

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
НЕФТИ И ГАЗА ИМ. И.М. ГУБКИНА

**Кафедра прикладной математики
и компьютерного моделирования**

Серия

Прикладная математика в инженерном деле

М.Г. Сухарев

МЕТОДЫ ПРОГНОЗИРОВАНИЯ

Учебное пособие для студентов специальности
230401 — «Прикладная математика» РГУ нефти и газа
им. И.М. Губкина

Москва, 2009 г.

РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
НЕФТИ И ГАЗА ИМ. И.М. ГУБКИНА

**Кафедра прикладной математики
и компьютерного моделирования**

Серия

Прикладная математика в инженерном деле

М.Г. Сухарев

МЕТОДЫ ПРОГНОЗИРОВАНИЯ

Учебное пособие для студентов специальности
230401 — «Прикладная математика» РГУ нефти и газа
им. И.М. Губкина

Москва, 2009 г.

УДК 62-192 – 519.2

М.Г.Сухарев. Методы прогнозирования. Учебное пособие — М.: РГУ нефти и газа, 2009 г., 208 с.

Издание является пособием для изучения дисциплины с тем же названием, включенной в учебные планы специальности «Прикладная математика» РГУ нефти и газа.

Его основная задача – помочь читателям в овладении методами прогнозирования при решении проблем научно-технического характера. Из широкого спектра методов особо выделены и даны в подробном изложении регрессионный анализ и временные ряды. Усвоение студентами предложенного материала возможно на разных уровнях – от рецептурного до глубокого понимания математического аппарата.

Помимо студентов специальности «Прикладная математика» пособие может быть рекомендовано студентам и магистрантам других специальностей при подготовке дипломных проектов и курсовых работ, может также оказаться полезным аспирантам, преподавателям и сотрудникам университета в научно-исследовательской работе.

Утверждено Советом факультета А и ВТ в качестве учебного пособия.

Рецензенты:

© Российский Государственный университет нефти и газа им. И.М. Губкина, 2009 г.

Содержание

1	Прогнозирование.....	8
1.1	О прогнозировании и прогнозах.....	8
1.2	Формальные и экспертные методы	9
2	Регрессионный анализ	22
2.1	Постановка задач, основные понятия.....	22
2.2	Функция регрессии и ее свойства (теоретический регрессионный анализ)	27
2.3	Прикладной регрессионный анализ. Линейные модели .	30
2.4	Линейная модель с одним переменным (регрессором)...	35
2.5	Множественная регрессия.....	37
2.6	Распределение выборочных функций. Коэффициент детерминации	42
2.7	Техника регрессионного анализа. Доверительные интервалы	46
2.8	Техника регрессионного анализа. Проверка гипотез.....	50
2.9	Полиномиальная регрессия.....	60
2.10	Нелинейные приближения. Подбор эмпирических зависимостей	63
3	Временные ряды.....	73
3.1	Основные обозначения и понятия. Примеры. Задачи исследования	73
3.2	Определение стационарного процесса. Белый шум.....	77
3.3	Детерминированные временные ряды	80
3.4	Сезонный тренд	87

3.5	Стохастические временные ряды. Случайное блуждание.	
	Операторы разности.....	88
3.6	Процессы скользящего среднего	91
3.7	Процессы авторегрессии.....	95
3.8	Марковский процесс — авторегрессия 1-го порядка.....	97
3.9	Процесс Юла — авторегрессия второго порядка.....	99
3.10	Общий процесс авторегрессии $AR(p)$. Коррелограмма.	
	Частные автокорреляции	102
3.11	Процессы авторегрессии — скользящего среднего	
	$ARCC(p, q)$	105
3.12	Модель авторегрессии — проинтегрированного	
	скользящего среднего $ARPCSS(p, d, q)$	107
3.13	Спектр стационарного случайного процесса с	
	дискретным временем.....	111
3.14	Спектры процессов $AR(1)$, $AR(2)$	117
3.15	Спектральная плотность процесса $ARCC(p, q)$	120
3.16	Критерии случайности.....	122
3.17	Процедуры обработки временных рядов. Модель	
	Бокса–Дженкинса.....	130
3.18	Оценка среднего, корреляционной функции и	
	спектральной плотности.....	136
3.19	Локальное сглаживание временных рядов (метод	
	скользящих средних).....	138
4	Экспертный логический анализ	144
4.1	Краткая справка	144

4.2	Техника экспертного логического анализа.....	145
4.3	Достоинства экспертного логического анализа. Некоторые рекомендации по проведению экспертизы	154
	Приложение А — Необходимые сведения из теории вероятности.....	159
A.1	Условное математическое ожидание	159
A.2	Двумерное нормальное распределение.....	164
A.3	Многомерное нормальное распределение	165
A.4	Свойства условных математических ожиданий	170
A.5	Свойства оптимального предиктора.....	171
A.6	Функции регрессии (пример обработки данных).....	174
	Приложение Б — Некоторые распределения, связанные с нормальным	176
B.1	Распределение χ^2	176
B.2	Распределение Стьюдента (t -распределение).....	179
B.3	Распределение Фишера	180
	Приложение В — Геометрический подход к регрессионному анализу	182
V.1	Геометрическая интерпретация метода наименьших квадратов.....	182
V.2	Операторы проектирования и их свойства.	183
V.3	Теорема о распределении квадратичной формы.....	184
V.4	Остаточная сумма квадратов	185
	Приложение Г — Теоремы о распределениях выборочных сумм.....	187
G.1	Теорема о распределении выборочной функции	

$\Sigma(Y_i - \bar{Y})^2$	187
Г.2 Теорема о распределении суммы квадратов величин из $N(\mathbf{0}, \mathbf{I}_n)$	188
Приложение Д — Изменение количества регрессоров	190
Приложение Е — Краткие сведения о математической модели, лежащей в основе экспертного логического анализа	196
Е.1 Свойства матрицы суждений и следствия из них.....	196
Е.2 Вычислительная процедура метода.....	199
Е.3 Согласованность суждений. Индекс согласованности .	202
Е.4 Способы выявления и пересмотра несогласованных суждений.....	204

1 ПРОГНОЗИРОВАНИЕ

1.1 О прогнозировании и прогнозах

Желание заглянуть в будущее свойственно каждому человеку в отдельности, группам людей и человечеству в целом. Прозреть детали грядущего нам не дано, и в этом наше величайшее счастье. Интрига сохраняется и постоянно подогревает интерес к жизни.

Но не думать о будущем мы не можем и стремимся так выстроить свои действия, чтобы они давали долговременный эффект. Для этого человечество изобрело различные способы прогнозирования, прорицания, прогностики и продолжает их совершенствовать и наращивать. С чего все началось, сейчас вряд ли можно установить. Но всем известно, в древнем Риме жрецы (авгуры) гадали на внутренностях жертвенных животных, затем люди научились гадать на картах, на кофейной гуще, предсказывать судьбы по положению планет, испробовали много других способов. Вопрос о том, как они это делали, в настоящем пособии не раскрывается.

Нас будут интересовать другие вопросы, например, такие: Как будет изменяться стоимость барреля нефти? Какие мероприятия целесообразно проводить в корпорации для укрепления ее финансового положения? Какие выплаты ожидают страховую компанию? Как будет изменяться потребление топлива в регионе, в населенном пункте, в стране? Как будет изменяться аварий-

ность производственного объекта? Какова будет загрузка производственных мощностей предприятия в предстоящий период?

Все эти вопросы очень непростые. Для определенного ответа на них надо знать множество конкретных обстоятельств. Даже профессионалы высокого уровня не застрахованы от просчетов, они ошибаются и иногда существенно. Будущее для всех скрыто во мгле. Не помогут получить точный ответ и методы прогнозирования. Но толковое применение этих методов поможет снизить вероятность ошибок, позитивно повлияет на уровень управленческих решений. Знакомство с подходами к прогнозированию нужно каждому специалисту, даже если он непосредственно не отвечает за прогноз. Эти знания пригодятся ему для общей оценки ситуации и выбора правильной линии в сложившейся конъюнктуре.

1.2 Формальные и экспертные методы

Люди придумали много способов прогноза как пассивного, с целью лишь предсказать грядущие изменения, так и активного, с целью выстроить свои усилия, чтобы воздействовать на события, добиваясь их развития в благоприятном для себя направлении. Даже простое перечисление этих методов является весьма трудоемким делом. Еще сложнее объяснить методы в одном издании. Поэтому такая задача в пособии не ставится.

На наш взгляд, естественнее всего очертить диапазон возможных методов и для большинства из них дать только общее представление о возможностях и сфере применения. Выделены и подробно изложены в пособии только формальные математиче-

ские методы: регрессионный анализ и временные ряды. Успешное применение этих методов невозможно без понимания математического аппарата. Следует надеяться, что усилия, затраченные читателем на овладение этим аппаратом, не окажутся напрасными: очень уж широк диапазон возможных применений.

Освоить методы, не имеющие глубокой математической подоплеки, обычно не представляет большого труда. Это можно сделать по мере возникновения такой потребности. А с продвинутыми математическими моделями имеет смысл знакомиться заблаговременно.

На рис. 1.1 представлена классификация методов прогнозирования. Перечислена малая часть известных на сегодняшний день методов. Они отобраны не по степени важности, а, прежде всего, для демонстрации их разнообразия, широты диапазона. Многие из этих методов (рис. 1.1) можно встретить иногда под другими названиями. Наряду с методами узкой направленности разработаны десятки комплексных методик, успешно применявшихся при решении крупных проблем на протяжении многих лет. На наш взгляд, эти комплексные методики следует отнести к наукам об управлении производством, тогда как спектр применения сингулярных методов значительно шире и охватывает информатику, вычислительную технику, инженерные науки и др. вплоть до социологии.

Эвристические	1. Экспертные оценки
Аналитические	2. Метод аналогий
Фактографические	3. Построение сценариев
	4. Метод дерева целей
	5. Сетевое планирование и управление
	6. Экспертный логический анализ
	7. Непосредственная интерполяция и экстраполяция
	8. Регрессионный анализ
	9. Временные ряды

Рис. 1.1 — Классификация методов прогнозирования

Применение понятия «классификация» к методам прогнозирования требует разъяснения. Рис. 1.1 выделяет три группы методов: эвристические, аналитические и фактографические. Но резкой границы между ними провести нельзя. Предложенную классификацию следует считать нечеткой, размытой, то есть методы 4, 5, например, следует считать не только эвристическими, но и аналитическими.

При прогнозировании реальных процессов трудно оставаться только на формальной основе, эвристические оценки так или иначе будут приняты во внимание. Аналитическими методами названы те, для которых предложены четко определенные процедуры, алгоритмы действий. Под фактографическими понимаются те методы, которые основаны на числовом материале (ретроспективной статистике).

Экспертные оценки

Экспертные оценки люди делают, опираясь на свой жизненный опыт, опыт производственной деятельности, а также на интуицию. Жизненный и производственный опыт приобретаются с годами, а интуиция дается каждому индивидууму от рождения, хотя у многих и развивается со временем. Природа интуитивных прозрений человека не полностью познана наукой. Принимая управленческие решения, люди в течение веков опирались на свою интуицию, а также на опыт и интуицию окружающих.

С развитием точных наук, в эпоху интенсивной информатизации большое значение стали придавать решениям, полученным с помощью математических моделей. Было время, когда всерьез надеялись с прогрессом методов математического моделирования и средств для реализации этих методов прийти к идеальному, безошибочному управлению производством, экономикой, обществом. Но то, что идея такой кибернетизированной системы управления утопична, осознали довольно быстро.

Оптимальным является сочетание формализованных методов и экспертных оценок. Формализованные методы – модели – позволяют обобщить накопленную информацию, провести необходимые расчеты, вычислив на их основе легко интерпретируемые показатели. Но принятие решений всегда было и в обозримом будущем остается за человеком. Те ли люди принимают решения, насколько они адекватны лежащей на них ответственности, зависит от сложившегося общественного устройства, подготовки кадров, принципам их продвижения по служебной лестнице.

це, от уровня среднего и высшего образования, в общем, от того, в каком состоянии находится общество, какие настроения в нем преобладают и куда оно движется.

Один думающий человек (Пауль Самуэльсон) сказал: «Наиболее существенная ошибка в прогнозах — не осознавать, насколько велики вероятности событий, не принятых во внимание». События, не принятые во внимание, иногда могут быть уловлены на интуитивном уровне. Процедура получения экспертных оценок может быть упорядочена, введена в такие рамки, которые помогают эксперту более осознанно сформулировать нужные ответы, учтя при этом все аспекты проблемы. Экспертный логический анализ (р. 4) является очень полезной процедурой с широким спектром возможных применений.

Методы аналогий

Метод аналогий является одним из наиболее общих приемов мыследеятельности. Изучение истории важно не только для того, чтобы иметь представление о событиях прошлого, но и для того, чтобы, заметив черты современности, сходные с каким-либо периодом прошлой жизни, угадать направление будущего движения. Метод аналогий широко применяется в общественных и естественных науках.

Врач, ставя диагноз, опирается на личный и коллективный опыт, полученный по наблюдениям за пациентами. В случае редких заболеваний он тщательно сверяет симптомы болезни с известными случаями, стремясь установить аналогию.

Предложены различные формализации, позволяющие отыскание аналогий проводить не вслепую, а, придерживаясь определенных правил, позволяющих улучшить процедуру сопоставления рассматриваемых явлений, объектов и событий с имевшими место ранее.

Построение сценариев

Прогнозы строятся для явлений и событий разных временных и пространственных масштабов. Чем более масштабным является событие, тем больше факторов влияет на его протекание и тем менее достоверна обычно бывает информация. К построению сценариев прибегают при определении риска техногенных аварий и катастроф, при формировании программ развития промышленного предприятия или корпорации, при исследовании глобальных геополитических изменений и т.д.

Сценарий развития аварийной ситуации предполагает перечисление возможных инцидентов, нарушающих нормальный ход производственного процесса, и отслеживание цепочки вызванных ими событий. Например: образование сквозного отверстия в трубе из-за внутренней или внешней коррозии – истечение газа – взрыв и возгорание газозвушной смеси – пожар и т.д.

Сценарии (программы) развития корпорации нефтегазового профиля могут включать такие мероприятия как приобретение лицензии на исследование участка недр, разработка месторождения, строительство трубопровода, строительство нового предприятия по переработке сырья и т.д. Разрабатывая сценарий развития корпорации, нельзя уйти от прогноза изменения внешних ус-

ловий. К ним относятся политика государства в области нефтегазового комплекса: налоги, таможенные пошлины и др., – стоимость сырья и продуктов его переработки на внешнем и внутреннем рынке, спрос на продукцию в обозримом периоде и др.

Метод дерева целей

Метод является одной из модификаций сценарного подхода. Метод состоит в том, что при формировании программы развития ставится желаемая цель, а затем формируются и выстраиваются во взаимосвязи по времени мероприятия, необходимые для достижения этой цели. Таким образом, метод дерева целей предполагает не только прогнозирование грядущих событий, но и целенаправленное воздействие на них. Ставя конечную цель, можно сосредоточить финансовые и кадровые ресурсы на главном направлении, своевременно реализовать мероприятия, сулящие хорошие перспективы в будущем.

Сетевое планирование и управление (СПУ)

Метод СПУ примыкает в определенной степени к методу дерева целей. Однако его назначение состоит не в формулировке цели и формировании путей ее достижения, а в организации и упорядочении работ, когда и конечная цель и необходимые работы вполне ясны.

Метод известен также в англоязычной аббревиатуре как PERT (Program Evaluation and Review Technique). Он был разработан и успешно применен в США для реализации важнейшей космической программы, перед которой ставились не только на-

учно-технические, но и пропагандистские, и внешнеполитические цели, программы, которая рассматривалась как одно из сражений холодной войны.

Поясним этот метод на простом, понятном каждому примере строительства жилого дома. При этом сознательно не будем вдаваться в детали технологического процесса. Профессионалы могли бы, конечно, обосновано раскритиковать этот пример, но мы не откажемся от хорошей иллюстрации из-за этой угрозы.

Итак, предположим, что строительство дома включает следующие работы:

- 1 — возведение фундамента, стен и перекрытий,
- 2 — сооружение крыши,
- 3 — проводка труб (вода, газ, канализация),
- 4 — электропроводка,
- 5 — монтаж сантехники,
- 6 — монтаж окон и дверей,
- 7 — окраска потолков,
- 8 — укладка полов,
- 9 — оклейка стен обоями,
- 10 — подготовка территории,
- 11 — сдача заказчику.

Работы эти должны быть упорядочены во времени: естественно, что крышу нельзя сооружать, пока нет фундамента и стен и т.д. Построим граф, вершины которого (кружочки) означают указанные работы, а дуги характеризуют упорядоченность работ во времени. В графе есть один вход — вершина 1 — и один выход — вершина 11 (рис. 1.2а).

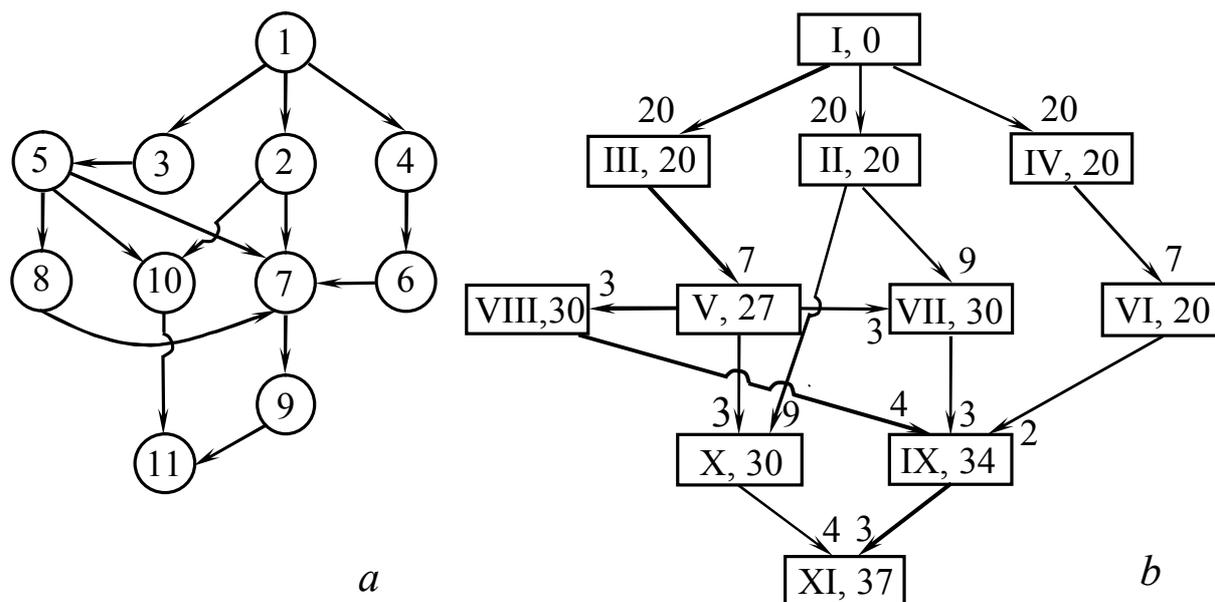


Рисунок 1.2 — Взаимосвязь работ при сооружении жилого дома:
a — вершины графа — работы, *b* — дуги графа — работы

Поскольку строительство жилья носит массовый характер, то время на проведение каждой работы, в общем-то, известно. Поэтому каждой вершине графа можно приписать определенное число — время выполнения работы, — а дуге, исходящей из вершины i , — минимально возможный срок окончания i -й работы. Граф строится для того, чтобы определить кратчайший срок окончания строительства.

Информацию о работах и их взаимосвязи можно передать несколько другим способом (рис. 1.2*b*). Вершины графа — прямо-

угольники – отвечают моменту начала работы. Работы пронумерованы, номер обозначен римскими цифрами. Работе I (рис. 1.2a) сопоставлен на рис. 1.2b верхний прямоугольник и все 3 дуги, исходящие из этого прямоугольника, следующей работе — прямоугольник II и две исходящие из него дуги и т.д.

Рядом с каждой дугой проставлена продолжительность соответствующей работы (в неделях). Момент, когда можно начинать работу, определяется длиной максимального пути из начала в соответствующую вершину. В вершину X, например, (начало работы 10) ведут два пути $I \rightarrow III \rightarrow V \rightarrow X$ длиной 30 и $I \rightarrow II \rightarrow X$ длиной 29. Таким образом, работу 10 нельзя начать ранее, чем через 30 временных единиц с момента начала строительства. В каждом прямоугольнике арабскими цифрами отмечен момент возможного начала соответствующей работы.

Цель СПУ — определение момента завершения всего строительства. В примере это эквивалентно определению пути максимальной длины в вершину XI. Такой путь называется критическим. На рис. 2b работы, составляющие этот путь ($I \rightarrow III \rightarrow V \rightarrow VIII \rightarrow IX \rightarrow XI$), отмечены жирными линиями. Длина критического пути равна 37 временным единицам.

Знание критического пути позволяет концентрировать внимание и усилия на входящих в него работах. Сосредоточение усилий может проявляться в разных формах: обеспечение работ наиболее квалифицированными специалистами, бесперебойное материально-техническое снабжение, поощрение и наказание ис-

полнителей и т.д. Для производства работ, не входящих в критический путь, имеется определенный резерв времени.

Описанная нами процедура является, пожалуй, наиболее простой модификацией СПУ. Разработано много приемов, позволяющих приблизить ее к учету жизненных реалий. Известно, что зачастую бывает сложно точно указать сроки выполнения работ. Например, при строительстве морских сооружений: подводных трубопроводов, нефтедобывающих платформ — возможность проведения работ зависит от погодных условий. Чтобы учесть это обстоятельство, при составлении графиков указывается интервальная оценка для срока каждой работы. Иногда задают 3 числа: минимальный, наиболее вероятный и максимально возможный сроки, а также их вероятности. В такого рода моделях критический путь определяется неоднозначно. Модифицированная процедура СПУ позволяет вычислить несколько критических путей и их вероятность.

Метод СПУ так же, как и метод дерева целей, является активным, позволяет воздействовать на будущий ход событий, правильно расставить приоритеты и стимулировать наиболее важные работы.

Экспертный логический анализ (ЭЛА)

ЭЛА представляет собой аппарат для организации и проведения экспертизы. Мы придаем методу большое значение, считая, что сфера его применения будет со временем расширяться. Именно по этой причине ЭЛА будет дан не в кратком, а в расширенном изложении, но помещен в финальном, четвертом разделе

пособия. Для этого есть определенные доводы. Основную часть курса — регрессионный анализ и временные ряды — следует рассматривать как формальный аппарат для обработки ретроспективной информации. Эта обработка проводится, как правило, для получения технологических и организационных выводов, а вот здесь особенно полезным может оказаться экспертный логический анализ.

Непосредственная интерполяция и экстраполяция являются простейшими методами обработки информации, которые можно рассматривать как предварительный этап исследования, подлежащий дальнейшему уточнению.

Методы *регрессионного анализа и временных рядов* позволяют по сути дела провести интерполяцию и экстраполяцию статистических данных, опираясь на современный аппарат теории вероятностей в расширенном понимании, то есть, включая математическую статистику и теорию случайных процессов.

Основным объектом обработки в моделях регрессии и временных рядах являются выборки $(\mathbf{x}_i, \mathbf{y}_i)$ ($i = 1, \dots, n$) двух переменных \mathbf{x} и \mathbf{y} . Обе переменные могут быть векторными величинами, поэтому для их обозначения использован полужирный шрифт. Вообще говоря, размерности этих векторов различны. Обозначим размерность \mathbf{x} через s , то есть $\mathbf{x}_i \in R^s$, а размерность \mathbf{y} через k , то есть $\mathbf{y}_i \in R^k$. Величины \mathbf{x}_i считаются реализациями s -мерного вектора \mathbf{X} , а \mathbf{y}_i — k -мерного вектора \mathbf{Y} . \mathbf{X}

рассматривается как аргумент, а Y как зависимая переменная, функция X .

Аппарат регрессионного анализа и временных рядов во многом основан на одних и тех же идеях. Принципиальное различие состоит в том, что временной ряд является случайным процессом, и в роли аргумента X выступает время $t = 1, 2, \dots$, а обрабатываемая выборка оказывается упорядоченной.

В регрессионном же анализе порядок следования членов выборки (x_i, y_i) не имеет значения.

Большими буквами здесь и далее обозначены случайные величины, малыми — их реализации. Полужирный шрифт используется для векторов и матриц.

2 РЕГРЕССИОННЫЙ АНАЛИЗ

2.1 Постановка задач, основные понятия

Регрессионный анализ применяется, когда надо установить зависимость одной переменной Y от другой переменной X или группы переменных $\mathbf{X} = \|X_1, \dots, X_s\|^T$.

Пример 1. Потребление газа населенным пунктом (Y) в зависимости от погодных и технологических факторов.

1а. Учитывается наиболее значимый погодный фактор — температура окружающего воздуха X_1 .

1б. Учитывается совокупность погодных факторов: температура воздуха в текущие сутки (X_1), сила ветра (X_2), облачность (X_3), влажность (X_4), температура за предыдущие сутки (X_5).

1в. К погодным факторам добавляются организационно технологические: превышение установленных лимитов потребления (X_6), снижение давления на входе питающего населенный пункт газопровода (из-за отказов в системе магистрального транспорта и хранения газа) (X_7) и др.

Пример 2. Выход целевого продукта Y на установке по переработке нефти. Влияющие факторы: температура в реакторе, расход сырья на входе, количество катализатора и др.

Пример 3. Урожайность Y злаковой культуры в регионе. Влияющие факторы: природные условия региона, количество внесенных удобрений, количество осадков, качество семян, каче-

ство подготовки почвы, соблюдение агротехнических требований и т.д. Предполагается, что статистические данные по всем факторам наличествуют и что качественные факторы переведены в количественную шкалу.

Во всех приведенных примерах зависимая переменная одна. Ограничимся в дальнейшем рассмотрением именно этого, наиболее часто встречающегося случая.

Независимые переменные – компоненты вектора X – принято называть регрессорами, зависимую переменную Y — откликом. Y называется также предиктором, а компоненты вектора X — предсказывающими переменными. Правильный выбор регрессоров является необходимой предпосылкой успеха исследования. В число регрессоров в начале исследования могут быть включены лишние, отвечающие незначимым или слабо влияющим факторам. Некоторые факторы могут оказаться зависимыми. Теория дает рецепты того, как по выборке можно исправить недочеты первоначального выбора. Однако если из-за недостаточной компетентности исследователя в число регрессоров не попали существенные факторы, то модель прогнозирования хорошей не будет.

В наиболее общей постановке задача регрессионного анализа не предполагает, что при установлении зависимости Y от X будут использоваться функции какого-либо заранее заданного класса. Для такого подхода на заре кибернетической эры был придуман термин «черный ящик». Рис. 2.1 иллюстрирует это понятие. Черный потому, что внутрь него невозможно заглянуть,

чтобы постичь механизм управления исследуемым объектом, природу процесса или явления.

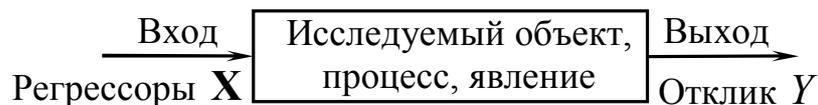


Рисунок 2.1 — «Черный ящик» — о природе связи отклика и регрессоров ничего заранее не известно

Задачей регрессионного анализа является построение модели, выявление зависимости Y от X на основании выборки — имеющихся эмпирических, экспериментальных данных. Мы говорим об экспериментальных данных независимо от того, был ли фактически подготовлен план эксперимента, то есть заранее намечена совокупность значений X , по которым вычислялся отклик, или не был.

