной информацией, а колонки с 7-й по 12-ю предназначены для выбора конкретного вида априорной информации. В колонках с 7-й по 12-ю перфорируются:

0 — в случае априорного многомерного бета-распределения;

9999 — в случае априорных независимых одномерных бета-распределений;

8888 — в случае априорного нормального распределения.

На каждой из последующих карт в первых двух случаях должны быть отперфорированы параметры I, J, a_{IJ} и $a_I - a_{IJ}$ в формате (2I6, 2F15.4), где I — номер строки, J — номер столбца матрицы переходных вероятностей, a_{IJ} и $a_I - a_{IJ}$ — параметры бета-распределения. В случае нормального распределения вместо a_{IJ} и $a_I - a_{IJ}$ перфорируются соответственно априорное среднее и априорная дисперсия.

Г.3. Исключение столбца из матрицы относительных частот

При вычислении оценки максимального правдоподобия и байесовской оценки требуется исключать один из столбцов матрицы относительных частот и соответствующее уравнение. Для этой цели предназначена 22-я колонка управляющей карты выбора варианта счета. Если не

требуется исключить какой-либо столбец, то в 22-й колонке перфорируется 0. Если число в этой колонке больше или равно числу марковских состояний r, то будет исключен r-й столбец (заметим, что в случае, когда в 22-й колонке имеется 0, оценка максимального правдоподобия не будет вычисляться, если не прибегнуть к рекуррентной процедуре, описанной в параграфе Г.5). Если можно удалить любой столбец, то рекомендуется исключить r-й столбец. При этом полученные результаты легко анализируются: величина оценки максимального правдоподобия не зависит от номера исключенного столбца (что характерно и для байесовской оценки).

Г.4. Присвоение веса

Весовая матрица M в целевой функции $\Phi = (y - Xp)' M (y - Xp)$ может быть определена пользователем. Для выбора типа взвешивания предназначена 23-я колонка карты выбора варианта счета. Большинство весовых матриц генерируется в программе. Однако при желании пользователь может назначить собственные веса. Для этого необходимо пробить число 3 в 23-й колонке и подготовить еще одну управляющую карту, в первых шести колонках которой перфорируется слово WEIGHT. Карта WEIGHT помещается вслед за колодой с выборочными данными. Непосредственно за картой WEIGHT размещаются перфокарты, на каждой из которых перфорируются величины I, J и VALUE в формате (2I3, F15.4), где I, J — соответственно номер строки и столбца матрицы M, VALUE — численная величина присваиваемого веса.

Г.5. Рекуррентное вычисление оценки максимального правдоподобия и байесовской оценки

Точная величина оценки максимального правдоподобия и байесовской оценки может быть найдена с помощью процедуры «с обратной связью» или итеративным методом квадратичного программирования. Обращение к этой процедуре осуществляется с помощью управляющей карты, в первых девяти колонках которой перфорируется слово RECURSIVE. Эта карта должна быть помещена вслед за колодой выборочных данных (в случае вычисления оценки максимального правдоподобия) или непосредственно за перфокартами с параметрами априорного распределения (в случае вычисления байесовской оценки). Вслед за картой RECURSIVE располагается карта, в 4—6-й колонках которой должно быть отперфорировано нужное число итераций, а в 12-й колонке — заданная толерантная граница (формат 2I6). Эта величина определяется как сумма абсолютных отклонений между двумя последовательными оценками. Итеративная процедура будет окончена при выполнении одного из требований. В колонках 28—34 перфорируется та же самая информация, что и в соответствующих колонках управляющей карты выбора варианта счета (см. параграф Г.8). Это дает возможность изменить объем и состав печатаемой информации, если возникает желание опустить печать некоторых из промежуточных результатов.

Интересно отметить, что в качестве начального приближения в итеративном процессе вычисления точной оценки необязательно выбирать первое приближение оценки максимального правдоподобия с учетом ограничений. Иначе говоря, в качестве начального приближения можно выбрать, например, классическую оценку, найденную по исходной матрице относительных частот (а не по редуцированной матрице). Независимо от выбора начального приближения итеративная процедура будет сходиться к одной и той же предельной оценке. Однако с целью уменьшения числа итераций можно рекомендовать первое приближение оценки максимального правдоподобия с учетом ограничений в качестве начального условия. Следует также отметить, что при вводе в машину карты RECURSIVE последний столбец матрицы переходных вероятностей исключается автоматически.

Г.6. Управляющие карты DITTO и DIT.

В случае, когда решается последовательность задач, в каждой из которых осуществляется вычисление байесовской оценки, карты с априорной информацией для первой задачи должны быть помещены непосредственно за картой PRIORI. При подготовке соответствующей информации для второй и последующих задач вместо карты PRIORI и карт с априорной информацией можно воспользоваться картами DITTO и DIT. (слова DITTO и DIT. перфорируются, начиная с 1-й колонки). В случае карты DITTO список априорных значений параметров выдаваться на печать не будет. В случае карты DIT. этот список будет отпечатан для второй, третьей и т. д. задач. Подобная возможность предусмотрена специально для выборочного эксперимента, описанного в этой книге.

Г.7. Управляющие карты CLEAR и SUMMARY

Эти карты также предназначены главным образом для обработки данных выборочного эксперимента. При решении последовательности нескольких задач за один выход на машину можно получить сводные результаты: средние оценки, средние дисперсии и стандартные отклонения оценок относительно истинных значений параметров (имеются в виду средние по всем решенным задачам). Для этого следует воспользоваться управляющими картами CLEAR и SUMMARY (оба слова перфорируются с первой колонки) и расположить их в общей колоде следующим образом. Карта CLEAR должна быть помещена в начале колоды перфокарт с последовательностью задач, по результатам решения которых будут вычисляться средние характеристики. Для каждой из этих задач управляющая карта выбора варианта счета должна содержать число 1 в 35-й колонке. Карту SUMMARY следует поместить непосредственно за перфокартами с последовательностью решаемых задач. Вслед за SUMMARY должны располагаться карты с истинными значениями параметров, перфорируемыми в формате 9F8.4. Общее число параметров не должно превышать 36, т. е. при формате 9F8.4 максимальное число карт с информацией об истинных значениях равно

четырем. Матрица истинных значений параметров перфорируется по столбцам.

Г.8. Управляющая карта выбора варианта счета

С помощью рассматриваемой в этом параграфе управляющей карты осуществляется выбор определенного режима работы программы при вычислении оценок, прогнозировании и проверке гипотез. В программе предусмотрены различные варианты выдачи результатов на печать. Эти варианты также задаются с помощью рассматриваемой здесь карты.

На управляющей карте выбора варианта счета перфорируется следующая информация:

Колонки	Идентификатор переменной	Описание переменной
1—3		Пробел
4—6	NTI	Число интервалов наблюдений
7—9	NS	Число марковских состояний
10—15	SS	Объем выборки, если объем одинаков для всех интервалов наблюдений, т. е. $N(t) \equiv N(t-1) \equiv N$. Если же объемы выборок для разных интервалов наблюдений отличаются, то в колонках 10—15 перфорируются пробелы. Кроме того, в этом случае объем выборки должен быть отперфорирован как последний элемент строки расширенной матрицы относительных частот (см. параграф Г.1). Соответствующий формат этих данных должен быть указан в колонках 37—72. Следует отметить, что если колонки 10—15 не заполнены пробелами в случае разных объемов выборок, то число, отперфорированное в этих колонках, заменит фактический объем выборки в каждом интервале. Вместо наблюдаемых относительных частот в программу может вводиться информация о наблюдаемом числе микрообъектов в каждом состоянии. См. также описание переменной KEY (1) в колонке 26
16—21	TOL	Допустимая погрешность при обращении матрицы и применении итеративного симплекс-метода. Если колонки 16—21 заполнены пробелами, то TOL=1.0E—5
22	KDROP	Номер столбца матрицы относительных частот, который следует исключить
23	KW	Присвоение весов, KW=0, 1, ..., 9. Перфорируются: 0 — в случае классической оценки по методу наименьших квадратов; 1 — в случае оценки максимального правдоподобия, оценки по обобщенному методу наименьших квадратов, оценки по критерию минимума χ^2; 2 — в случае взвешивания на $N(t)/W_j(t)$; 3 — в случае произвольной весовой матрицы, задаваемой пользователем; 4 — в случае взвешивания на средние относительные частоты; 5 — в случае взвешивания на произведения средних относительных частот; 6 — в случае взвешивания в соответствии с весами, вычисленными с помощью псевдообратной матрицы дисперсии ошибок наблюдений;

7 — в случае взвешивания на $N(t)/(w_j(t)(1-w_j(t))$;
8 — не задействована;
9 — в случае двухступенчатого оценивания при взвешивании в соответствии с матрицей дисперсии ошибок наблюдений

24 KV Выбор конкретного варианта выдачи ковариационной матрицы $(X'MX)^{-1}$.
Перфорируются:
0 — для пропуска выдачи;
1 — для печати матрицы на всех шагах;
2 — для печати матрицы только на первом шаге (без учета ограничений);
3 — для печати матрицы только на втором шаге (с учетом ограничений);
4 — для печати матрицы в случае, когда решение вычислено без учета ограничений, но ограничениям удовлетворяет

25 KP Вычисление прогноза и выбор критерия проверки гипотез.
Перфорируются:
0 — если прогноз не осуществляется;
1 — для вычисления прогноза;
2 — для прогноза на первом шаге;
3 — для прогноза на втором шаге;
4 — для прогноза в случае приемлемой оценки

26 KEY (1) Выбор варианта печати исходной информации.
Перфорируются:
0 — печать исходных данных в случае, когда данные представляют собой наблюдаемые микрообъекты;
1 — печать наблюдаемых относительных частот

27 KEY (2) Выбор варианта печати весовой матрицы.
Перфорируются:
0 — если матрицу печатать не надо;
1 — для печати весовой матрицы;
2 — для печати весовой матрицы только на втором шаге (печать оценки ковариационной матрицы)

28 KEY (3) Выбор варианта печати матричных произведений $X'MX$ и $X'My$.
Перфорируются:
0 — если матрицы печатать не надо;
1 — для печати на всех шагах;
2 — для печати только на первом шаге;
3 — для печати только на втором шаге

29 KEY (4) Выбор варианта печати симплекс-таблицы.
Перфорируются:
0 — ни одну из симплекс-таблиц печатать не надо;
1 — для печати исходной симплекс-таблицы;
7 — для печати окончательной симплекс-таблицы;
8 — для печати таблицы в «аварийном» случае, например при линейной зависимости, зацикливании, вырождении и т. д.;
9 — в случае малых задач для печати симплекс-таблицы на каждом шаге (этот режим требует много машинного времени, поэтому им не рекомендуется пользоваться в больших задачах)

30	KEY (5)	Выбор режима вычисления определителя, итерационного шага и левой части симплекс-таблицы (столбца B_0)*

Перфорируются	Определитель	Итерация	Столбец
0	—	—	—
1	—	—	+
2	—	+	+
3	+	—	—
4	+	—	+
5	+	+	+

31—32	KEY (6)	Масштабный коэффициент. Если перфорируется 1, то масштабный коэффициент равен 1,0E01. Если перфорируется 2, то этот параметр равен 1,0E02, и т. д.
33	KEY (7)	Выбор варианта вычисления оценки с учетом ограничений. Перфорируются: 0 — если оценку с ограничениями вычислять не надо; 1 — для вычисления оценки с ограничениями только на первом шаге; 2 — для вычисления оценки с ограничениями только на втором шаге; 3 — для вычисления оценки с ограничениями на всех шагах (нельзя применять в сочетании с KW=9); 4 — для вычисления оценки с ограничениями на всех шагах и с печатями согласно KEY (4)
34	KEY (8)	Выбор итеративного метода решения. Перфорируются: 0 — для решения задачи методом линейного программирования при взвешивании на величину квадратного корня из суммы элементов в столбце матрицы M; 1 — для решения задачи методом линейного программирования при взвешивании на величину, равную сумме элементов в столбце матрицы M; 2 — для решения задачи методом квадратичного программирования; 9 — только для вычисления байесовской оценки методом квадратичного программирования
35	KEY (9)	Перфорируется 1 для учета результатов решения задачи при вычислении сводных результатов.
36		Пробел
37—72	FMT	Формат, описывающий расположение исходных выборочных данных на перфокартах, заключенный в круглые скобки. Этот формат в различных задачах может быть разным

* В приведенной таблице знак «—» в столбце означает, что соответствующие вычисления не проводятся. Знак «+» означает выполнение вычислений. — *Прим. пер.*

Г.9. Пример расположения входной информации на перфокартах

Рассмотрим пример из глав 7 и 9. Распечатка управляющих карт и карт с выборочными данными этого примера имеет следующий вид:

```
SAMPLE          AN ABSORBING CHAIN
  005002   100      21041008200320  (4X, 2F8.4)
     0.5000   0.5000
     0.7500   0.2500
     0.8800   0.1200
     0.9400   0.0600
     0.9700   0.0300
RECURSIVE
     6        3       041008200320
PRIORI
     4
     1        1       99.0          1.0
     1        2        1.0         99.0
     2        1       50.0         50.0
     2        2       50.0         50.0
RECURSIVE
     6        3       041008200320
FINIS
```

Г.10. Пример выдачи результатов

Для рассмотренного в параграфе Г.9 примера приведем распечатку полученных результатов *.

```
ПОГЛОЩАЮЩАЯ ЦЕПЬ
НАБЛЮДАЕМЫЕ ПРОПОРЦИИ И ОБЪЕМЫ ВЫБОРОК

          1                  2                  3
1    0.5000000E  00   0.5000000E  00   0.1000000E  03
2    0.7500000E  00   0.2500000E  00   0.1000000E  03
3    0.8800000E  00   0.1199999E  00   0.1000000E  03
4    0.9700000E  00   0.3000000E  00   0.1000000E  03

СОГЛАСНО ВАШЕМУ ЖЕЛАНИЮ СТОЛБЕЦ 2 НЕ УЧИТЫВАЛСЯ ПРИ
ФОРМИРОВАНИИ СИСТЕМЫ
СЛЕДУЮЩИЕ МАТРИЦЫ ЧИТАТЬ В ПОСЛЕДОВАТЕЛЬНОСТИ 1 2
ВЕСОВАЯ МАТРИЦА ПОЛУЧЕНА ПО ОЦЕНКАМ МАКСИМАЛЬНОГО ПРАВ-
ДОПОДОБИЯ И ИМЕЕТ ЭЛЕМЕНТЫ N/WR(T)+Z*(N/WJ(T)),  Z=1  ДЛЯ
ДИАГОНАЛЬНЫХ ЭЛЕМЕНТОВ И Z=0 В ПРОТИВНОМ СЛУЧАЕ
ОЦЕНКА МАТРИЦЫ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ БЕЗ УЧЕТА ОГРА-
НИЧЕНИЙ
0.99999785     0.00000215
0.50514352     0.49485648
```

* В этой распечатке заголовки даны в переводе на русский язык. — *Прим. пер.*

ПРЕДСКАЗАННЫЕ ОТНОСИТЕЛЬНЫЕ ЧАСТОТЫ

	1		2	
2	0.7525707E	00	0.2474293E	00
3	0.8762842E	00	0.1237157E	00
4	0.9406152E	00	0.5938463E	—01
5	0.9703066E	00	0.2969340E	—01
6	0.9851522E	00	0.1484777E	—01