При постановке и обсуждении задач регрессионного анализа надо попытаться найти ответы на следующие принципиальные вопросы:

В каком классе функций целесообразно искать зависимость Y от X ?

Как получить наилучшее приближение в выбранном классе функций?

Как оценить качество полученной модели? В частности, действительно ли имеет место зависимость Y от X ?

Нельзя ли упростить полученную зависимость, опираясь лишь на совокупность обрабатываемых экспериментальных данных?

В регрессионном анализе зависимая переменная считается случайной величиной. Пусть мы получили некоторый результат, проведя эксперимент с некоторыми значениями регрессоров $\mathbf{X} = \mathbf{x}^0$. Зафиксировав эти значения \mathbf{x}^0 , проведем еще эксперимент. Из-за случайного характера зависимости Y от \mathbf{X} при этом, вообще говоря, получится другой результат. Это означает, что точно определить Y при фиксированном $\mathbf{X} = \mathbf{x}^0$ нельзя, но можно определить некоторую величину $\hat{Y} = \hat{Y}(\mathbf{x})$, оценивающую Y , в каком-то смысле близкую к Y . Разность $Y - \hat{Y}$ представляет собой ошибку оценивания

$$Y - \hat{Y}(\mathbf{x}) = \varepsilon(\mathbf{x}).$$

Конечно, следует стремиться, чтобы ошибка оценивания была мала. Но, какую ошибку можно считать малой? Это надо строго определить, введя меру близости — норму $\|\varepsilon\|$ величины $\varepsilon = Y - \hat{Y}(\mathbf{x})$.

В качестве нормы используется среднеквадратическое отклонение $(\hat{Y}(\mathbf{X}) - Y)^2$. Так как \hat{Y} величина случайная, то следует брать математическое ожидание отклонения

$$\mathbf{M}[(\hat{Y} - Y)^2 | \mathbf{X} = \mathbf{x}].$$

Таким образом, если поиск оптимального предиктора производится на заданном классе функций Φ , то оптимальный предиктор определяется как решение задачи

$$\hat{Y}^* = \arg \min_{\hat{Y}(\mathbf{x}) \in \Phi} \mathbf{M}(\hat{Y} - Y)^2. \quad (2.1.1)$$

Качество предиктора (качество предсказания) характеризуется его среднеквадратической ошибкой

$$\Delta^* = \Delta(\hat{Y}^*) = \mathbf{M}(\hat{Y}^*(\mathbf{x}) - Y)^2 = \min_{\hat{Y}(\mathbf{x}) \in \Phi} \mathbf{M}(\hat{Y}(\mathbf{x}) - Y)^2. \quad (2.1.2)$$

Попробуем пояснить, почему в качестве меры расхождения принято брать квадрат $(\hat{Y} - Y)^2$, а не какую-нибудь другую норму, например, модуль $|\hat{Y} - Y|$ или четвертую степень разности $(\hat{Y} - Y)^4$. Одним из аргументов в пользу среднеквадратического критерия является то, что при его использовании математический аппарат оказывается наиболее простым. Существуют также и другие, более глубокие причины, свидетельствующие о целесообразности такого выбора. Так, в математической статистике единственным оправданным критерием при оценке неизвестных параметров модели является метод максимального правдоподобия. В случае, когда распределение Y близко к нормальному, минимум среднеквадратической ошибки будет следствием метода максимального правдоподобия. А с нормальным законом при обработке выборок приходится сталкиваться чаще, чем с другими распределениями, что следует из центральной предельной теоремы

В заключение этого постановочного раздела обсудим, почему $\hat{Y}(\mathbf{x})$ нельзя считать точным значением Y . Во-первых, Y иногда может рассматриваться как величина, получающаяся в результате измерений, и тогда $\varepsilon(\mathbf{x})$ будет являться ошибкой изме-

рения. Во-вторых, во многих ситуациях трудно назвать все влияющие факторы. Достаточно обратиться к приведенным выше примерам 1-3. Читатель легко может расширить перечень указанных там факторов. Регрессионная зависимость устанавливается для наиболее значимых факторов, а остальные факторы (они называются скрытыми или латентными) трудно учесть, их влияние определяет ошибку модели.

Сведения о регрессионном анализе можно найти в изданиях [2, 6, 9, 11, 13, 16, 19, 20] и др.

2.2 Функция регрессии и ее свойства (теоретический регрессионный анализ)

Переходим к формальному построению регрессионного анализа. В этом подразделе будем считать случайными не только Y , но и X . Функцией регрессии Y на X называется условное математическое ожидание Y при фиксированном $X = x$

$$\mathbf{M}(Y|X = x) = m_Y(x). \quad (2.2.1)$$

Важную роль в регрессионном анализе играет также условная дисперсия

$$\mathbf{D}(Y|X = x) = \mathbf{M}[(Y - m_Y(x))^2 | X = x]. \quad (2.2.2)$$

Наряду с этими функциями рассматриваются случайные величины

$$m_Y(\mathbf{X}), \mathbf{D}(Y|\mathbf{X}),$$

которые при $X = x$ принимают значения (2.2.1) и (2.2.2) соответственно.

Если вспомнить, что такое математическое ожидание, то легко понять, почему (2.2.1) следует считать естественной характеристикой для прогнозирования Y как функции \mathbf{X} , то есть предиктором.

Из простых соображений (теорема А.1, см. приложение А.5) вытекает, что этот предиктор $\hat{Y}^* = m_Y(\mathbf{X})$ является оптимальным, то есть по критерию наименьших квадратов не хуже любого другого предиктора \hat{Y}

$$\Delta(\hat{Y}) = \mathbf{M}(\hat{Y}(\mathbf{X}) - Y)^2 \geq \mathbf{M}(m_Y(\mathbf{X}) - Y)^2.$$

«Крышка» над Y является символом предиктора, а значок * используется как символ оптимальности.

Качество оптимального прогноза (предиктора) характеризуется остаточной суммой квадратов

$$\begin{aligned} \Delta^* &= \Delta(\hat{Y}^*) = \mathbf{M}(m_Y(\mathbf{X}) - Y)^2 = \\ &= \mathbf{M}\left\{\mathbf{M}\left[(Y - m_Y(\mathbf{X}))^2 \mid \mathbf{X}\right]\right\} = \mathbf{M}\mathbf{D}(Y \mid \mathbf{X}) = \sigma_{YX}^2. \end{aligned} \quad (2.2.3)$$

Третье равенство в этой цепочке следует из свойства условного математического ожидания (А.17). Обозначение σ_{YX}^2 используется для среднего значения $\mathbf{M}\mathbf{D}(Y \mid \mathbf{X})$ условной дисперсии $\mathbf{D}(Y \mid \mathbf{X})$ (см. (А.18), (А.18')).

Оказывается, что функция регрессии имеет максимальную ковариацию с Y . Представим этот результат как теорему 2.1, доказав предварительно лемму 2.1.

Лемма 2.1. Для любой функции $\hat{Y}(\mathbf{X})$

$$\text{cov}(\hat{Y}(\mathbf{X}), Y) = \text{cov}(\hat{Y}(\mathbf{X}), m_Y(\mathbf{X})).$$

Для доказательства достаточно воспользоваться соотношением (А.17)

$$\begin{aligned} \text{cov}(\hat{Y}(\mathbf{X}), Y) &= \mathbf{M}\left[\left(\hat{Y}(\mathbf{X}) - \mathbf{M}\hat{Y}\right)(Y - m_Y)\right] = \\ &= \mathbf{M}\left\{\left[\left(\hat{Y}(\mathbf{X}) - \mathbf{M}\hat{Y}\right)\mathbf{M}(Y - m_Y)\right]|\mathbf{X}\right\} = \\ &= \mathbf{M}\left[\left(\hat{Y}(\mathbf{X}) - \mathbf{M}\hat{Y}\right)(m_Y(\mathbf{X}) - m_Y)\right] = \text{cov}(\hat{Y}(\mathbf{X}), m_Y(\mathbf{X})). \end{aligned} \quad (2.2.4)$$

Теорема 2.1. Оптимальный предиктор $\hat{Y}^* = m_Y(\mathbf{X})$ имеет максимальную корреляцию с Y .

Из (2.2.4) при $\hat{Y}(\mathbf{X}) = m_Y(\mathbf{X})$ получаем $\text{cov}(m_Y(\mathbf{X}), Y) = \mathbf{D}m_Y(\mathbf{X}) = \sigma_{m_Y(\mathbf{X})}^2$. Значит,

$$\begin{aligned} \rho(m_Y(\mathbf{X}), Y) &= \frac{\sigma_{m_Y(\mathbf{X})}^2}{\sigma_{m_Y(\mathbf{X})}\sigma_Y} = \frac{\sigma_{m_Y(\mathbf{X})}}{\sigma_Y}, \\ \rho^2(\hat{Y}(\mathbf{X}), Y) &= \frac{\text{cov}^2(\hat{Y}, Y)}{\sigma_{\hat{Y}}^2\sigma_Y^2} = \frac{\text{cov}^2(\hat{Y}(\mathbf{X}), m_Y(\mathbf{X}))}{\sigma_{\hat{Y}}^2\sigma_{m_Y(\mathbf{X})}^2} \frac{\sigma_{m_Y(\mathbf{X})}^2}{\sigma_Y^2} = \\ &= \rho^2(\hat{Y}, m_Y(\mathbf{X}))\rho^2(m_Y(\mathbf{X}), Y) \leq \rho^2(m_Y(\mathbf{X}), Y). \end{aligned} \quad (2.2.5)$$

Квадрат коэффициента корреляции Y с функцией регрессии $m_Y(\mathbf{X})$, то есть $\rho^2(m_Y(\mathbf{X}), Y)$, называется корреляционным отношением и обозначается η_{YX}^2

$$\eta_{YX}^2 = \sup_{\hat{Y}} \rho^2(\hat{Y}(\mathbf{X}), Y) = \frac{\sigma_{m_Y(\mathbf{X})}^2}{\sigma_Y^2} = 1 - \frac{\sigma_{YX}^2}{\sigma_Y^2} = 1 - \frac{\Delta^*}{\sigma_Y^2}. \quad (2.2.6)$$

При выводе (2.2.6) использованы формулы (А.18') и (2.2.3).

Для двумерного нормального распределения с учетом (А.11) получаем

$$\eta_{YX}^2 = \rho^2. \quad (2.2.7)$$

В случае многомерного нормального распределения функция регрессии

$$m_Y(\mathbf{X}) = m_Y - \sum_{i=1}^s (X_i - m_i) \sigma^{iY} / \sigma^{YY} \quad (2.2.8)$$

распределена по нормальному же закону, причем, согласно (А.16) $\sigma_{YX}^2 = 1/\sigma^{YY}$. Следовательно, корреляционное отношение

$$\eta_{YX}^2 = \frac{\sigma_{m_Y(\mathbf{X})}^2}{\sigma_Y^2} = \frac{\sigma_Y^2 - \sigma_{YX}^2}{\sigma_Y^2} = 1 - 1/\sigma_Y^2 \sigma^{YY}. \quad (2.2.9)$$

Корреляционное отношение является характеристикой качества прогноза. Если ошибка Δ^* близка к нулю, то η_{YX}^2 близко к 1. И, наоборот, если остаточная сумма квадратов σ_{YX}^2 близка к дисперсии σ_Y^2 , то качество прогноза низкое и корреляционное отношение близко к нулю.

2.3 Прикладной регрессионный анализ. Линейные модели

Изложенная в предыдущем разделе теория, к сожалению, далеко не всегда может пригодиться при решении практических задач. Дело в том, что функцию регрессии (2.2.1) можно найти, если известно совместное распределение Y и \mathbf{X} . Но такой информации обычно у исследователя нет, а имеются только наблюдаемые значения регрессоров $\mathbf{X} = \|X_1, \dots, X_s\|^T$ и отклика Y :

$$\mathbf{X}_i, Y_i \quad (i = 1, \dots, n).$$

Эта совокупность является выборкой. Чтобы избежать в дальнейшем ошибок и непонимания, приведем толкование этого

понятия, одного из важнейших в математической статистике. Все рассуждения сначала будем вести для простейшего случая одного регрессора X , а затем обобщим на многомерный случай $\mathbf{X} = \|X_1, \dots, X_s\|$.

Положим, что проводятся n испытаний. В i -м испытании регрессор принимает значение X_i , а результат испытания (отклик) является случайной величиной Y_i . Условия проведения испытаний не меняются, поэтому результаты испытаний независимы. Совокупность пар (X_i, Y_i) $i=1, \dots, n$ называется выборкой. Если же испытания проведены и результаты $Y_i = y_i$ зафиксированы, то полученная совокупность (x_i, y_i) $i=1, \dots, n$ будет реализацией выборки. Значит, выборка — это совокупность независимых случайных величин, а реализация — это совокупность проведенных наблюдений, то есть совокупность чисел.

Таким образом, исходной информацией для регрессионного анализа служит выборка, информация о виде зависимости Y от x отсутствует. Подбор этой зависимости — задача исследователя. От того, насколько удачно подбор осуществлен, определяется во многом качество результатов анализа.

Чаще других пользуются линейными зависимостями. Это обусловлено рядом причин.

Класс линейных функций приводит к наиболее простым и глубоким аналитическим результатам. В случае нормального распределения, как мы видели, функция регрессии (2.2.1) является линейной. Следовательно, ее использование оправдано, когда

распределения величин близки к нормальному закону, то есть в широком диапазоне практических ситуаций.

К тому же линейная зависимость может в некотором интервале рассматриваться как хорошее приближение нелинейной.

Линейная модель для одного регрессора имеет вид

$$Y = \eta + \varepsilon, \quad \eta = \beta_0 + \beta_1 x. \quad (2.3.1)$$

Такая запись предполагает, что зависимая переменная η измеряется с ошибкой. Y — замеренное значение, η — точное, но неизвестное значение, ε — ошибка. Величина η линейно зависит от регрессора x , β_0, β_1 — неизвестные коэффициенты линейной функции. Снабжая величины Y, x, ε нижним индексом i , получим n соотношений

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (i = 1, \dots, n), \quad (2.3.2)$$

которые удобно записать с использованием векторно-матричных обозначений в виде

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.3.3)$$

где

$$\mathbf{Y} = \begin{Bmatrix} Y_1 \\ \vdots \\ Y_n \end{Bmatrix}, \quad \mathbf{X} = \begin{Bmatrix} 1 & x_1 \\ \cdots & \cdots \\ 1 & x_n \end{Bmatrix}, \quad \boldsymbol{\beta} = \begin{Bmatrix} \beta_0 \\ \beta_1 \end{Bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{Bmatrix}.$$

Запись (2.3.3) остается справедливой и для общего случая линейной зависимости с s регрессорами

$$\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_s x_s. \quad (2.3.4)$$

При этом вектор $\boldsymbol{\beta}$ будет иметь размер $(s+1) \times 1$, а матрица \mathbf{X} — $n \times (s+1)$

$$\mathbf{X} = \begin{vmatrix} 1 & x_{11} & \cdots & x_{1s} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{ns} \end{vmatrix}, \quad \boldsymbol{\beta} = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_s \end{vmatrix}. \quad (2.3.5)$$

Следует обратить внимание, что модель должна быть линейной только по параметрам β . Линейность по регрессорам x_1, \dots, x_s , вообще говоря, не требуется. Более детально этот вопрос будет рассмотрен в дальнейшем, после изложения основных положений теории.

Относительно ошибок ε_i делаются следующие предположения.

Предположение 1. Отсутствие систематической ошибки

$$\mathbf{M}\varepsilon_i = 0 \quad (i=1, \dots, n), \quad (2.3.6)$$

или иначе

$$\mathbf{M}Y_i = \eta_i.$$

Предположение 2. Постоянство дисперсии

$$\mathbf{D}\varepsilon_i = \sigma^2 \quad (i=1, \dots, n). \quad (2.3.7)$$

Предположение 3. Независимость. Из независимости, в частности, следует некоррелированность

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad (i, j = 1, \dots, n; i \neq j). \quad (2.3.8)$$

Для того чтобы получить более широкие результаты вводится также предположение о распределении ошибок по нормальному закону.

Предположение 4.

$$\varepsilon_i \in N(0, \sigma^2). \quad (2.3.9)$$

Если матрицу ковариаций случайного вектора ε обозначить $D\varepsilon$, то соотношения (2.3.6) – (2.3.8) представятся в виде

$$M\varepsilon = \mathbf{0}, D\varepsilon = \sigma^2 \mathbf{I}, \quad (2.3.10)$$

где \mathbf{I} — единичная матрица.

В модели (2.3.1) – (2.3.2) параметры β_0, β_1 являются детерминированными, но неизвестными величинами. Эти величины надо оценить по выборке. Оценки будут функциями выборки, то есть случайными величинами.

В математической статистике для оценок (точечных) вводятся определенные требования. Чтобы оценки были «хорошими» они должны обладать свойствами состоятельности, несмещенности и минимума дисперсии. Все эти свойства весьма естественны. Состоятельность означает, что чем больше членов в выборке, тем лучше оценка. Более строго: при $n \rightarrow \infty$, оценка стремится к истинному значению параметра.

Требование несмещенности состоит в том, чтобы математическое ожидание оценки было равно оцениваемому параметру. Если оценка получается смещенной, ее подправляют, корректируют, добиваясь несмещенности.

Минимум дисперсии — разброса вокруг истинного значения оцениваемого параметра — также является очевидным требованием. Чем меньше разброс, тем меньше, вообще говоря, ошибка при оценке по имеющейся выборке.

Критерием для оценивания неизвестных параметров β_0, β_1 выдвигается минимум суммы квадрата ошибок

$$\sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \rightarrow \min. \quad (2.3.11)$$

Целесообразность использования критерия (2.3.11) или, другими словами, метода наименьших квадратов обсуждалась ранее. Отметим, что в предположении 4 о нормальности ошибок критерий (2.3.11) является следствием метода максимального правдоподобия — наиболее обоснованного метода точечной оценки неизвестных параметров.

2.4 Линейная модель с одним переменным (регрессором)

Применительно к модели (2.3.2) критерий (2.3.11) имеет вид

$$R_1^2 \equiv \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 \rightarrow \min. \quad (2.4.1)$$

Нижний индекс 1 в обозначении R_1^2 свидетельствует о том, что модель содержит один регрессор. Приравняв нулю производные по β функции $R_1^2(\beta_0, \beta_1)$

$$\frac{\partial R_1^2}{\partial \beta_0} = -2 \sum (Y_i - \beta_0 - \beta_1 x_i), \quad \frac{\partial R_1^2}{\partial \beta_1} = -2 x_i \sum (Y_i - \beta_0 - \beta_1 x_i),$$

получим 2 линейных уравнения с двумя неизвестными

$$\begin{aligned} \hat{\beta}_0 n + \hat{\beta}_1 \sum x_i &= \sum Y_i, \\ \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 &= \sum x_i Y_i. \end{aligned} \quad (2.4.2)$$

Суммирование везде проводится от 1 до n , поэтому пределы суммирования в записи опущены. Над β_0, β_1 поставлены «га-

лочки». Это указывает на то, что если ранее величины β_j рассматривались как параметры, то решение системы (2.4.2) будет давать оценки этих параметров, которые, следовательно, зависят от выборки (вектора \mathbf{Y}) и являются случайными величинами.

Определитель системы (2.4.2) отличен от нуля, если только не все x_i одинаковы. Этот вырожденный случай не представляет интереса, поэтому можно утверждать, что система (2.4.2) имеет единственное решение. Это решение является единственной экстремальной точкой функции $R_1^2(\beta_0, \beta_1)$ и определяет ее минимум. Минимум, по крайней мере, один должен существовать, так как функция R_1^2 неотрицательна.

Оценки $\hat{\beta}_0, \hat{\beta}_1$ как решения системы (2.4.2) имеют вид

$$\hat{\beta}_1 = \frac{n \sum x_i Y_i - \sum x_i \sum Y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\sum (Y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \quad (2.4.3)$$

$$\hat{\beta}_0 = \frac{\sum Y_i \sum x_i^2 - \sum x_i y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\bar{Y} \sum (x_i - \bar{x})^2 - \bar{x} \sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}. \quad (2.4.4)$$

Чертой сверху обозначаются средние значения $\bar{x} = \frac{1}{n} \sum x_i, \bar{Y} = \frac{1}{n} \sum Y_i$. Формула (2.4.4) может быть представлена в виде

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}. \quad (2.4.5)$$

Оценки параметров β дают возможность оценить прямую регрессии (2.3.1)

$$\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 x. \quad (2.4.6)$$

Метод наименьших квадратов позволяет получить оценки для других линейных моделей с помощью выкладок, подобных проведенным выше для модели (2.3.1). Пусть, например, мы хотим воспользоваться однопараметрической моделью

$$Y_i = \beta x_i + \varepsilon_i. \quad (2.4.7)$$

Из условия минимума суммы квадратов (2.3.11) получаем оценку $\hat{\beta}$ параметра β

$$\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}. \quad (2.4.8)$$

Модель (2.4.7) естественно возникла бы при определении зависимости разности потенциалов на концах проводника (напряжения) Y от силы тока x . Закон Ома устанавливает, что эти величины пропорциональны, то есть имеет место соотношение (2.4.7). Величина $\hat{\beta}$ из (2.4.8) является оценкой сопротивления проводника, учитывающей ошибки при измерении напряжения.

2.5 Множественная регрессия

В общем случае произвольного числа s регрессоров (2.3.4), (2.3.5) будем пользоваться векторно-матричной записью (2.3.3). Величина $\beta^T X^T Y$ — скаляр, значит, $\beta^T X^T Y = (\beta^T X^T Y)^T = Y^T X \beta$. С учетом этого запишем сумму квадратов в виде

$$\varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X \beta.$$

Дифференцируя по β , получим линейное уравнение для нахождения оценки $\hat{\beta}$

$$-2X^T Y + 2X^T X \hat{\beta} = 0. \quad (2.5.1)$$

Если матрица $\mathbf{X}^T \mathbf{X}$ имеет обратную, то решение этого векторного уравнения существует, единственно и записывается в виде

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (2.5.2)$$

Вопрос о существовании обратной матрицы заслуживает обсуждения. Количество регрессоров – столбцов матрицы \mathbf{X} – в задачах регрессионного анализа должно быть не больше числа наблюдений

$$s + 1 \leq n. \quad (2.5.3)$$

На самом деле обычно n бывает существенно больше, чем $s + 1$. Будем считать, что $\text{rank}(\mathbf{X}) = s + 1$. Из линейной алгебры известно [3], что в этом случае матрица $\mathbf{X}^T \mathbf{X}$ размера $(s + 1) \times (s + 1)$ положительно определена и, следовательно, невырождена ($\text{rank}(\mathbf{X}^T \mathbf{X}) = s + 1$) и имеет обратную.

Когда $\text{rank}(\mathbf{X}) < s + 1$, столбцы матрицы \mathbf{X} линейно зависимы. Это значит, что зависимы, то есть неудачно определены, влияющие факторы — регрессоры. Радикальный способ исправления ситуации состоит в пересмотре состава регрессоров и исключении зависимых.

Случается, что $|\mathbf{X}^T \mathbf{X}|$ — определитель матрицы $\mathbf{X}^T \mathbf{X}$ — хотя и не равен, но близок к нулю, иными словами, матрица $\mathbf{X}^T \mathbf{X}$ плохо обусловлена. В этом случае малые ошибки измерения, то есть малые изменения вектора \mathbf{Y} весьма существенно влияют на результаты оценивания. Для преодоления трудностей, связанных с плохой обусловленностью задачи, разработаны методы регуля-

ризации [18]. Кроме того, предложен специальный аппарат — рекуррентный метод наименьших квадратов [19] —, позволяющий существенно улучшить вычислительный процесс.

Оценку (2.5.2) можно получить другим методом, опирающимся на геометрические представления (см. Приложение Б).

Минимальное значение $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$, то есть величина

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

называется остаточной суммой квадратов

$$\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}. \quad (2.5.4)$$

Метод наименьших квадратов дает возможность получить оценки параметров в любых линейных моделях.

Пример. Положим Y_1, Y_2 — две независимые величины со средними β и 2β соответственно. Требуется оценить β по единственному измерению величин Y_1, Y_2 .

Записываем модель в виде $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$, что совпадает с

(2.3.3), если положить $\mathbf{Y} = \|Y_1, Y_2\|^T$, $\mathbf{X} = \|1, 2\|^T$, $\boldsymbol{\varepsilon} = \|\varepsilon_1, \varepsilon_2\|^T$.

Оптимальная оценка $\hat{\beta}$ согласно (2.5.2) имеет вид

$$\hat{\beta} = \left(\|1, 2\| \left\| \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\| \right)^{-1} \|1, 2\| \mathbf{Y} = \frac{1}{5} (Y_1 + 2Y_2).$$

При этом остаточная сумма квадратов равна

$$\hat{R}_1^2 = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} = Y_1^2 + Y_2^2 - \frac{1}{5} (Y_1 + 2Y_2)^2.$$

Свойства оценок наименьших квадратов

а) Среднее

Оценка (2.5.2) является несмещенной. Действительно,

$$\mathbf{M}\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{M}\mathbf{Y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}. \quad (2.5.5)$$

При выводе (2.5.5) использовано первое из соотношений (2.3.10) и тот факт, что оператор математического ожидания \mathbf{M} линеен.

б) Дисперсия

Матрицей ковариаций оценок $\hat{\boldsymbol{\beta}}$ является с точностью до множителя матрица $(\mathbf{X}^T \mathbf{X})^{-1}$.

Пусть \mathbf{A} и \mathbf{B} произвольные матрицы, а \mathbf{X} и \mathbf{Y} — случайные векторы. Легко проверяется, что

$$\text{cov}(\mathbf{A}\mathbf{X}, \mathbf{B}\mathbf{Y}) = \mathbf{A} \text{cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}^T. \quad (2.5.6)$$

Конечно, предполагается, что размеры матриц и векторов таковы, что все операции в (2.5.6) выполнимы. Из (2.5.6) следует

$$\mathbf{D}(\mathbf{A}\mathbf{Y}) = \text{cov}(\mathbf{A}\mathbf{Y}, \mathbf{A}\mathbf{Y}) = \mathbf{A}(\mathbf{D}\mathbf{Y})\mathbf{A}^T. \quad (2.5.7)$$

Значит, с учетом второго равенства (2.3.10) получаем

$$\begin{aligned} \mathbf{D}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{D}\mathbf{Y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \end{aligned} \quad (2.5.8)$$

Что и требовалось доказать.

Параметр σ^2 в равенстве (2.5.8) является константой, которая, однако, обычно неизвестна. Оказывается, что несмещенной оценкой σ^2 служит (с точностью до множителя) остаточная сумма квадратов. Для доказательства этого, пожалуй, лучше всего

воспользоваться формулой (В.7). Преобразуем остаточную сумму квадратов

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} &= \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{I}_n - \mathbf{P}) (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \\ &+ \boldsymbol{\beta}^T \mathbf{X}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y} - \boldsymbol{\beta}^T \mathbf{X}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{X}\boldsymbol{\beta} + \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{X}\boldsymbol{\beta}. \end{aligned} \quad (2.5.9)$$

Первое равенство в этих выкладках следует из (В.7), второе получается путем алгебраических преобразований. Два последних слагаемых в правой части равны нулю вследствие (В.4). По той же причине равно 0 второе слагаемое

$$\boldsymbol{\beta}^T \mathbf{X}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y} = \left[\boldsymbol{\beta}^T \mathbf{X}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y} \right]^T = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{X}\boldsymbol{\beta} = 0.$$

Значит,

$$\begin{aligned} \mathbf{M} \left[\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} \right] &= \mathbf{M} \left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{I}_n - \mathbf{P}) (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right] = \mathbf{M} \left[\boldsymbol{\varepsilon}^T (\mathbf{I}_n - \mathbf{P}) \boldsymbol{\varepsilon} \right] = \\ &= \sum_i \sum_j q_{ij} \mathbf{M} (\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j) = \sigma^2 \sum_i q_{ii} = \sigma^2 \text{tr} (\mathbf{I}_n - \mathbf{P}) = \sigma^2 (n - s - 1). \end{aligned} \quad (2.5.10)$$

Здесь через q_{ij} обозначены элементы матрицы $\mathbf{I}_n - \mathbf{P}$. Приведенные выше выкладки используют соотношения (2.3.9), (2.3.10) и (В.4) (см. также п.В.2 приложения В).

Таким образом, несмещенной оценкой параметра σ^2 является величина

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{n - s - 1} = \frac{\| \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \|^2}{n - s - 1} = \frac{\sum_{i=1}^n \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_s x_{is} \right)^2}{n - s - 1}. \quad (2.5.11)$$

Регрессионная модель имеет $s + 1$ параметров, число $n - s - 1$ называется числом степеней свободы. Этот термин можно трактовать следующим образом. Выборка содержит n независимых наблюдений, то есть «имеет n степеней свободы». По выборке

оценивается $s+1$ параметров, что накладывает на выборку $s+1$ связей. Остается, следовательно, $n-s-1$ степеней свободы.

2.6 Распределение выборочных функций.

Коэффициент детерминации

Если пользоваться предположением (2.3.9) о распределении ошибок по нормальному закону, то возможности статистического анализа существенно расширяются.

В этом случае можно сделать вполне определенные выводы о распределении оценок коэффициентов $\hat{\beta}$ и дисперсии $\hat{\sigma}^2$, а именно

$$\hat{\beta} \in N\left(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\right), \quad (2.6.1)$$

$$(n-s-1)\hat{\sigma}^2/\sigma^2 \in \chi^2(n-s-1). \quad (2.6.2)$$

Оценка $\hat{\beta}$ является линейной функцией \mathbf{Y} (2.5.2) и, следовательно, также подчинена нормальному закону. Среднее и дисперсия $\hat{\beta}$ вычислены ранее (2.5.5), (2.5.8).

Остаточная сумма квадратов ортогональным преобразованием может быть приведена к сумме квадратов независимых нормальных величин (см. п. В.4 приложения В), отсюда, учитывая Б.4, следует справедливость (2.6.2).

Коэффициент детерминации

Ключевым соотношением регрессионного анализа служит равенство (2.6.3)

$$\Sigma(Y_i - \bar{Y})^2 = \Sigma(Y_i - \hat{Y}_i)^2 + \Sigma(\hat{Y}_i - \bar{Y})^2. \quad (2.6.3)$$

Проверим его сначала для случая одного регрессора

$$\begin{aligned}
\Sigma(Y_i - \hat{Y}_i)^2 &= \Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \Sigma[Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})]^2 = \\
&\Sigma(Y_i - \bar{Y})^2 + \hat{\beta}_1^2 \Sigma(x_i - \bar{x})^2 - 2\hat{\beta}_1 \Sigma(Y_i - \bar{Y})(x_i - \bar{x}) = \\
\Sigma(Y_i - \bar{Y})^2 + \hat{\beta}_1^2 \Sigma(x_i - \bar{x})^2 - 2\hat{\beta}_1 \Sigma \hat{\beta}_1 (x_i - \bar{x})^2 &= \Sigma(Y_i - \bar{Y})^2 - \Sigma[\hat{\beta}_1(x_i - \bar{x})]^2 = \\
&\Sigma(Y_i - \bar{Y})^2 - \Sigma(\hat{Y}_i - \bar{Y})^2.
\end{aligned}$$

При выводе использовались формулы (2.4.6), (2.4.5), (2.4.3).

Последнее равенство доказывается соотношением

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{Y} = \hat{\beta}_1 (x_i - \bar{x}),$$

которое следует из (2.4.5) и (2.4.6).

Перейдем теперь к доказательству (2.6.3) в общем случае. Примем во внимание 2 равенства. Первое получается приравниванием к нулю производной по β_0 выражения $\Sigma(Y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_s x_{is})^2$ и имеет вид $\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_s x_{is}) = 0$, то есть

$$\Sigma(Y_i - \hat{Y}_i) = 0. \quad (2.6.4)$$

Второе равенство следует из соотношений

$$\hat{\mathbf{Y}}^T \hat{\mathbf{Y}} = \mathbf{Y}^T \mathbf{P}^2 \mathbf{Y} = \mathbf{Y}^T \mathbf{P} \mathbf{Y} = \mathbf{Y}^T \hat{\mathbf{Y}},$$

в которых используется формула (B.2) $\hat{\mathbf{Y}} = \mathbf{P} \mathbf{Y}$.

Теперь легко выводим (2.6.3)

$$\begin{aligned}
\Sigma(Y_i - \bar{Y})^2 &= \Sigma(Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 = \Sigma(Y_i - \hat{Y}_i)^2 + \Sigma(\hat{Y}_i - \bar{Y})^2 + \\
&+ 2\Sigma(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}).
\end{aligned} \quad (2.6.5)$$

Последний член равен нулю. Чтобы в этом убедиться, представим его в виде суммы двух слагаемых

$$\Sigma(Y_i - \hat{Y}_i) \hat{Y}_i - \bar{Y} \Sigma(Y_i - \hat{Y}_i) = (\mathbf{Y}^T \hat{\mathbf{Y}} - \hat{\mathbf{Y}}^T \hat{\mathbf{Y}}) - \bar{Y} \Sigma(Y_i - \hat{Y}_i).$$

Как показано выше, оба слагаемых равны нулю. Соотношение (2.6.3) является эмпирическим аналогом формулы (А.18'). С точностью до множителя $1/(n-1)$ суммы в (2.6.3) являются оценками величин σ_Y^2 , σ_{YX}^2 , $\sigma_{m_Y(X)}^2$ соответственно.

Сумма в левой части равенства (2.6.3) характеризует разброс эмпирических данных около среднего значения \bar{Y} . Этот разброс представляется в виде суммы двух слагаемых правой части. Первое из них является характеристикой разброса Y_i относительно оценки прямой регрессии, а второе — разброса точек прямой регрессии относительно среднего \bar{Y} .

Качество регрессионного приближения определяется остаточной суммой квадратов $\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \Sigma (Y_i - \hat{Y}_i)^2$: чем меньше доля этого слагаемого в общей сумме $\Sigma (Y_i - \bar{Y})^2$, тем лучше приближение.

Определение. Величина

$$R^2 = \frac{\Sigma (\hat{Y}_i - \bar{Y})^2}{\Sigma (Y_i - \bar{Y})^2} = 1 - \frac{\Sigma (Y_i - \hat{Y}_i)^2}{\Sigma (Y_i - \bar{Y})^2} \quad (2.6.6)$$

называется коэффициентом детерминации.

R^2 является эмпирическим аналогом (оценкой) корреляционного отношения η_{YX}^2 (2.2.6). Для величины R^2 употребляется также термин выборочный множественный коэффициент корреляции.

Если все точки корреляционного поля (\mathbf{X}_i, Y_i) лежат на прямой (2.3.4), то $R^2 = 1$. Если же $R^2 = 0$, то $\Sigma (\hat{Y}_i - \bar{Y})^2 = 0$, что можно

интерпретировать как независимость наблюдений Y_i от регрессоров x_1, \dots, x_s .

Заметим также, что если за функцию $\eta(x)$ в (2.3.1) или в (2.3.4) взять константу $\eta = \beta_0$, то в качестве оценки β_0 по методу наименьших квадратов получим

$$\hat{\beta}_0 = \bar{Y}. \quad (2.6.7)$$

Значит, левая часть (2.6.3) служит характеристикой регрессионного приближения эмпирических данных при отсутствии регрессоров x_1, \dots, x_s .

Соотношения (2.6.3), (2.6.6) могут быть также записаны с помощью векторных обозначений. Введем вектор $\mathbf{e}_n^T = \|1, 1, \dots, 1\|$, все компоненты которого равны 1. Нижний индекс равен количеству компонент. Тогда (2.6.3) примет вид

$$\|\mathbf{Y} - \mathbf{e}_n \bar{Y}\|^2 = \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 + \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{e}_n \bar{Y}\|^2 = \|\hat{\boldsymbol{\varepsilon}}\|^2 + \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{e}_n \bar{Y}\|^2, \quad (2.6.8)$$

а коэффициент детерминации запишется как

$$R^2 = 1 - \frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{\|\mathbf{Y} - \mathbf{e}_n \bar{Y}\|^2}. \quad (2.6.9)$$

В случае нормальной выборки все суммы в (2.6.3) имеют распределение χ^2 (с точностью до множителя σ^2). В соответствии с (2.6.2) $\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} / \sigma^2 \in \chi^2(n-s-1)$. В приложении Г доказано, что $\Sigma(\hat{Y}_i - \bar{Y})^2 / \sigma^2 \in \chi^2(n-1)$. Следствие Б.1 позволяет утверждать, что $\Sigma(\hat{Y}_i - \bar{Y})^2 / \sigma^2 \in \chi^2(s)$. То, что число степеней свободы рассматри-

ваемой величины $\Sigma(\hat{Y}_i - \bar{Y})^2$ определено правильно, можно подкрепить таким рассуждением. Для оценки \hat{Y}_i надо оценить $s+1$ параметров $\beta_0, \beta_1, \dots, \beta_s$, но величины $\hat{Y}_i - \bar{Y}$ связаны одной линейной зависимостью $\Sigma \hat{Y}_i - n\bar{Y} = 0$. Таким образом, число степеней свободы $\Sigma(\hat{Y}_i - \bar{Y})^2$ должно быть равно $(s+1) - 1 = s$.

При проверке гипотез в дальнейшем будем использовать распределение Фишера для отношения выборочных функций, входящих в соотношение (2.6.3). Например,

$$F = \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{s} : \frac{\Sigma(Y_i - \hat{Y}_i)^2}{n-s-1} \in F(s, n-s-1). \quad (2.6.10)$$

2.7 Техника регрессионного анализа. Доверительные интервалы

Доверительные интервалы для коэффициентов β . Интервальные оценки для коэффициентов β строятся на основании соотношений (2.6.1), (2.6.2), (2.5.10) в предположении (2.3.9) о распределении ошибок ϵ по нормальному закону.

Для случая одного регрессора матрица ковариаций $\mathbf{D}\hat{\beta}$ из (2.5.8) имеет вид

$$\mathbf{D}[\hat{\beta}] = \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2} \begin{vmatrix} (1/n)\Sigma x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{vmatrix}.$$

Построим доверительный интервал для коэффициента β_1 (2.4.3). При известной дисперсии σ^2 с этой целью использовалось бы нормальное распределение, так как

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\beta_1}} = \frac{\hat{\beta}_1 - \beta_1}{\sigma / \sqrt{\sum(x_i - \bar{x})^2}} \in N(0,1). \quad (2.7.1)$$

Поскольку величина σ^2 обычно бывает неизвестна, в знаменателе (2.7.1) вместо σ ставится ее оценка (2.5.11), которая в этом случае имеет вид $\hat{\sigma}^2 = R_1^2 / (n-2)$. Нормальное распределение при этом заменяется t -распределением Стьюдента

$$(\hat{\beta}_1 - \beta_1) \sqrt{\sum(x_i - \bar{x})^2} : R_1 / \sqrt{n-2} \in t(n-2). \quad (2.7.2)$$

В общем случае для доверительного интервала коэффициента β_j вместо (2.7.2) следует воспользоваться соотношением

$$\frac{\hat{\beta}_j - \beta_j}{d_{jj} \sqrt{R_{n-s-1}^2 / (n-s-1)}} \in t(n-s-1), \quad (2.7.3)$$

где d_{jj} — элемент матрицы $(\mathbf{X}^T \mathbf{X})^{-1}$, стоящий на диагонали на пересечении строки и столбца с номерами $j+1$.

Доверительный интервал, построенный в соответствии с (2.7.2), (2.7.3), может понадобиться для проверки гипотезы $H_0 : \beta_j = c$. Если интервал с доверительной вероятностью γ не покрывает значение c , то гипотеза H_0 не отвергается с уровнем значимости $\alpha = 1 - \gamma$ при альтернативе $H_1 : \beta_j \neq c$.

Следует обратить внимание на то, что оценки коэффициентов β_i, β_j не являются независимыми. Поэтому прямое произведение доверительных интервалов для β_i, β_j не будет совпадать с совместной доверительной областью для этих параметров. Ситуация иллюстрируется рис. 2.2, где совместная доверительная

область является эллипсом, а прямое произведение доверительных интервалов для β_i и β_j — прямоугольником.

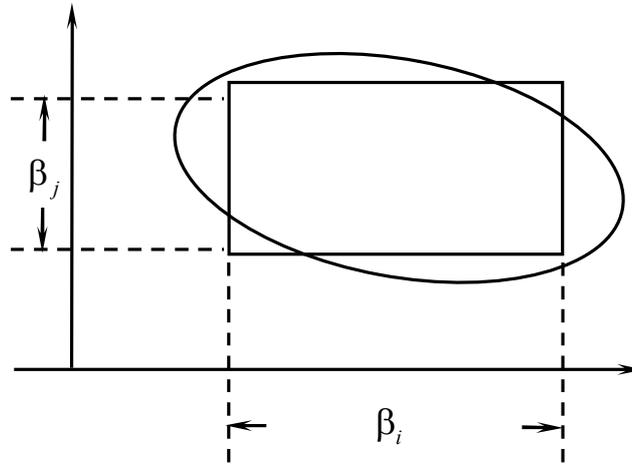


Рисунок 2.2 — Доверительные интервалы для параметров β_i , β_j и совместная доверительная область

Доверительная полоса для прямой регрессии

Оценка прямой регрессии в одномерном случае (регрессор x) имеет вид

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x. \quad (2.7.4)$$

При фиксированном x \hat{Y} является нормально распределенной величиной с математическим ожиданием $\mathbf{M}\hat{Y} = \beta_0 + \beta_1 x$ и дисперсией

$$\mathbf{D}\hat{Y} = \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) = \sigma^2 v(x), \quad (2.7.5)$$

что следует из (2.5.8) как частный случай. Очевидно, что $\mathbf{D}\hat{Y}$ принимает минимальное значение при $x = \bar{x}$. Величина

$$\frac{\hat{Y} - \mathbf{M}\hat{Y}}{\hat{\sigma} \sqrt{v(x)}} \in t(n-2). \quad (2.7.6)$$

Следовательно, γ -доверительный интервал для $\beta_0 + \beta_1 x$ имеет вид

$$\hat{Y} \pm t_\gamma (n-2) \hat{\sigma} \sqrt{v(x)}, \quad (2.7.7)$$

где t_γ — γ -квантиль распределения Стьюдента.

На рис. 2.3 изображена прямая регрессии и доверительная полоса (пунктиром), получающаяся из (2.7.7), где функция $v(x)$ определена в соответствии с (2.7.5).

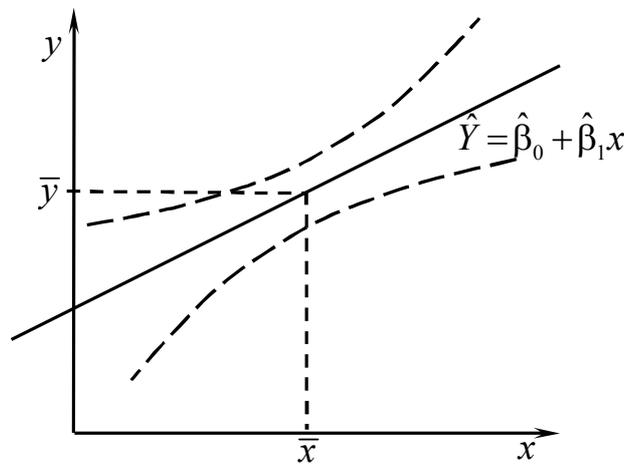


Рисунок 2.3 — Доверительная полоса для прямой регрессии

В случае более чем одного регрессора следует говорить не о прямой регрессии, а о плоскости

$$\eta(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_s x_s = \mathbf{x}^T \boldsymbol{\beta}, \quad (2.7.8)$$

оценкой которой служит

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_s x_s = \mathbf{x}^T \hat{\boldsymbol{\beta}}, \quad (2.7.9)$$

где \mathbf{x} — $(s+1)$ -мерный вектор $\|\mathbf{x}\| = \|1, x_1, \dots, x_s\|^T$.

Здесь опять-таки \hat{Y} в предположении (2.3.9) является нормальной величиной со средним $\eta(x)$. Дисперсию величины \hat{Y} найдем с использованием (2.5.7), (2.5.8)

$$\mathbf{D}\hat{Y} = \mathbf{D}(\mathbf{x}^T \hat{\boldsymbol{\beta}}) = \mathbf{x}^T \mathbf{D}\hat{\boldsymbol{\beta}}\mathbf{x} = \sigma^2 \mathbf{x}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}^T = \sigma^2 v(\mathbf{x}). \quad (2.7.10)$$

Таким образом, соотношения (2.7.6), (2.7.7) имеют место и в этом случае с заменой числа степеней свободы на $n - s - 1$ при $v(\mathbf{x}) = \mathbf{x}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}^T$ то есть доверительная полоса имеет вид

$$\hat{Y} \pm t_\gamma (n - s - 1) \hat{\sigma} \sqrt{v(\mathbf{x})}. \quad (2.7.11)$$

Доверительный интервал для отклика

Соотношения (2.7.7), (2.7.11) характеризуют полосу, в которой с вероятностью γ находится прямая (плоскость) регрессии. Однако прямая регрессии нужна обычно не сама по себе, а для прогнозирования. Прогноз Y_* представляется в виде суммы \hat{Y} и ошибки ε_* . Предполагая, что $\varepsilon_* \in N(0, \sigma^2)$ и $\text{cov}(\varepsilon_*, \varepsilon_i) = 0$ ($i = 1, \dots, n$), получаем для Y_*

$$\mathbf{M}Y_* = \mathbf{M}\hat{Y} = \eta(\mathbf{x}); \quad \mathbf{D}Y_* = \mathbf{D}\hat{Y} + \mathbf{D}\varepsilon_* = \sigma^2(v(\mathbf{x}) + 1). \quad (2.7.12)$$

В соответствии с этим доверительный интервал для отклика будет шире, чем доверительный интервал для прямой регрессии при том же \mathbf{x}

$$\hat{Y} \pm t_\gamma (n - s - 1) \hat{\sigma} \sqrt{v(\mathbf{x}) + 1}. \quad (2.7.13)$$

2.8 Техника регрессионного анализа. Проверка гипотез

Проверка гипотезы о значимости регрессии

Вопрос о значимости регрессии можно сформулировать в одномерном случае так: имеет ли место тенденция к увеличению или уменьшению отклика Y при увеличении независимого пере-

менного x . Для ответа на этот вопрос полезно изобразить корреляционное поле, то есть построить точки выборки (x_i, Y_i) ($i = 1, \dots, n$) в декартовых координатах. На рис. 2.4 изображены 4 примера.

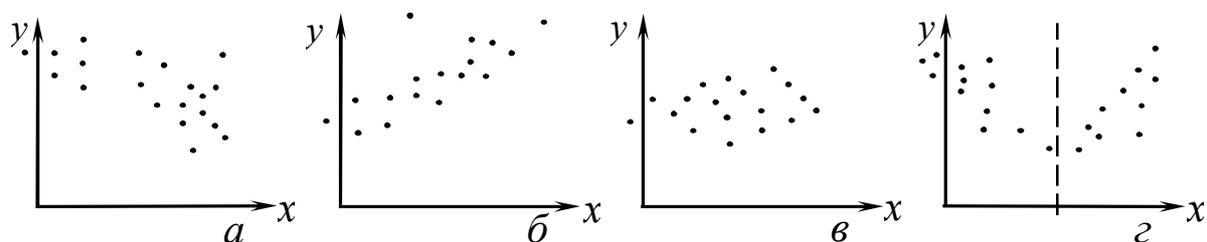


Рисунок 2.4 — Примеры корреляционных полей

Визуально представляется, что в случаях a и $б$ регрессия должна быть значимой. В первом из них наблюдается тенденция к уменьшению Y с ростом x , а во втором — тенденция к увеличению Y . В случае $в$ ответ о значимости регрессии должен быть, скорее всего, отрицательным. В случае $г$ те части выборки, которые изображаются точками слева и справа от разделяющей пунктирной вертикали имеют определенную тенденцию: слева к уменьшению Y , а справа — к увеличению, но корреляционное поле в целом свидетельствует, скорее всего, о незначимости регрессии.

Для формальной проверки значимости регрессии используется соотношение (2.6.3). При нормальном распределении ошибок каждая из трех сумм в (2.6.3) распределена по χ^2 с числом степеней свободы $n-1, n-2, 1$ соответственно. Отметим, что сумма степеней свободы величин в правой части (2.6.3) равна числу степеней свободы величины в левой части.

Если регрессия значима, то члены выборки (точки корреляционного поля) сгруппированы вокруг прямой регрессии и доля первого слагаемого в правой части (2.6.3) должна быть невелика. Гипотеза о незначимости регрессии формулируется как $H_0: \beta_1 = 0$. Рассмотрим ее при альтернативе $H_1: \beta_1 \neq 0$. Для проверки гипотезы используем распределение Фишера, которому подчинена статистика

$$F = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{1} : \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}. \quad (2.8.1)$$

Гипотезу о незначимости регрессии следует отвергнуть, когда статистика F велика, то есть доля слагаемого $\sum(Y_i - \hat{Y}_i)^2$ в разложении (2.6.3) мала. В соответствии со стандартной техникой проверки гипотез поступаем следующим образом.

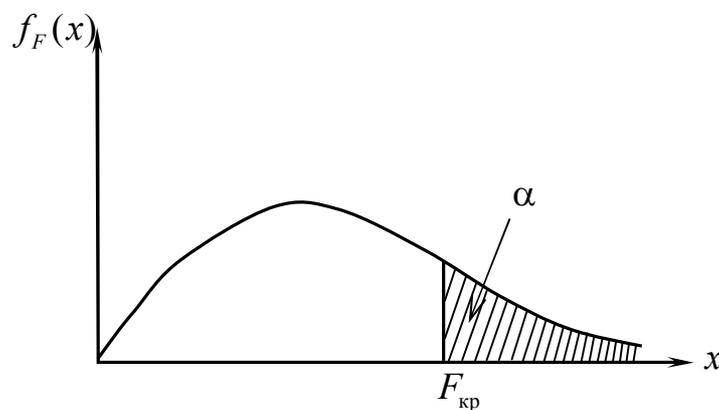


Рисунок 2.5 — Определение критического значения F -распределения Фишера

Задаемся уровнем значимости α , по таблицам F -распределения с 1 и $n-2$ степенями свободы находим отвечающую вероятности $1-\alpha$ квантиль $F_{кр} = F_{кр}(1, n-2, 1-\alpha)$. Заштрихованная площадь на рис. 2.5 равна α . Критическая область для гипотезы

H_0 при альтернативе H_1 состоит из больших значений критерия $F > F_{кр}$. При попадании в эту область гипотеза H_0 отвергается.

В общем случае модели с s регрессорами гипотеза о незначимости регрессии имеет вид $H_0 : \beta_1 = \beta_2 = \dots = \beta_s = 0$. Проверка гипотезы производится по критерию

$$F = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{1} : \frac{\sum(Y_i - \hat{Y}_i)^2}{n - s - 1}. \quad (2.8.2)$$

Статистика F распределена по Фишеру с числом степеней свободы 1 и $n - s - 1$, то есть $F \in F(1, n - s - 1)$.

Проверка гипотезы о значимости группы регрессоров

Первоначальный выбор регрессоров может быть неудачным. Если есть сомнения в целесообразности включения в модель одного или нескольких регрессоров, то формальную проверку можно провести примерно так же, как проверку значимости регрессии.

Рассмотрим общую модель регрессии (2.3.3), (2.3.5) с s регрессорами и свободным членом β_0 . Проверим гипотезу о значимости p ($p < s + 1$) регрессоров. Заново перенумеруем регрессоры так, чтобы проверяемые на значимость регрессоры получили наибольшие номера. Тем самым матрица \mathbf{X} разобьется на 2 блока $\mathbf{X} = \|\mathbf{X}^{(1)}, \mathbf{X}^{(2)}\|$ с $s + 1 - p$ и p столбцами соответственно. В число последних p столбцов можно включить также столбец, все элементы которого равны 1, то есть к числу проверяемых на значи-

мость параметров модели отнести свободный член β_0 . Техника проверки при этом не изменится.

Таким образом, задача о значимости группы факторов свелась к сопоставлению двух моделей с матрицами \mathbf{X} и $\mathbf{X}^{(1)}$ (модели Д.1, Д.3 приложения Д). Качество моделей определяется остаточными суммами квадратов, которые связаны соотношением (Д.17). Это соотношение можно записать также в форме

$$\Sigma(Y_i - \hat{Y}_i^{(1)})^2 - \Sigma(Y_i - \hat{Y}_i)^2 = \Sigma(\hat{Y}_i - \hat{Y}_i^{(1)})^2, \quad (2.8.3)$$

а левую часть равенства в соответствии с (Д.16) обозначить $R_{s+1-p}^2 - R_{s+1}^2$. Нижний индекс в обозначении остаточной суммы квадратов R^2 соответствует числу параметров модели.

Если факторы $X_{s-p+1}, \dots, X_{s+1}$ не оказывают существенного влияния на отклик, то их добавление к модели (Д.3) не приведет к значимому уменьшению остаточной суммы квадратов. В этом случае величина $R_{s+1-p}^2 - R_{s+1}^2$ будет мала. В приложении Д показано, что каждая сумма в равенстве (2.8.3) (с точностью до множителя σ^2) имеет распределение χ^2 с числами степеней свободы $n - (s + 1 - p)$, $n - (s + 1)$ и p соответственно. Для чисел степеней свободы выполняется то же соотношение, что и для соответствующих сумм в (2.8.3) $[n - (s + 1 - p)] - [n - (s + 1)] = p$.

Гипотеза о незначимости факторов $X_{s-p+1}, \dots, X_{s+1}$ формулируется в виде $H_0 : \beta_{s-p+1} = \dots = \beta_{s+1} = 0$. Для ее проверки (при аль-

тернативе H_1 , состоящей в том, что эти равенства не выполняются) используется статистика

$$F = \frac{R_{s+1-p}^2 - R_{s+1}^2}{p} : \frac{R_{s+1}^2}{n - (s + 1)} \in F(p, n - s - 1), \quad (2.8.4)$$

которая имеет распределение Фишера с p и $n - s - 1$ степенями свободы. Критической областью здесь будут большие значения статистики F . Если же величина $R_{s+1-p}^2 - R_{s+1}^2$ относительно невелика, то оснований для того, чтобы отвергнуть гипотезу H_0 , нет.