СУММА КВАДРАТОВ
ОШИБОК ПРОГНОЗА

0.000041776

СРЕДНЕКВАДРАТИЧЕСКАЯ
ОШИБКА ПРОГНОЗА

0.000005222

ВЕЛИЧИНА
КРИТЕРИЯ χ^2

0.017288841

ВЕЛИЧИНА МОДИФИЦИРОВАННОГО КРИТЕРИЯ χ^2

0.017593563

ОЦЕНКА БЕЗ УЧЕТА ОГРАНИЧЕНИЙ ПРИЕМЛЕМА
ВЫХОД ИТЕРАТИВНОЙ ПРОЦЕДУРЫ, НОМЕР ШАГА $K=1$
ОЦЕНКА МАТРИЦЫ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ БЕЗ УЧЕТА ОГРАНИЧЕНИЙ

0.99999774	0.00000226
0.50501567	0.49498433

ПРЕДСКАЗАННЫЕ ОТНОСИТЕЛЬНЫЕ ЧАСТОТЫ

	1		2	
2	0.7525067E	00	0.2474933E	00
3	0.8762522E	00	0.1237477E	00
4	0.9405998E	00	0.5940008E	—01
5	0.9702988E	00	0.2970118E	—01
6	0.9851482E	00	0.1485172E	—01

СУММА КВАДРАТОВ
ОШИБОК ПРОГНОЗА

0.000041557

СРЕДНЕКВАДРАТИЧЕСКАЯ
ОШИБКА

0.000005195

ВЕЛИЧИНА КРИТЕРИЯ χ^2

0.017281160

ВЕЛИЧИНА МОДИФИЦИРОВАННОГО КРИТЕРИЯ χ^2

0.017597124

ОЦЕНКА БЕЗ УЧЕТА ОГРАНИЧЕНИЙ ПРИЕМЛЕМА
ОШИБКА ИТЕРАЦИИ = 0.000256
БАЙЕСОВСКИЕ ОЦЕНКИ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ
АПРИОРНЫЕ ПАРАМЕТРЫ A (I, J)

99.000000	1.000000
50.000000	50.000000

СРЕДНИЕ ЗНАЧЕНИЯ СОВОКУПНОСТИ АПРИОРНЫХ БЕТА-РАСПРЕДЕЛЕНИЙ

0.990000	0.010000
0.500000	0.500000

КОВАРИАЦИОННАЯ МАТРИЦА СОВОКУПНОСТИ АПРИОРНЫХ БЕТА-РАСПРЕДЕЛЕНИЙ

	1	2
1	0.9801979E—04	0.0
2	0.0	0.2475247E—02

ОЦЕНКА МАТРИЦЫ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ БЕЗ УЧЕТА ОГРАНИЧЕНИЙ

1.00015163	— 0.00015163
0.50168413	0.49831587

ОЦЕНКА БЕЗ УЧЕТА ОГРАНИЧЕНИЙ НЕ ОБЛАДАЕТ СВОЙСТВАМИ ВЕРОЯТНОСТИ, ПОЭТОМУ НЕОБХОДИМО ВОСПОЛЬЗОВАТЬСЯ ИТЕРАТИВНОЙ ПРОЦЕДУРОЙ.
СЛЕДУЮЩИЕ РЕЗУЛЬТАТЫ ЯВЛЯЮТСЯ РЕШЕНИЕМ ЗАДАЧИ КВАДРАТИЧНОГО ПРОГРАММИРОВАНИЯ

НОМЕР ИТЕРАЦИИ	РАЗРЕШАЮЩАЯ СТРОКА	НОВЫЙ БАЗИС	СТАРЫЙ БАЗИС	ЗНАЧЕНИЕ ЦЕЛЕВОЙ ФУНКЦИИ
1	1	3	5	0.1635901E 05
2	4	4	0	0.6617109E 03
3	3	1	0	0.2175781E 01

ТРЕБУЕМОЕ ЧИСЛО ЦИКЛОВ = 3
ЗНАЧЕНИЕ ЦЕЛЕВОЙ ФУНКЦИИ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ ИЛИ ОШИБКА ОКРУГЛЕНИЯ В ЗАДАЧЕ КВАДРАТИЧНОГО ПРОГРАММИРОВАНИЯ = —0.2685547E—02

I	J	$B(I)$
1	3	1.0000000000
2	6	0.4981471896
3	1	2.1784667969
4	4	0.5018528104

СЛЕДОВАТЕЛЬНО, БАЙЕСОВСКИЕ ОЦЕНКИ ЭЛЕМЕНТОВ МАТРИЦЫ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ РАВНЫ

1.00000000	0.0
0.50185281	0.49814719

ВЫХОД ИТЕРАТИВНОЙ ПРОЦЕДУРЫ, НОМЕР ШАГА K=1
КОВАРИАЦИОННАЯ МАТРИЦА СОВОКУПНОСТИ АПРИОРНЫХ БЕТА-РАСПРЕДЕЛЕНИЙ

	1	2
1	0.9801979E—04	0.0
2	0.0	0.2475247E—02

ОЦЕНКА МАТРИЦЫ ПЕРЕХОДНЫХ ВЕРОЯТНОСТЕЙ БЕЗ УЧЕТА ОГРАНИЧЕНИЙ

1.00000000	0.0
0.50184596	0.49815404

ПРЕДСКАЗАННЫЕ ОТНОСИТЕЛЬНЫЕ ЧАСТОТЫ

	1	2
2	0.7509230E 00	0.2490770E 00
3	0.8754615E 00	0.1245385E 00
4	0.9402214E 00	0.2988924E—01
5	0.9701107E 00	0.2988924E—01
6	0.9850553E 00	0.1494462E—01

СУММА КВАДРАТОВ ОШИБОК

0.000043023

СРЕДНЕКВАДРАТИЧЕСКАЯ ОШИБКА

0.000005378

ВЕЛИЧИНА КРИТЕРИЯ χ^2

0.019477647

ВЕЛИЧИНА МОДИФИЦИРОВАННОГО КРИТЕРИЯ χ^2

0.020089485

ОЦЕНКА БЕЗ УЧЕТА ОГРАНИЧЕНИЙ ПРИЕМЛЕМА.
ОШИБКА ИТЕРАЦИИ = 0.000014

Г.11. Текст программы на Фортране *

Приведенная далее программа[1] позволяет оценивать переходные вероятности марковской цепи с учетом и без учета ограничений и взвешивания, с помощью метода максимального правдоподобия, метода Байеса, метода наименьших абсолютных отклонений, метода наименьших квадратов с псевдообратной матрицей. Кроме того, вычисляются прогноз относительных частот и критерий согласия χ^2.

```
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96), FMT(9), IS(17), PP(16)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      DATA SAMPLE/'SAMP'/
      DATA PRIOR1/'PRIO'/
      DATA DITTO/'DITT'/
      DATA DITO/'DIT'/
      DATA CLEAR/'CLEA'/
      DATA SUMMAR/'SUMM'/
      DATA RECURS/'RECU'/
      DATA FINIS/'FINI'/
```

* В конце этого параграфа дан перевод на русский язык комментариев, имеющихся в тексте программы. — *Прим. пер.*

[1] Публикуемый здесь текст программы не является фотокопией с оригинала. Кроме того, в каждой строке этой книги помещается менее 80 символов. Поэтому значения счетчиков символов в спецификациях холлеритовских полей буквенно-цифровой информации (типа *H*) не точны и при отладке программы возможны ошибки в операторах формата. Однако логика программы верна.

```
C1    READ SAMPLE INFORMATION
  105 FORMAT (18A4)
  110 FORMAT (1H1)
  112 FORMAT (51H0THIS PROBLEM EXCEEDS THE CAPACITY OF
     1THIS PROGRAM./)
  114 FORMAT (3X, 2I3, F6.0, E6.0, 9I1, I2, 3I1, 1X, 9A4)
  115 READ (5, 105) TESTA, (IS(I), I = 1, 17)
  116 IF (TESTA-SAMPLE) 117, 125, 117
  117 IF (TESTA-CLEAR)    120, 121, 120
  120 IF (TESTA-FINIS)    115,  99, 115
  121 NS = 6
      NS1 = 5
      CALL SMRY(1, 1)
      GO TO 115
  125 READ (5, 114) NTI, NS, SS, TO, KDROP, KW, KV, KP, (KEY(I), I = 1,9),
     1(FMT(I), I = 1,9)
      IF (TO) 126, 126, 127
  126 TOL=1.0E-6
      GO TO 128
  127 TOL=TO
  128 WRITE (6, 110)
      WRITE (6, 105) (IS(I), I=2, 17)
      IF (KDROP) 129, 129, 130
  129 NS1=NS
      GO TO 131
  130 NS1=NS - 1
  131 NT= NT1 - 1
      MM=NS1*NT
      IF (NS-6) 132, 132, 133
  132 IF (MM-96) 134, 134, 133
  133 WRITE (6, 112)
      GO TO 115
C2    GENERALIZATION OF N INTO N(T)
  134 NS2=NS+1
      IF (SS) 80, 80, 135
   80 K=NS+1
      GO TO 136
  135 K=NS
  136 DO 140 I=1,NT1
      READ (5,FMT) (A(I,J), J=1,K)
      IF (KEY(1) - 1) 140, 137, 137
  137 ROWSUM=0.0
      DO 138 J = 1,NS
  138 ROWSUM = ROWSUM+A(I,J)
      DO 139 J = 1,NS
  139 A(I,J) = A(I,J)/ROWSUM
  140 CONTINUE
      IF (SS) 143, 143, 141
  141 DO 142 I = 1,NT1
  142 A(I,NS2) = SS
  143 IF (KEY(1) - 1) 145, 160, 160
  145 DO 146 I = 1,NT1
  146 A(I,NS2) = A(I,NS2)*A(I,NS2)
      WRITE (6, 156)
  156 FORMAT (36H0THE OBSERVED UNITS AND SAMPLE SIZES/)
      GO TO 165
  160 WRITE (6, 163)
```

```
  163 FORMAT (42H0THE OBSERVED PROPORTIONS AND SAMPLE
     1SIZES/)
  165 CALL PRINT (1, 1, NT1, NS2)
      IF (KEY(8) — 1) 166, 166, 170
C3    SOLVED BY LINEAR PROGRAMMING
  166 LC = 2*NS1*NT + NS*NS
      IF (LC-96) 167, 167, 133
  167 CALL LP
      GO TO 115
  170 IF (KDROP) 180, 180, 190
C4    TAKING GARE OF THE WEIGHT AND SOLVING FOR P.
  180 CALL OMEGA(2)
      GO TO 191
  190 CALL OMEGA(1)
  191 IF (KW-9) 500, 192, 500
  192 WRITE (6, 193)
  193 FORMAT (48H0THE FOLLOWINGS ARE THE SECOND STAGE
     1ESTIMATION.)
      CALL PREDIC (10)
      DO 197 J = 1,NS1
      X(1,J) = 0.0
      DO 197 I = 2,NT1
  197 X(1,J) = X(1,J) + (A(I,J) — E(I,J))**2
      DO 198 I = 1,NT1
      DO 198 J = 1,NS
  198 A(I,J) = E(I,J)
      IF (KEY(7) — 2) 202, 200, 202
  200 WRITE (6, 201)
  201 FORMAT (50H0THE DISTURBANCES ARE FROM UNRESTRICT-
     1ED ESTIMATOR.)
      GO TO 205
  202 WRITE (6, 203)
  203 FORMAT (48H0THE DISTURBANCES ARE FROM RESTRICTED
     1ESTIMATOR.)
  205 KEY(7) = KEY(7) — 1
      KW = 10
      CALL OMEGA (3)
C5    READ BETA PRIOR OR NORMAL PRIOR
  500 READ (5, 105) TESTA, (IS(I), I = 1, 17)
      IF (TESTA-RECURS) 700, 703, 700
  700 IF (TESTA-PRIORI) 507, 511, 507
  507 IF (TESTA-DITTO) 508, 510, 508
  508 IF (TESTA-DITO) 600, 510, 600
  510 CALL BETA(TESTA, 1)
      GO TO 701
  511 CALL BETA(TESTA,2)
  701 READ (5, 105) TESTA, (IS(I), I = 1, 17)
      IF (TESTA-RECURS) 700, 702, 700
C6    RECURSIVE ML AND BAYESIAN ESTIMATOR
  702 NBB = 1
      GO TO 704
  703 NBB = 0
  704 READ (5, 799) NREC, NST, KV, KP, (KEY(I), I = 1, 9)
  799 FORMAT (5X, I1, 5X, I1, 11X, 711, I2, 311)
      IF (KDROP) 800, 800, 801
  800 NS1 = NS — 1
      MM = NS1*NT
      KDROP = NS
```

```
  801 KW = 1
      TT = 0.1**NST
      L = NS*NS
      DO 705 I = 1, L
  705 PP(I) = B(I,1)
      DO 752 KREC = 1, NPEC
      WRITE (6, 706) KREC
  706 FORMAT (21H1RECURSIVE OUTPUT, K = , I5/)
      CALL PREDIC(10)
      CALL SIGMA(7)
      IF (NBB-1) 707, 710, 707
  707 CALL OMEGA(4)
      GO TO 740
  710 CALL OMEGA(5)
C7    AMENDING, STORE 0.00001 FOR P(I,J) = 0.0
      DO 733 I = 1, NS
      DO 733 J = 1, NS
      K = I + NS*(J - 1)
      IF (PP(K)) 732, 731, 732
  731 XS(I,J) = 0.00001
      GO TO 733
  732 XS(I,J) = PP(K)
  733 CONTINUE
      CALL BETA (TESTA, 3)
  740 RR = 0.0
      DO 748 I = 1, L
      RR = RR + ABS(PP(I) - B(I,1))
  748 PP(I) = B(I, 1)
      WRITE (6, 750) RR
  750 FORMAT (23H0RESURSIVE DIFFERENCE = , F9.6)
      IF (RR-TT) 500, 500, 752
  752 CONTINUE
      GO TO 500
  600 IF (TESTA-SUMMAR) 116, 610, 116
  610 CALL SMRY (3, 1)
      CALL SMRY(3, 2)
      GO TO 500
   99 CALL EXIT
      END
C
C
C8    THIS SUBROUTINE INCORPORATES UNIVARIATE BETA, OR
C9    MULTIVARIATE BETA,  OR INDEPENDENT  NORMAL WITH
C10   SAMPLE OBSERVATION TO OBTAIN POSTERIOR ESTIMATES.
C
      SUBROUTINE BETA(TESTA, IB)
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96), PR(6, 7), PQ(6, 6)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      DATA DITTO/'DITT'/
      DATA DITO/'DIT.'/
      GO TO (510, 511, 834), IB
  510 IF (KALT-8888) 826, 901, 826
  511 WRITE (6, 501)
      ARBI = 10.0**KEY(6)
      READ (5,503) NBETA, KALT
      IF(KALT-9999) 800, 522, 800
  800 IF(KALT-8888) 801, 512, 801
```

```
C11   FROM UNIVARIATE BETA PRIOR TO MULTIVARIATE BETA
C12   PRIOR PDF
  801 SN = NS
      DO 810 I = 1, NS
  810 PR(I, 7) = 0.0
      DO 820 NPR = 1, NBETA
  820 READ (5, 503) I, J, PR(I,J), PQ(I,J)
      DO 825 I = 1, NS
      DO 825 J = 1, NS
  825 PR(I, 7) = PR(I, 7) + PR(I,J)
  826 DO 828 I = 1, NS
      DO 828 J = 1, NS
  828 XS(I,J) = PR(I,J)/PR(I, 7)
      IF (TESTA-DITTO) 831, 834, 831
  831 WRITE (6, 870)
      DO 829 I = 1, NS
  829 WRITE (6, 873) (PR(I,J), J = 1, NS)
      WRITE (6, 871)
      DO 830 I = 1, NS
  830 WRITE (6, 873) (XS(I,J), J = 1, NS)
  834 L = NS*NS
      DO 832 I = 1,L
      DO 832 J = 1,L
  832 A(I,J) = 0.0
      DO 850 I = 1, NS
      DO 850 J = 1, NS1
      K = I + NS*(J - 1)
  835 C(K, K) = C(K, K) + (PR(I, 7) - SN)/(XS(I,J)*ARBI) + (PR(I, 7) - SN)/
     1(XS(I, NS)*ARBI)
      D(K) = D(K) + (PR(I,J) - 1.0)/(XS(I,J)*ARBI) + (PR(I, 7) -
     1PR(I, NS) - SN + 1.0)/(XS(I, NS)*ARBI)
      DO 850 JK = J, NS
      IF (JK-J) 850, 850, 837
  837 L = I + NS*(JK - 1)
      IF (JK - NS) 839, 845, 845
  839 C(K, L) = C(K, L) + (PR(I, 7) - SN)/(XS(I, NS)*ARBI)
  840 C(L, K) = C(K, L)
  845 IF (TESTA-DITTO) 846, 850, 846
  846 A(K, K) = PR(I,J)*(PR(I, 7) - PR(I,J))/(PR(I, 7)**2*(PR(I,7) + 1.0))
      A(K, L) = - PR(I,J)*PR(I,JK)/(PR(I, 7)**2*(PR(I, 7) + 1.0))
      A(L, K) = A(K, L)
  850 CONTINUE
      IF (TESTA-DITTO) 855, 526, 855
  855 WRITE (6, 872)
      CALL PRINT (1, 1, K, K)
  870 FORMAT (24H0PRIOR PARAMETERS A(I,J)/)
  871 FORMAT (23H0MULTI-BETA PRIOR MEANS/)
  872 FORMAT (35H0MULTI-BETA PRIOR COVARIANCE MATRIX/)
  873 FORMAT (1H, 6F12.6)
      GO TO 526
  901 WRITE (6, 501)
      WRITE (6, 509)
      DO 521 M = 1, NS
      DO 521 N = 1, NS1
      R = PR(M, N)
      Q = PQ(M, N)
      ARBI = 10.0**KEY(6)
      K = M + NS*(N - 1)
      GO TO 514
  512 WRITE (6, 502)
      DO 520 NPR = 1, NBETA
```

```
      READ 503, I, J, R, Q
      PR(I,J) = R
      PQ(I,J) = Q
      PM = R/(R+Q)
      V = (R*Q)/((R+Q+1.0)*(R+Q)**2.0)
      IF (J - NS1) 513, 513, 520
  513 K = I+NS*(J - 1)
  514 IF (KEY(6)) 516, 515, 516
  515 C(K, K) = C(K, K) + (R+Q)**2.0*(R+Q - 2.0)/(R*Q)
      D(K) = D(K) + (R+Q)**2.0*(R - 1.0)/(R*Q)
      GO TO 517
  516 C(K, K) = C(K, K) + (R+Q)**2.0*(R+Q - 2.0)/(R*Q*ARBI)
      D(K) = D(K) + (R+Q)**2.0*(R - 1.0)/(R*Q*ARBI)
  517 IF (TESTA-DITTO) 518, 521, 518
  518 IF (TESTA-DITO) 520, 519, 520
  519 I = M
      J = N
  520 WRITE (6, 503) I, J, R, Q, PM, V
  521 CONTINUE
      GO TO 526
C13   ASSUMING NORMAL PRIOR
  522 WRITE (6, 505)
      DO 525 NPR = 1, NBETA
      READ (5, 506) I, J, AVE, VAR
      IF (J - NS1) 555, 555, 525
  555 K = I+NS*(J - 1)
      IF (KEY(6)) 524, 523, 524
  523 C(K,K) = C(K,K) + 1.0/VAR
      D(K) = D(K) + AVE/VAR
      GO TO 525
  524 C(K, K) = C(K, K) + 1.0/VAR*ARBI)
      D(K) = D(K) + AVE/(VAR*ARBI)
  525 WRITE (6, 506) I, J, AVE, VAR
C14   OBTAIN POSTERIOR
  526 DO 530 I = 1, NN
      B(I, 1) = D(I)
      DO 530 J = 1, NN
  530 A(I,J) = C(I,J)
  501 FORMAT (52H1BAYESIAN ESTIMATION OF THE TRANSITION
     1PROBABILITIES//)
  502 FORMAT (35H0 THE LIST OF THE PRIOR PARAMETERS/72H0 I        J
     1          R(I,J)          S(I,J)                MEAN        VARIANCE/)
  503 FORMAT (2I6, 2F15.4, 2F15.8)
  504 FORMAT (72H0THE FOLLOWING  XSX  AND  XSY  DENOTE
     1XSX+S(0)  AND XSY+S(0) P(0) RESPECTIVELY//)
  505 FORMAT (32H0THE LIST OF THE PRIOR KNOWLEDGE//42H  I        J
     1          MEAN          VARIANCE/)
  506 FORMAT (2I6, 2F15.8)
  509 FORMAT (69H0THE PRIOR PARAMETERS ARE THE SAME AS
     1LISTED IN THE PREVIOUS PROBLEM./)
  540 WRITE (6, 504)
      CALL GLSQP(2)
      RETURN
      END
C
C
C15   THIS SUBROUTINE INITIATES THE PROBLEM CONCERNING
C16   THE WEIGHT MATRIX AND THE COLUMN DROPPED TO FORM
C17   THE  MULTIVARIATE  STRUCTURE  WITH  NON-SINGULAR
C18   WEIGHT MATRIX.
```

```
C
      SUBROUTINE OMEGA(KO)
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      GO TO (700, 743, 790, 800, 800), KO
C19   DROP A COLLUMN IF REQUIRED
  700 IF (KDROP-NS) 705,1702, 701
  701 KDROP=NS
  702 DO 703 I=1, NS
  703 NP(I)=I
      GO TO 735
  705 DO 710 I=1, NT1
      DO 710 J=1, NS
  710 E(I,J)=A(I,J)
      DO 720 I=1, NT1
      K=0
      DO 720 J=1, NS
      IF (J-KDROP) 715, 720, 715
  715 K=K+1
      NP(K)=J
      A(I, K)=E(I,J)
  720 CONTINUE
      DO 730 I=1, NT1
  730 A(I, NS)=E(I, KDROP)
      NP(NS)=KDROP
  735 WRITE (6, 740) KDROP
      WRITE (6, 741) (NP(I), I=1, NS)
      GO TO 744
  740 FORMAT (23H0AS YOU DESIRED, COLUMN, I2, 34H IS DROPPED
     1IN FORMING THE SYSTEM.)
  741 FORMAT (44H READ THE FOLLOWING MATRICES IN THE
     1SEQUENCE, 6I3/)
  742 FORMAT (50H0NO COLUMN HAS BEEN DROPPED IN FORMING
     1THE SYSTEM.)
C20   CHOICE OF WEIGHT
  743 WRITE (6, 742)
  744 NS2=NS+1
      DO 745 I=1, NT1
      DO 745 J=1, NS2
  745 E(I,J)=A(I,J)
      IF (KW) 753, 753, 746
  746 GO TO (755, 764, 770, 758, 760, 747, 748, 748, 750, 790), KW
  747 IF (KDROP) 764, 764, 762
  748 WRITE (6, 749)
  749 FORMAT (65H0THE WEIGHT MATRIX IS DIAGONAL WITH
     1ELEMENTS N/(WJ(T)*(1-WJ(T))).)
      CALL SIGMA(5)
      GO TO 800
  750 WRITE (6, 751)
  751 FORMAT (37H0THIS WILL BE A TWO STAGE ESTIMATION.)
  753 WRITE (6, 754)
  754 FORMAT (54H0THE WEIGHT MATRIX IS AN IDENTITY MATRIX
     (UNWEIGHTED).)
      CALL SIGMA(1)
      GO TO 800
  755 IF (KDROP) 764, 764, 756
  756 WRITE (6, 757)
```

```
757 FORMAT (116H0THE WEIGHT MATRIX IS DERIVED FROM THE
   1MLE WITH ELTMENTS N/WR(T)+Z*(N/WJ(T)), Z=1 FOR
   2DIAGONALS AND Z=0 OTHERWISE./)
    CALL SIGMA(7)
    GO TO 800
758 WRITE (6, 759)
759 FORMAT (72H0THE WEIGHT MATRIX IS DIAGONAL WITH THE
   1INVERSE OF THE MEAN PROPORTIONS./)
    CALL SIGMA(2)
    GO TO 800
760 WRITE (6, 761)
761 FORMAT (87H0THE WEIGHT MATRIX IS DIAGONAL WITH THE
   1INVERSE OF THE PRODUCT OF THE MEAN PROPORTIONS./)
    CALL SIGMA(3)
    GO TO 800
762 WRITE (6, 763)
763 FORMAT (111H0THE WEIGHT MATRIX IS DERIVED FROM THE
   1GENERALIZED INVERSE. ONE COLUMN IS DROPPED TO PRE-
   2VENT MULTICOLINEARITY./)
    CALL SIGMA(8)
    GO TO 800
764 WRITE (6, 765)
765 FORMAT (53H0THE WEIGHT MATRIX IS DIAGONAL WITH
   1ELEMENTS N/WJ(T)./)
    CALL SIGMA(4)
    GO TO 800
770 READ 771, TESTA, NUMBER
771 FORMAT (A6, 6X, I6)
    DATA WEIGHT/'WEIG'/
    IF (TESTA-WEIGHT) 772, 774, 772
772 WRITE (6, 773)
773 FORMAT (22H0ERROR IN INPUT GARDS./)
    CALL EXIT
774 CALL SIGMA(9)
    DO 778 K=1, NUMBER
    READ (5, 776) I, J, WEIGH
776 FORMAT (2I6, F15.4)
778 S(I,J)=WEIGH
    WRITE (6, 779)
779 FORMAT (43H0THE WEIGHT MATRIX IS ASSIGNED BY THE
   1USER./)
    IF (KEY(2) - 1) 800, 780, 800
780 WRITE (6, 784)
784 FORMAT (26H0THE ASSIGNED WEIGH MATRIX/)
    CALL TABLE (1, MM, 2)
    GO TO 800
790 WRITE (6, 791)
791 FORMAT (115H0THE WEIGHT MATRIX IS DIAGONAL WITH
   1ELEMENTS THE INVERSE OF THE ESTIMATE OF THE I-TH
   2EQUATION DISTURBANCE VARIANCE./)
    CALL SIGMA(6)
C21 IN ORDER TO REDUCE RANGES BETWEEN COUNTER PARTS,
C22 THE WEIGH MATRIX WILL BE DIVIDED BY THE NUMBER AS-
C23 SIGNED.
800 IF (KEY(6)) 805, 820, 805
805 ARBI=10.0**KEY(6)
    DO 810 I=1, MM
    DO 810 J=1, MM
810 S(I,J)=S(I,J)/ARBI
    WRITE (6, 815) ARBI
```

```
  815 FORMAT (48HTHE WEIGHT MATRIX WILL BE DIVIDED BY
     1THE SCALAR, F10.2)
  820  IF (KEY(1) — 1) 825, 826, 826
  825 SS=1.0
  826 IF (KEY(8) — 1) 835, 835, 836
  836 IF (KO-5) 830, 827, 830
  827 CALL GLSQP(5)