Замечание. Если гипотеза о незначимости факторов $X_{s-p+1}, \dots, X_{s+1}$ не была отвергнута, то из этого не следует, что эти факторы надо сразу же удалять из модели. Целесообразно продолжить исследование, проанализировав каждый фактор из этой совокупности в отдельности. Для этого можно, как показано выше, построить доверительный интервал для фактора и проверить, накрывает ли он значение 0. Процедуры статистического анализа в какой-то степени неустойчивы. Может оказаться, что группа факторов отвергается, но из них один (или несколько) факторов имеет все-таки смысл оставить в модели.

Проверка гипотезы о наличии зависимостей между факторами

Проверка того, имеют ли место линейные взаимосвязи между параметрами модели, основывается на той же идее, что и проверка значимости факторов. А именно, наряду с первоначальной моделью рассматривается модель, в которой учтены взаимосвязи. Число степеней свободы в этой модели меньше, чем в первоначальной.

чальной. Гипотеза принимается или отбрасывается в зависимости от того, существенно или несущественно увеличивается остаточная сумма квадратов при введении ограничений. Для формальной проверки используется F -критерий, статистика строится по аналогии с (2.8.4).

Пример. В модели $Y_1 = \beta_1 + \varepsilon_1$, $Y_2 = 2\beta_1 - \beta_2 + \varepsilon_2$, $Y_3 = \beta_1 + 2\beta_2 + \varepsilon_3$ проверим гипотезу $H_0 : \beta_1 = \beta_2$.

В матричной форме модель записывается как

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}, \text{ или } \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Отсюда получаем $\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix}$, $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{6} (Y_1 + 2Y_2 + Y_3) \\ \frac{1}{5} (-Y_2 + 2Y_3) \end{pmatrix}$. Оста-

точная сумма квадратов $R_2^2 = Y_1^2 + Y_2^2 + Y_3^2 - 6\hat{\beta}_1^2 - 5\hat{\beta}_2^2$.

Если же $\beta_1 = \beta_2 = \beta$, то в соответствии с методом наименьших квадратов для оценки неизвестного параметра β надо найти минимум функции

$$(Y_1 - \beta)^2 + (Y_2 - \beta)^2 + (Y_3 - 3\beta)^2.$$

Приравнивая к нулю производную этой функции по β , получаем оценку $\hat{\beta} = (Y_1 + Y_2 + 3Y_3)/11$. Остаточная сумма квадратов при этом будет равна $R_1^2 = (Y_1 - \hat{\beta})^2 + (Y_2 - \hat{\beta})^2 + (Y_3 - 3\hat{\beta})^2$. Для

проверки гипотезы $H_0 : \beta_1 = \beta_2$ следует воспользоваться статистикой $F = \frac{R_1^2 - R_2^2}{1} : \frac{R_2^2}{1}$, которая распределена по Фишеру $F(1;1)$.

Проведем расчет при $Y^T = \|1, 1; 1, 5; 3, 1\|$. Оценки в общей модели оказываются равными $\hat{\beta}_1 = 1, 2; \hat{\beta}_2 = 0, 94$. При этом остаточная сумма квадратов $R_2^2 = 0, 012$. Если же гипотеза H_0 справедлива, то $\hat{\beta} = 11, 9/11 = 1, 082$, а остаточная сумма квадратов $R_1^2 = 0, 1964$. Таким образом, статистика $F = \frac{0, 1964 - 0, 012}{1} : \frac{0, 012}{1} = 15, 4$. По таблицам F -распределения находим при 5% уровне значимости $F_{кр}(1; 1; 0, 95) = 161$. Значит, гипотеза отвергается, а эмпирическое значение критерия далеко отстоит от границы области принятия гипотезы. Конечно, этот пример носит чисто иллюстративный характер, так как всего по трем наблюдениям оценивается 2 параметра.

Проверка адекватности регрессионной модели

Такую проверку можно осуществить в случае наличия повторных наблюдений при некоторых значениях регрессоров. Регрессионная модель строится зачастую, как уже говорилось, по принципу «черного ящика», при отсутствии убедительных данных, обосновывающих ее вид. Поэтому любые доказательства адекватности обычно бывают полезными.

Итак, пусть имеется n различных наборов регрессоров, при которых проводятся наблюдения. Эти значения представляются строками матрицы X . При фиксированных значениях регрессо-

ров выполняется несколько экспериментов (наблюдений). Y_{ij} — результат j -го наблюдения при i -м наборе регрессоров. Регрессионная модель записывается в виде

$$Y_{ij} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_s x_{is} + \varepsilon_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, k_i. \quad (2.8.5)$$

Общее количество наблюдений равно $k_1 + \dots + k_n = K$.

Оценки коэффициентов регрессии, как и ранее, находятся по методу наименьших квадратов. Запишем (2.8.5) в виде

$$Y_{ij} = \eta_i + \varepsilon_{ij}, \quad i = 1, \dots, n; \quad j = 1, \dots, k_i. \quad (2.8.6)$$

Применяя метод наименьших квадратов к модели (2.8.6), приходим к задаче минимизации суммы $\sum_i \sum_j (Y_{ij} - \eta_i)^2$. Отсюда получаем оценки

$$\hat{\eta}_i = \frac{1}{k_i} \sum_{j=1}^{k_i} Y_{ij} = \bar{Y}_{i\bullet}$$

и остаточную сумму квадратов $\sum_i \sum_j (Y_{ij} - \bar{Y}_{i\bullet})^2$.

Далее, для определения оценок коэффициентов получаем задачу минимизации суммы

$$\sum_i (\bar{Y}_{i\bullet} - \beta_0 - \beta_1 x_{i1} - \dots - \beta_s x_{is})^2.$$

Как нам известно, эта задача имеет решение

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{Y}}_{\bullet},$$

при котором остаточная сумма квадратов равна

$$\sum_i (\bar{Y}_{i\bullet} - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_s x_{is})^2 = \|\bar{\mathbf{Y}}_{\bullet} - \mathbf{X} \hat{\beta}\|^2 = \|\bar{\mathbf{Y}}_{\bullet} - \hat{\mathbf{Y}}\|^2.$$

Имеет место следующая формула

$$\sum_{i=1}^n \sum_{j=1}^{k_i} (Y_{ij} - \hat{Y}_i)^2 = \sum_{i=1}^n \sum_{j=1}^{k_i} (Y_{ij} - \bar{Y}_{i\bullet})^2 + \sum_{i=1}^n k_i (\bar{Y}_{i\bullet} - \hat{Y}_i)^2. \quad (2.8.7)$$

Действительно, представим $(Y_{ij} - \hat{Y}_i)^2 = [(Y_{ij} - \bar{Y}_{i\bullet}) + (\bar{Y}_{i\bullet} - \hat{Y}_i)]^2 =$
 $= (Y_{ij} - \bar{Y}_{i\bullet})^2 + (\bar{Y}_{i\bullet} - \hat{Y}_i)^2 + 2(Y_{ij} - \bar{Y}_{i\bullet})(\bar{Y}_{i\bullet} - \hat{Y}_i).$

Суммируя удвоенные произведения, получаем
 $\sum_{i=1}^n \sum_{j=1}^{k_i} (Y_{ij} - \bar{Y}_{i\bullet})(\bar{Y}_{i\bullet} - \hat{Y}_i) = \sum_{i=1}^n (\bar{Y}_{i\bullet} - \hat{Y}_i) \sum_{j=1}^{k_i} (Y_{ij} - \bar{Y}_{i\bullet}) = 0$, откуда и вытекает справедливость (2.8.7).

В предположении $\varepsilon_{ij} \in N(0, \sigma^2)$ каждая сумма в соотношении (2.8.7) имеет с точностью до множителя σ^2 распределение χ^2 с числами степеней свободы $K - (s + 1)$, $K - n$ и $n - (s + 1)$ соответственно.

Равенство (2.8.7) устанавливает, что разброс наблюдений вокруг прямой (плоскости) регрессии представляется в виде суммы двух слагаемых

$$R_s^2 = R_{\text{ист}}^2 + R_{\text{н.а.}}^2. \quad (2.8.8)$$

Первое из них, $R_{\text{ист}}^2$ характеризует разброс наблюдений при фиксированных значениях регрессоров и определяется точностью замеров (наблюдений). Второе, $R_{\text{н.а.}}^2$ характеризует степень близости наблюдений к прямой (плоскости) регрессии. Если бы все средние $\bar{Y}_{i\bullet}$ лежали на этой прямой (плоскости), то величина $R_{\text{н.а.}}^2$ была бы равна нулю. Таким образом, $R_{\text{н.а.}}^2$ служит мерой неадекватности модели.

Для проверки адекватности воспользуемся F -статистикой

$$F = \frac{R_{\text{н.а.}}^2}{n-s-1} : \frac{R_{\text{ист}}^2}{K-n} \in F(n-s-1, K-n). \quad (2.8.9)$$

При уровне значимости α гипотезу об адекватности следует отвергнуть, если вычисленное по выборке значение F превышает критическое значение, то есть $(1-\alpha)$ -квантиль распределения Фишера $F_{\text{кр}}(n-s-1, K-n, 1-\alpha)$.

2.9 Полиномиальная регрессия

Полиномиальная регрессия получается из общей модели регрессии, если положить регрессоры x_1, \dots, x_s равными степеням некоторого переменного x

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_s x_i^s + \varepsilon_i \quad (i = 1, \dots, n). \quad (2.9.1)$$

Ранг матрицы \mathbf{X} должен быть равен $s+1$.

С увеличением степени полинома повышается точность приближения облака точек (x_i, Y_i) , однако эта точность должна быть соразмерна с точностью исходной информации. При повышении степени полинома остаточная сумма квадратов уменьшается. Значимо ли это снижение, проверяется по F -критерию (2.8.4).

Пусть R_{s+1}^2 — остаточная сумма квадратов при использовании полинома степени s , а R_s^2 — в предположении $\beta_s = 0$. Тогда статистика F

$$F = \frac{R_s^2 - R_{s+1}^2}{1} : \frac{R_{s+1}^2}{n-(s+1)} \in F(1, n-s-1) \quad (2.9.2)$$

распределена по Фишеру с 1 и $n-s-1$ степенями свободы. Гипотеза $H_0 : \beta_s = 0$ отвергается при больших значениях критерия

$F > F_{\text{кр}}(1, n - s - 1, 1 - \alpha)$, где α — уровень значимости.

Точно так же можно проверять целесообразность отбрасывания сразу нескольких коэффициентов β_k , причем необязательно наивысших степеней. Рассмотрим, например, модель $\eta(x) = \beta_0 + \beta_1 x + \beta_2 x^2$. Может оказаться, что отбрасывать последний член не имеет смысла, а при отбрасывании линейного по x слагаемого $\beta_1 = 0$ остаточная сумма квадратов уменьшится незначительно. В этом случае оправдан переход от трехчленной модели к двухчленной $\eta(x) = \beta_0 + \beta_2 x^2$. Эта модель проще исходной, она содержит всего 2 параметра и линейна по x^2 .

Пример. При обработке данных о суточном потреблении газа крупным городом ($n = 20$) в зависимости от температуры x окружающей среды были получены результаты, представленные в табл. 2.1.

Примем уровень значимости равным 0,05. Последовательно ответим на вопросы о целесообразности включения в модель параметров $\beta_1, \beta_2, \beta_3$ из модели (2.9.1) при $s = 3$.

а. Значима ли регрессия? В модели $Y_i = \beta_0 + \beta_{1i} x_i$ проверяем гипотезу $H_0: \beta_1 = 0$. Находим критерий (2.9.2) при $s = 1$

$$F = \left[\frac{(\hat{R}_1^2 - \hat{R}_2^2)}{1} \right] : \left[\frac{\hat{R}_2^2}{(n - 2)} \right] = (3,91/1) : (5,94/18) = 11,85.$$

$F_{\text{кр}}(1; 18; 0,95) = 4,41 < F$, значит, гипотеза $H_0: \beta_1 = 0$ отвергается.

Гипотеза была бы отвергнута и при уровне значимости 0,01

$$F_{\text{кр}}(1; 18; 0,99) = 8,28.$$

б. Достаточна ли линейная регрессия? В модели $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ проверяем гипотезу $H_0 : \beta_2 = 0$. Находим критерий

$$(2.9.2) \quad \text{при } s = 2. \quad F = \left[\left(\hat{R}_2^2 - \hat{R}_3^2 \right) / 1 \right] : \left[\hat{R}_3^2 / (n - 3) \right] = (1,28/1) :$$

$$: (4,66/17) = 4,67. \quad F_{\text{кр}}(1;17;0,95) = 4,45 < F, \quad \text{значит, гипотеза}$$

$H_0 : \beta_2 = 0$ отвергается. При уровне значимости 0,01 гипотеза не была бы отвергнута, так как $F_{\text{кр}}(1;17;0,99) = 8,40$.

в. Достаточна ли квадратичная (трехчленная) регрессия? Проверяем гипотезу $H_0 : \beta_3 = 0$. Находим критерий (2.9.2) при

$$s = 3. \quad F = \left[\left(\hat{R}_3^2 - \hat{R}_4^2 \right) / 1 \right] : \left[\hat{R}_4^2 / 16 \right] = 3,32. \quad F_{\text{кр}}(1;16;0,95) = 4,49,$$

следовательно, гипотеза $\beta_3 = 0$ не отвергается, для приближения рекомендуется пользоваться полиномом 2-й степени. Оценка кривой регрессии имеет вид $\hat{Y} = 15,53 - 0,1641x - 0,0064x^2$.

Таблица 2.1 — Пример: подбор степени полиномиальной регрессии

s	R_{s+1}^2	$R_s^2 - R_{s+1}^2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0	9,85		15,70			
1	5,94	3,91	15,48	-0,0878		
2	4,66	1,28	15,53	-0,1641	-0,0064	
3	3,86	0,80	15,74	-0,2290	-0,0316	-0,00129

Использованная выше логика рассуждений может давать сбои. Может оказаться, что при увеличении степени полинома на 1 эффект оказывается незначительным, тогда как при дальнейшем увеличении s остаточная сумма квадратов $R_s^2 - R_{s+1}^2$ уменьшится существенно. Поэтому иногда берут максимальное

значение s (в приведенном примере можно было взять $s = 4$) и последовательно отбрасывают члены с наибольшими степенями, пока снижение остаточной суммы квадратов не станет значимым.

2.10 Нелинейные приближения. Подбор эмпирических зависимостей

Модель регрессионного анализа должна быть линейной относительно неизвестных параметров β . Однако каноническая запись модели в форме (2.3.3), где \mathbf{X} и β представлены соотношениями (2.3.5), не является единственно возможной. Регрессоры x_1, \dots, x_s выбираются первоначально в соответствии с содержательной постановкой проблемы. Но при этом линейность по \mathbf{x} в модели (2.3.4) не является обязательным условием. Естественным обобщением модели (2.3.4) является модель

$$\eta(\mathbf{x}) = \beta_1 \varphi_1(\mathbf{x}) + \dots + \beta_q \varphi_q(\mathbf{x}) \quad (2.10.1)$$

Функции $\varphi_1(\mathbf{x}), \dots, \varphi_q(\mathbf{x})$ компонент вектора могут быть приняты за новые регрессоры, которые, однако, не обязательно будут иметь естественную интерпретацию. Модель (2.10.1), расширяя границы регрессионного анализа, не вносит новых проблем, техника исследования не изменяется.

Еще один прием состоит в том, что вместо Y пользуются некоторыми функциями от Y . Рассмотрим примеры нелинейных моделей, сводящихся к моделям линейным, при этом ограничимся пока случаем одного регрессора. Вид используемой функции может быть подсказан структурой корреляционного поля, физическими, технологическими соображениями или, наконец, сло-

жившимися традициями. Рассмотрим степенную, показательную, дробно-линейную функцию и некоторые их комбинации.

Степенная функция

$$y = \alpha_1 x^{\alpha_2} \quad (2.10.2)$$

Перейдем к логарифмам в этом равенстве $\ln y = \ln \alpha_1 + \alpha_2 \ln x$ и обозначим $y' = \ln y$; $\beta_0 = \ln \alpha_1$; $\beta_1 = \alpha_2$; $x' = \ln x$. Тогда с точностью до обозначений получим модель регрессионного анализа (2.3.2)

$$y' = \beta_0 + \beta_1 x' + \varepsilon. \quad (2.10.3)$$

Возникает вопрос о законности применимости этой модели. Неотрицательность переменных y и x обычно обеспечивается, так как иначе степенная функция просто бы не использовалась. Сложнее дело обстоит с остатками ε_i , которые должны удовлетворять условиям (2.3.6 – 2.3.9). Если первоначальную модель выбрать в виде $y = \varepsilon_1 \alpha_1 x^{\alpha_2}$, то будут ли остатки $\varepsilon = \ln \varepsilon_1$ удовлетворять предположениям 1 – 4 п. 2.3? Анализ выполнения этих предположений проводится на основании статистического материала. Некоторые рекомендуемые методы анализа изложены ниже.

Показательная функция

$$y = \alpha_1 e^{\alpha_2 x} \quad (2.10.4)$$

Прологарифмировав, вводим новые координаты и параметры $y' = \ln y$, $\beta_0 = \ln \alpha_1$, $\beta_1 = \alpha_2$, $x' = x$. Как и в предыдущем случае, приходим к модели (2.10.3).

Дробно-линейная функция (частный случай)

$$y = 1/(\beta_0 + \beta_1 x) \quad (2.10.5)$$

Положив $y' = 1/y$, $x' = x$, приходим к модели (2.10.3). То же преобразование годится, если в знаменателе вместо $\beta_0 + \beta_1 x$ стоит полином степени $s > 1$.

Полином по степеням x^{-1}

$$y = \beta_0 + \frac{\beta_1}{x} + \dots + \frac{\beta_s}{x^s} \quad (2.10.6)$$

Эта зависимость является частным случаем формулы (2.10.1). Положив $y' = y$, $x' = 1/x$, получаем в новых переменных стандартную полиномиальную модель.

Комбинация степенной и показательной функции

$$y = \alpha_0 x^{\alpha_1} e^{\alpha_2 x} \quad (2.10.7)$$

Новые переменные и параметры вводим по формулам $y' = \ln y$, $x'_1 = \ln x$, $x'_2 = x$, $\beta_0 = \ln \alpha_0$, $\beta_1 = \alpha_1$, $\beta_2 = \alpha_2$. При этом получим стандартную модель с регрессорами 1 , x , $\ln x$.

Использованные выше преобразования могут оказаться также полезными в случае, когда имеется не один, а несколько регрессоров.

Степенная функция многих переменных

Пусть имеем

$$y = \alpha_0 x_1^{\alpha_1} \dots x_s^{\alpha_s} . \quad (2.10.8)$$

Для линеаризации достаточно, перейдя к логарифмам, положить $y' = \ln y$, $x'_k = \ln x_k$, $\beta_0 = \ln \alpha_0$, $\beta_k = \alpha_k$.

Полуэмпирические модели

В начале этого раздела регрессионный анализ был представлен как естественный метод установления зависимости между входными x_k и выходной Y переменными в модели «черного ящика», то есть в ситуации, когда нет априорной информации о взаимосвязи между переменными.

Но иногда случается, что форма функциональной зависимости между переменными, включенными в исследование, известна, например, определена физическими или другими законами природы. Тогда, руководствуясь этими законами, пишут формулу (модель), в которую включают неизвестные константы (параметры). Если параметры входят в формулу линейно, то их оценки получаются методами регрессионного анализа по результатам наблюдений или специально проведенного эксперимента. Такие модели называются полуэмпирическими, или моделями «серого ящика».

Пример. Процессы течения флюидов в трубопроводах хорошо изучены. Так, для течений газа в магистральном газопроводе пользуются следующей взаимосвязью между значениями давления в начале (P), в конце (p) трубопровода и расходом (q) по нему

$$P^2 - p^2 = \lambda c q^2.$$

Здесь c — константа, определяемая длиной и диаметром трубопровода, свойствами газа, его температурой и некоторыми другими физическими величинами, а λ — коэффициент гидрав-

лического сопротивления. Сопротивление зависит от внутреннего покрытия трубы, состояния ее поверхности, наличия твердых и жидких отложений в полости. Коэффициент λ меняется в процессе эксплуатации и называется при таком подходе фактическим коэффициентом гидравлического сопротивления λ_{ϕ} . Определить коэффициент λ_{ϕ} можно по замерам давления и расхода, которые производятся со случайными ошибками и, следовательно, должны рассматриваться как случайные величины. Эта модель может быть исследована методами регрессионного анализа. Полагая $Y = cq^2$, $x = P^2 - p^2$, $\beta = 1/\lambda_{\phi}$, получаем

$$Y = \beta x. \quad (2.10.9)$$

В качестве независимой выбрана переменная x , так как давление замеряется с большей точностью, чем расход. Запись модели в форме (2.10.9) ближе всего подходит к канонической модели регрессионного анализа, где x считается величиной детерминированной, а Y — случайной.

Сопоставление линеаризируемых моделей

На вопрос о том, каким приближением, из числа предложенных выше, лучше пользоваться, нельзя ответить однозначно. Выбор наиболее удачной зависимости не в последнюю очередь определяется интуицией и опытом исследователя. Установить формально преимущество той или иной модели довольно трудно. Например, остаточная сумма квадратов в каноническом случае определяется как $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$, где остатки $\boldsymbol{\varepsilon}$ имеют ту же размерность, что и Y . Если использовать показательную функцию (2.10.4)

$y = \alpha_1 e^{\alpha_2 x}$, то придется иметь дело с корреляционным полем в координатах $(x, \ln y)$. Таким образом, остатки этой модели должны характеризовать ошибки в определении $\ln y$. Стандартных методов для сопоставления этой модели с моделью (2.10.5) или (2.10.6) нет, поэтому в каждом случае следует учитывать специфику проблемы и проводить детальный анализ.

Проверка адекватности модели

Все результаты, полученные методом регрессионного анализа, будут обоснованными в случае, если остатки ε удовлетворяют предположениям 1 – 4 подраздела 2.3 (см. соотношения (2.3.6) – (2.3.9)). Проверку этих предположений можно произвести, рассматривая оценки остатков $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ ($i = 1, \dots, n$), или в векторной записи

$$\hat{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{P})\mathbf{Y}. \quad (2.10.10)$$

Остатки $\hat{\varepsilon}$ в модели регрессионного анализа удовлетворяют соотношениям (см. п. 2.5 и приложение В)

$$\mathbf{M}\hat{\varepsilon} = \mathbf{0}; \mathbf{D}\hat{\varepsilon} = \sigma^2 (\mathbf{I}_n - \mathbf{P})^2 = \sigma^2 (\mathbf{I}_n - \mathbf{P}). \quad (2.10.11)$$

Несмещенной оценкой σ^2 служит $\hat{\sigma}^2 = \sum \hat{\varepsilon}_i^2 / (n - s - 1)$.

В предположении (2.3.9) вектор $\hat{\varepsilon}$ имеет многомерное нормальное распределение, а маргинальное распределение $\hat{\varepsilon}_i$ имеет вид

$$\hat{\varepsilon}_i \in N(0, \sigma^2(1 - p_{ii})). \quad (2.10.12)$$

Дисперсия $\hat{\varepsilon}_i$ в (2.10.12) определяются согласно (2.10.11), p_{ii} — диагональный элемент матрицы \mathbf{P} .

Остатки целесообразно привести к дисперсии, близкой к 1, то есть ввести соответствующие масштабы для $\hat{\varepsilon}_i$. Для этого пользуются средней дисперсией

$$\frac{1}{n} \Sigma \mathbf{D} \hat{\varepsilon}_i = \frac{1}{n} \text{tr}(\mathbf{D} \hat{\varepsilon}) = \frac{1}{n} \sigma^2 \text{tr}(\mathbf{I}_n - \mathbf{P}) = \frac{\sigma^2(n-s-1)}{n}, \quad (2.10.13)$$

что приводит к остаткам

$$c_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma} \sqrt{(n-s-1)/n}}, \quad (2.10.14)$$

распределение которых близко к $N(0,1)$. Можно рассматривать также остатки, пользуясь соотношением (2.10.12)

$$d_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma} \sqrt{1-p_{ii}}}, \quad (2.10.15)$$

которые так же, как и c_i , имеют распределение $N(0,1)$.

Проверку остатков на принадлежность их распределения к нормальному закону целесообразно проводить с помощью нормальной вероятностной бумаги. В случае, когда гипотеза о нормальности выполняется, все точки d_i (или c_i) лежат недалеко от биссектрисы 1-го и 3-го координатных углов на вероятностной бумаге.

Индекс i зачастую представляет собой момент времени. Так бывает, когда рассматриваются значения некоторого показателя по годам, кварталам или месяцам. В этом случае выборка представляет собой временной ряд (см. р. 3, там вопрос об остатках регрессионных моделей рассматривается более основательно). Тем не менее, уже сейчас приведем некоторые соображения, по-

звояющие получить первое представление о предмете, не прибегая к формальному аппарату.

Рассмотрим остатки в примере А.6 (см. приложение А.6). На рис. 2.6 представлены остатки регрессионной модели в предположении, что замеры проведены последовательно во времени. Как видно из рисунка, остатки регрессии хаотичны. Это приводит к выводу, что отклонения от модели, по всей видимости, отсутствуют. К тому же специальная проверка показывает, что остатки можно считать некоррелированными.

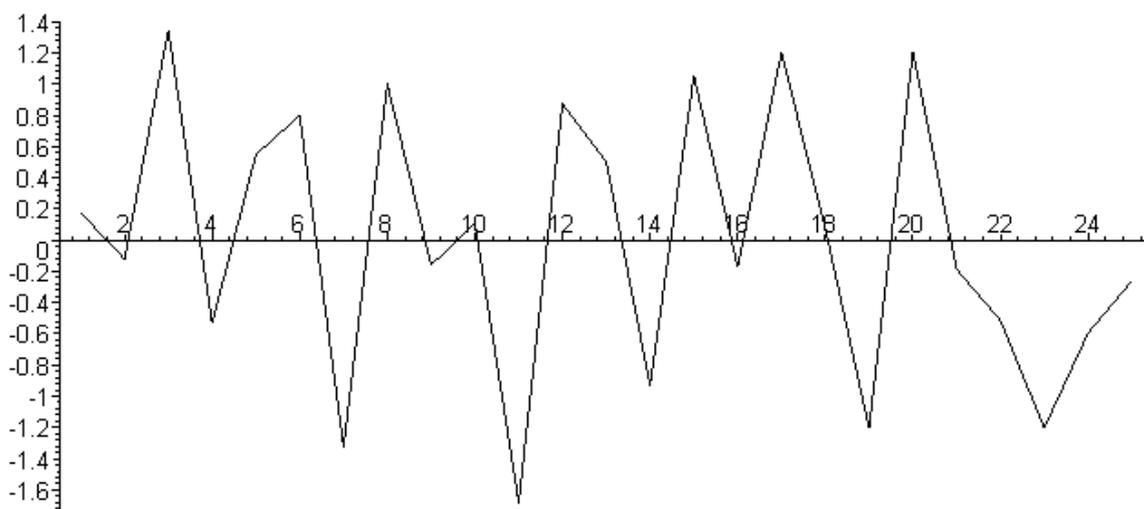


Рисунок 2.6 — Остатки регрессионной модели в примере А.6

При анализе остатков могут встретиться различные случаи, ситуацию иллюстрирует рис. 2.7. Рис. 2.7а показывает удовлетворительное согласие с предположениями модели, рисунки 2.7б, в, г некоторые случаи расхождения. На рис. 2.7 б явно проявляется рост дисперсии с увеличением времени. Это свидетельствует о том, что предположение (2.3.7) не выполняется. Неизменность дисперсии во времени называется гомоскедастичностью. В случае, изображенном на рис. 2.7 б, имеет место гетеро-

скедаксичность. В случае z остатки имеют линейный (положительный) тренд. Поскольку регрессионная модель должна снять линейный тренд, картина остатков на рис. z свидетельствует об ошибках, допущенных в расчетах.