      RETURN
  830 CALL GLSQP(1)
  835 RETURN
      END
C
C
C24   THIS SUBROUTINE FORMS THE MATRICES S, X, AND Y FOR THE
C25   MULTIVARIATE STRUCTURE
C
      SUBROUTINE SIGMA(MAP)
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 69), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      TN=NT
      SN=NS
      DO 100 I=1, MM
      DO 100 J=1, MM
  100 S(I,J)=0.0
      GO TO (120, 130, 130, 160, 160, 170, 210, 210, 300, 350), MAP
C26   UNWEIGHTED
  120 DO 121 K=1, MM
  121 S(K, K)=1.0
      GO TO 300
C27   HOMOSKEDASTICITY WITHIN EQUATIONS  BUT  HETEROSKE-
C28   DASTICITY AMONG EQUATIONS
  130 DO 131 J=1, NS1
  131 X(1, J)=0.0
      DO 136 J=1, NS1
      DO 135 I=2, NT1
  135 X(1,J)=X(1,J)+A(I,J)
  136 X(1,J)=TN/X(1,J)
      IF (MAP-2) 120, 140, 150
C29   HETEROSKEDASTICITY BY THE INVERSE  OF  THE  MEAN
C30   PROPORTIONS
  140 DO 145 J=1, NS1
      LL=1+NT*(J — 1)
      LS=LL+NT — 1
      DO 145 K=LL, LS
  145 S(K, K)=X(1,J)
      GO TO 290
C31   HETEROSKEDASTICITY BY THE INVERSE OF THE PRODUCT
C32   OF THE AVERAGE PROPORTION IN STATE I AND THE AVER-
C33   AGE PROPORTION NOT IN STATE I
  150 DO 155 J=1, NS1
      SIGMA 1=X(1,J)*X(1,J)/(X(1,J) — 1.0)
      LL=1+NT*(J — 1)
      LS=LL+NT — 1
      DO 155 K=LL, LS
  155 S(K, K)=SIGMA 1
      GO TO 290
C34   DIAGONAL HETEROSKEDASTICITY BY N/WJ(T) OR BY N/(WJ
```

```
C35     (T)*(1 — WJ(T)))
  160 K=0
      DO 166 J=1, NS1
      DO 166 I=2, NT1
      K=K+1
      IF (A(I,J)) 162, 162, 161
  161 IF (A(I,J) — 1.0) 163, 162, 163
  162 S(K, K)=10.0*E(I,NS+1)*E(I,NS+1)
      GO TO 166
  163 IF (MAP — 4) 164, 164, 165
  164 S(K, K)=E(I, NS+1)/A(I,J)
      GO TO 166
  165 S(K, K)=E(I, NS+1)/(A(I,J)*(1.0 — A(I,J)))
  166 CONTINUE
      GO TO 290
C36   HETEROSKEDASTICITY BY THE INVERSE OF THE I-TH EQUA-
C37   TION DIST. VARIANCE.
  170 DO 171 J=1, NS1
  171 X(I, J)=(TN — SN)/X(I,J)
      IF (KEY(2) — 2) 140, 172, 140
  172 KEY(2)=1
      GO TO 140
C38   THE WEIGHT DERIVED FROM  THE  MAXIMUM  LIKELIHOOD
C39   FUNCTION
  210 DO 280 I=1, MM
      DO 280 L=1, NS1
C40   J IS DETERMINED BY L, AND I MINUS THE MULTIPLE OF NT
      IR=I
  215 IF (IR — NT) 225, 225, 220
  220 IR=IR — NT
      GO TO 215
  225 J=IR+NT*(L — 1)
C41   K IS DETERMINED BY I MINUS THE MULTIPLE OF NT
      K=IR
C42   CHECK WHETHER IT IS A DIAGONAL ELEMENT
      IF (I — J) 245, 250, 245
  245 Z=0.0
      GO TO 255
  250 Z=1.0
  255 CONTINUE
C43   STORE THE APPROPRIATE VALUE. IF A IS ZERO, REPLACED
C44   BY 0.1/SS
      SIGMA 1=A(K+1, NS)
      SIGMA 2=A(K+1, L)
      IF (SIGMA 1) 256, 256, 257
  256 SIGMA 1=0.1/E(K+1, NS+1)
  257 IF (SIGMA 2) 258, 258, 259
  258 SIGMA 2=0.1/E(K+1, NS+1)
  259 IF (MAP — 7) 265, 260, 265
  260 S(I,J)=(E(K+1, NS+1)/SIGMA 1)+Z*(E(K+1, NS+1)/
     1SIGMA 2)
      GO TO 280
  265 S(I,J)=(E(K+1, NS+1)/(SIGMA 1*SN)+Z*(E(K+1, NS+1)/
     1SIGMA 2) — E(K+1, NS+1)/(SIGMA 2*SN)
  280 CONTINUE
C45.  FOR PROGRAMMING CHECKING-OPTION
  290 IF (KEY(2) — 1) 300, 294, 300
  294 WRITE (6, 295)
  295 FORMAT (18H0THE WEIGH MATRIX/)
      CALL TABLE (1, MM, 2)
C46   MVS-FORMING A KRONECKER PRODUCT
```

```
  300 NN=NS*NS1
      DO 310 I=1, MM
      DO 310 J=1, NN
  310 X(I,J)=0.0
      I=0
      DO 320 KT=1, NS1
      DO 320 K=1, NT
      I=I+1
      J=(KT-1)*NS
      DO 320 L=1, NS
      J=J+1
  320 X(I, J)=E(K, L)
      I=0
      DO 330 L=1, NS1
      DO 330 K=2, NT1
      I=I+1
  330 Y(I)=E(K, L)
  350 RETURN
      END
C
C
C47   GENERALIZED LEAST SQUARES AND QUADRATIC  PROGRAM-
C48   MING.
C
      SUBROUTINE GLSQP(NCASE)
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      GO TO (400, 420, 616, 625, 400), NCASE
  400 DO 410 I=1, NN
      DO 410 J=1, MM
      XS(I,J)=0.0
      DO 410 K=1, MM
  410 XS(I,J)=XS(I,J)+X(K,I)*S(K,J)
      DO 416 I=1, NN
      B(I,1)=0.0
      DO 416 L=1, MM
  415 B(I, 1)=B(I, 1)+XS(I, L)*Y(L)
      DO 416 J=1, NN
      A(I,J)=0.0
      DO 416 K=1, MM
      A(I,J)=A(I,J)+XS(I, K)*X(K,J)
  416 CONTINUE
  420 DO 421 I=1, NN
      D(I)=B(I, 1)
      DO 421 J=1, NN
  421 C(I,J)=A(I,J)
      IF (NCASE-5) 417, 700, 417
  417 IF (KEY(8)-NCASE-8) 422, 700, 422
  422 IF (KEY(3)-1) 426, 424, 423
  423 IF (KEY(3)-NCASE-1) 426, 424, 426
  424 WRITE (6, 490)
      CALL PRINT (1, 1, NN, NN)
      WRITE (6, 491)
      DO 425 I=1, NN
  425 WRITE (6, 495) I, B(I, 1)
  426 IF (KEY(5)-3) 427, 428, 428
  427 CALL SWPMAT (NN, 1, TOL, DETERM, 0)
```

```
      IF (DETERM) 461, 480, 461
  428 CALL SWPMAT (NN, 1, TOL, DETERM, 1)
      IF (DETERM) 460, 480, 460
  460 WRITE (6, 494) DETERM
C49   PRINTING P AND REGISTERING NEGATIVE P
  461 NEG=0
      KL=NN
      WRITE (6, 492)
      DO 470 K=1, NS
      L=0
      PR=1.0
      N=NN-NS+K
      DO 465 I=K, N, NS
      L=L+1
      Y(L)=B(I,1)
      IF (B(I,1)) 464, 465, 465
  464 NEG=NEG+1
  465 PR=PR-B(I, 1)
      KL=KL+1
      B(KL, 1)=PR
      IF (PR) 466, 467, 467
  466 NEG=NEG+1
  467 CONTINUE
      IF (KDROP) 468, 468, 469
  468 WRITE (6, 496) (Y(L), L=1, NS1)
      GO TO 470
  469 WRITE (6, 496) (Y(L), L=1, NS1), PR
  470 CONTINUE
      IF (KEY(9)-1) 472, 471, 472
  471 CALL SMRY(2, 1)
  472 IF (KV-1) 481, 474, 473
  473 IF (KV-NCASE-1) 481, 474, 475
  474 WRITE (6, 488)
      CALL PRINT (1, 1, NN, NN)
      GO TO 481
  475 IF (NEG) 472, 472, 481
  480 WRITE (6, 493)
      GO TO 530
  481 IF (KP-1) 487, 486, 482
  482 IF (KP-NCASE-1) 487, 486, 483
  483 IF (NEG) 487, 486, 487
  486 CALL PREDIC (1)
  487 IF (KEY(7)-1) 700, 501, 501
  488 FORMAT (102H0THE DISPERSION MATRIX OF THE ML (GLS,
     1MCS) ESTIMATOR P(I,J), OR THE INVERSE OF XSX FOR ANY-
     2THING ELSE./)
  489 FORMAT (39H0THE UNRESTRICTED ESTIMATOR IS PERFECT./)
  490 FORMAT (29H0GENERALIZED XSX FOR QP INPUT/)
  491 FORMAT (29H0GENERALIZED XSY FOR QP INPUT/)
  492 FORMAT (48H0UNRESTRICTED ESTIMATOR OF THE TRANSI-
     1TION MATRIX/)
  493 FORMAT (33H0XSX IS SINGULAR, TRY QP SOLUTION/)
  494 FORMAT (21H0DETERMINANT OF XSX=, E14.7)
  495 FORMAT (1H, I3, 3X, E14.7)
  496 FORMAT (8F15.8)
  497 FORMAT (67H0THE UNRESTRICTED ESTIMATOR VIOLATES
     1THE PROPERTIES OF PROBABILITY. /44H HENCE, THE ITERA-
     2TION PROCEDURE IS REQUIRED./)
C50   IF P IS NOT IN THE RANGE BETWEEN ONE AND ZERO CALL
C51   FOR QP SOLUTION
  501 IF (KEY(7)-NCASE) 700, 510, 502
```

```
  502 IF (KEY(7)-3) 700, 510, 510
  510 IF (NEG) 511, 511, 512
  511 WRITE (6, 489)
      GO TO 700
  512 WRITE (6, 497)
      IF (KEY(7)-4) 530, 520, 520
  520 K4=KEY(4)
      IF (KEY(7)-NCASE-3) 525, 530, 525
  525 KEY(4)=0
  530 CALL QP(I)
      LX=LR
      IF (KEY(7)-4) 540, 531, 531
  531 KEY(4)=K4
C52   OUTPUTING THE QP SOLUTIONS IN THE MATRIX FORM
  540 IF (KW-1) 548, 542, 541
  541 GO TO (542, 546, 546, 546, 546, 549, 546, 546, 548, 547), KW
  542 IF (NCASE-1) 544, 544, 545
  544 WRITE (6, 560)
      GO TO 616
  545 WRITE (6, 561)
      GO TO 616
  546 WRITE (6, 562)
      GO TO 616
  547 WRITE (6, 563)
      GO TO 616
  548 WRITE (6, 564)
      GO TO 616
  549 WRITE (6, 565)
      GO TO 616
  560 FORMAT (63H0HENCE, THE ML (GLS, MCS) ESTIMATOR OF THE
     1TRANSITION MATRIX IS/)
  561 FORMAT (58H0HENCE, THE BAYESIAN ESTIMATOR OF THE
     1TRANSITION MATRIX IS/)
  562 FORMAT (58H0HENCE, THE WEIGHTED ESTIMATOR OF THE
     1TRANSITION MATRIX IS/)
  563 FORMAT (59H0HENCE, THE TWO-STAGE ESTIMATOR OF THE
     1TRANSITION MATRIX IS/)
  564 FORMAT (70H0HENCE, THE UNWEIGHTED CLASSICAL ESTI-
     1MATOR OF THE TRANSITION MATRIX IS/)
  565 FORMAT (92H0HENCE, THE FIRST STAGE RESTRICTED GENE-
     1RALIZED INVERSE ESTIMATOR OF THE TRANSITION MATRIX
     2IS/)
  616 DO 617 I=1, LR
  617 B(I, 1)=0.0
      DO 620 I=1, LR
      K=NP(I) — LD
      IF (K) 620, 620, 618
  618 IF (K-LX) 619, 619, 620
  619 B(K, 1)=Y(I)
  620 CONTINUE
      IF (KEY(9)-1) 625, 621, 625
  621 CALL SMRY(2,2)
  625 DO 650 L=1, NS
      LL=0
      N=NN+L
      DO 630 I=L, N, NS
      LL=LL+1
  630 Y(LL)=B(I, 1)
  650 WRITE (6, 496) (Y(LL), LL=1, NS)
      IF (NCASE-4) 651, 700, 651
  651 IF (KP-1) 700, 660, 652
```

```
  652 IF (KP-NCASE-1) 700, 660, 653
  653 IF (KP-4) 700, 660, 660
  660 CALL PREDIC (1)
  700 RETURN
      END
C
C
C53   THIS SUBROUTINE FORMS A SIMPLEX TABLEAU FOR A LI-
C54   NEAR PROGRAMMING FITTING A SET OF REGRESSION HY-
C55   PERPLANES BY MINIMIZING THE SUM OF ABSOLUTE DIS-
C56   TANCES, WEIGHTED OR UNWEIGHTED.
C
      SUBROUTINE LP
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      WRITE (6, 100)
  100 FORMAT (75H0YOU HAVE INDICATED THAT THIS PROBLEM
     1WILL BE SOLVED BY LINEAR PROGRAMMING./)
      IF (KDROP) 101, 101, 102
  101 CALL OMEGA(2)
      GO TO 104
  102 CALL OMEGA(1)
  104 NNS=NS*NS
      NN1=NN+1
      MM1=MM+1
      LR=MM+NS
      LX=LC
      LD=0
      NOCYCL=0
C57   FORM A SIMPLEX TABLEAU FOR LP
      DO 105 J=1, LC
  105 XS(1, J)=0.0
      DO 110 K=1, MM
      J=NNS+K
      L=MM+J
      DO 106 I=1, MM
  106 XS(I,J)=XS(I,J)+S(I,K)
  110 XS(I,L)=XS(I,J)
      DO 111 I=1, MM
      DO 111 J=1, NN
  111 S(I,J)=X(I,J)
      DO 112 J=1, LC
  112 X(J,1)=XS(1,J)
      IF (KEY(8)) 113, 113, 115
  113 K=NNS+1
      DO 114 J=K, LC
  114 X(J,1)=SQRT(X(J,1))
  115 IF (KDROP-1) 120, 116, 116
  116 DO 117 I=1, MM
      DO 117 J=NN1, NNS
  117 S(I,J)=0.0
  120 IF (KEY(4)-1) 122, 123, 122
  122 IF (KEY(2)-1) 129, 123, 123
  123 K=NNS+1
      WRITE (6, 124) K, LC
  124 FORMAT (40H0THE COST VECTOR FOR ERROR TERMS (CO-
     1LUMN, I3, 10H TO COLUMN, I3, 1H)/)
```

```
      WRITE (6, 125) (X(J,1),J=K, LC)
  125 FORMAT (1H, 9E14.7)
C58   FORM ERROR TERMS  IN THE  FORM  OF  COUNTER  PARTS  AND
C59   INTRODUCE INTO BASIS
  129 DO 130 I=1, MM
      DO 130 J=NN1, LC
  130 S(I,J)=0.0
      DO 135 I=1, MM
      J=NNS+I
      NP(I)=J
      S(I,J)=1.0
      L=MM+J
  135 S(I,L)= — 1.0
C60   FORM ROW SUM CONSTRAINTS
      DO 140 I=MM1, LR
      NP(I)=0
  140 Y(I)=1.0
      DO 145 I=MM1, LR
      DO 145 J=1, LC
  145 S(I,J)=0.0
      DO 146 K=1, NS
      I=MM+K
      L=NNS — NS+K
      DO 146 J=K, L, NS
  146 S(I,J)=1.0
      IF (KDROP) 150, 150, 147
  147 J=NN
      DO 148 I=MM1, LR
      J=J+1
  148 NP(I)=J
      IF (KEY(4)-1) 156, 149, 156
  149 CALL TABLE(1,LC,1)
      GO TO 156
C61   PHASE I ITERATION
  150 IF (KEY(4)-1) 152, 151, 152
  151 CALL TABLE (1,LC,1)
  152 CALL SOLVE (1)
      IF (KEY(5)-2) 156, 154, 153
  153 IF (KEY(5)-5) 156, 154, 156
  154 WRITE (6, 155)
  155 FORMAT (42HEND OF PHASE I AND BEGINNING OF PHASE II.)
C62   PHASE II ITERATION
  156 DO 160 I=1, LR
      M=NP(I)
  160 X(I, 2)=X(M, 1)
      KS=LR+1
      DO 165 J=1, LC
      S(KS,J)=0.0
      DO 164 I=1, LR
  164 S(KS,J)=S(KS,J)+X(I,2)*S(I,J)
  165 S(KS,J)=S(KS,J) — X(J,1)
      Y(KS)=0.0
      DO 170 I=1, LR
  170 Y(KS)=Y(KS)+X(I,2)*Y(I)
      IF (KDROP) 172, 172, 175
  172 CALL SOLVE (3)
      GO TO 176
  175 CALL SOLVE (2)
  176 Y(KS)=0.0
      DO 180 I=1, LR
      M=NP(I)
```

```
      X(1,2)=X(M,1)
  180 Y(KS)=Y(KS)+X(1,2)*Y(I)
      CALL QP(2)
C63   OUTPUT THE TRANSITION MATRIX
      WRITE (6, 201)
  201 FORMAT (66H0THE MINIMUM ABSOLUTE DEVIATION ESTI-
     1MATOR OF THE TRANSITION MATRIX/)
      LX=NNS
      CALL GLSQP(3)
      RETURN
      END
C
C
      SUBROUTINE QP(KOUT)
C
C64   THIS SUBROUTINE FOR MS SIMPLEX TABLEAUX FOR QUADRA-
C65   TIC PROGRAMMING PROBLEMS, WITH EQUALITY OR IN-
C66   EQUALITY CONSTRAINTS.
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      GO TO (3, 100), KOUT
C67   DEFINE CONSTANTS
    3 LD=NS
      LR=NS*NS
      IF (KDROP) 5, 5, 8
    5 LR=LR+NS
    8 LC=2*LR
      LX=2*LC
C68   STORE THE VALUE FOR P-ZERO COLUMN, OR RIGHT HAND
C69   SIDE.
      DO 10 I=1, NS
   10 Y(I)=1.0
      K=0
      NS2=NS+1
      DO 15 I=NS2, LR
      K=K+1
   15 Y(I)=D(K)
C70   CLEAR THE CORE LOCATIONS FOR LEFT HAND SIDE.
      DO 20 I=1, LR
      DO 20 J=1, LC
   20 S(I, J)=0.0
C71   FORM G-PRIME AND G MATRICES (ALSO — G-PRIME FOR
C72   EQUALITY TYPE).
      DO 30 I=1, LD
      KR=LD+I
      KC=NN+I
      DO 30 J=KR, KC, NS
      S(I,J)=1.0
      IF (KDROP) 25, 25, 30
   25 L=LR+I
      S(J,L)=-1.0
   30 S(J,I)=1.0
C73   FORM BETA-MATRIX
      DO 40 K=1, NN
      I=LD+K
      DO 40 L=1, NN
      J=LD+L
   40 S(I,J)=C(K,L)
```

```
C74   FORM SLACK VARIABLES, W(I).
      IF (KDROP) 55, 55, 48
   48 DO 50 I=1, NS
      J=LR+I
   50 S(I,J)=1.0
   55 DO 60 I=NS2, LR
      J=LR+I
   60 S(I,J)= — 1.0
C75   INITIALIZATION
      DO 70 K=1, LR
   70 NP(K)=0
      IF (KDROP) 85, 85, 75
C76   INTRODUCE ALL THE ELEMENTS OF THE DROPPED COLUMN
C77   OF P INTO THE BASIS INITIALLY AS A SHORT-CUT.
   75 DO 80 K=1, LD
   80 NP(K)=LR+K
C78   REDEFINE CONSTANTS SO THAT P(I)S ARE COUNTERPARTS
C79   TO W(I)S.
   85 LD=LD+NN
      NOCYCL=0
C80   OUTPUT THE INITIAL TABLEAU — OPTION
      IF (KEY(4)-1) 95, 90, 95
   90 CALL TABLE (1, LC, 1)
   95 WRITE (6, 190)
      CALL SOLVE(1)
      LD=NS
C81   OUTPUT ROUTINE
  100 KS=LR+1
      WRITE (6, 191) NOCYCL, Y(KS)
      IF (KEY(5)) 200, 200, 105
  105 IF (KEY(5)-3) 106, 200, 106
  106 WRITE (6, 192)
      DO 110 I=1, LR
  110 WRITE (6, 193) I, NP(I), Y(I)
  190 FORMAT (32H1THE FOLLOWINGS ARE QP SOLUTIONS//)
  191 FORMAT (28H0NUMBER OF CYCLES REQUIRED=I4, 49H,
     1OBJ. VALUE FOR LP OR ROUNDING ERROR FOR QP=E15.7/)
  192 FORMAT (44H0              I              J              B(I)/)
  193 FORMAT (2I10, F30.10)
  200 RETURN
      END
C
C
C82   THIS SUBROUTINE SOLVES SIMPLEX TABLEAUX FOR LP AND
C83   ALSO FOR QP. BY T. C. LEE.
C
      SUBROUTINE SOLVE (LQP)
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      ASSIGN 135 TO KK
      KS=LR+1
      IF (LQP-2) 16, 16, 380
   16 IF (KEY(5)-2) 80, 19, 17
   17 IF (KEY(5)-5) 80, 19, 80
   19 WRITE (6, 20)
   20 FORMAT (84H0NO. OF ITERATIONS PIVOT ROW NEW BASIS
     1OLD BASIS OBJ. VALUE BEFORE ITERATION)
```

```
   80 IF (LQP-1) 90, 90, 380
   90 NOCYCL=0
C84   OBTAIN ALTERNATIVE COSTS
  100 LOOP=0
      DO 110 J=1, LC
  110 S(KS,J)=0.0
      DO 120 J=1, LC
      DO 120 I=1, LR
      IF (NP(I)) 120, 119, 120
  119 S(KS,J)=S(KS,J)+S(I,J)
  120 CONTINUE
      Y(KS)=0.0
      DO 125 I=1, LR
      IF (NP(I)) 125, 124, 125
  124 Y(KS)=Y(KS)+Y(I)
  125 CONTINUE
C85   SHALL PROBLEM MAY WISH TO KNOW ITERATION STEPS
  131 IF (KEY(4)-9) 135, 132, 135
  132 CALL TABLE (I, LC, 1)
      GO TO KK, (135, 400)
C86   CHECK WHETHER TO GO OR TO STOP
  135 IF (NOCYCL-LX) 140, 136, 400
  136 WRITE (6, 137) Y(KS)
  137 FORMAT (82H0I THINK I AM LOOPING, SO I WILL TERMINATE.
     1THE VALUE OF THE OBJECTIVE FUNCTION=, E20.10/)
  138 IF (KEY(4)-8) 400, 139, 400
  139 NOCYCL=LX+1
      GO TO 132
C87   FIND BASIS AMONG ACTIVE COLUMNS
  140 KMAX=0
      DO 160 J=1, LC
      IF (KMAX) 501, 501, 500
  500 IF (S(KS,J)) 160, 160, 143
  501 IF (S(KS,J)) 160, 141, 143
  141 DO 142 I=1, LR
      IF (NP(I)-J) 142, 160, 142
  142 CONTINUE
  143 IF (LD) 150, 150, 144
  144 IF (J-LD) 147, 147, 145
  145 IF (J-LR) 150, 150, 146
  146 L=J-LR
      GO TO 148
  147 L=J+LR
  148 DO 149 K=1, LR
      IF (NP(K)-L) 149, 160, 149
  149 CONTINUE
  150 IF (KMAX) 151, 151, 152
  151 KMAX=J
      GO TO 160
  152 IF(S(KS,KMAX)-S(KS,J)) 153, 160, 160
  153 KMAX=J
  160 CONTINUE
      IF (KMAX) 161, 161, 210
  161 IF (LD) 235, 235, 172
C88   TAKE CARE OF LOOPING OR LINEAR DEPENDENCY
  165 DO 169 I=1, LR
      IF (NP(I)) 169, 167, 169
  167 SUM=0.0
      DO 168 J=1, LC
  168 SUM=SUM+ABS(S(I,J))
      IF (SUM) 169, 170, 169
```

```
  169 CONTINUE
      GO TO 100
  170 WRITE (6, 171) I
  171 FORMAT (88H FORCED TERMINATION DUE TO LINEAR DE-
     1PENDENCE OR INCONSISTENCY OF CONSTRAINTS. CHECK
     2ROW, I5)
      GO TO 138
C89   ERASE THE COUNTER PARTS OF THE POSSIBLE NEW BASES
C90   FROM THE CURRENT BASES.
  172 WRITE (6, 173)
  173 FORMAT (54H I CAN NOT FIND A BASIS AND  WILL  NOW  ALTER
     1ACTIVITIES.)
      KAN=0
      DO 181 I=1, LR
      IF (NP(I)) 181, 174, 181
  174 DO 181 J=1, LC
     IF (S(I,J)) 181, 181, 175
  175 IF (J-LD) 177, 177, 176
  176 IF (J-LR) 181, 181, 178
  177 L=J+LR
      GO TO 179
  178 L=J — LR
  179 DO 181 K=1, LR
      IF (NP(K)-L) 181, 180, 181
  180 KAN=KAN+1
      NP(K)=0
  181 CONTINUE
      IF (KAN) 235, 235, 182
  182 WRITE (6, 183) KAN
  183 FORMAT (10H THERE ARE, I6, 14H BASES ERASED.)
      GO TO 100
C91   FIND PIVOT RATIOS
  210 DO 220 I=1, LR
      IF (S(I,KMAX)) 215, 215, 212
  212 XS(I,I)=Y(I)/S(I,KMAX)
      GO TO 220
  215 XS(I,I)= — 1.0
  220 CONTINUE
C92   FIND THE SMALLEST NONNEGATIVE PIVOT RATIO
      KPIVOT=0
      DO 225 I=1, LR
      IF (XS(I,I)) 225, 221, 221
  221 IF (KPIVOT) 222, 222, 223
  222 KPIVOT=1
      GO TO 225
  223 IF (XS(I,KPIVOT)-XS(I, I)) 225, 225, 224
  224 KPIVOT=1
  225 CONTINUE
      IF (KPIVOT) 235, 235, 248
C93   IF NO NONNEGATIVE PIVOT RATIO, NO FEASIBLE SOLUTION
C94   EXISTS
  235 WRITE (6, 236)
  236 FORMAT (39H0THIS PROBLEM HAS NO FEASIBLE SOLUTION./)
      GO TO 138
C95   RECORD POSSIBLE LOOPING AND PRINT LOCATION OF BASIS
  248 IF (LQP-1) 249, 249, 260
  249 IF (NP(KPIVOT)) 255, 255, 250
  250 LOOP=LOOP+1
      IF (LOOP-9) 260, 260, 251
  251 WRITE (6, 252)
  252 FORMAT (62H I AM CHECKING FOR LOOPING, AND RECOM-
```

```
     1PUTING ALTERNATIVE COSTS.)
      GO TO 165
  255 LOOP=0
  260 NOCYCL=NOCYCL+1
      IF (KEY(5)-2) 275, 265, 262
  262 IF (KEY(5)-5) 275, 265, 275
  265 WRITE (6, 270) NOCYCL, KPIVOT, KMAX,  NP(KPIVOT),  Y(KS)
  270 FORMAT (2H, 4I12, 10X, E14.7)
  275 NP(KPIVOT)=KMAX
C96   ITERATION THROUGH MATRIX AND ALTERNATIVE COSTS
      PIVOT=1.0/S(KPIVOT, KMAX)
      DO 340 J=1, LC
  340 S(KPIVOT,J)=S(KPIVOT,J)*PIVOT
      Y(KPIVOT)=Y(KPIVOT)*PIVOT
      S(KPIVOT,KMAX)=1.0
      DO 360 I=1, KS
      IF (I-KPIVOT) 345, 360, 345
  345 IF (S(I,KMAX)) 346, 360, 346
  346 ELEMT=S(I,KMAX)
      DO 350 J=1, LC
      IF (S(KPIVOT, J) 347, 350, 347
  347 S(I,J)=S(I,J)-ELEMT*S(KPIVOT,J)
      IF (ABS(S(I,J))-TOL) 349, 349, 350
  349 S(I,J)=0.0
  350 CONTINUE
      Y(I)=Y(I) — ELEMT*Y(KPIVOT)
      IF (ABS(Y(I))-TOL) 353, 353, 354
  353 Y(I)=0.0
  354 S(I,KMAX)=0.0
  360 CONTINUE
C97   FIND OUT WHETHER BASES ARE FILLED
      IF (LQP-1) 365, 365, 380
  365 DO 370 I=1, LR
      IF (NP(I)) 131, 131, 370
  370 CONTINUE
      IF (KEY(4)-7) 371, 375, 371
  371 IF (KEY(4)-9) 400, 375, 400
  375 ASSIGN 400 TO KK
      GO TO 132
  380 DO 385 J=1, LC
      IF (S(KS,J)) 385, 385, 131
  385 CONTINUE
C98   CHECKING FOR MULTIPLE SOLUTIONS
      DO 390 J=1, LC
      IF (S(KS,J)) 390, 386, 390
  386 DO 387 I=1, LR
      IF (NP(I)-J) 387, 390, 387
  387 CONTINUE
      DO 388 K=1, LR
      IF (S(K,J)) 388, 388, 394
  388 CONTINUE
  390 CONTINUE
      GO TO 370
  394 WRITE (6, 395) J
  395 FORMAT (69H0THIS PROBLEM HAS MULTIPLE SOLUTIONS. AT
     1LEAST THE VARIABLE OF COLUMN, I4, 24H COULD BE IN THE
     2BASIS./)
      GO TO 138
  400 RETURN
      END
C
```

```
C
      SUBROUTINE TABLE (K, L, LINE)
C99   THIS SUBROUTINE PRINTS SIMPLEX TABLEAU AT ANY STAGE.
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      KS=L
      NE=K
      GO TO (400, 510), LINE
  400 WRITE (6, 500) NOCYCL
  499 FORMAT (10H  Z-C,        9E13.7)
  500 FORMAT (16H0SIMPLEX TABLEAU, I6)
      KS=LR
      IF (LC-8) 501, 501, 502
  501 LS=LC
      GO TO 503
  502 LS=8
  503 WRITE (6, 504) (J, J=1, LS)
  504 FORMAT (18H0ROW BASIS B(I), 8I13)
  505 FORMAT (1H, I3, I5, 1X, 9E13.7)
  506 FORMAT (1H, I3, 6X, 9E13.7)
  507 FORMAT (1H0)
      DO 508 I=1, KS
  508 WRITE (6, 505) I, NP(I), Y(I), (S(I,J), J=1, LS)
      IF (KEY(4)-1) 600, 610, 600
  600 WRITE (6, 499) Y(LR+1), (S(LR+1,J), J=1, LS)
  610 WRITE (6, 507)
      IF (LC-8) 540, 540, 509
  509 NE=9
C
  510 ASSIGN 512 TO KK
      LL=NE
  512 IF ((L-LL)-8) 515, 515, 520
  515 ASSIGN 540 TO KK
      LS=L
      GO TO 525
  520 LS=LL+8
  525 WRITE (6, 526) (J, J=LL,LS)
  526 FORMAT (1H, 5X, 9I13)
      DO 530 I=1, KS
  530 WRITE (6, 506) I, (S(I,J), J=LL, LS)
      IF (LINE-1) 531, 531, 533
  531 IF (KEY(4)-1) 533, 533, 532
  532 WRITE (6, 499) (S(LR+1, J), J=LL, LS)
  533 WRITE (6, 507)
      LL=LL+9
      GO TO KK, (512, 540)
  540 RETURN
      END
C
C
      SUBROUTINE SWPMAT (N, M, TOL, DETERM, KASE)
C
C
C100  THIS SUBROUTINE INVERTS THE MATRIX A AND SOLVES THE
C101  EQUATION AX = B. THE ITERATIVE METHOD SEE B. W. ARDEN,
C102  AN INTRODUCTION TO DIGITAL COMPUTING, PP. 215—223.
C
      DIMENSION A(30, 30), B(30, 2)
```

```
      COMMON A, B
C
      KOVER=KASE
      DETERM=1.0
      DO 60 K=1,N
C103  TEST FOR SIGNIFICANCE OF PIVOT ELEMENT.
      IF (ABS(A(K,K))-TOL) 80, 80, 10
   10 IF (KOVER-1) 19, 15, 19
   15 DETERM=DETERM*A(K, K)
   19 PIVOT=1.0/A(K,K)
C104  SCALE KTH ROW.
      DO 20 J=1, N
   20 A(K,J)=A(K,J)*PIVOT
      DO 25 J=1, M
   25 B(K,J)=B(K,J)*PIVOT
      A(K,K)=PIVOT
C105  REDUCE ALL ROWS BUT KTH ROW, ALL ELEMENTS.
      DO 50 I=1, N
      IF (I-K) 30, 50, 30
C106  AUTOMATICALLY REPLACES A(I,K) WITH -A(I,K)*A(K,K)
   30 ELEMT=A(I,K)
      A(I,K)=0.0
      DO 40 J=1, N
   40 A(I,J)=A(I,J) - ELEMT*A(K,J)
      DO 45 J=1, M
   45 B(I,J)=B(I,J) - ELEMT*B(K,J)
   50 CONTINUE
   60 CONTINUE
   99 RETURN
   80 WRITE (6, 90) K
   90 FORMAT (31H0THE MATRIX IS SINGULAR, ON K=, I8/)
      DETERM=0.0
      GO TO 99
      END
C
C
      SUBROUTINE PRINT (K, L, M, N)
C107  THIS SUBROUTINE PRINTS MATRIX A WITH INDICES.
      DIMENSION A(30, 30)
      COMMON A
C
   05 FORMAT (1H0)
   20 FORMAT (1H , 9I14)
   30 FORMAT (1H, I3, 2X, 9E14.7)
      INT=9
      INT1=INT-1
   50 ASSIGN 60 TO KK
      LL=L
   60 IF ((N-LL)-INT1) 70, 70, 75
   70 ASSIGN 180 TO KK
      LS=N
      GO TO 80
   75 LS=LL+INT1
   80 WRITE (6, 20) (J, J=LL, LS)
      DO 90 I=K, M
   90 WRITE (6, 30) I, (A(I,J), J=LL, LS)
      WRITE (6, 5)
      LL=LL+INT
      GO TO KK, (60, 180)
  180 RETURN
      END
```

```
C
C
C108 THIS SUBROUTINE PREDICTS THE FUTURE OUTCOMES AND
C109 EVALUATES THE SQUARED ERROR AS THE VALUE OF THE
C110 LOSS FUNCTION AND ALSO COMPUTES THE CHI-SQUARE
C111 VALUE FOR THE TEST OF THE GOODNESS-OF-FIT.
C
      SUBROUTINE PREDIC (KN)
C
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      DO 20 I=1, NT1
      L=0
      DO 20 J=1, NS
      A(I+1, J)=0.0
      DO 20 K=1, NS
      L=L+1
   20 A(I+1, J)=A(I+1, J)+E(I, K)*B(L, 1)
      IF (KN-10) 24, 60, 24
   24 WRITE (6, 25)
   25 FORMAT (26H1THE PREDICTED PROPORTIONS/)
      NT2=NT1+1
      CALL PRINT (2, 1, NT2, NS)
      SSQ=0.0
      CHI=0.0
      DO 40 J=1, NS
      DO 40 I=2, NT1
      IF (A(I,J)) 38, 35, 38
   35 IF (E(I,J)) 39, 40, 39
   38 CHI=CHI+(E(I, NS+1)*(A(I,J) - E(I,J))**2)/A(I,J)
      GO TO 40
   39 CHI=CHI+(E(I, NS+1)*(A(I,J) - E(I,J))**2)/E(I,J)
   40 SSQ=SSQ+(A(I,J) - E(I,J))**2
      TN=NT*NS
      ASSQ=SSQ/TN
      CHIR=0.0
      DO 45 J=1, NS
      DO 45 I=2, NT1
      IF (E(I,J)) 43, 42, 43
   42 IF (A(I,J)) 44, 45, 44
   43 CHIR=CHIR+(E(I, NS+1)*(A(I,J) - E(I,J))**2)/E(I,J)
      GO TO 45
   44 CHIR=CHIR+(E(I, NS+1)*(A(I,J) - E(I,J))**2)/A(I,J)
   45 CONTINUE
      WRITE (6, 50) SSQ, ASSQ, CHI, CHIR
   50 FORMAT (91H0 SUM OF SQUARED ERROR MEAN SQUARED
     1ERROR CHI SQUARE VALUE MODIFIED CHI SQUARE/7X, 4F20.9)
   60 RETURN
      END
C
C
      SUBROUTINE SMRY (MARY, KT)
C112 THIS SUBROUTINE SUMMARIZES THE ESTIMATES FOR THE
C113 SAMPLING EXPERIMENT ONLY.
      DIMENSION A(30, 30), B(30, 2), C(30, 30), D(30), E(30, 30), X(96, 30),
     1Y(96), S(96, 96), XS(30, 96), KEY(9), NP(96), AA(2), SUM(36, 2), SUMSQ
     2(36, 2)
      COMMON A, B, C, D, E, X, Y, XS, S, SS, TOL, KDROP, KW, KV, KP,
```

```
     1KEY, LD, LR, LC, LX, MM, NN, NP, NS, NS1, NT, NT1, NOCYCL
C
      NNS=NS*NS
      GO TO (10, 25, 35), MARY
   10 DO 20 J=1, 2
      AA(J)=0.0
      DO 20 I=1, NNS
      SUM(I,J)=0.0
   20 SUMSQ(I,J)=0.0
      GO TO 100
   25 DO 30 I=1, NNS
      SUM(I,KT)=SUM(I,KT)+B(I,1)
   30 SUMSQ(I,KT)=SUMSQ(I,KT)+B(I,1)**2
      AA(KT)=AA(KT)+1.0
      WRITE (7, 31) (S(I, 1), I=1, NNS)
   31 FORMAT (8X, 8F8.4)
      GO TO 100
   35 IF (KT-1) 37, 37, 39
   37 WRITE (6, 38)
   38 FORMAT (56H1THIS IS A SUMMARY SHEET FOR THE UNRE-
     1STRICTED ESTIMATOR./)
      GO TO 41
   39 WRITE (6, 40)
   40 FORMAT (54H1THIS IS A SUMMARY SHEET FOR THE RE-
     1STRICTED ESTIMATOR./)
   41 DO 42 I=1, NNS
   42 B(I,1)=SUM(I,KT)/AA(KT)
      WRITE (6, 44) AA(KT)
   44 FORMAT (14H0SAMPLE SIZE=F4.0)
      WRITE (6, 45)
   45 FORMAT (6H0MEANS/)
      CALL GLSQP (4)
      DO 47 I=1, NNS
   47 B(I,1)=SUMSQ(I,KT)/AA(KT)-B(I,1)**2
      WRITE (6, 48)
   48 FORMAT (10H0VARIANCES/)
      CALL GLSQP(4)
      DO 49 I=1, NNS
   49 B(I,1)=SQRT(B(I,1))
      WRITE (6, 50)
   50 FORMAT (20H0STANDARD DEVIATIONS/)
      CALL GLSQP(4)
   55 READ (5, 56) (B(I,1), I=1, 36)
   56 FORMAT (9F8.4/9F8.4/9F8.4/9F8.4)
      WRITE (6, 57)
   57 FORMAT (12H0TRUE VALUES/)
      CALL GLSQP(4)
      DO 70 I=1, NNS
   60 Y(I)=SUMSQ(I,KT)-2.0*B(I, 1)*SUM(I, KT)+AA(KT)*B(I,1)**2
   70 B(I, 1)=SQRT(Y(I)/AA(KT))
      WRITE (6, 75)
   75 FORMAT (41H0ROOT MEAN SQUARE ERRORS FROM TRUE
     1VALUES/)
   80 CALL GLSQP(4)
      S1=0.0
      DO 85 I=1, NNS
   85 S1=S1+B(I,1)
      WRITE (6, 89) S1
   89 FORMAT (59H0SUM OF THE ELEMENTS OF THE ROOT MEAN
     1SQUARE ERROR MATRIX=, F15.8)
  100 RETURN
      END
```