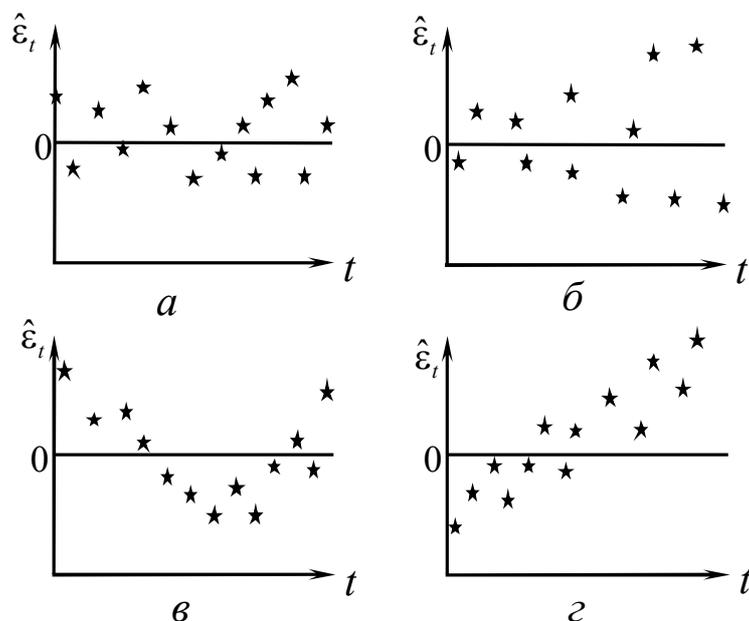


Рисунок 2.7 — Возможные нарушения предположений модели регрессионного анализа. a — отклонения от модели, по всей видимости, отсутствуют; $б$ — нарушено свойство (2.3.7) (гомоскедаксичность); $в$ — криволинейный тренд; z — линейный тренд — ошибки в вычислениях

Дефекты модели могут быть обнаружены путем анализа зависимости остатков от регрессоров. Если предположить, что зависимость остатков $\hat{\epsilon}_1, \dots, \hat{\epsilon}_n$ от некоторого фактора s $\hat{\epsilon}_{1s}, \dots, \hat{\epsilon}_{ns}$ имеет такой вид, как на рис. $в$, то следует предположить, что зависимость Y от x_s нелинейна, и попытаться улучшить модель, включив, например, в число регрессоров наряду с (или вместо) x_s квадрат фактора x_s^2 .

Еще один полезный прием для заключений об адекватности модели дает построение графиков зависимости регрессоров x_k и x_r . Визуальное изучение графиков иногда позволяет сделать вывод о коррелированности регрессоров или ее отсутствии. Если группа двух или более факторов сильно коррелирована, то некоторые из них следует выбросить, стремясь при этом, чтобы оставшиеся были независимы и в то же время достаточно полно отражали влияние всех факторов группы.

3 ВРЕМЕННЫЕ РЯДЫ

3.1 Основные обозначения и понятия. Примеры.

Задачи исследования

Основные понятия

Временным рядом называют упорядоченную во времени конечную совокупность произведенных в последовательные моменты времени t_1, t_2, \dots, t_n наблюдений некоторой величины $x(t_1), x(t_2), \dots, x(t_n)$. Члены временного ряда будем обозначать x_1, x_2, \dots, x_n . Моменты t_1, t_2, \dots, t_n , когда производятся наблюдения, обычно бывают равноудаленными $t_i - t_{i-1} = \tau$ ($i = 2, \dots, n$). Тогда можно положить $\tau = 1$, хотя это предположение не является обязательным.

Временные ряды бывают детерминированными и случайными (стохастическими). В первом случае члены временного ряда рассматриваются как значения некоторой неслучайной функции, хотя и допускаются ошибки наблюдений, во втором — как реализации случайной величины

$$X_1, X_2, \dots, X_n. \quad (3.1.1)$$

Иногда различие между обозначениями ряда (3.1.1) и его реализации

$$x_1, x_2, \dots, x_n \quad (3.1.2)$$

стирается.

Наряду с конечным рядом (3.1.1) бывает полезно рассматривать бесконечную случайную последовательность

$$X_1, X_2, \dots, X_n, \dots \quad (3.1.3)$$

или случайную последовательность вида X_i ($i = 0, \pm 1, \pm 2, \dots$). В обоих вариантах мы имеем случайный процесс с дискретным временем.

Между понятиями выборка и временной ряд имеют место различия, отраженные в табл. 3.1.

Таблица 3.1 — Отличие временного ряда от выборки

Характеристика	Тип данных	
	Выборка	Временной ряд
Порядок следования	Не существенен	Существенен
Зависимость членов	Не имеет места	Вообще говоря, имеется
Распределение членов	Одинаково	Вообще говоря, неодинаково

Таким образом, временной ряд — это упорядоченная последовательность данных. Область применения временных рядов — изучение функций, зависящих от времени, по имеющимся результатам наблюдений. Примерами могут служить ежемесячный объем продукции предприятия, суточное потребление бензина в регионе, давление газа в сечении газопровода, фиксируемое через определенный промежуток времени.

Аргументом ряда не обязательно является календарное время. Индекс может характеризовать наработку, километраж и т.д. Приведем такой пример: в последовательности (3.1.1) X_1 означает количество выпущенных заводом машин некоторой марки пробег которых до первой поломки составил не более 1000 км, X_2 — от 1000 до 2000 км и т.д.

Еще один пример. Рассмотрим шоссе с односторонним движением и зафиксируем на нем положение автомашин в некото-

рый момент времени. Выберем начало и направление отсчета. Обозначим X_1 — расстояние от начала отсчета до первой машины, X_2 — расстояние от первой машины до второй и т.д.

На рис. 3.1 – 3.4 приведены различные примеры (реализации) временных рядов.

Временной ряд зависит от дискретного аргумента, и изображать его в декартовых координатах естественно было бы точками с целочисленными абсциссами. Однако принято соединять эти точки отрезками, поэтому получаются графики в виде ломаных, что визуальнее лучше воспринимается.

Чаще всего целью исследования временных рядов является прогнозирование поведения процесса в будущем. Как показывают приведенные примеры, временные ряды могут существенно отличаться друг от друга.

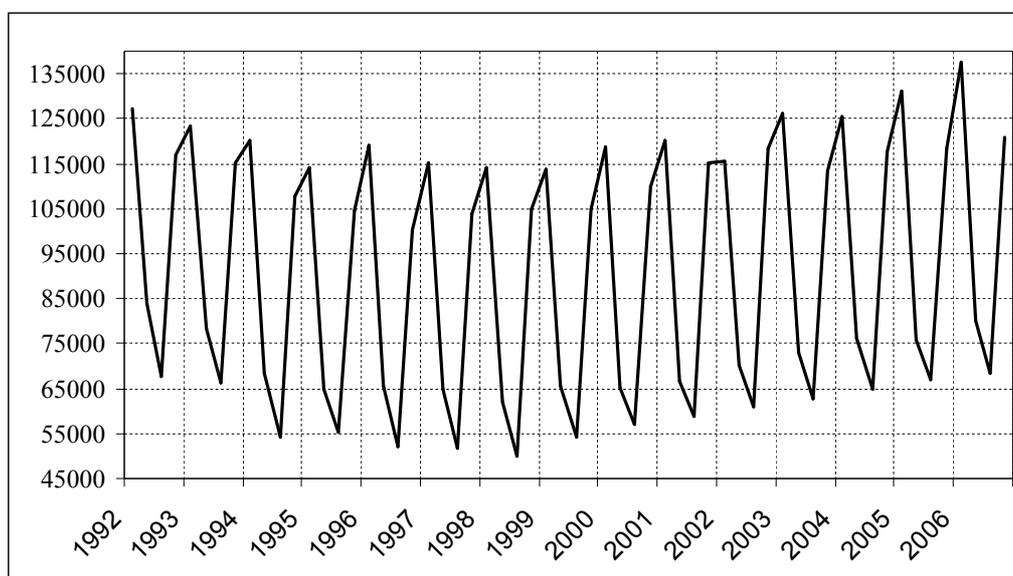


Рисунок 3.1 — Поквартальное потребление газа в млрд. куб.м в РФ за период 1992 – 2006 гг.

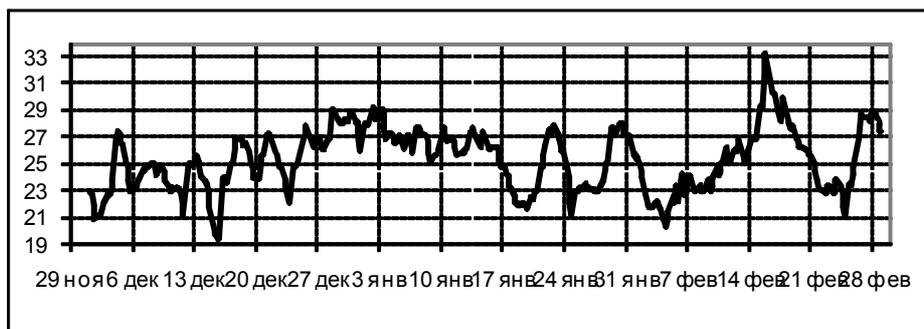


Рисунок 3.2 — Расход газа через газораспределительную станцию (в тыс. куб. м/час) за декабрь – февраль



Рисунок 3.3 — Курс акций ОАО «Газпром»

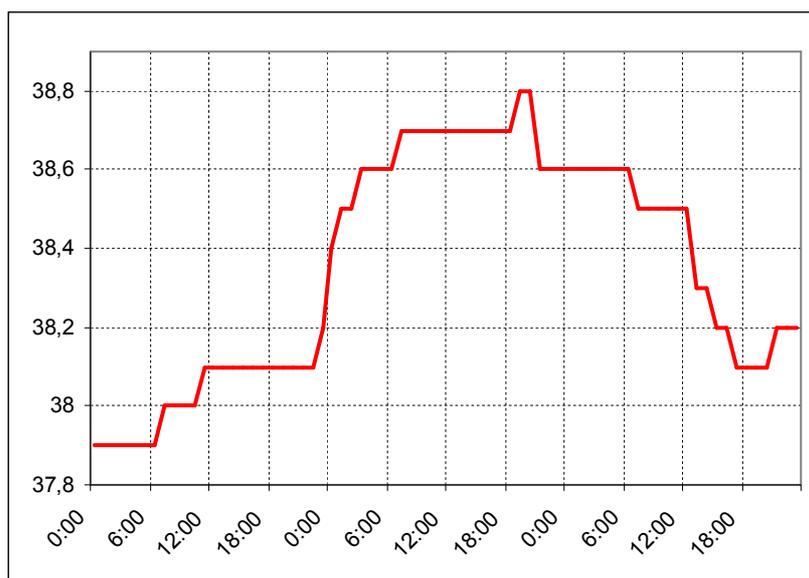


Рисунок 3.4 — Давление газа (бар) на входе в компрессорную станцию по часовым замерам за 3 суток

Ряд, изображенный на рис. 3.1, по своей природе обязан иметь циклическую составляющую с периодом равным году. Эта составляющая обычно называется сезонной. Нетрудно привести примеры, когда период равен не году, а суткам, неделе, месяцу... В некоторых случаях ряды имеют нерегулярную периодичность, как например ряд, характеризующий солнечную активность в многолетнем разрезе.

Важной характеристикой поведения ряда является наличие или отсутствие тренда. Тенденция к равномерному возрастанию или равномерному убыванию называется линейным трендом. Когда тенденция немонотонна или не подчиняется линейному закону, то говорят о нелинейном тренде. Если на графике рис. 3.1 отвлечься от сезонной составляющей, то можно заметить нелинейный тренд: сначала идет снижение потребления до 1999 года, а затем рост. Взяв выборку за 1999 – 2006 гг., получим пример ряда с трендом, близким к линейному.

3.2 Определение стационарного процесса. Белый шум

Случайные процессы с дискретным и непрерывным временем имеют много общего. Они подразделяются на стационарные и нестационарные. Под стационарностью имеют в виду неизменность во времени вероятностного механизма, порождающего процесс. Существуют понятия строгой стационарности (стационарности в узком смысле) и слабой стационарности (стационарности в широком смысле).

Случайный процесс с дискретным временем является строго стационарным, если сдвиг во времени не меняет ни одну из опре-

деляющих его плотностей вероятности, $f_m(X_{t_1}, \dots, X_{t_m}) = f_m(X_{t_1+\tau}, \dots, X_{t_m+\tau})$ для любой совокупности моментов времени t_1, \dots, t_m и любых целочисленных τ .

Для стационарности в широком смысле требуется, чтобы среднее и дисперсия не зависели от времени, а автоковариационная (автокорреляционная) функция $K(t_1, t_2)$ зависела только от разности аргументов $\tau = t_2 - t_1$: $K(t_1, t_2) = K(\tau)$.

Практическая проверка свойства строгой стационарности наталкивается на трудно преодолимые препятствия. Удостовериться же, что процесс стационарен в широком смысле, можно, применяя известные процедуры математической статистики. Всякий строго стационарный процесс является слабо стационарным. Обратное утверждение не верно, но если процесс гауссовский, то из стационарности в широком смысле следует стационарность в узком смысле.

Говоря о стационарных процессах, мы будем иметь в виду слабую стационарность.

Коррелограмма

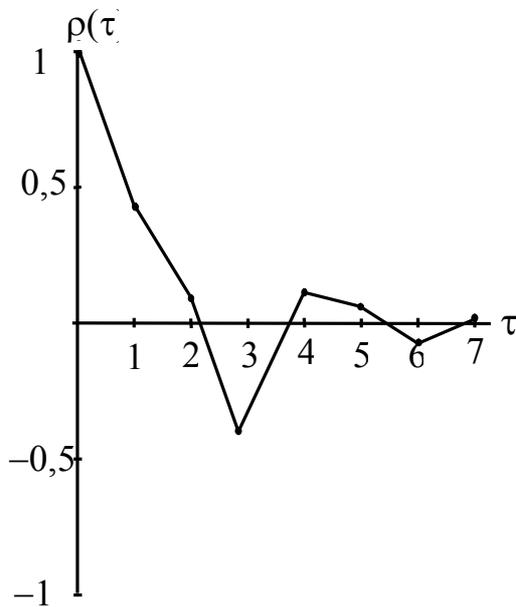
Стационарный процесс $X(t) = X_t$ характеризуется математическим ожиданием $\mathbf{M}[X(t)] = \mu$ и автоковариационной функцией $K(\tau)$.

Далее вплоть до п. 3.18 мы всегда будем считать $\mathbf{M}X_t = 0$, переходя при $\mathbf{M}X_t = \mu$, к процессу $X_t - \mu$. Функция $K(\tau)$ является четной

$$K(-\tau) = K(\tau). \quad (3.2.1)$$

Эту функцию называют также просто ковариационной функцией, поскольку мы имеем дело пока с одним случайным процессом. Следует учитывать, что дисперсия стационарного процесса выражается через значение ковариационной функции в нуле $\mathbf{D}X(t) = K(0)$.

Корреляционная функция $\rho(\tau) = K(\tau)/K(0)$ является функ-



цией целочисленного аргумента τ . Однако обычно точки, изображающие на плоскости значения этой функции, соединяют отрезками, а полученный график называют коррелограммой (рис. 3.5).

Значения корреляционной функции не могут быть произвольными. Во-первых, при любом τ коэффициент корреляции должен удовлетворять неравенству

Рисунок 3.5 — Коррелограмма случайного процесса

$|\rho(\tau)| \leq 1$. Кроме того, для любого натурального m необходимо, чтобы квадратичная форма

$$Q(z_1, \dots, z_m) = \sum_{i,j=1}^m z_i \rho(i-j) z_j \quad (3.2.2)$$

была положительно определенной. Действительно, эта форма представляет собой (с точностью до множителя) дисперсию ли-

нейной комбинации с.в. X_i ($i = 1, \dots, m$) $\sum_{i=1}^m z_i X_i$ и, следовательно,

не может принимать отрицательные значения.

Белый шум

Дискретным белым шумом называется случайный процесс ε_t , если он удовлетворяет условиям

$$\mathbf{M}\varepsilon_t = 0, \mathbf{D}\varepsilon_t = \sigma^2, \text{cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ при } i \neq j. \quad (3.2.3)$$

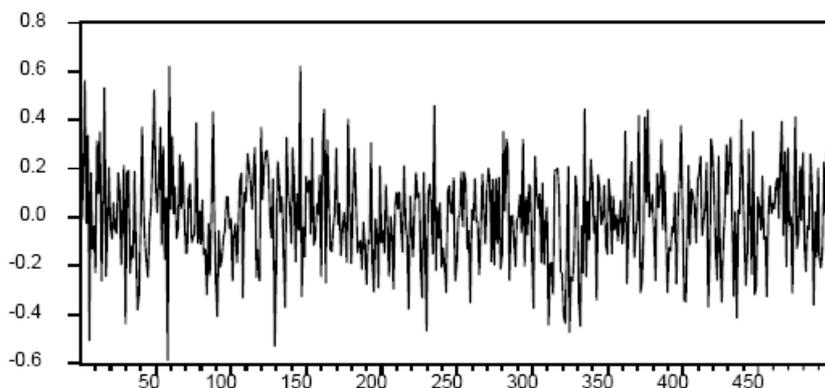


Рисунок 3.6 — Реализация гауссовского белого шума ($\sigma = 0,04$)

Дискретный белый шум является аналогом процесса с непрерывным временем, имеющего то же название. Однако дискретный белый шум поддается значительно более простой интерпретации. Это некоррелированная последовательность одинаково распределенных случайных величин с нулевым средним. Дисперсия σ^2 этих величин является единственным параметром белого шума. Если любая совокупность с.в. ε_t распределена по нормальному закону, то процесс называется гауссовским белым шумом. На рис. 3.6 приведен пример реализации гауссовского белого шума.

3.3 Детерминированные временные ряды

Одна из моделей временных рядов строится следующим образом. Временной ряд представляется в виде суммы двух слагае-

мых, первым из которых служит определенная с точностью до одного или нескольких параметров детерминированная (неслучайная) функция, а вторым — последовательность независимых центрированных случайных величин — белый шум

$$X_t = f(t) + \varepsilon_t, \quad t = 1, \dots, T. \quad (3.3.1)$$

Модель (3.3.1) называют детерминированным временным рядом, а составляющую $f_t = f(t)$ — трендом. Обычно $f(t)$ представляют в виде линейной комбинации некоторых функций

$$f(t) = \sum_{k=0}^K a_k \varphi_k(t). \quad (3.3.2)$$

Функции $\varphi_k(t)$ подбирают специальным образом, чтобы отразить воздействие на процесс различных факторов. Среди факторов следует выделить те, которые

- а) определяющие тенденцию долговременного изменения,
- б) сезонные, определяющие периодические изменения в годовом разрезе,
- в) циклические, определяющие изменения анализируемого процесса под воздействием циклов экономической, природной или другой активности.

Долговременные тренды свидетельствуют о наличии тенденции к возрастанию или убыванию рассматриваемого временного ряда. Для выявления тенденции целесообразно в качестве $\varphi_k(t)$ использовать простейшие алгебраические функции — полиномы, то есть полагать (3.3.2) в виде

$$f(t) = \sum_{k=0}^K a_k t^k. \quad (3.3.3)$$

Ортогональные полиномы

Определение коэффициентов a_k в моделях (3.3.2), (3.3.3) представляет собой типичную задачу линейного регрессионного анализа. Выбор степени полинома осуществляется стандартными методами (см. р. 2). Однако следует обратить внимание на одну особенность, отличающую рассматриваемый случай от общей задачи регрессионного анализа. Независимая переменная (в наших обозначениях t) принимает не произвольные, а лишь целочисленные значения. Это позволяет в качестве базовых функций вместо степеней ($\varphi_k(t) = t^k$) взять ортогональные полиномы. Ортогональность определяется на дискретном множестве $t = 1, \dots, n$.

Покажем, как строится система ортогональных полиномов. Обозначим эти полиномы $\varphi_{0n}, \varphi_{1n}(t), \dots, \varphi_{kn}(t)$. Первый индекс показывает степень полинома, второй — длину временного ряда. В качестве полинома нулевой степени возьмем $\varphi_{0n} \equiv 1$. Полином 1-й степени будем искать в виде $\varphi_{1n}(t) = t + \alpha_{10}$, где коэффициент α_{10} определим из условия ортогональности $\varphi_{1n}(t)$ и φ_{0n}

$$\sum_{t=1}^n \varphi_{0n} \varphi_{1n}(t) = 0. \quad (3.3.4)$$

Получаем $\sum t + \sum \alpha_{10} = 0$, то есть $\frac{n(n+1)}{2} + n\alpha_{10} = 0$, откуда $\alpha_{10} = -\frac{n+1}{2}$, $\varphi_{1n}(t) = t - \frac{n+1}{2}$. Здесь и далее в этом пункте суммирование по t производится от 1 до n , пределы суммирования в обозначениях опущены.

Полином 2-й степени ищем в виде $\varphi_{2n}(t) = t^2 + \alpha_{21}t + \alpha_{20}$, коэффициенты α_{20} и α_{21} находим из условия ортогональности $\varphi_{2n}(t)$ с φ_{0n} и $\varphi_{1n}(t)$. Имеем $\sum \varphi_{0n} \varphi_{2n}(t) = 0 \Rightarrow \sum t^2 + \alpha_{21} \sum t + \sum \alpha_{20} = 0$.

Для того, чтобы выполнялось равенство $\sum \varphi_{1n}(t) \varphi_{2n}(t) = 0$ достаточно выполнение условия $\sum t \varphi_{2n}(t) = 0 \Rightarrow \sum t^3 + \alpha_{21} \sum t^2 + \alpha_{20} \sum t \equiv 0$. Учитывая, что $\sum t^2 = \frac{n(n+1)(2n+1)}{6}$, $\sum t^3 = \left[\frac{n(n+1)}{2} \right]^2$,

для определения α_{20} , α_{21} получаем систему уравнений

$$\begin{aligned} \frac{(n+1)(2n+1)}{6} + \frac{n+1}{2} \alpha_{21} + \alpha_{20} &= 0 \\ \frac{n(n+1)}{2} + \frac{2n+1}{3} \alpha_{21} + \alpha_{20} &= 0. \end{aligned}$$

Отсюда находим $\varphi_{2n}(t) = t^2 - (n+1)t + \frac{(n+1)(n+2)}{6}$. Теперь можно определить $\varphi_{3n}(t)$ и так же продолжать процедуру далее.

Регрессионная задача при использовании ортогональных полиномов будет иметь вид

$$\sum \left[x_t - \gamma_0 - \gamma_1 \varphi_{1n}(t) - \dots - \gamma_k \varphi_{kn}(t) \right]^2 \rightarrow \min, \quad (3.3.5)$$

где $x_t = f(t)$ ($t = 1, \dots, n$), $\gamma_0, \dots, \gamma_k$ — неизвестные параметры, которые находятся как обычно. А именно, вычислим производную от суммы в (3.3.5) по γ_0 и приравняем ее к нулю:

$$\sum x_t - \gamma_0 n - \gamma_1 \sum \varphi_{1n}(t) - \dots - \gamma_k \sum \varphi_{kn}(t) = 0. \quad (3.3.6)$$

Отсюда получаем оценку $\hat{\gamma}_0$ параметра γ_0 $\hat{\gamma}_0 = (1/n) \sum x_t$, поскольку все последние суммы в (3.3.6) равны нулю из условий ортогональности.

Приравнивая к нулю производную по γ_1 , получим

$$\Sigma \varphi_{1n}(t) [x_t - \gamma_0 - \gamma_1 \varphi_{1n}(t) - \dots - \gamma_k \varphi_{kn}(t)] = 0,$$

то есть $\hat{\gamma}_1 = \Sigma x_t \varphi_{1n}(t) / \Sigma \varphi_{1n}^2(t)$. Так же находятся оценки остальных параметров $\hat{\gamma}_k = \Sigma x_t \varphi_{kn}(t) / \Sigma \varphi_{kn}^2(t)$. Таким образом, для определения оценок $\hat{\gamma}_0, \dots, \hat{\gamma}_k$ коэффициентов регрессии при использовании ортогональных полиномов не требуется решать систему уравнений общего вида: система распадается на n уравнений с одним неизвестным.

Для определения ортогональных полиномов есть специальные таблицы. Однако пользоваться ими сейчас вряд ли имеет смысл, так как все вычисления лучше выполнять на компьютере.

Выбор степени полинома

Ортогональные полиномы имеют еще одно важное преимущество. Преобразуем остаточную сумму квадратов S_k

$$\begin{aligned} S_k^2 &= \Sigma [x_t - \hat{\gamma}_0 - \hat{\gamma}_1 \varphi_{1n}(t) - \dots - \hat{\gamma}_k \varphi_{kn}(t)]^2 = \\ &= \Sigma x_t^2 + \Sigma \hat{\gamma}_0^2 + \dots + \hat{\gamma}_k^2 \Sigma \varphi_{kn}^2(t) - 2\hat{\gamma}_0 \Sigma x_t - 2\hat{\gamma}_1 \Sigma x_t \varphi_{1n}(t) - \dots - 2\hat{\gamma}_k \Sigma x_t \varphi_{kn}(t) = \\ &= \Sigma x_t^2 - n\hat{\gamma}_0^2 - \hat{\gamma}_1^2 \Sigma \varphi_{1n}^2(t) - \dots - \hat{\gamma}_k^2 \Sigma \varphi_{kn}^2(t). \end{aligned}$$

Отсюда виден вклад каждого параметра $\gamma_0, \dots, \gamma_k$ в остаточную сумму квадратов S_k^2 . При введении в модель очередного коэффициента γ_k S_k^2 уменьшается на величину $\hat{\gamma}_k^2 \Sigma \varphi_{kn}^2(t)$. Вопрос о целесообразности перехода от модели с $k-1$ параметрами к модели с k параметрами решается по критерию Фишера (приложение В.3)

$$F(1, k-1) = \frac{\hat{\gamma}_k^2 \sum \varphi_{kn}^2(t)}{1} : \frac{S_{k-1}^2}{n-k}, \quad (3.3.7)$$

где S_{k-1}^2 — остаточная сумма квадратов в модели с параметрами $\gamma_0, \gamma_1, \dots, \gamma_{k-1}$. Гипотеза о незначимости параметра γ_k отвергается при уровне значимости α , если значение критерия (3.3.7) превышает α -квантиль распределения $F(1, n-k)$.

Аналогично можно проверить гипотезу $H_0 : \gamma_q = \gamma_{q+1} = \dots = \gamma_k = 0$ при альтернативе $H_1 : H_0$ не имеет места.