Перевод комментариев, имеющихся в тексте программы

C1 — Ввод выборочной информации

C2 — Преобразование N в $N(I)$

C3 — Решение с помощью метода линейного программирования

C4 — Присвоение веса и решение задачи оценивания

C5 — Ввод параметров априорного бета-распределения или априорного нормального распределения

C6 — Итеративное вычисление оценки максимального правдоподобия и байесовской оценки

C7 — Внесение исправлений, засылка 0.00001 в элементы $P(I,J)=0.0$

C8—C10 — Подпрограмма объединяет априорную информацию (одномерное бета-распределение, многомерное бета-распределение или нормальное распределение) с выборочной информацией с целью вычисления апостериорных оценок

C11—C12 — Формирование плотности вероятности многомерного априорного бета-распределения из одномерного априорного бета-распределения

C13 — Предположение о нормальности априорного распределения

C14 — Вычисление апостериорного распределения

C15—C18 — Подпрограмма осуществляет исключение столбца, выбор веса и преобразование исходной задачи к многомерной задаче с невырожденной весовой матрицей

C19 — Требуется исключить столбец

C20 — Выбор веса

C21—C23 — Для уменьшения масштаба величин, участвующих в вычислениях, весовая матрица будет разделена на заданное число

C24—C25 — Подпрограмма формирует матрицы S, X и Y для многомерной задачи

C26 — Без учета взвешивания

C27—C28 — Гомоскедастичность уравнений, но гетероскедастичность между различными уравнениями

C29—C30 — Введение гетероскедастичности путем взвешивания на величины, обратные к средним относительным частотам

C31—C33 — Введение гетероскедастичности путем взвешивания на произведение средней относительной частоты состояния I и средней относительной частоты всех других состояний

C34—C35 — Введение гетероскедастичности диагональных элементов посредством $N/WJ(T)$ или $N/(WJ(T)*(1-WJ(T))$

C36—C37 — Введение гетероскедастичности путем взвешивания на матрицу, обратную к матрице дисперсий ошибок I-го уравнения

C38—C39 — Вес, вычисленный по функции правдоподобия

C40 — J определяется через L и I минус произведение NT

C41 — K определяется через I минус произведение NT

C42 — Проверка того, является ли найденный элемент диагональным

C43—C44 — Запоминание соответствующей величины. Если A равно нулю, то A заменяется на $0.1/SS$

C45 — Реализация выбранного варианта печати

C46 — Формирование кронеккеровского произведения

C47—C48 — Обобщенный метод наименьших квадратов и квадратичное программирование

C49 — Печать P и проверка P на отрицательность

C50—C51 — Если P не принадлежит интервалу [0, 1], то необходимо решать задачу методом квадратичного программирования

C52 — Выдача решения задачи квадратичного программирования в матричной форме

C53—C56 — Подпрограмма формирует симплекс-таблицу для оценки параметров совокупности регрессионных гиперплоскостей по критерию наименьших абсолютных отклонений с учетом и без учета взвешивания методом линейного программирования

C57 — Формирование симплекс-таблицы для задачи линейного программирования

C58—C59 — Формирование вектора ошибок в виде дополнительной переменной и включение этого вектора в базис

C60 — Формирование ограничений нормировки по строке

C61 — Первый этап итерации

C62 — Второй этап итерации

C63 — Печать матрицы оценок переходных вероятностей

C64—C66 — Подпрограмма формирует симплекс-таблицы для задачи квадратичного программирования с ограничениями типа равенств и неравенств

C67 — Вычисление констант

C68—C69 — Запоминание значений элементов столбца P_0 или вектора правых частей

C70 — Очистка массива цен правых частей

C71—C72 — Формирование матриц G' и G (а также матрицы G' для случая ограничений типа равенств)

C73 — Формирование бета-матрицы

C74 — Формирование дополнительных переменных $W(I)$

C75 — Задание начальных условий

C76—C77 — Введение всех элементов исключенного столбца матрицы P в базис

C78—C79 — Повторное вычисление констант так, чтобы $P(I)$ и $W(I)$ стали взаимодополняющими переменными

C80 — Выбор печати начальной симплекс-таблицы

C81 — Выдача результатов

C82—C83 — Подпрограмма по симплекс-таблице осуществляет решение задач линейного и квадратичного программирования методом Ц. Ли

C84 — Вычисление различных цен

C85 — В случае задач малой размерности выдача промежуточных результатов на каждом итерационном шаге

C86 — Проверка на продолжение счета или останов

C87 — Отыскание базиса среди ненулевых столбцов

C88 — Принятие мер в случае зацикливания или линейной зависимости

C89—C90 — Исключение из текущего базиса дополнительных переменных любого из возможных новых базисов

C91 — Отыскание разрешающих коэффициентов

C92 — Отыскание наименьшего неотрицательного разрешающего коэффициента

C93—C94 — Если отсутствует неотрицательный разрешающий коэффициент, то нет допустимого решения

C95 — Проверка на зацикливание и печать базиса

C96 — Итерация по матрице и различным ценам

C97 — Установление факта заполнения базисов

C98 — Проверка наличия нескольких решений

C99 — Подпрограмма печатает симплекс-таблицу на любом шаге

C100—C102 — Подпрограмма обращает матрицу A и решает уравнение $AX=B$ итеративным методом Ардена (ARDEN B. W. An Introduction to digital Computing, Addison, Wefley, 1963, с. 215—223)

C103 — Проверка значимости разрешающего элемента

C104 — Пересчет K-й строки

C105 — Редуцирование всех строк, кроме K-й строки, которая исключается целиком

C106 — Автоматическая замена $A(I, K)$ на $A(I, K)*A(K, K)$

C107 — Подпрограмма печатает матрицу с индексами

C108—C111 — Подпрограмма прогнозирует будущие исходы, оценивает квадратическую ошибку как величину функции потерь и вычисляет также значение критерия согласия χ^2

C112—C113 — Подпрограмма суммирования оценок предназначена только для выборочного эксперимента, описанного в книге

ЛИТЕРАТУРА

Далее приводится список литературы, который не является исчерпывающим, но включает широкий перечень работ, посвященных рассмотренным в книге методам. Часть этих работ в тексте книги не упоминается и приводится в списке литературы для того, чтобы более полно охватить затронутый круг вопросов.

1. Adelman I. G. 1958. A Stochastic Analysis of the Size Distribution of Farms. *Journal of the American Statistical Association*, *53*, 893—904.
2. Aitken A. C. 1934—35. On Least Squares and Linear Combination of Observations. *Proceedings of the Royal Society of Edinburgh*, *55*, 42—48.
3. Anderson T. W. 1955. Probability Models for Analyzing Time Changes and Attitudes in: P. F. Lazarsfeld, ed. *Mathematical Thinking in the Social Sciences* (The Free Press, Glencoe), p. 17—66.
4. Anderson T. W. and Goodman L. A. 1957. Statistical Inference About Markov Chains. *The Annals of Mathematical Statistics*, *28*, *89—110*.
5. Arrow K. J. 1967. Applications of Control Theory to Economic Growth. Technical Report. No 2. Institute of Mathematical Studies, Stanford.
6. Ashar V. G. and Wallace T. D. 1963. A Sampling Study of Minimum Absolute Deviations Estimators. *Operations Research*, *11*, *747—758*.
7. Bailey N. T. J. 1964. The Elements of Stochastic Processes with Applications to the Natural Sciences (John Wiley and Sons, New York).
8. Barten A. P. and Kloek T. 1965. A Generalization of Generalized Least Squares. Econometric Institute Report 6508, Netherlands School of Economics.
9. Bartlett M. S. 1951. The Frequency Goodness-of-Fit Test for Probability Chains. *Proceedings of the Cambridge Philosophical Society*, *47*, 86—95.
10. Bartlett M. S. 1955. An Introduction to Stochastic Processes (Cambridge University Press, Cambridge), p. 24—26, 240—242.
Русский перевод: Бартлет М. С. Введение в теорию случайных процессов. М., ИЛ, 1958.
11. Beale E. M. L. 1955. On Minimizing a Convex Function Subject to Linear Inequalities. *Journal of the Royal Statistical Society* (B), *17*, 173—184.
12. Blackwell D. and Girschick M. A. 1954 Theory of Games and Statistical Decision (John Wiley and Sons, New York), p. 143—178.
Русский перевод: Блекуэлл Д., Гиршик М. А. Теория игр и статистических решений. М., ИЛ, 1958.
13. Blumen I., Kogan M. and McCarthy P. J. 1955. The Industrial Mobility of Labor as a Probability Process (Cornell University Press, Ithaca).
14. Boot J. C. G. 1963. The Computation of the Generalized Inverse of Singular or Rectangular Matrices. *American Mathematical Monthly*, *70*, 302—303.
15. Bush R. R. and Mosteller F. 1955. Stochastic Models for Learning (John Wiley and Sons, New York). Chapters 3 and 5.
Русский перевод: Буш Р. Р., Мостеллер Ф. Стохастические модели обучаемости. М., Физматгиз, 1962.
16. Champernowne D. B. 1953. A Model of Income Distribution. *Economic Journal*, *63*, 318—351.
17. Charnes A. and Cooper W. W. 1961. *Management Models and Industrial Applications of Linear Programming*. Vol. 1 (John Wiley and Sons, New York), p. 389 — 393.

18. Chipman J. C. 1964. On Least Squares with Insufficient Observations. *Journal of the American Statistical Association 59*, 1078—1111.
19. Cline R. E. 1964. Note on the Generalized Inverse of the Product of Matrices. *SIAM Review, 6*, 57—58.
20 Coleman J. S. 1964. Introduction to Mathematical Sociology (Collier-Macmillan, London).
21. Collins N. R. and Preston L. E. 1961. The Size Structure of the Largest Industrial Farms, 1900—1958. *American Economic Review, 51*, 986—1011.
22. Cootner P. H. 1964. The Random Character of Stock Market Prices (Massachusetts Institute of Technology Press, Cambridge).
23. Cornish E. A. 1954. The Multivariate *t*-Distribution Associated with a Set of Normal Sample Deviates. *Australian Journal of Physics, 7*, 531—542.
24. Cox D. R. and Miller H. D. 1965. The Theory of Stochastic Processes (Methuen, London).
25. Cragg J. G. 1964. Small-Sample Properties of Various Simultaneous Equation Estimators: The Results of Some Monte Carlo Experiments. Econometric Research Program Research Memorandum. No 68, Princeton University.
26. Cramer H. 1946. Mathematical Methods of Statistics (Princeton University Press, Princeton), p. 247—248.
Русский перевод: Крамер Г. Методы математической статистики. М., ГИНЛ, 1948.
27. Dantzig G. B. and Orden A. 1953. Duality Theorems. *RAND Report R. M. — 1265*. Santa Monica, California. The RAND Corporation.
28. Dean G. W., Johnson S. S. and Carter H. O. 1963. Supply Functions for Cotton in Imperical Valley, California. *Agricultural Economics Research, 15*, 1—14.
29. Deutsch R. 1965. Estimation Theory (Prentice-Hall, New York).
30. Dorn W. S. 1960. Duality in Quadratic Programming. *Quarterly of Applied Mathematics, 18*. 155—162.
31. Dryden M. M. 1967. Share Price Movements: A Markovian Approach, unpublished paper, Department of Economics, Edinburgh.
32. Dunnett C. S. and Sobel M. 1954. A Bivariate Generalization of Student's *t*-Distribution, with Tables for Certain Special Cases. *Biometrika, 41*, 153—169.
33. Durbin J. 1953. A Note on Regression When There is Extraneous Information About One of the Coefficients. *Journal of the American Statistical Association, 48*, 700—708.
34. Edwards W., Lindman H. and Savage L. J. 1963. Bayesian Statistical Inference for Psychological Research. *Psychological Review, 70*, 193—242.
35. Farris P. L. and Padberg S. I. 1964. Measures of Market Structure Change in the Florida Fresh Citrus Packing Industry. *Agricultural Economics Research, 16*, 93—102.
36. Ferguson T. S. 1967. Mathematical Statistics: A Decision Theoretic Approach (Academic Press, New York).
37. Fisher W. D. 1961. A Note on Curve Fitting with Minimum Deviations by Linear Programming. *Journal of the American Statistical Association, 56*, 359—363.
38. Frank M. and Wolfe P. 1956. An Algorithm for Quadratic Programming. *Naval Research Logistic Quarterly, 3*, 97—98.
39. Gale D. 1960. The Theory of Linear Economic Models (McGraw-Hill, New York), p. 260 — 290.
Русский перевод: Гейл Д. Теория линейных экономических моделей. М., ИЛ, 1963.
40. Goldberger A. S. 1964. Econometric Theory (John Wiley and Sons, New York), p. 246—248.
41. Goodman L. A. 1953. A Further Note on Miller's' Finite Markov Processes in Psychology'. *Psychometrika, 18*, 245—248.
42. Goodman L. A. 1965. On Statistical Analysis of Mobility Tables. *The American Journal of Sociology, 70*, 564—585.

43. H a d l e y G. 1964. Nonlinear and Dynamic Programming (Addison-Wesley Publishing Company, Boston), p. 212—250.
Русский перевод: Х е д л и Дж. Нелинейное и динамическое программирование. М., «Советское радио», 1967.
44. H a r t P. E. and P r a i s S. J. 1956. The Analysis of Business Concentration, A Statistical Approach. *Journal of the Royal Statistical Society* (A), *119*, 150—175.
45. H a r t l e y H. O. 1963. The Application of Mathematical Programming to Statistical Least Squares Estimation with Constraints, unpublished mimeographed report, University of Texas.
46. H a w k i n s D. and S i m o n H. A. 1949. Note: Some Conditions on Macro Economic Stability, *Econometrica*, *17*, 245—248.
47. H e a d y E. O. and C a n d l e r W. 1963. Linear Programming Methods (The Iowa State University Press, Ames), p. 98—99, 131, 136.
48. H o e k i n g R. R. 1963. The Distribution of a Projected Least Squares Estimator. *Annals of the Institute of Statistical Mathematics*, *17*, 357—362.
49. H o e l P. E. 1954. A Test for Markov Chains. *Biometrika*, *41*, 430—433.
50. H o g g R. and C r a g g A. 1965. Introduction to Mathematical Statistics. 2nd Edition (Macmillan, New York).
51. H o w a r d R. A. 1960. Dynamic Programming and Markov Processes (John Wiley and Sons, New York).
Русский перевод: Х о в а р д Р. А. Динамическое программирование и марковские процессы решения. М., «Советское радио», 1964.
52. J e f f r e y s H. 1961 Theory of Probability. 3rd Edition (Clarendon Press, Oxford).
53. J o h n s t o n J. 1963. Econometric Methods (McGraw-Hill, New York).
54. J u d g e G. G. and S w a n s o n E. R. 1962. Markov Chains: Basic Concepts and Suggested Uses in Agricultural Economics. *Australian Journal of Agricultural Economics*, *6*, 49—61.
55. J u d g e G. G. and T a k a y a m a T. 1966. Inequality Restrictions in Regression Analysis. *Journal of the American Statistical Association*, *61*, 166—181.
56. K a o R. C. W. 1953. Note on Miller's Finite Markov Processes in Psychology'. *Psychometrika*, *18*, 241—243.
57. K a r l i n S. 1959. Mathematical Methods and Theory in Games, Programming and Economics (Addison — Wesley, London), p. 241—243.
Русский перевод: К а р л и н С. Математические методы и теории игр, программирование и экономике. М., «Мир», 1964.
58. K a r s t O. J. 1958. Linear Curve Fitting Using Least Deviations. *Journal of the American Statistical Association*, *53*, 118—132.
59. K e m e n y J. G., M i r k i l H., S n e l l J. L. and T h o m p s o n G. L. 1959. Finite Mathematical Structures (Prentice-Hall, New York), p. 348—438.
60. K e m e n y J. G. and S n e l l J. L. 1960. *Finite Markov Chains* (De van Nostrand Co., Inc., Princeton), p. 200—206.
Русский перевод: К е м е н и Дж., С н е л л Дж. Конечные цепи Маркова. М., «Наука», 1970.
61. K e n d a l l M. G. and S t u a r t A. 1958. The Advanced Theory of Statistics. Vol. 1 (Charles Griffin and Co., Ltd., London).
Русский перевод: К е н д а л л М. Дж., С т ь ю а р т А. Теория распределений. М., «Наука», 1966.
62. K e n d a l l M. G. and S t u a r t A. 1961. The Advanced Theory of Statistics. Vol. 2 (Charles Griffin and Co., Ltd., London).
Русский перевод: К е н д а л л М. Дж., С т ь ю а р т А. Статистические выводы и связи. М., «Наука», 1973.
63. K i s l e y Y. and A m a i d A. 1968. Linear and Dynamic Programming in Markov Chains. *American Journal of Agricultural Economics*, *50*, 111—129.
64. K o t t k e M. W. 1964. Patterns of Dairy Farm Exit and Growth. Connecticut Agricultural Experiment Station Bulletin No. 382.
65. K r e n z R. D. 1964. Projection of Farm Numbers for North Dakota with Markov Chains. *Agricultural Economics Research*, *16*, 77—83.
66. K u h n H. and T u c k e r A. 1951. Non Linear Programming in: J. Neyman,

ed. *Proceedings of the Second Berkeley Symposium* (The University of California Press, Berkeley), p. 481—492.
67. Lee T. C., Judge G. G. and Cain R. L. 1969. A Sampling Study of the Properties of Estimators of Tiansition Probabilities. *Management Science* (A), *15*, 374 — 398.
68. Lee T. C., Judge G. G. and Takayama T. 1965. On Estimating the Transition Probabilities of a Markov Process. *Journal of Farm Economics*, *47*, 742—762.
69. Lee T. C., Judge G. G. and Zellner A. 1968. Maximum Likelihood and Bayesian Estimation of Transition Probabilities. *Journal of the American Statistical Association*, *63*, 1162—1179.
70. Lindgren B. W. 1962. Statistical Theory (MacMillan Company, New York), p. 300—304.
71. Lindley D. V. 1965. Introduction to Probability and Statistics. Parts I and II (Cambridge University Press, Cambridge).
72. Madansky A. 1959. Least Squares Estimation in Finite Markov Processes. *Psychometrika*, *24*, 137—144.
73. Marschak J. 1950. Statistical Inference in Economics in: T. C. Koopmans, ed. *Studies in Econometric Method* (John Wiley and Sons, New York).
74. Marschak J. 1953. Economic Measurements for Policy and Prediction in: W. C. Hood and T. C. Koopmans eds. *Studies in Econometric Method* (John Wiley and Sons, New York), p. 10—26.
75. Martin J. J. 1967. Bayesian Decision Problems and Markov Chains (John Wiley and Sons, New York).
76. Matras J. 1960. Differential Fertility, Intergenerational Occupational Mobility and Change in Occupational Distribution, *Population Studies*, *15*, 187—197.
77. Mauldon J. G. 1961. A Generalization of the Beta-Distribution. *Annals of Mathematical Statistics*, *32*, 509—520.
78. Meyer J. and Glauber R. G. 1964. Investment Decision, Economic Forecasting and Public Policy (Harvard University, Boston).
79. Miller G. A. 1952. Finite Markov Processes in Psychology. *Psychometrika*, *17*, 149—167.
80. Moses L. E. and Oakford R. V. 1963. Tables of Random Permutations (Stanford University Press, Stanford), p. 67—77, 190—229.
81. Neudecker H. 1966. A Critique of the Charnes-Cooper Procedure of Local Aggregation in Input — Output Analysis. Discussion paper 74, University of Birmingham.
82. Neyman J. 1949. Contribution to the Theory of the χ^2 Test. *The First Proceedings of Berkeley Symposium in Math. Stat. and Prob.*, 239 (University of California Press, Berkeley).
83. Nicholls W. H. 1951. Price Policies in the Cigarette Industry (The Vanderbilt University Press, Nashville).
84. Odell P. L. and Lewis T. O. 1966. A Generalization of the Gauss — Markov Theorem. *Journal of the American Statistical Association*, *61*, 1063—1066.
85. Padberg D. I. 1962. The Use of Markov Processes in Measuring Changes in Market Structure. *Journal of Farm Economics*, *44*, 189—199.
86. Panne van de and Whinston A. 1963. The Simplex and the Dual Method for Quadratic Programming. Research Report, International Center for Management Science, Rotterdam.
87. Penrose R. 1955. A Generalized Inverse for Matrices. *Proceedings of the Cambridge Philosophical Society*, *51*, 406—413.
88. Prais S. J. 1955. Measuring Social Mobility. *Journal of the Royal Statistical Society* (A), *118*, 56—66.
89. Preston L. E. and Bell E. J. 1961. The Statistical Analysis of Industry Structure: An Application to the Food Industry. *Journal of the American Statistical Association*, *56*, 925—932.
90. Price C. M. 1964. The Matrix Pseudo Inverse and Minimal Variance Estimates, *SIAM Review*, *6*, 115—120.
91. Rado R. 1956. Note on Generalized Inverse of Matrices. *Proceedings of the Cambridge Philosophical Society*, *52*, 600—601.

92. R a o C. R. 1962. A. Note on the Generalized Inverse of a Matrix with Applications to Problems in Mathematical Statistics. *Journal of the Royal Statistical Society* (B), *24*, 152—158.
93. R e i t z H. L. et al. 1934. Handbook of Mathematical Statistics (Houghton Mifflin Co., Boston).
94. R e y G. 1964. A Markov Chain Prediction of Value Added and Expenditure Shares, Italy, 1861—1956. Research Report, International Center for Management Science, Rotterdam.
95. R o b e r t s H. V. 1963. Bayesian Inference. *American Statistical Association, Proceedings of the Social Statistical Section*, p. 76—80.
96. S a a t y T. L. and B r a m J. 1964. Nonlinear Mathematics. (McGraw-Hill Book Company), p. 113—133.
97. S a m u e l s o n P. A. 1947. Foundations of Economic Analysis (Harvard University Press, Cambridge), p. 348—349.
98. S a v a g e L. J. 1962. Bayesian Statistics in: *Recent Development in Decision and Information Processes* (Macmillan and Co., New York), p. 161—194.
99. S h u p p F. 1968. Stabilization Policies in Non-linear Stochastic Macro Economic Systems. Unpublished paper, University of Illinois.
100. S i e g e l S. 1956. Nonparametric Statistics for the Behavior Sciences (McGraw-Hill Book Company, New York), p. 68—83, 229—238.
101. S m i t h P. E. 1961. Markov Chains, Exchange Matrices and Regional Development. *Journal of Regional Science*, *3*, 27—36.
102. S o l o w R. 1951. Some Long-Run Aspects of the Distribution of Wage Incomes. *Econometrica*, *19*, 333—334.
103. S p a r k s W. R. 1960. On Markov Chains in Demand Analysis. Unpublished paper, Michigan State University.
104. S t e i n d l J. 1965. Random Processes and the Growth of Firms (Hafner Publishing Company, New York).
105. S u m m e r s R. 1965. A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators. *Econometrica*, *35*, 1—4 .
106. Taiwan Provincial Government, Department of Agriculture and Forestry, *Taiwan Agricultural Yearbook*, 1941—1967.
107. T a k a y a m a T., J u d g e G. G. and L e e T. C. 1969. An Additional Note on Miller's'Finite Markov Processes in Psychology'. Workshop Paper, University of Illinois.
108. T e l s e r L. G. 1962a. The Demand for Branded Goods as Estimated from Consumer Panel Data. *Review of Economic Statistics*, *44*, 300—324.
109. T e l s e r L. G. 1962b. Advertising and Cigarettes. *Journal of Political Economy*, *70*, 471—499.
110. T e l s e r L. G. 1963. Least Squares Estimates of Transition Probabilities in: *Measurement of Economics* (Stanford University Press, Stanford).
111. T h e i l H. 1961. Economic Forecasts and Policy. 2nd Edition (North-Holland Publishing Company, Amsterdam), p. 331—333.