Остаточная сумма квадратов позволяет получить несмещенную оценку дисперсии

$$s^2 = [\sum x_i^2 - \hat{\gamma}_{0n}^2 - \dots - \hat{\gamma}_k^2 \sum \varphi_{kn}^2(t)] / (n-k-1). \quad (3.3.8)$$

Для проверки гипотезы $H_0 : \gamma_k = 0$ можно также воспользоваться t -критерием Стьюдента. Величина $(n-k-1)s^2/\sigma^2$ имеет распределение χ^2 с $n-k-1$ степенями свободы. Следовательно, гипотеза $H_0 : \gamma_k = 0$ при альтернативе $H_1 : \gamma_k \neq 0$ и уровне значимости α отвергается, если

$$\frac{|\hat{\gamma}_k| \{ \sum \varphi_{kn}^2(t) \}^{1/2}}{s} > t_{n-k-1} \left(1 - \frac{\alpha}{2} \right) \quad (3.3.9)$$

где через $t_{n-k-1}(1-\alpha/2)$ обозначена $\left(1 - \frac{\alpha}{2}\right)$ -квантиль t -распределения (приложение Б.2).

Нелинейные приближения и их сопоставление

Наряду с полиномиальным трендом могут быть использованы другие модели (см. п. 2.10), например степенная $y = \alpha_1 x^{\alpha_2}$

(2.10.2), показательная $y = \alpha_1 e^{\alpha_2 x}$ (2.10.4), полином по отрицательным степеням t (2.10.6). Сопоставление моделей рекомендуется проводить с помощью коэффициента детерминации (см. п. 2.6). Приведем пример, заимствованный из [20]. Исходный ряд содержит 10 членов, представляющих собой темпы роста номинальной заработной платы за январь–октябрь 1999 г. в процентах к уровню декабря 1998 г. $x_1 = 82,9$; $x_2 = 87,3$; $x_3 = 99,4$; $x_4 = 104,8$; $x_5 = 107,2$; $x_6 = 121,6$; $x_7 = 118,5$; $x_8 = 114,1$; $x_9 = 123,0$; $x_{10} = 127,3$.

Для исследования этого ряда использовано 5 моделей. Оценки параметров моделей и коэффициенты детерминации R^2 приведены в табл. 3.2.

Таблица 3.2 — Сопоставление моделей для оценки тренда временного ряда

№	Формула	Уравнение	R^2
1	(3.3.3) $K = 1$	$\hat{x}_t = 82,66 + 4,72t$	0,873
2	(3.3.3) $K = 2$	$\hat{x}_t = 72,9 + 9,59t - 0,444t^2$	0,920
3	(2.10.2)	$\ln \hat{x}_t = 4,39 + 0,193 \ln t$	0,931
4	(2.10.4)	$\ln \hat{x}_t = 4,43 + 0,045t$	0,856
5	(2.10.6) $s = 1$	$\hat{x}_t = 122,58 - 47,63/t$	0,728

В данном примере преимущества степенной модели № 3 оказываются бесспорными. Она обеспечивает более высокое значение R^2 даже по сравнению с трехпараметрической моделью № 2.

3.4 Сезонный тренд

Модель с сезонным трендом представляется формулой (3.3.1), где $f(t)$ — периодическая функция с известным периодом τ

$$f(t + \tau) = f(t). \quad (3.4.1)$$

Оценка сезонного тренда также производится методами регрессионного анализа. Функция $f(t)$ ищется в виде разложения по тригонометрическим функциям

$$f(t) = a_0 + \sum_{j=1}^r (a_j \cos \omega_j t + b_j \sin \omega_j t), \quad (3.4.2)$$

где a_0, a_j, b_j ($j=1, \dots, r$) — искомые параметры, а частоты ω определяются в зависимости от периода τ . Оценки параметров ищутся, как и положено в регрессионном анализе, методом наименьших квадратов. Подробные сведения о сезонном тренде можно найти в монографии [1], а в приложении к прогнозу газопотребления в [17].

Не вдаваясь в теорию и технику регрессионного анализа применительно к модели (3.4.2), рассмотрим частный случай, который, однако, поможет понять суть дела и важен сам по себе. Положим, что период τ равен одному году и заданы фактические поквартальные значения x_t показателя X_t из (3.3.1). Будем считать, что длина временного ряда n кратна τ $n = \tau H$. Для квартала i года h значение индекса t равно $t = 4(h-1)\tau + i$ ($h=1, \dots, H$; $i=1, 2, 3, 4$). Задача состоит в том, чтобы оценить сезонный тренд, то есть 4 величины $d_i = f[4(h-1)\tau + i]$, ($i=1, \dots, 4$).

Критерием для оценки служит, как обычно, минимальное значение остаточной суммы квадратов

$$\sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n [x_t - f(t)]^2 = \sum_{h=1}^H \sum_{i=1}^4 (x_{hi} - d_i)^2 \rightarrow \min. \quad (3.4.3)$$

где $x_{hi} = x_{4(h-1)\tau+i}$, ($h = 1, \dots, H$) — члены временного ряда, относящиеся к i -му кварталу года.

Приравнивая к нулю производную критериальной функции по d_i , получим уравнение для вычисления оценки \hat{d}_i :

$$\sum_{h=1}^H (x_{hi} - \hat{d}_i) = 0 \Rightarrow \hat{d}_i = \frac{1}{H} \sum_{h=1}^H x_{hi}. \quad (3.4.4)$$

Таким образом, оценкой сезонного значения за квартал i служит среднее арифметическое значений рассматриваемого показателя по всем периодам (годам), охватываемым временным рядом. Если предположение $n = \tau H$ не выполнено, то формулу (3.4.4) легко скорректировать для учета того, что количество наблюдений по разным кварталам не одинаково.

3.5 Стохастические временные ряды. Случайное блуждание. Операторы разности

Случайное блуждание

Случайным блужданием (в одномерном пространстве, или на одномерной решетке) называется процесс

$$X_t = X_{t-1} + \varepsilon_t, \quad (3.5.1)$$

где ε_t — белый шум.

Этот процесс называется также броуновским движением на одномерной решетке. Легко обобщить процесс (3.5.1) на случай двух и трех измерений и представить случайное блуждание на

двух- и трехмерной решетке. Термин «случайное блуждание» обязан, пожалуй, двумерному варианту. Здесь траекторию процесса можно интерпретировать как движение пьяного по ровной плоскости с нанесенной координатной сеткой. Он с одинаковой вероятностью делает шаг в одном из 4-х направлений: вперед, назад, влево или вправо, независимо от предыдущих перемещений. Процесс (3.5.1), где \mathbf{X}_t и $\boldsymbol{\varepsilon}_t$ двумерные векторы $\mathbf{X}_t = \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$, представляет траекторию его движения.

Пусть начальная точка процесса (3.5.1) X_0 зафиксирована, тогда

$$X_t = X_{t-1} + \varepsilon_t = X_{t-2} + \varepsilon_t + \varepsilon_{t-1} = \dots = X_0 + \sum_{h=0}^{t-1} \varepsilon_{t-h}.$$

Математическое ожидание X_t равно $\mathbf{M}X_t = \mathbf{M}X_0 + \sum_{h=0}^{t-1} \mathbf{M}\varepsilon_{t-h} = X_0$, то есть не меняется во времени. Подсчитаем дисперсию процесса X_t $\mathbf{D}X_t = \mathbf{D} \sum_{h=0}^{t-1} \varepsilon_{t-h} = \mathbf{M} \left(\sum_{h=0}^{t-1} \varepsilon_{t-h} \right)^2 = t\sigma^2$. При этом использованы свойства белого шума (3.2.3).

Мы видим, что дисперсия рассматриваемого процесса изменяется во времени, увеличиваясь от 0 до ∞ . Это значит, что процесс случайного блуждания нельзя считать стационарным.

Операторы разности

При изучении временных рядов широко используются операторные обозначения. Оператор сдвига назад B определяется как $BX_t = X_{t-1}$. Полагая $B^q X_t = B(B^{q-1} X_t)$, получаем

$$B^q X_t = X_{t-q}. \quad (3.5.2)$$

Удобно также считать, что $B^0 X_t = X_t$, то есть B^0 — единичный оператор. Через оператор B выражаются операторы взятия разностей. Оператор первой разности вперед записывается как

$$\Delta X_t = X_{t+1} - X_t. \quad (3.5.3)$$

Второй разностью называется оператор

$$\Delta^2 X_t = \Delta(\Delta X_t) = \Delta(X_{t+1} - X_t) = X_{t+2} - 2X_{t+1} + X_t. \quad (3.5.4)$$

Аналогично определяются разности высших порядков

$$\Delta^r X_t = X_{t+r} - C_r^1 X_{t+r-1} + C_r^2 X_{t+r-2} - \dots + (-1)^r X_t. \quad (3.5.5)$$

Здесь C_r^k число сочетаний из r элементов по k
 $C_r^k = r! / k!(r-k)!$

Иногда оказываются полезными разности назад

$$\nabla X_t = X_t - X_{t-1} = (1-B)X_t, \quad \nabla^r X_t = \nabla(\nabla^{r-1} X_t). \quad (3.5.6)$$

Заметим сразу, что разностные операторы для функций целочисленного аргумента во многом схожи с операторами дифференцирования, определенными для функций непрерывного аргумента. Рассмотрим, к примеру, как действуют разностные операторы на многочлены. В соответствии с определениями (3.5.3) – (3.5.5), получаем

$$\begin{aligned} \Delta t &= (t+1) - t = 1, & \Delta^2 t &= 0; \\ \Delta t^2 &= (t+1)^2 - t^2 = 2t+1, & \Delta^2 t^2 &= 2, & \Delta^3 t^2 &= 0. \end{aligned}$$

Легко видеть, что

$$\Delta^r t^p = 0 \quad \text{при } r > p. \quad (3.5.7)$$

Возвратимся к процессу случайного блуждания. Переходя к первым разностям, запишем (3.5.1) в виде

$$\Delta X_t = \varepsilon_t.$$

Значит, первая разность для процесса случайного блуждания является белым шумом. Это рассуждение показывает, что первая разность нестационарного процесса может являться процессом стационарным.

Возьмем еще пример, рассмотрим детерминированный временной ряд (3.3.1), в котором $f(t)$ — линейная функция $X_t = at + b + \varepsilon_t$. Применяя к обеим частям этого равенства оператор Δ , получаем

$$\Delta X_t = a(t+1) + b + \varepsilon_{t+1} - (at + b + \varepsilon_t) = a + \varepsilon_{t+1} - \varepsilon_t,$$

то есть опять-таки с помощью оператора Δ удалось от нестационарного временного ряда перейти к стационарному.

3.6 Процессы скользящего среднего

Мы переходим к изучению группы важных моделей стационарных стохастических временных рядов [10, 11, 13]. В их число входят

- модель скользящего среднего (СС), в английской терминологии moving average model (MA);
- модель авторегрессии (АР), в английской терминологии autoregressive model (AR);
- модель авторегрессии — скользящего среднего (АРСС), в английской терминологии autoregressive moving average model (ARMA);

➤ модель авторегрессии — проинтегрированного скользящего среднего (АРПСС), в английской терминологии autoregressive integrated moving average model (ARIMA).

Приведенные выше английские аббревиатуры встречаются также и в русскоязычных изданиях.

Процесс скользящего среднего определяется соотношением

$$X_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}, \quad (3.6.1)$$

где ε_t , как и ранее, белый шум. Термин скользящее среднее объясняется тем, что текущее значение процесса X_t определяется одномоментным значением белого шума ε_t и взвешенным средним q предыдущих значений белого шума.

Для процесса (3.6.1) непосредственно находим

$$\mathbf{M}X_t = 0; \quad \text{cov}(X_t, X_{t+\tau}) = \gamma(\tau) = \begin{cases} \sigma^2 \sum_{i=0}^{q-\tau} \beta_i \beta_{i+\tau}, & \tau \leq q \\ 0 & \tau > q \end{cases}, \quad (3.6.2)$$

где надо положить $\beta_0 = 1$.

В частности, при $\tau = 0$ из (3.6.2) получаем $\mathbf{D}X_t = \sigma^2 \sum_{i=0}^q \beta_i^2$.

Соотношение (3.6.1) можно записать с использованием оператора сдвига B в форме

$$X_t = (B^0 + \beta_1 B + \dots + \beta_q B^q) \varepsilon_t = \beta_q(B) \varepsilon_t, \quad (3.6.3)$$

которая служит определением операторного полинома $\beta_q(B)$.

Если имеется два операторных полинома $\varphi(B)$ и $\psi(B)$, то, применяя их последовательно к X_t , получим тот же результат,

что и действуя на X_t оператором $\varphi(B)\psi(B)$, где произведение получается по обычным правилам умножения полиномов

$$\varphi(B)[\psi(B)X_t] = [\varphi(B)\psi(B)]X_t. \quad (3.6.4)$$

В частности, из (3.6.4) следует, что любой операторный полином $\varphi(B)$ может быть разложен на множителя по корням уравнения $\varphi(B) = 0$. Обозначая корни уравнения $1 + \beta_1 z + \dots + \beta_q z^q = 0$ через z_1, \dots, z_q , получим

$$\beta_q(B) = 1 + \beta_1 B + \dots + \beta_q B^q = \beta_q \prod_{i=1}^q (B - z_i). \quad (3.6.5)$$

Для оператора первого порядка $1 - \alpha B$ естественно положить $(1 - \alpha B)^{-1} = 1 + \alpha B + \alpha^2 B^2 + \dots$, так как $(1 - \alpha B) \cdot (1 + \alpha B + \alpha^2 B^2 + \dots) = 1$, если выражение во второй скобке имеет смысл. Вместо бесконечного ряда рассмотрим конечную сумму $(1 - \alpha B)(1 + \alpha B + \alpha^2 B^2 + \dots + \alpha^p B^p) = 1 - \alpha^{p+1} B^{p+1}$. По определению $(1 - \alpha^{p+1} B^{p+1})X_t = X_t - \alpha^{p+1} X_{t-p-1}$. Если значения случайного процесса ограничены и $|\alpha| < 1$, то при $p \rightarrow \infty$ правая часть этого равенства стремится к X_t . Следовательно, обратный оператор $[1 - \alpha B]^{-1}$ существует, если $|\alpha| < 1$.

Из (3.6.3) и (3.6.5), получим

$$X_t = \beta_q \prod_{i=1}^q (B - z_i) \varepsilon_t = \beta_q \prod_{i=1}^q \left(B - \frac{1}{\zeta_i} \right) \varepsilon_t = \prod_{i=1}^q (1 - \zeta_i B) \varepsilon_t, \quad (3.6.6)$$

где $\zeta_i = 1/z_i$, корни уравнения

$$\zeta^q + \beta_1 \zeta^{q-1} + \dots + \beta_q = 0. \quad (3.6.7)$$

Формально из (3.6.6) получаем $\varepsilon_t = \frac{1}{\prod_{i=1}^q (1 - \zeta_i B)} X_t$. Если все

корни ζ_i различны, то функцию от B $\frac{1}{\prod_{i=1}^q (1 - \zeta_i B)}$ можно пред-

ставить в виде суммы элементарных дробей

$$\frac{1}{\prod_{i=1}^q (1 - \zeta_i B)} = \frac{A_1}{1 - \zeta_1 B} + \frac{A_2}{1 - \zeta_2 B} + \dots + \frac{A_q}{1 - \zeta_q B}. \quad (3.6.8)$$

Те дробь, для которых выполняется соотношение

$$|\zeta_i| < 1, \quad (3.6.9)$$

могут быть представлены степенным рядом

$$\frac{A_i}{1 - \zeta_i B} = A_i (1 + \zeta_i B + \zeta_i^2 B^2 + \dots). \quad (3.6.10)$$

Если неравенство (3.6.9) выполнено для всех корней уравнения (3.6.7), то ε_t можно выразить в виде бесконечной суммы значений ряда X_t , уходящих в прошлое. Для этого надо вместо каждой дроби в правой части (3.6.8) подставить ее выражение (3.6.10) и собрать коэффициенты при одинаковых степенях B . Тогда получим выражение вида

$$\varepsilon_t = (\alpha_0 + \alpha_1 B + \alpha_2 B^2 + \dots) X_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots, \quad (3.6.11)$$

где α_j выражаются через известные величины ζ_i .

Доказывается, что представление (3.6.11) справедливо и в случае комплексных и кратных корней, если для них выполняется неравенство (3.6.9).

3.7 Процессы авторегрессии

Общая модель АР имеет вид

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} = \varepsilon_t, \quad (3.7.1)$$

p называется порядком авторегрессии. Значение процесса АР в момент t определяется p предыдущими значениями с добавлением белого шума.

Так же, как и в случае скользящего среднего, модель (3.7.1) можно представить в операторном виде $\alpha_p(B)X_t = \varepsilon_t$, где $\alpha_p(B)$ операторный полином степени p $\alpha_p(B) = 1 + \alpha_1 B + \dots + \alpha_p B^p$.

В моделях АР и СС процессы X_t и ε_t как бы меняются местами. Это позволяет для исследования АР применить те же подходы, что и в п. 3.6.

Но, прежде всего, заметим, что (3.7.1) можно рассматривать как разностное уравнение относительно X_t . Отмеченное выше сходство функций дискретного и континуального переменного позволяет разностные уравнения исследовать по аналогии с уравнениями дифференциальными. Общее решение неоднородного уравнения (3.7.1) представляется в виде суммы общего решения однородного $\alpha_p(B)X_t = 0$ и частного решения неоднородного. Общее решение однородного уравнения может быть записано в виде

$$X_t = \sum P_i(t) \zeta_i^t, \quad (3.7.2)$$

где сумма берется по корням ζ_i уравнения

$$1 + \alpha_1 \zeta + \dots + \alpha_p \zeta^p = 0, \quad (3.7.3)$$

$P_i(t)$ — полиномы, степень которых на единицу меньше кратности корней ζ_i .

Решение уравнения (3.7.1) остается конечным с ростом t , если все корни характеристического полинома (3.7.2) удовлетворяют условию

$$|\zeta_i| < 1. \quad (3.7.4)$$

Но, то же условие означает существование обратного оператора $\alpha_p^{-1}(B)$. То есть в случае выполнения (3.7.4)

$$X_t = \frac{1}{\alpha_p(B)} \varepsilon_t = \sum_{\tau=0}^{\infty} \psi_{\tau} \varepsilon_{t-\tau}, \quad (3.7.5)$$

при этом ряд $\sum_{\tau=0}^{\infty} |\psi_{\tau}|$ сходится, а процесс X_t будет стационарным.

Если же хотя бы один корень уравнения (3.7.3) не удовлетворяет условию (3.7.4), то есть не лежит внутри единичного круга, то решение уходит в бесконечность, а ряд X_t не может быть стационарным.

Обратим внимание на то, что процесс СС всегда стационарен, а условие (3.6.9) необходимо только для его представления в виде (3.6.11). В случае же процесса АР условие (3.7.4) куда как более существенно: только при его выполнении модель АР является стационарной.

Умножим соотношение (3.7.1) на $X_{t-\tau}$, применим операцию математического ожидания к обеим частям и разделим на дисперсию X_t . Получим

$$\rho(\tau) + \alpha_1 \rho(\tau - 1) + \dots + \alpha_p \rho(\tau - p) = 0. \quad (3.7.6)$$

При этом мы учли, что, во-первых, $\mathbf{M}X_t = 0$ и, следовательно, $\mathbf{M}(X_t, X_{t-\tau}) = \text{cov}(X_t, X_{t-\tau}) = K(\tau)$, и, во-вторых, $\text{cov}(X_{t-\tau}, \varepsilon_t) = 0$ при $\tau > 0$, так как согласно (3.7.5) $X_{t-\tau}$ представляется через $\varepsilon_{t-\tau}$ и предшествующие ему значения $\varepsilon_{t-\tau-1}, \varepsilon_{t-\tau-2}, \dots$, откуда и вытекает некоррелированность $X_{t-\tau}$ и ε_t .

Мы видим, что корреляционная функция $\rho(\tau)$ при $\tau > p$ удовлетворяет разностному уравнению (3.7.6), совпадающему с однородным уравнением для X_t . Взяв достаточное количество соотношений (3.7.6) при различных τ , получим линейную систему уравнений для определения $\rho(\tau)$ через параметры авторегрессионной модели $\alpha_1, \dots, \alpha_p$. Посмотрим, какие при этом получаются результаты на примерах моделей авторегрессии 1-го и 2-го порядков AR(1) и AR(2).

3.8 Марковский процесс — авторегрессия 1-го порядка

Этот процесс определяется соотношением

$$X_t + \alpha_1 X_{t-1} = \varepsilon_t. \quad (3.8.1)$$

Текущее значение X_t зависит только от предыдущего значения X_{t-1} и белого шума ε_t . Марковский процесс не зависит от предыстории, забывает прошлое. Обозначим $\alpha_1 = -\rho$. Тогда (3.8.1) примет вид

$$X_t = \rho X_{t-1} + \varepsilon_t. \quad (3.8.2)$$

Если (3.8.2) умножить на $X_{t-\tau}$, перейти к математическим ожиданиям и разделить на σ^2 , то получим для коэффициентов корреляции $\rho(\tau) = \rho\rho(\tau-1)$. В частности, $\rho(1) = \rho$, что оправдывает введенное нами обозначение. Оказывается, что ρ действительно является коэффициентом корреляции 1-го порядка для марковского процесса. Применяя полученную формулу последовательно, имеем

$$\rho(\tau) = \rho\rho(\tau-1) = \rho^2\rho(\tau-2) = \dots = \rho^\tau. \quad (3.8.3)$$

Проведем теперь исследование марковского процесса, используя операторную запись. Операторный полином $\alpha_1(B)$ и характеристическое уравнение (3.7.2) имеют вид $\alpha_1(B) = 1 + \alpha_1 B$ и, соответственно, $1 + \alpha_1 \zeta = 0$.

Теперь видно, насколько существенно требование (3.7.4), чтобы корни характеристического уравнения лежали внутри единичного круга. Если условие $|\zeta_1| < 1$ не выполняется, процесс (3.8.1) не будет стационарным.

Соотношение (3.7.5) для рассматриваемого процесса AP(1) имеет вид

$$X_t = \frac{1}{1 + \alpha_1 B} \varepsilon_t \equiv \frac{1}{1 - \rho B} \varepsilon_t = (1 + \rho B + \rho^2 B^2 + \dots) \varepsilon_t = \sum_{\tau=0}^{\infty} \rho^\tau \varepsilon_{t-\tau}. \quad (3.8.4)$$

На рис. 3.7 построены коррелограммы марковского процесса при $\rho = 0,6$ и $\rho = -0,6$.

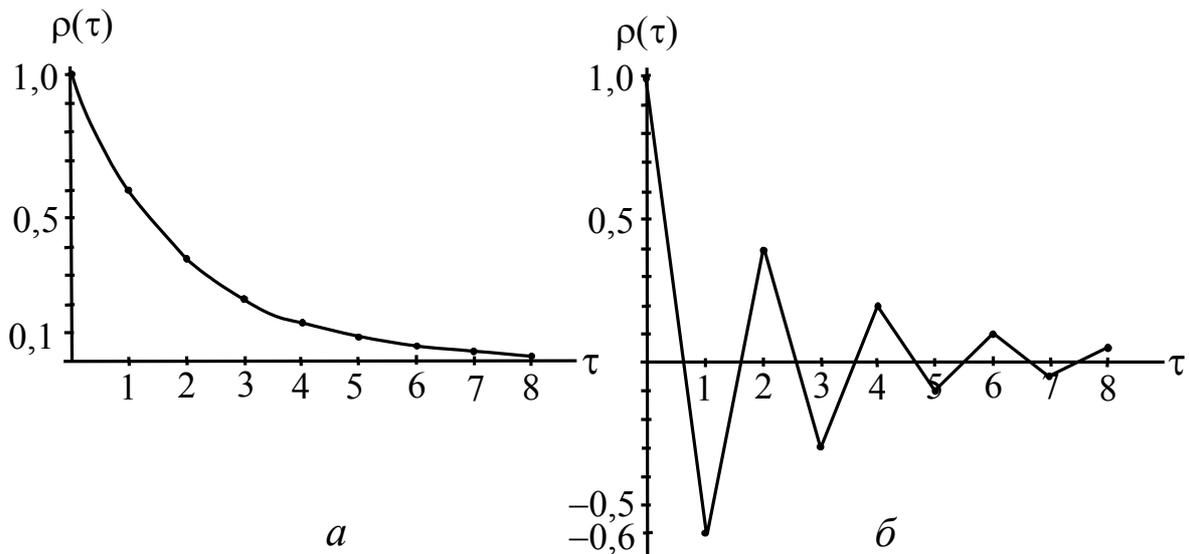


Рисунок 3.7 — Коррелограмма марковского процесса
 а) при $\rho = 0,6$, б) при $\rho = -0,6$

То же самое можно получить, последовательно применяя (3.8.2)

$$X_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2 X_{t-2} = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \rho^3 X_{t-3}. \quad (3.8.5)$$

Чтобы прийти к ряду (3.8.4) надо потребовать выполнения условия $\lim_{\tau \rightarrow \infty} \rho^\tau X_{t-\tau} = 0$, что имеет место, если значения X_t ограничены по модулю и $|\rho| < 1$.

Заметим еще, что $\mathbf{DX} = \text{cov}(\rho X_{t-1} + \varepsilon_t, \rho X_{t-1} + \varepsilon_t) = \rho^2 \mathbf{DX} + \sigma^2$.

Откуда

$$\mathbf{DX} = \frac{\mathbf{D}\varepsilon}{1 - \rho^2}.$$

При $\rho \rightarrow 1$ \mathbf{DX} возрастает неограниченно и авторегрессионный процесс приближается к рассмотренному ранее процессу случайного блуждания, не являющемуся стационарным.

3.9 Процесс Юла — авторегрессия второго порядка

Процесс AP(2) определяется соотношением

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t. \quad (3.9.1)$$

Умножая (3.9.1) на X_{t-1} и X_{t-2} и переходя к средним, получаем систему для определения коэффициентов автокорреляции 1-го и 2-го порядка $\rho_1 = \rho(1)$, $\rho_2 = \rho(2)$

$$\begin{aligned} \rho_1 + \alpha_1 + \alpha_2 \rho_1 &= 0 \\ \rho_2 + \alpha_1 \rho_1 + \alpha_2 &= 0. \end{aligned} \quad (3.9.2)$$

Откуда

$$\rho_1 = -\frac{\alpha_1}{1 + \alpha_2}, \quad \rho_2 = -\alpha_2 + \frac{\alpha_1^2}{1 + \alpha_2}. \quad (3.9.3)$$

Характеристическое уравнение для процесса (3.9.1) имеет вид

$$\zeta^2 + \alpha_1 \zeta + \alpha_2 = 0. \quad (3.9.4)$$

Условие $|\zeta_{1,2}| < 1$ для корней этого уравнения

$$\zeta_{1,2} = \frac{-\alpha_1 \pm \sqrt{\alpha_1^2 - 4\alpha_2}}{2} \quad (3.9.5)$$

можно свести к следующим двум неравенствам

$$|\alpha_1| < 2, \quad -\alpha_2 < 1 - |\alpha_1|. \quad (3.9.6)$$

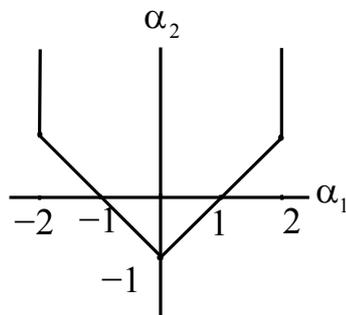


Рисунок 3.8 — Область, отвечающая условиям существования стационарного процесса AP(2)

Область, отвечающая условиям (3.9.6), изображена на рис. 3.8. При выполнении соотношений (3.9.6), как нетрудно проверить, $|\rho_1| < 1$, $|\rho_2| < 1$.