Русский перевод: Т э й л Г. Экономические прогнозы и принятие решений. М., «Статистика», 197 .

112. T h e i l H. 1963. On the Use of Incomplete Prior Information in Regression Analysis. *Journal of the American Statistical Association*, *58*, 401—414.
113. T h e i l H. and G o l d b e r g e r A. S. 1961. On Pure and Mixed Statistical Estimation in Economics. *International Economic Review*, *2*, 65—78.
114. T h e i l H. and R e y G u i d o. 1966. A Quadratic Programming Approach to the Estimation of Transition Probabilities. *Management Science*, *12*, 714—721.
115. T h o r n b e r H. 1967. Finite Sample Monte Carlo Studies; An Auto-Regressive Illustration. *Journalof the American Statistical Association*, *62*, 801—818.
116. T i a o G. C. and Z e l l n e r A. 1964. On the Bayesian Estimation of Multivariate Regression. *The Journal of the Royal Statistical Society* (B), *26*, 277—285.
117. T i a o G. C. and Z e l l n e r A. 1965. Bayes' Theorem and the Use of Prior Knowledge in Regression Analysis. *Biometrika*, *51*, 219—230.

118. Wagner H. H. 1959. Linear Programming for Regression Analysis. *Journal of the American Statistical Association*, *54*, 206—212.
119. Wolfe P. 1959. The Simplex Method for Quadratic Programming. *Econometrica*, *27*, 382—398.
120. Zellner A. 1961. Linear Regression With Inequality Constraints on the Coefficients. Report 6109, Econometric Institute, Rotterdam.
121. Zellner A. 1962. An Efficient Method of Estimating Seemingly Unrelated Regression and Tests for Aggregation Bias. *Journal of the American Statistical Association*, *57*, 348—368.
122. Zellner A. 1969. Introduction to Bayesian Inference in Econometrics (University of Chicago, Chicago). Manuscript.
123. Zellner A. and Lee T. H. 1965. Joint Estimation of Relationships Involving Discrete Random Variables. *Econometrica*, *33*, 382—394.
124. Zellner A. and Tiao G. C. 1964. Bayesian Analysis of the Regression Model with Autocorrelated Errors. *Journal of the American Statistical Association*, *59*, 763—778.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1*. Аркин В. И., Пресман Э. Л., Сонин И. М. Некоторые задачи оптимального выбора в условиях неопределенности. Тезисы докладов. *Первая конференция по оптимальному планированию и управлению народным хозяйством*. Секция 3. М., АН СССР, ЦЭМИ, 1971.
2*. Аркин В. И., Пресман Э. Л., Сонин И. М. Модели экономической динамики с учетом управляемого научно-технического прогресса. *Рефераты докладов VI Всесоюзного совещания по проблемам управления*. М., «Наука», 1974.
3*. Берг А. И., Черняк Ю. И. Информация и управление. М., «Экономика», 1968.
4*. Бусленко Н. П. Моделирование сложных систем. М., «Наука», 1968.
5*. Бусленко Н. П., Калашников В. В., Коваленко И. П. Лекции по теории сложных систем. М., «Советское радио», 1973.
6*. Вентцель Е. С. Теория вероятностей. М., «Наука», 1964.
7*. Вентцель Е. С., Овчаров Л. А. Теория вероятностей. М., «Наука», 1969.
8*. Гнеденко Б. В., Беляев Ю. К., Соболев А. Д. Математические методы в теории надежности. М., «Наука», 1965.
9*. Гнеденко Б. В. Курс теории вероятностей. М., «Наука», 1965.
10*. Дынкин Е. Б. Марковские процессы. М., Физматгиз, 1963.
11*. Закс Ш. Теория статистических выводов. М., «Мир», 1975.
12*. Кобринский Н. Е. Основы экономической кибернетики. М., «Экономика», 1969.
13*. Колмогоров А. Н. Цепи Маркова со счетным числом возможных состояний. Бюллетень МГУ, 1, 1973, № 3.
14*. Канторович Л. В. Экономический расчет наилучшего использования ресурсов. М., АН СССР, 1963.
15*. Ланге О. Введение в экономическую кибернетику. М., «Прогресс», 1968.
16*. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки данных. М., Физматгиз, 1962.
17*. Липтцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. М., «Наука», 1974.
18*. Математика и кибернетика в экономике. Словарь-справочник. М., «Экономика», 1975.
19*. Немчинов В. С. Экономико-математические методы и модели. М., «Мысль», 1965.
20*. Острем К. Введение в стохастическую теорию управления. М., «Мир», 1973.
21*. Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. М., «Наука», 1967.

22*. Пугачев В. С. Теория случайных функций и ее применения к задачам автоматического управления. М., Физматгиз, 1962.
23*. Пугачев В. С. Введение в теорию вероятностей. М., «Наука», 1968.
24*. Райбман Н. С. Что такое идентификация. М., «Наука», 1970.
25*. Розанов Ю. А. Случайные процессы. М., «Наука», 1971.
26*. Романовский В. И. Дискретные цепи Маркова. Л., Гостехиздат, 1949.
27*. Рыжиков Ю. И. Управление запасами. М., «Наука», 1969.
28*. Сарымсаков Т. А. Основы теории процессов Маркова. М., Гостехиздат, 1954.
29*. Сейдж Э. П., Мелса Д. Л. Идентификация систем управления. М., «Наука», 1974.
30*. Сираждинов С. Х. Предельные теоремы для однородных цепей Маркова. Ташкент, АН УзССР, 1955.
31*. Френкель А. А. Математические методы анализа динамики и прогнозирования производительности труда. М., «Экономика», 1972.
32*. Хедли Дж., Уайтин Т. Анализ систем управления запасами. М., «Наука», 1969.
33*. Эйкхофф П. Основы идентификации систем управления. М., «Мир», 1975.
34*. Юдин Д. Б., Гольштейн Е. Г. Линейное программирование. М., «Наука», 1969.
35*. Fisher F. M. The Identification in Econometrics. McGraw-Hill, New York, 1966.
36*. Forrester J. W. World Dynamics. Wright-Allen Press, Cambridge, 1971.
37*. Johnston J. Econometric Methods. Mc Graw-Hill, New York, 1972.
38*. Naylor T. H. Computer Simulation Experiments with Models of Economic Systems. Wiley, New York, 1971.
39*. Young P., Noughton J., Neethling C., Shellswell S. Macro-economic modelling; a case study. *Proceedings of the 3-rd IFAC Symposium on Identification and System Parameter Estimation*. Part 1, North-Holland Publishing Company, Amsterdam, 1973.
40*. Лексис В. О теории стабильности статистических рядов. — В сб.: О теории дисперсии. М., «Статистика», 1968.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

УКАЗАТЕЛЬ ОБОЗНАЧЕНИЙ

a_{ij} — параметры априорного распределения 19
α — вектор дополнительных переменных прямой задачи 31
β — вектор дополнительных переменных двойственной задачи 31
δ — оценка переходных вероятностей неоднородной марковской цепи 167
η_r — $(r \times 1)$-вектор-столбец единиц 24
η_T — $(T \times 1)$-вектор-столбец единиц 24
g_* — обобщенная линейная оценка 157
g_*^c — обобщенная линейная оценка при учете ограничений 159
G — $(r \times r^2)$-матрица, содержащая r единичных подматриц I_r 25
Γ — гамма-функция 19
H — весовая матрица размера $(rT \times rT)$ для системы уравнений регрессии 50
$H'H$ — весовая матрица размера $(rT \times rT)$ для системы нормальных уравнений 50
I_r — единичная $(r \times r)$-матрица 25
J — диагональная $(r(r-1) \times (r-1))$-матрица 90
λ — вектор множителей Лагранжа 25
Λ — диагональная матрица собственных чисел 138
$n_j(t)$ — число микрообъектов в состоянии j в момент времени t 15
$N(t)$ — полное число микрообъектов в момент времени t 37
Ω — область допустимых значений в пространстве параметров 17
$\mathbf{p}$ — вектор-столбец, составленный из векторов $\mathbf{p}_j$, $j = 1,2,.., r$ 24
$\mathbf{p}_0$ — априорное среднее значение вектора p_* 92
$\mathbf{p}_*$ — вектор-столбец, составленный из векторов $\mathbf{p}_j$, $j = 1, 2, \ldots, r-1$ 59
$\mathbf{p}_j$ — j-й столбец матрицы P 23
p_{ij} — вероятность перехода (элемент матрицы P)12
$\hat{p}_{ij}$ — микрооценка переходной вероятности по методу максимального правдоподобия 17
$\ddot{p}_{ij}$ — байесовская микрооценка переходной вероятности 21
$\hat{\mathbf{p}}$ — оценка по методу наименьших квадратов 24
$\tilde{\mathbf{p}}_*$ — байесовская оценка 91
$\tilde{\mathbf{p}}_*^c$ — байесовская оценка при учете ограничений 94
$\tilde{\mathbf{p}}^c$ — оценка по методу наименьших квадратов при учете ограничений 30
$\hat{\mathbf{p}}_*$ — оценка по обобщенному методу наименьших квадратов, оценка по методу минимума критерия χ^2 или оценка максимального правдоподобия 62
$\hat{\mathbf{p}}_*^c$ — оценка по обобщенному методу наименьших квадратов при учете ограничений, оценка по методу минимума критерия χ^2 при учете ограничений или макрооценка максимального правдоподобия 63
$\bar{\mathbf{p}}_*$ — оценка Тейла — Гольдбергера 93
P — $(r \times r)$-матрица вероятностей перехода 13
π — вектор стационарных вероятностей 13
$q_j(t)$ — истинное значение относительной частоты состояния j в момент времени t 22
r — число состояний 12
R — $(r \times (r-1)r)$-матрица, содержащая $r-1$ единичных подматриц I_r 62
R_y — выборочное пространство 18
Σ — ковариационная матрица вектора ошибки и 24

Σ_* — ковариационная матрица вектора ошибки $\mathbf{u}_*$ 59
Σ_0 — априорная ковариационная матрица 90
$\hat{\Sigma}_*$ — выборочная оценка матрицы Σ_* 69
Σ^+ — матрица, псевдообратная к матрице Σ 137
T — длительность интервала моделирования (число шагов) 12
Θ — вектор параметров 17
$\mathbf{u}$ — вектор-столбец с элементами $\mathbf{u}_j$, $j = 1,2,\ldots, r$ 24
$\mathbf{u}_*$ — вектор-столбец с элементами $\mathbf{u}_j$, $j = 1, 2,\ldots, r-1$ 59
$\mathbf{u}_j$ — вектор-столбец с элементами $\mathbf{u}_j(t)$, $t = 1, 2, \ldots, T$ 23
$u_j(t)$ — величина ошибки в состоянии j в момент времени t 22
x_t — дискретная случайная величина 12
X — $(rT \times r^2)$-матрица с r одинаковыми блоками X_j, $j = 1, 2,\ldots, r$ на диагонали 24
X_* — $((r-1)T \times (r-1)T)$-матрица с $(r-1)$ одинаковыми блоками X_j, $j = 1,2,\ldots, r-1$ на диагонали 59
X_j — $(T \times r)$-матрица наблюдаемых значений относительных частот y_t $(t=1)$ 23
$\mathbf{y}$ — вектор-столбец, составленный из векторов y_j, $j = 1, 2,\ldots, r$ 24
$\mathbf{y}_*$ — вектор-столбец, составленный из векторов y_j, $j = 1, 2,\ldots, r-1$ 59
$\mathbf{y}_j$ — вектор-столбец, составленный из чисел $y_j(t)$, $t = 1,2\ldots, T$ 23
$y_j(t)$ — наблюдаемое значение относительной частоты состояния j в момент времени t 22
Z — матрица базовых или определяющих переменных 164

ОГЛАВЛЕНИЕ

Ли Ц. и др.

Л55 Оценивание параметров марковских моделей по агрегированным временным рядам. Пер. с англ. А. Д. Касавина, В. А. Лотоцкого, А. С. Манделя. Под ред. и с предисл. Н. С. Райбмана. М., «Статистика», 1977.

221 с. с ил. (Мат.-стат. методы за рубежом).
Перед загл. авт.: Ц. Ли, Д. Джадж, А. Зельнер.

Книга знакомит читателей с современными методами определения переходных вероятностей марковских цепей по статистическим данным. Монография включает ряд приложений, в которых сконцентрированы математические сведения, необходимые для понимания материала основных глав книги, а также программу на Фортране, реализующую изучаемые методы оценивания.

Книга представляет интерес для специалистов, занимающихся вопросами управления и идентификации в различных областях науки и техники, и особенно для специалистов в области экономики. Она будет полезна также аспирантам и студентам экономических вузов в качестве дополнительного учебного пособия.

Л $\frac{10805^{1}\text{-}033}{008(01)\text{-}77}$ 43-77 517.8

---

[1] Второй индекс 20204.

Ц. Ли, Д. Джадж, А. Зельнер

Оценивание параметров марковских моделей по агрегатированным временным рядам

Перевод с английского

А. Д. Касаина, В. А. Лотоцкого, А. С. Манделя

Под редакцией и с предисловием *Н. С. Райбмана*

Редактор *Е. В. Крестьянинова*
Мл. редактор *О. Б. Степанченко*
Техн. редактор *В. А. Чуракова*
Корректоры *Я. Б. Островский, Н. П. Сперанская*
Худ. редактор *Т. В. Стихно*
Переплет художника *Т. Н. Погореловой*

ИБ № 247

Сдано в набор 9/IX 1976 г.
Подписано к печати 11/I 1977 г.
Формат бумаги 60×90 1/16. Бумага книжн.-журн.
Объем 14,0 печ. л. Уч.-изд. л. 14,31 Усл. п. л. 14.
Тираж 6000 экз. (Тематич. план 1977 г. № 43)
Заказ 1113 Цена 1 р. 59 к.

Издательство «Статистика», Москва, ул. Кирова, 39.

Московская типография № 4 Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли, Москва, Н-41, Б. Переяславская, 46

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

Во втором квартале 1977 года издательство «СТАТИСТИКА» выпустит в свет сборник

«Математико-статистические методы исследования взаимосвязей в экономике» (ГДР, перевод с немецкого, 9 л. с ил.).

Сборник составлен докт. экон. наук, проф. В. В. Швырковым (СССР) и проф. К. Отто (ГДР). Авторами статей являются известные ученые ГДР: К. Отто, Э. Фёрстер, Ю. Вильке, Г. Штрохе и др. В сборнике две части. В часть I «Статический анализ» входят статьи, посвященные мультипликационному анализу, основам факторного анализа, статистическим оценкам, множественной регрессии и др. Часть II «Динамический анализ» включает статьи об анализе временных рядов, о прогнозировании и др.

Особенный интерес у читателя вызовет то, что все авторы рассматривают статистические методы на реальных примерах экономических исследований.

Книга может быть полезна широкому кругу экономистов, статистиков. Она может служить учебным пособием для аспирантов и студентов вузов.

Сборник можно приобрести в книжных магазинах, продающих общественно-политическую литературу.