Умножая (3.9.1) на $X_{t-\tau}$ и переходя к средним, получим

разностное уравнение для коэффициентов корреляции $\rho(\tau)$

$$\rho(\tau) + \alpha_1 \rho(\tau - 1) + \alpha_2 \rho(\tau - 2) = 0 \quad (3.9.7)$$

Это уравнение полностью определяет все коэффициенты корреляции через два первые ρ_1, ρ_2 .

Общее решение уравнения (3.9.6) имеет вид

$$\rho(\tau) = C_1 \zeta_1^\tau + C_2 \zeta_2^\tau. \quad (3.9.8)$$

Здесь ζ_1, ζ_2 — корни (3.9.5), а C_1 и C_2 — постоянные коэффициенты, которые можно найти из следующей системы уравнений

$$1 = C_1 + C_2, \quad C_1 \zeta_1^{-1} + C_2 \zeta_2^{-1} = C_1 \zeta_1 + C_2 \zeta_2.$$

Первое уравнение получается из условия $\rho(0) = 1$, а второе из условия $\rho(-1) = \rho(1)$. Решая эту систему, получаем

$$\rho(\tau) = \frac{1}{(\zeta_1 - \zeta_2)(1 + \zeta_1 \zeta_2)} \left\{ \zeta_1^{\tau+1} (1 - \zeta_2^2) - \zeta_2^{\tau+1} (1 - \zeta_1^2) \right\}. \quad (3.9.9)$$

Если корни (3.9.5) — комплексные, то, записывая их в тригонометрической форме $\zeta_{1,2} = r e^{\pm i\theta}$, получим

$$r = \sqrt{\alpha_2}, \quad \cos \theta = -\frac{\alpha_1}{2\sqrt{\alpha_2}}, \quad \rho(\tau) = r^\tau \frac{\sin(\tau\theta + \psi)}{\sin \psi},$$

где $\operatorname{tg} \psi = \frac{1+r^2}{1-r^2} \operatorname{tg} \theta$.

На рис. 3.9б представлена реализация процесса AP(2) с параметрами $\alpha_1 = -1, \alpha_2 = 0,5$ ($\rho_1 = 0,73, \rho_2 = 0,31$). Полезно сопоставить с ним гармонический ряд с наложенным белым шумом ε_t (рис. 3.9а).

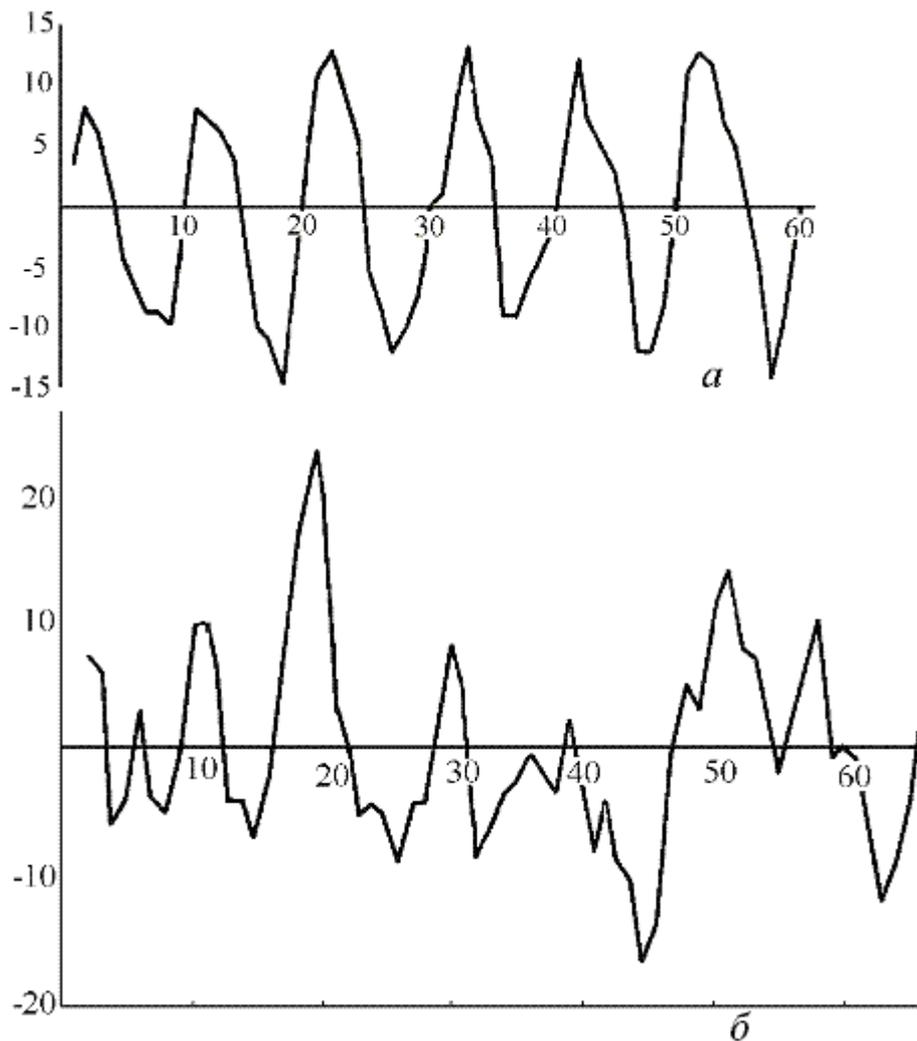


Рисунок 3.9 — Реализация *а*) гармонического ряда (3.9.10) и *б*) процесса $AR(2)$

$$Y_t = \sin \frac{\pi t}{5} + \varepsilon_t. \quad (3.9.10)$$

Дисперсия белого шума $\sigma^2 = 10$, то есть ε_t — равномерно распределенная величина на множестве $-5, -4, \dots, 5$.

3.10 Общий процесс авторегрессии $AR(p)$.

Коррелограмма. Частные автокорреляции

По аналогии с рассмотренными процессами $AR(1)$, $AR(2)$ можно исследовать особенности поведения общего процесса авторегрессии (3.7.1)

$$X_t + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} = \varepsilon_t.$$

Для определения первых p коэффициентов автокорреляции по аналогии с (3.9.2) получим

$$\begin{aligned} \rho_1 + \alpha_1 + \alpha_2 \rho_2 + \dots + \alpha_p \rho_{p-1} &= 0 \\ \rho_2 + \alpha_1 \rho_1 + \alpha_2 + \dots + \alpha_p \rho_{p-2} &= 0 \\ &\dots \\ \rho_p + \alpha_1 \rho_{p-1} + \alpha_2 \rho_{p-2} + \dots + \alpha_p &= 0. \end{aligned} \tag{3.10.1}$$

Коэффициенты $\rho(\tau) = \rho_\tau$ при $\tau > p$ получаются из решения разностного уравнения $\rho(\tau) + \alpha_1 \rho(\tau - 1) + \dots + \alpha_p \rho(\tau - p) = 0$ в форме аналогичной (3.9.8), где, возможно, появятся также произведения полиномов на синусы и косинусы, что, однако, не повлияет на выполнение соотношения $\lim_{\tau \rightarrow \infty} \rho(\tau) = 0$. Таким образом, для АР-процесса характерно затухание автокорреляционной функции с ростом τ при $\tau > p$. Если этот признак не выполняется, процесс не может быть авторегрессионным.

Частные автокорреляции

Наряду с коэффициентами автокорреляции случайные процессы характеризуются коэффициентами частной корреляции. Например, коэффициент $\rho(2)$ характеризует корреляционную связь между X_t и X_{t-2} . Но эта связь определяется, в частности, тем, что X_t зависит от X_{t-1} , а эта величина, в свою очередь, связана с X_{t-2} . Частный коэффициент корреляции $\rho_{13|2}$ характеризует непосредственно зависимость X_t от X_{t-2} , если влияние X_{t-1} элиминировано.

В регрессионном анализе (р. 2, см. также [6, 9, 11, 16]) вводятся частные коэффициенты корреляции $\rho_{12|3\dots n}$, характеризующие степень корреляции переменных с номерами 1 и 2 при фиксированных остальных переменных

$$\rho_{12|3\dots n} = \frac{\sigma^{12}}{\sigma^{11}\sigma^{22}}, \quad (3.10.2)$$

где σ^{ij} — элементы матрицы, обратной матрице ковариаций.

В частности, для интересующего нас случая формула (3.10.2) дает

$$\rho_{13|2} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{1 - \rho_{12}^2}\sqrt{1 - \rho_{23}^2}}. \quad (3.10.3)$$

Для марковского процесса $\rho_{13} = \rho(2) = \rho^2$, $\rho_{12} = \rho_{23} = \rho$. В числителе (3.10.3) получается 0, что и должно быть, так как в марковском процессе X_t зависит лишь от X_{t-1} , и корреляционная связь X_t и X_{t-2} должна отсутствовать.

В процессе AP(2) согласно (3.10.3) и (3.9.3)

$$\rho_{13|2} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = -\alpha_2. \quad (3.10.4)$$

Частные корреляции более высокого порядка в AP(2) равны нулю. Таким образом, частная коррелограмма процесса AP(2) содержит только 2 ненулевых члена.

Аналогично, для AP(p) частные коэффициенты корреляции порядка $\tau > p$ равны нулю. В связи с этим следует вспомнить, что для процессов СС все коэффициенты корреляции, начиная с некоторого τ , оказываются нулевыми.

3.11 Процессы авторегрессии — скользящего среднего АРСС (p, q)

Процесс авторегрессии — скользящего среднего с параметрами p, q АРСС(p, q) определяется соотношением

$$X_t + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}. \quad (3.11.1)$$

Вводя, как и ранее, операторы $\alpha(B) = 1 + \alpha_1 B + \dots + \alpha_p B^p$ и $\beta(B) = 1 + \beta_1 B + \dots + \beta_q B^q$, соотношение (3.11.1) запишем в виде

$$\alpha_p(B) X_t = \beta_q(B) \varepsilon_t. \quad (3.11.2)$$

Если все корни ζ_j характеристического уравнения (3.7.3) для оператора $\alpha_p(B)$ удовлетворяют неравенству $|\zeta_j| < 1$, то существует оператор $\alpha_p^{-1}(B)$ и

$$X_t = \alpha_p^{-1}(B) \beta_q(B) \varepsilon_t, \quad (3.11.3)$$

то есть процесс АРСС может быть представлен как процесс скользящего среднего с бесконечно большим количеством членов.

Практически это представление можно найти методом неопределенных коэффициентов. Продемонстрируем это на примере процесса АРСС(1,1)

$$X_t + \alpha X_{t-1} = \varepsilon_t + \beta \varepsilon_{t-1}. \quad (3.11.4)$$

Ищем X_t в виде $X_t = \sum_{\tau=0}^{\infty} \psi_{\tau} \varepsilon_{t-\tau}$ с неопределенными коэффициентами ψ_0, ψ_1, \dots . Согласно (3.11.3) должно тождественно вы-

полняться операторное равенство $\alpha_1^{-1}(B)\beta_1(B) = \sum_{\tau=0}^{\infty} \psi_{\tau} B^{\tau}$ или же

$\beta_1(B) = \alpha_1(B) \sum_{\tau=0}^{\infty} \psi_{\tau} B^{\tau}$, что в развернутой записи примет вид

$$1 + \beta B = (1 + \alpha B)(\psi_0 + \psi_1 B + \psi_2 B^2 + \dots). \quad (3.11.5)$$

Приравнивая коэффициенты при одинаковых степенях B в правой и левой частях тождества (3.11.5), получаем соотношения $1 = \psi_0$; $\beta = \alpha\psi_0 + \psi_1$; $0 = \alpha\psi_1 + \psi_2$, из которых последовательно находим $\psi_0 = 1$, $\psi_1 = \beta - \alpha$, $\psi_2 = -\alpha(\beta - \alpha)$, ..., $\psi_k = (-\alpha)^{k-1}(\beta - \alpha)$.

Действуя по этому принципу, в общем случае процесса $APCC(p, q)$ мы также получили бы систему линейных уравнений и нашли последовательно неизвестные коэффициенты ψ_k .

Представление процесса $APCC(1, 1)$ (3.11.4) в виде

$$X_t = \varepsilon_t + (\beta - \alpha)\varepsilon_{t-1} - \alpha(\beta - \alpha)\varepsilon_{t-2} + \dots \quad (3.11.6)$$

позволяет получить его коррелограмму. Найдем сначала дисперсию

$$\mathbf{D}X = [1 + (\beta - \alpha)^2 + \alpha^2(\beta - \alpha)^2 + \dots] \sigma^2 = \left[1 + \frac{(\beta - \alpha)^2}{1 - \alpha^2} \right] \sigma^2 = \frac{1 + \beta^2 - 2\alpha\beta}{1 - \alpha^2} \sigma^2$$

и ковариацию первого порядка

$$\text{cov}(X_t, X_{t-1}) = [(\beta - \alpha) - \alpha(\beta - \alpha)^2 - \alpha^3(\beta - \alpha)^2 - \dots] \sigma^2 = \frac{(\beta - \alpha)(1 - \alpha\beta)}{1 - \alpha^2} \sigma^2.$$

Отсюда выводим, что коэффициент корреляции первого порядка ρ_1 равен

$$\rho_1 = \frac{(\beta - \alpha)(1 - \alpha\beta)}{1 + \beta^2 - 2\alpha\beta}. \quad (3.11.7)$$

Ковариация 2-го порядка находится следующим образом $\text{cov}(X_t, X_{t-2}) = \text{cov}(-\alpha X_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}, X_{t-2}) = -\alpha \text{cov}(X_t, X_{t-1})$. Отсюда $\rho_2 = -\alpha \rho_1$. И точно так же $\rho_{\tau+1} = -\alpha \rho_\tau$. Это соотношение, как мы знаем, имеет место для марковского процесса. Отличие заключается в том, что для АРСС(1,1) $\rho_1 \neq -\alpha$, а выражается формулой (3.11.7). Но для следующих коэффициентов автокорреляции ρ_τ , $\tau \geq 2$ имеем $\rho_{\tau+1} = (-\alpha)^\tau \rho_\tau$, то есть коэффициенты автокорреляции убывают со скоростью $|\alpha|^\tau$ так же, как и для марковского процесса.

Для процесса АРСС(p, q) первые $p + q$ коэффициентов автокорреляции выражаются через параметры модели $\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q$, а, начиная с номера равного $\max(p, q + 1)$, «затухают» в том же смысле, что и для процесса авторегрессии.

3.12 Модель авторегрессии — проинтегрированного скользящего среднего АРПСС(p, d, q)

Перейдем теперь к описанию модели нестационарного временного ряда АРПСС. Здесь весьма полезным окажется оператор взятия разности назад $\nabla = 1 - B$. Мы с ним уже встречались. Этот оператор переводит нестационарный процесс (3.5.1) случайного блуждания $X_t = X_{t-1} + \varepsilon_t$ в стационарный процесс $\nabla X_t = \varepsilon_t$. В моделях детерминированных временных рядов ∇ позволяет элиминировать полиномиальный тренд.

Познакомимся с этим оператором поближе. В табл. 3.3 приведены реализация абсолютного случайного ряда (белого шума)

из 20 членов и разности этого ряда от 1-го до 4-го порядка. Члены исходного ряда имеют равномерное распределение на множестве $0,1,2,\dots,99$ ($\sigma^2 = 833,25$). И без специального исследования видно, что дисперсия рядов разностей возрастает с увеличением порядка. Пусть исходный ряд $\varepsilon_1, \dots, \varepsilon_n$ абсолютно случайный. Ряд из разностей порядка r имеет вид

$$\nabla^r \varepsilon_t = \varepsilon_t - C_r^1 \varepsilon_{t-1} + C_r^2 \varepsilon_{t-2} - \dots + (-1)^r \varepsilon_{t-r}.$$

Дисперсия этого ряда имеет вид

$$\mathbf{D}(\nabla^r \varepsilon_t) = \{1 + (C_r^1)^2 + (C_r^2)^2 + \dots + 1\} \mathbf{D}\varepsilon = C_{2r}^r \mathbf{D}\varepsilon. \quad (3.12.1)$$

Покажем, что выражение в фигурных скобках действительно равно C_{2r}^r . Рассмотрим тождество $(x+1)^r (1+x)^r = (x+1)^{2r}$, коэффициент при x^r в левой части тождества равен выражению в фигурных скобках, а в правой части — C_{2r}^r .

Таблица 3.3 — Реализация ряда ε_t и разностей $\nabla \varepsilon_t$ от 1-го до 4-го порядков

t	ε_t	$\nabla^r \varepsilon_t$	$\nabla^2 \varepsilon_t$	$\nabla^3 \varepsilon_t$	$\nabla^4 \varepsilon_t$
1	04	—	—	—	—
2	28	24	—	—	—
3	50	22	-02	—	—
4	13	-37	-59	-57	—
5	92	79	116	175	232
6	17	-75	-154	-270	-445
7	97	80	155	309	579
8	41	-46	-126	-281	-589
9	50	09	55	181	462
10	77	27	18	-37	-218
11	90	13	-14	-32	05
12	71	-19	-32	-18	14
13	22	-49	-30	02	20

t	ε_t	$\nabla^r \varepsilon_t$	$\nabla^2 \varepsilon_t$	$\nabla^3 \varepsilon_t$	$\nabla^4 \varepsilon_t$
14	67	45	94	124	122
15	69	02	-43	-137	-261
16	21	-48	-50	-07	130
17	77	56	104	154	161
18	83	06	-50	-154	-308
19	09	-74	-80	-30	124
20	76	67	141	221	251

Таким образом, дисперсия ряда $\nabla^r \varepsilon_t$ возрастает, но величина

$$V_r = \mathbf{D}(\nabla^r \varepsilon_t) / C_{2r}^r \quad (3.12.2)$$

не зависит от r и равна σ^2 . Отсюда вытекает следующий способ элиминирования тренда временного ряда X_t . Наряду с исходным рядом последовательно при $r = 1, 2, \dots$ рассматриваем разности $\nabla^r X_t$ и оцениваем их дисперсии. Как только величина $\mathbf{D}(\nabla^r \varepsilon_t) / C_{2r}^r$ обнаруживает признаки стабильности, прерываем процесс, полагая, что соответствующую разность можно считать белым шумом.

Приведем примеры, показывающие, как работает представленная методика. Для обработки возьмем 2 ряда по 20 членов в каждом $X_t^{(1)} = 1,5^{t-1} + \varepsilon_t$, $X_t^{(2)} = t^2 - 15t + 10 + \varepsilon_t$ ($t = 1, \dots, 20$), где ε_t — независимые равномерно распределенные на множестве $-9, -8, \dots, 9$ случайные величины. В табл. 3.4 представлены результаты расчетов.

Таблица 3.4 — Примеры: элиминирование тренда с помощью оператора ∇

r	C_{2r}^r	$\mathbf{D}(\nabla^r X_t^{(1)})$	$V_r^{(1)}$	$\mathbf{D}(\nabla^r X_t^{(2)})$	$V_r^{(2)}$
0	1	333920	333920	2109	2109
1	2	38180	19090	160	80
2	6	4522	754	191	32
3	20	1000	50	5755	29
4	70	1496	21	1875	28
5	252	5146	20	6353	25

Из таблицы 3.4 видно, что для ряда $X_t^{(1)}$ существенное снижение показателя V_r наблюдается вплоть до 3-х разностей (а, может быть, и 4-х разностей). Это означает, что полином 3-й степени достаточно хорошо приближает детерминированную составляющую ряда $1,5^{t-1}$ на рассматриваемом множестве $t = 1, \dots, 20$ и 3-я разность практически элиминирует тренд.

Для ряда $X_t^{(2)}$ разность 2-го порядка должна была полностью элиминировать тренд, что и наблюдается в проведенном эксперименте: изменение величины V_r для этого ряда при $r > 2$ незначительно.

Процесс АРПСС(p, d, q) определяется соотношением

$$\alpha_p(B)\nabla^d X_t = \beta_q(B)\varepsilon_t, \quad (3.12.3)$$

то есть это процесс АРСС(p, q) для разностей $\nabla^d X_t$ порядка d . При этом считается, что разность порядка d элиминирует тренд, то есть процесс $Y_t = \nabla^d X_t$ — стационарный. Термин проинтегрированного (скользящего среднего) обязан следующему рассуждению. Операцией, обратной взятию разности, является суммиро-

вание. Аналогами этих двух понятий для функций непрерывного аргумента служат дифференцирование и интегрирование. Обратный переход от разностей к исследуемому процессу и определяется как интегрирование, хотя точнее было бы говорить о суммировании.

Широкому применению к практическим задачам модель АРПСС обязана Боксу и Дженкинсу [4]. Идеи их метода будут изложены в п. 3.17.

3.13 Спектр стационарного случайного процесса с дискретным временем

Определения

Закон распределения случайной величины может быть задан различными способами, в том числе с помощью характеристической функции

$$\varphi(t) = \int_{-\infty}^{\infty} f(x)e^{itx} dx, \quad (3.13.1)$$

где i — мнимая единица, $f(x)$ — плотность распределения с.в. Характеристическая функция $\varphi(t)$ является преобразованием Фурье от плотности. Она может рассматриваться также как производящая функция моментов случайной величины. Действительно, разложив e^{itx} в ряд Маклорена, из (3.13.1) получаем

$$\varphi(t) = \sum_{k=0}^{\infty} v_k \frac{(it)^k}{k!}, \quad (3.13.2)$$

где v_k — начальный момент порядка k .

В некотором смысле аналогом $\varphi(t)$ для случайного процесса с дискретным временем служит функция

$$w(\alpha) = \sum_{k=-\infty}^{\infty} \rho_k e^{ik\alpha}, \quad (3.13.3)$$

где ρ_k — коэффициент корреляции порядка k (то же самое, что $\rho(k)$).

Использование функции $w(\alpha)$ иногда приводит к нужным результатам проще, чем другие формы задания случайного процесса. Полагая $z = e^{i\alpha}$, функцию (3.13.3) можно представить как производящую функцию автокорреляций

$$w(\alpha) = \sum_{k=-\infty}^{\infty} \rho_k z^k. \quad (3.13.4)$$

Функция $w(\alpha)$ называется спектральной плотностью. В силу четности коэффициентов автокорреляции $\rho_{-k} = \rho_k$ из (3.13.3) получаем

$$w(\alpha) = 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos k\alpha. \quad (3.13.5)$$

Это означает, что величины ρ_k являются коэффициентами в разложении $w(\alpha)$ в ряд Фурье по косинусам.

Умножая (3.13.5) на $\cos k\alpha$ и интегрируя от 0 до π , находим

$$\rho_k = \frac{1}{\pi} \int_0^{\pi} w(\alpha) \cos k\alpha d\alpha. \quad (3.13.6)$$

Соотношения (3.13.5) и (3.13.6) свидетельствуют о том, что корреляционная функция $\rho(k)$ и спектральная плотность однозначно определяются друг через друга.

График функции $w(\alpha)$ называется спектром. Из определения (3.13.5) следует, что $w(\alpha)$ имеет период 2π и обладает симметрией $w(2\pi - \alpha) = w(\alpha)$. Поэтому спектр обычно представляют

только на интервале $[0, \pi]$. Часто вместо функции $w(\alpha)$ изображают $\ln w(\alpha)$.

Еще одной характеристикой стационарного случайного процесса является спектральная функция

$$W(\alpha) = \int_0^{\alpha} w(\alpha) d\alpha = \alpha + 2 \sum_{k=1}^{\infty} \rho_k \frac{\sin k\alpha}{k}. \quad (3.13.7)$$

Интерпретация спектральной плотности

Как известно (см. п. 3.2), исчерпывающей характеристикой случайного процесса служит совокупность совместных распределений n случайных величин – значений процесса в моменты времени t_1, t_2, \dots, t_n при любом n и любых t_1, t_2, \dots, t_n . Оценить эти распределения, пользуясь одной или несколькими реализациями случайного процесса, практически невозможно.

Именно поэтому стационарные процессы с дискретным временем описываются более простыми, но зато практически реализуемыми способами:

1) моментами низких порядков: средним, дисперсией и ковариационной функцией;

2) параметрическим представлением в виде линейной комбинации текущего и предыдущих значений процесса, текущего и предыдущих значений белого шума, то есть в виде модели АРСС и ее частных случаев АР и СС моделей;

3) разложением по гармоникам, то есть с помощью спектральной плотности.

Все 3 способа однозначно взаимосвязаны, от любого из них можно перейти к двум другим. Но каждый способ привносит по-

нимание некоторых особенностей процесса, позволяет почувствовать такие его свойства, которые не открываются, если пользоваться другими способами.

Первый из способов не нуждается в комментариях. Кстати, одно из определений стационарности случайного процесса (стационарность в широком смысле) состоит в требовании, чтобы среднее, дисперсия и ковариационная функция не зависели от времени.

Второй способ дает возможность генерировать реализации случайного процесса с помощью метода Монте-Карло.

Осталось прокомментировать третий способ, то есть объяснить, как следует подходить к интерпретации спектральной плотности процесса.

Понять, какие свойства случайного процесса характеризует спектральная плотность, помогают следующие выкладки. Рассмотрим функции

$$a(\alpha) = \frac{1}{\sqrt{\pi n}} \sum_{t=1}^n X_t \cos \alpha t, \quad b(\alpha) = \frac{1}{\sqrt{\pi n}} \sum_{t=1}^n X_t \sin \alpha t. \quad (3.13.8)$$

Эти функции имеют вид скалярного произведения и характеризуют силу взаимосвязи временного ряда с гармоникой периода $2\pi/\alpha$.

Введем функцию $I(\alpha) = a^2(\alpha) + b^2(\alpha)$. Имеем

$$\begin{aligned}
I(\alpha) &= \frac{1}{\pi n} \left[\left(\sum_{t=1}^n X_t \cos \alpha t \right)^2 + \left(\sum_{t=1}^n X_t \sin \alpha t \right)^2 \right] = \\
&= \frac{1}{\pi n} \left\{ \sum_{t=1}^n X_t^2 + 2 \sum_{k=1}^{n-1} \sum_{t=1}^{n-k} \left(\cos \alpha t \cos \alpha (t+k) + \sin \alpha t \sin \alpha (t+k) \right) X_t X_{t+k} \right\} = \\
&= \frac{1}{\pi n} \left\{ \sum_{t=1}^n X_t^2 + 2 \sum_{k=1}^{n-1} \sum_{t=1}^{n-k} X_t X_{t+k} \cos k\alpha \right\} = \frac{s^2}{\pi} \left\{ 1 + 2 \sum_{k=1}^{n-1} r_k \cos k\alpha \right\}.
\end{aligned}$$

Здесь s^2 — оценка дисперсии, r_k — оценка коэффициента автокорреляции порядка k , $s^2 = \frac{1}{n} \sum_{t=1}^n X_t^2$, $r_k = \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k}$. Если $n \rightarrow \infty$, то $s^2 \rightarrow \sigma^2$, $r_k \rightarrow \rho_k$ и в пределе получается

$$\mathbf{M}[I(\alpha)] = \frac{\sigma^2}{\pi} \left\{ 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos k\alpha \right\} = \frac{\sigma^2}{\pi} w(\alpha). \quad (3.13.9)$$

Таким образом, значение $w(\alpha)$ тем больше, чем сильнее взаимосвязь временного ряда с гармоникой периода $2\pi/\alpha$.

Еще более прозрачную интерпретацию интенсивности $I(\alpha)$ можно получить путем следующих рассуждений. Найдем эту функцию для ряда

$$X_t = c \sin \beta t + Y_t, \quad (3.13.10)$$

где c — постоянная, Y_t — некоторый стационарный процесс, не коррелированный с гармоническим неслучайным процессом $c \sin \beta t$.

Для вычисления функций $a(\alpha)$ и $b(\alpha)$ нам понадобятся формулы

$$\sum_{t=1}^n \sin(\gamma t) = \frac{\sin \frac{n\gamma}{2} \sin \frac{(n+1)\gamma}{2}}{\sin \frac{\gamma}{2}}, \quad (3.13.11)$$

$$\sum_{t=1}^n \cos(\gamma t) = -1 + \frac{\cos \frac{n\gamma}{2} \sin \frac{(n+1)\gamma}{2}}{\sin \frac{\gamma}{2}},$$

которые следуют из формулы для суммы членов геометрической прогрессии $\sum_{t=1}^n e^{i\gamma t} = \frac{e^{i\gamma(n+1)} - e^{i\gamma}}{e^{i\gamma} - 1}$.

Построим функцию $I(\alpha)$ и рассмотрим ее при малых $\beta - \alpha$. Сумма $\sum Y_t e^{i\alpha t}$ мала по предположению, поэтому члены, содержащие Y_t , в выражениях для $a(\alpha)$ и $b(\alpha)$ можно не принимать во внимание. Имеем

$$\begin{aligned} a(\alpha) &\sim \frac{c}{\sqrt{\pi n}} \sum_{t=1}^n \sin(\beta t) \cos(\alpha t) = \\ &= \frac{c}{\sqrt{\pi n}} \sum_{t=1}^n \frac{1}{2} [\sin(\beta + \alpha)t + \sin(\beta - \alpha)t], \\ b(\alpha) &\sim \frac{c}{\sqrt{\pi n}} \sum_{t=1}^n \sin(\beta t) \sin(\alpha t) = \\ &= \frac{c}{\sqrt{\pi n}} \sum_{t=1}^n \frac{1}{2} [\cos(\beta - \alpha)t - \cos(\beta + \alpha)t]. \end{aligned} \quad (3.13.12)$$

В силу формул (3.13.11) при $n \rightarrow \infty$ все члены в выражениях (3.13.12) для $a(\alpha)$ и $b(\alpha)$ стремятся к нулю за исключением слагаемого $\frac{c}{\sqrt{\pi n}} \sum_{t=1}^n \frac{\cos(\beta - \alpha)t}{2}$. При $\gamma = \beta - \alpha \rightarrow 0$ получаем

$$\sum_{t=1}^n \cos(\gamma t) \sim \frac{\cos \frac{n\gamma}{2} \sin \frac{(n+1)\gamma}{2}}{\sin \frac{\gamma}{2}} \sim \frac{(n+1)\gamma}{2} / \frac{\gamma}{2} \sim n. \quad (3.13.13)$$

Следовательно, $b(\alpha) \sim \sqrt{n}$, то есть в точке $\alpha = \beta$ существует пик функции $I(\alpha)$, высота которого имеет порядок \sqrt{n} .

Обычно спектр наряду с пиком на основной частоте α имеет боковые пики на кратных частотах $\frac{1}{2}\alpha, \frac{1}{3}\alpha, \dots$. Например, если пик спектра отвечает году, то с большой вероятностью будут также пики, отвечающие 6, 4, 3, 2, 1 месяцам.

Таким образом, спектральная плотность $w(\alpha)$ характеризует степень взаимосвязи между временным рядом и гармоникой с периодом $2\pi/\alpha$ (частотой α).

3.14 Спектры процессов AP(1), AP(2)

Спектр марковского процесса

Для марковского процесса имеют место следующие соотношения а) при $j > 0$ $\rho_j = \rho^j$, б) при $j < 0$ $\rho_j = \rho^{|j|}$. Спектральную плотность найдем, пользуясь ее представлением в виде производящей функции (3.13.4), где $z = e^{i\alpha}$,

$$\begin{aligned} w(\alpha) &= \sum_{k=-\infty}^{\infty} \rho_k z^k = \sum_{k=0}^{\infty} \rho^k z^k + \sum_{k=-1}^{-\infty} \rho^{|k|} z^k = \frac{1}{1-\rho z} + \frac{1}{1-\rho z^{-1}} - 1 = \\ &= \frac{1}{1-\rho e^{i\alpha}} + \frac{1}{1-\rho e^{-i\alpha}} - 1 = \frac{1-\rho^2}{1-2\rho \cos \alpha + \rho^2}. \end{aligned} \quad (3.14.1)$$

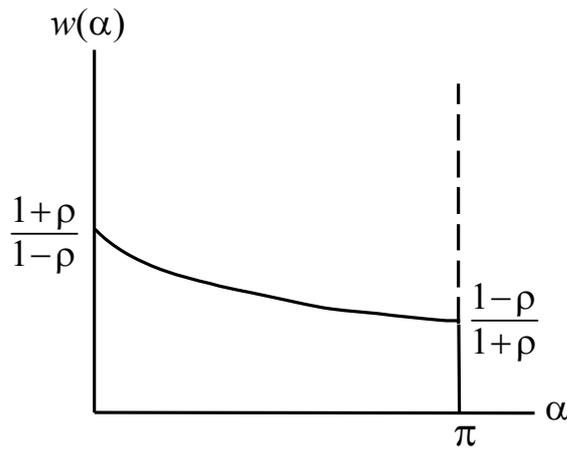


Рисунок 3.10 — Спектр марковского процесса при $\rho > 0$

Спектр марковского процесса для случая $\rho > 0$ изображен на рис. 3.10. Рисунок показывает, что при $\rho > 0$ в спектре преобладают низкие частоты, то есть гармоники с большими периодами. При $\rho < 0$ спектр будет возрастающей на

отрезке $[0, \pi]$ функцией, что свидетельствует о преобладании высоких частот – гармоник с малыми периодами.

Спектр процесса $AP(2)$

Рассмотрим процесс $AP(2)$

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t. \quad (3.14.2)$$

Для определения коэффициентов автокорреляции процесса (3.14.2) рассмотрим соотношение

$$\begin{aligned} \text{cov}(\varepsilon_t, \varepsilon_{t-\tau}) = \text{cov}(X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2}, \\ X_{t+\tau} + \alpha_1 X_{t+\tau-1} + \alpha_2 X_{t+\tau-2}). \end{aligned} \quad (3.14.3)$$

Левая часть равенства (3.14.3) равна

$$\text{cov}(\varepsilon_t, \varepsilon_{t-\tau}) = \begin{cases} \sigma^2 & \text{при } \tau=0 \\ 0 & \text{при } \tau \neq 0 \end{cases}.$$

Правая же часть после деления на \mathbf{DX} записывается как $\alpha_2 \rho_{\tau-2} + (\alpha_1 + \alpha_1 \alpha_2) \rho_{\tau-1} + (1 + \alpha_1^2 + \alpha_2^2) \rho_{\tau} + (\alpha_1 + \alpha_1 \alpha_2) \rho_{\tau+1} + \alpha_2 \rho_{\tau+2}$.

Отсюда видно, что производящая функция автоковариаций процесса ε_t может быть представлена в виде

$$(1 + \alpha_1 z + \alpha_2 z^2) \cdot (1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}) A(z),$$

где $A(z)$ — производящая функция автоковариаций процесса X_t .

Следовательно, спектральная плотность процесса (3.14.2) равна

$$\begin{aligned} w(\alpha) &= \frac{\mathbf{D}\varepsilon}{\mathbf{D}X} / \left\{ 1 + \alpha_1^2 + \alpha_2^2 + \alpha_1(1 + \alpha_2)(e^{i\alpha} + e^{-i\alpha}) + \alpha_2(e^{2i\alpha} + e^{-2i\alpha}) \right\} = \\ &= \frac{\mathbf{D}\varepsilon}{\mathbf{D}X} / \left\{ 1 + \alpha_1^2 + \alpha_2^2 + 2\alpha_1(1 + \alpha_2)\cos\alpha + 2\alpha_2\cos 2\alpha \right\}. \end{aligned} \quad (3.14.4)$$

Отношение дисперсий найдем следующим образом. Полагая в соотношении (3.14.3) $\tau = 0$, получим

$$\begin{aligned} \mathbf{D}\varepsilon &= \mathbf{M}(X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2})^2 = \\ &= \mathbf{D}X(1 + \alpha_1^2 + \alpha_2^2 + 2\alpha_2\rho_2 + 2\alpha_1\rho_1 + 2\alpha_1\alpha_2\rho_1). \end{aligned}$$

Если подставить сюда вместо ρ_1 и ρ_2 их выражения через α_1, α_2 (3.9.3), то получим

$$\mathbf{D}(X)/\mathbf{D}\varepsilon = \frac{1 + \alpha_2}{(1 - \alpha_2)\left[(1 + \alpha_2)^2 - \alpha_1^2\right]}. \quad (3.14.5)$$

На рис. 3.11 изображен спектр процесса (3.14.2) при $\alpha_1 = -1, \alpha_2 = 0,5$. Из графика видно преобладание частот в области максимума $\alpha^* \approx 0,23\pi \quad w(\alpha^*) = 19,2$.

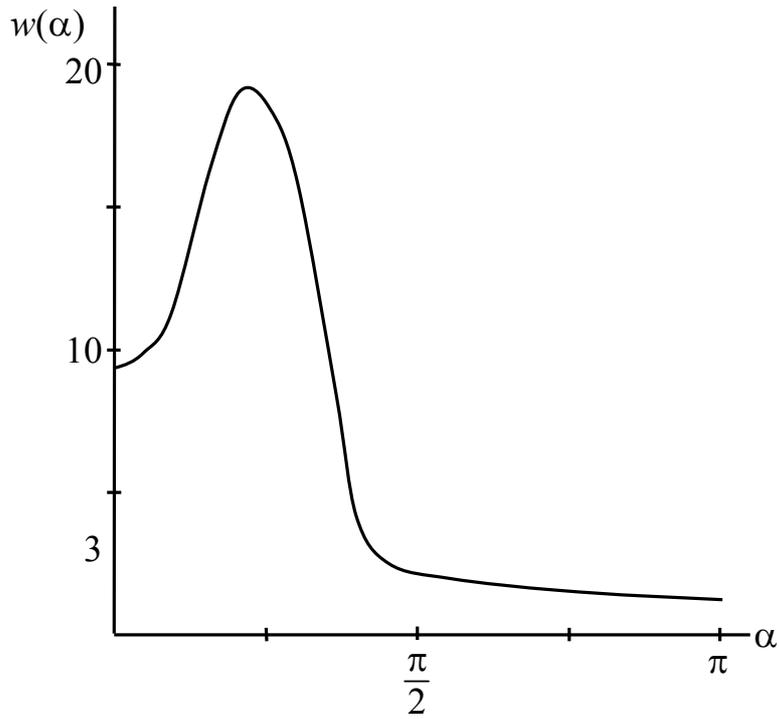


Рисунок 3.11 — Спектр процесса $AR(2)$ при $\alpha_1 = -1$, $\alpha_2 = 0,5$

3.15 Спектральная плотность процесса $ARCC(p, q)$

Использованная выше процедура для построения производящей функции процесса $AR(2)$ может быть обобщена. Рассмотрим процесс Y_t , линейно связанный с процессом X_t

$$Y_t = \sum_{k=0}^{\infty} \alpha_k X_{t-k}. \quad (3.15.1)$$

Заметим, что производящую функцию автоковариаций $B(z)$ процесса Y_t можно выразить через производящую функцию $A(z)$ порождающего процесса X_t . Чтобы это показать, проведем преобразования

$$\begin{aligned} \mathbf{M}(Y_t Y_{t+\tau}) &= \mathbf{M}\left(\sum_{k=0}^{\infty} \alpha_k X_{t-k} \sum_{k=0}^{\infty} \alpha_k X_{t+\tau-k}\right) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha_j \alpha_k \mathbf{M}(X_{t-j} X_{t+\tau-k}) = \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha_j \alpha_k \rho_{|j+\tau-k|} \mathbf{D}X = \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \alpha_j \alpha_{j+\tau-l} \rho_l \mathbf{D}X. \end{aligned}$$

Отсюда видно, что коэффициент в выражении для автоковариации порядка τ ряда Y_t равен коэффициенту при z^τ в выражении $(1 + \alpha_1 z + \alpha_2 z^2 + \dots)(1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots)A(z)$, где $A(z)$ — производящая функция процесса X_t . Это значит, что

$$B(z) = \left(\sum_{k=0}^{\infty} \alpha_k z^k \sum_{k=0}^{\infty} \alpha_k z^{-k} \right) A(z). \quad (3.15.2)$$

Формула (3.15.2) дает также возможность построить производящую функцию автоковариаций процесса $\alpha_p(B)X_t = \beta_q(B)\varepsilon_t$

в виде $\frac{\beta_q(z)\beta_q(z^{-1})}{\alpha_p(z)\alpha_p(z^{-1})}\sigma^2$. Разделив это выражение на $\mathbf{D}X$ и подста-

вив $e^{i\alpha}$ вместо z , получим спектральную плотность процесса $\alpha_p(B)X_t = \beta_q(B)\varepsilon_t$

$$w(\alpha) = \frac{\mathbf{D}\varepsilon}{\mathbf{D}X} \frac{\beta_q(e^{i\alpha})\beta_q(e^{-i\alpha})}{\alpha_p(e^{i\alpha})\alpha_p(e^{-i\alpha})}. \quad (3.15.3)$$

Формула (3.14.4) следует из (3.15.3) как частный случай.

Замечание. Рассуждения, которые привели нас к построению спектральной плотности (3.14.4) процесса Юла (3.14.1) и общего процесса авторегрессии — скользящего среднего $ARCC(p, q)$, могут быть получены с помощью так называемого z -преобразования. Это преобразование является дискретным аналогом преобразования Лапласа. Преобразование Лапласа применяется к функциям непрерывного переменного, определенным на положительной полуоси, а z -преобразование — к решетчатым функциям, определенным на множестве $0, 1, 2, \dots$. z -

преобразование является наиболее удобным аппаратом для решения разностных уравнений (по этому вопросу см. [7, 19]).

3.16 Критерии случайности

Теоретические модели временных рядов, изложенные в пп. 3.5 – 3.15, хотя и содержат конструктивную информацию, но не дают конкретных рекомендаций для анализа процессов по ретроспективным данным. Наибольшее практическое значение имеет процедура, предложенная Боксом и Дженкинсом (см. п. 3.17). Однако прежде, чем переходить к описанию этой процедуры, изложим некоторые полезные для дальнейшего соображения, а именно, напомним о том, что такое эргодические процессы, и приведем ряд критериев случайности.

Эргодические процессы

Стохастический процесс является как бы функцией двух величин: случая и времени $X(\omega, t)$. При каждом фиксированном t $X(\omega, t)$ является случайной величиной. Фиксируя же ω , мы получаем реализацию процесса. Можно ли по одной реализации оценить числовые характеристики случайного процесса, например, среднее $\mathbf{M}X_t$? Строго говоря, для определения этих характеристик следовало бы иметь возможность в каждый момент времени оперировать с генеральной совокупностью случайной величины. Однако обычно для анализа имеется именно одна реализация, а не их совокупность. Тогда, чтобы оценить $\mathbf{M}X_t$, можно оперировать только с этой реализацией.

Результат оценки можно будет считать обоснованным, если осреднение по времени одной реализации в каком-то смысле эквивалентно осреднению по множеству реализаций. Процессы, обладающие таким свойством, называются эргодическими. Эргодические процессы являются стационарными, но обратное не всегда верно. Эргодичность следует понимать как свойство хорошего перемешивания. Взятые из одной реализации значения случайного процесса должны в статистическом смысле составить такую же выборку, как и значения, взятые из разных реализаций в один и тот же момент времени.

Пример стационарного процесса, не являющегося эргодическим

Значение процесса в начальный момент времени берется как реализация случайной величины ω . А далее процесс не меняется во времени $X_t = \text{const}$.

Стационарные процессы $APCC(p, q)$ свойством эргодичности обладают.

Критерии случайности

Критерии случайности [11] предназначены для проверки того, можно ли временной ряд x_1, \dots, x_n считать абсолютно случайным, то есть реализацией белого шума. Вопрос этот решается методами статистической проверки гипотез. Альтернативой к основной гипотезе H_0 о случайности могут служить различные гипотезы:

- ряд нельзя считать абсолютно случайным,

- ряд имеет тенденцию к возрастанию,
- ряд имеет тенденцию к убыванию.

Как всегда, при проверке гипотез назначается уровень значимости α . С уверенностью делается только отрицательный вывод: гипотезу H_0 следует отвергнуть. В случае, когда значение критерия не попадает в критическую область, вывод делается более осторожный: рассматриваемый ряд не дает оснований отвергнуть гипотезу. Принимая во внимание эту идеологию проверки гипотез, целесообразно пользоваться не одним, а несколькими критериями случайности, чтобы с большей уверенностью говорить о случайности рассматриваемого ряда.

Если ряд абсолютно случайный, то есть является белым шумом, то проводить его дальнейшее исследование не имеет смысла. Дополнительной информации (кроме той, что он абсолютно случайный) из него уже не извлечешь: для прогнозирования ряд не годится. Если же ряд не является абсолютно случайным, то целесообразно выявить, имеет ли он тренд (тенденцию к изменению), подобрать подходящий вид тренда, выделить сезонную составляющую, а после элиминирования тренда и сезонной составляющей исследовать остатки методами стационарных случайных процессов.

Существует довольно много критериев случайности. Все их мы перечислять не будем, ограничимся лишь несколькими. Доказательство же проведем только для критерия поворотных точек, чтобы дать представление об идеях, на которых методы основаны.

Критерий поворотных точек

Поворотной точкой называется такой член ряда, который больше двух соседних или меньше двух соседних. Обозначим члены исходного ряда x_1, \dots, x_n и для каждой тройки последовательно расположенных членов x_t, x_{t+1}, x_{t+2} введем счетную переменную

$$\eta_t = \begin{cases} 1, & \text{если } x_{t+1} > \max(x_t, x_{t+2}) \text{ или } x_{t+1} < \min(x_t, x_{t+2}) \\ 0, & \text{в противном случае.} \end{cases}$$

Проверку случайности будем проводить по критерию $\mathfrak{Q} = \sum_{t=1}^{n-2} \eta_t$.

В случае, когда ряд является абсолютно случайным, все возможные варианты взаимного расположения 3-х точек равновероятны. Если упорядочить точки по возрастанию, то возможными вариантами упорядочения будут (1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1). Вариант (1,2,3) означает, что наименьшая из величин 1-я наибольшая 3-я, а 2-я занимает промежуточное положение. Аналогично интерпретируются и остальные варианты. Величина η_t равна нулю только в двух вариантах из 6: первом и последнем. Значит, $\mathbf{P}\{\eta_t = 1\} = \frac{4}{6} = \frac{2}{3}$ и

$$\mathbf{M}\mathfrak{Q} = \frac{2}{3}(n-2). \quad (3.16.1)$$

Если в рассматриваемом временном ряде поворотных точек слишком много или слишком мало, то гипотезу H_0 о том, что ряд абсолютно случайный, придется отвергнуть.

Чтобы определить, что такое «слишком много» или «слишком мало», подсчитаем дисперсию \mathfrak{D}

$$\mathbf{M}\mathfrak{D}^2 = \mathbf{M}\left(\sum_{t=1}^{n-2} \eta_t\right)^2 = \mathbf{M}\left\{\sum \eta_t^2 + 2\sum \eta_t \eta_{t+1} + 2\sum \eta_t \eta_{t+2} + \sum \eta_t \eta_{t+\tau}\right\},$$

где в последней сумме $\tau \neq 0, \pm 1, \pm 2$. Количество членов в 4-х суммах, оказавшихся в фигурных скобках, равно соответственно $n-2$, $n-3$, $n-4$ и $(n-4)(n-5)$. Для первых трех сумм это очевидно, для 4-й проверяем: $(n-2) + 2(n-3) + 2(n-4) + (n-4)(n-5) = (n-2)^2$, то есть действительно, общее количество членов в фигурных скобках равно, как и положено, $(n-2)^2$.

$$\text{Далее, так как } \eta_t^2 = \eta_t, \text{ имеем } \mathbf{M}\sum \eta_t^2 = \mathbf{M}\mathfrak{D} = \frac{2}{3}(n-2).$$

$\mathbf{M}(\eta_t \eta_{t+1})$ находим перебором всех возможных вариантов расположения четырех последовательных точек $x_t, x_{t+1}, x_{t+2}, x_{t+3}$. Равенство $\eta_t \eta_{t+1} = 1$ имеет место, когда точки x_{t+1}, x_{t+2} — поворотные.

Возможные варианты расположения 4-х точек следующие $(1,2,3,4)$, $(1,2,4,3)$, $(1,3,2,4)$, $(1,3,4,2)$, $(1,4,2,3)$, $(1,4,3,2)$ и т.д., всего 24 варианта. Из них $\eta_t \eta_{t+1} = 1$ в 10 случаях. Например, в выписанных выше 6 вариантах только в 3-м и 4-м $\eta_t \eta_{t+1} = 1$. Итак,

$$\mathbf{M}\sum \eta_t \eta_{t+1} = \frac{10}{24}(n-3).$$

Аналогично подсчитывается $\mathbf{M}\sum \eta_t \eta_{t+2} = \frac{9}{20}(n-4)$. Для доказательства этого соотношения приходится перебрать все 120

вариантов возможного расположения пяти последовательных членов временного ряда.

Величины $\eta_t, \eta_{t+\tau}$ в последней сумме независимы, так как тройки, по которым они определяются, не имеют общих членов.

Окончательно получаем

$$D\vartheta = M\vartheta^2 - (M\vartheta)^2 = \frac{16n - 29}{90}. \quad (3.16.2)$$

Доказывается, что распределение с.в. ϑ с ростом n довольно быстро приближается к нормальному. Поэтому для проверки гипотезы H_0 рекомендуется пользоваться нормальным распределением. Подсчитываем по временному ряду число поворотных точек ϑ^* . Если величина $(\vartheta^* - M\vartheta) / \sqrt{D\vartheta}$ попадает вне интервала $[-1,96; 1,96]$, то гипотеза H_0 отвергается при уровне значимости $\alpha = 0,05$.

Приведем пример применения критерия. Временной ряд из 50 членов содержит $\vartheta^* = 38$ поворотных точек. Проверим гипотезу о случайности при уровне значимости $\alpha = 0,05$. Находим моменты – среднее и дисперсию – величины ϑ $M\vartheta = \frac{2}{3}(50 - 2) = 32$, $D\vartheta = \frac{16 \cdot 50 - 29}{90} = 8,57$. Значение критерия $\frac{\vartheta^* - M\vartheta}{\sqrt{D\vartheta}} = 2,05$ попадает в критическую область, то есть вне интервала $\pm 1,96$. Гипотеза о случайности временного ряда отвергается. Число поворотных точек слишком велико, чтобы ряд можно было считать абсолютно случайным.

Критерий ранговой корреляции Кендалла

Критерий ранговой корреляции учитывает ранги каждой пары $x_t, x_{t+\tau}$ членов ряда. Подсчитаем число случаев, когда $x_{t+\tau} > x_t$ при $\tau > 0$. Обозначим это число через P . Общее число пар равно C_n^2 . Вероятность того, что число с большим номером больше, для абсолютно случайного ряда равна $1/2$. Значит, $\mathbf{M}P = \frac{1}{4}n(n-1)$.

Вместо P принято рассматривать коэффициент Кендалла $v = \frac{4P}{n(n-1)} - 1$; $\mathbf{M}v = 0$, $-1 \leq v \leq 1$. При справедливости гипотезы

H_0 (ряд абсолютно случайный) — $\mathbf{D}v = \frac{2(2n+5)}{9n(n-1)}$. Распределение

v с ростом n стремится к нормальному, поэтому проверка гипотезы, как и в случае критерия поворотных точек, проводится с помощью нормального распределения $v/\sqrt{\mathbf{D}v} \in N(0,1)$.

Критерий Фостера-Стюарта

Критерий служит для выявления того, можно ли считать, что ряд имеет тенденцию к убыванию или возрастанию.

Для каждого t подсчитываются 2 величины

$$k_t = \begin{cases} 1, & \text{если } x_t > \max(x_{t-1}, x_{t-2}, \dots, x_1) \\ 0 & \text{в противоположном случае,} \end{cases}$$
$$l_t = \begin{cases} 1, & \text{если } x_t < \min(x_{t-1}, x_{t-2}, \dots, x_1) \\ 0 & \text{в противоположном случае.} \end{cases}$$

По этим величинам вычисляются критерии $S = \sum_{t=2}^n S_t$,

$S_t = k_t + l_t$ и $d = \sum_{t=2}^n d_t$, $d_t = k_t - l_t$. Очевидно, $0 \leq S \leq n-1$. Наиболь-

шее значение $S = n - 1$ достигается для монотонно возрастающего ряда, а наименьшее $S = 0$ — для монотонно убывающего. Критерий d заключен в границах $-(n - 1) \leq d \leq n - 1$. Правая граница опять-таки достигается при монотонном возрастании, а левая — при монотонном убывании ряда.

Гипотезы о наличии тенденции к возрастанию или убыванию проверяются по значениям статистик $(S - \mu)/\sigma_1$ и d/σ_2 , которые распределены по Стьюденту. Величины μ, σ_1, σ_2 табулированы.

Двухвыборочные критерии

Гипотезы о наличии тенденций к возрастанию или убыванию могут быть проверены также с помощью двухвыборочных критериев. При этом ряд делится на 2 части, и проверяется, имеются ли основания считать, что одна из них в среднем больше, чем другая. Конечно, двухвыборочные критерии используют не всю информацию, которую содержит временной ряд: ведь порядок следования членов в каждой из частей двухвыборочный критерий не учитывает.

Из двухвыборочных критериев опишем только технику критерия Вилкоксона, для которого составлены легкодоступные таблицы. При проверке по Вилкоксоу надо подсчитать сумму рангов членов обеих выборок. Покажем на примере, как это делается.

Пример. Временной ряд 3, 2, 4, 4, 6, 4, 2, 7, 7 разбит на 2 части. В первую 3, 2, 4, 4 включено 4 члена, во вторую 6, 4, 2, 7, 7

— пять. Упорядочим все члены ряда по возрастанию и назовем рангом члена его номер в объединенной выборке. Если 2 или более членов равны, припишем им равные ранги — среднееарифметическое их номеров. В примере получим: упорядоченная общая выборка — 2, 2, 3, 4, 4, 4, 6, 7, 7; ранги членов — 1,5; 1,5; 3; 5; 5; 5; 7; 8,5; 8,5. Сумма рангов второй выборки $W = 7 + 5 + 1,5 + 8,5 + 8,5 = 30,5$.

Если сумма рангов 2-й выборки велика, то можно говорить о тенденции к возрастанию, если мала, то о тенденции к убыванию. Граничные значения критической области приведены в специальных таблицах [напр., 5].

3.17 Процедуры обработки временных рядов.

Модель Бокса–Дженкинса

Бокс и Дженкинс в деталях разработали процедуру выбора и идентификации модели стационарного процесса по последовательности фактических наблюдений (временному ряду). Поиск подходящей модели ведется на совокупности процессов АРПСС(p, d, q)

$$\nabla^d X_t + \alpha_1 \nabla^d X_{t-1} + \dots + \alpha_p \nabla^d X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}. \quad (3.17.1)$$

Модель (3.17.1) содержит структурные параметры p, d, q и параметры $\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q$. В соответствии с рекомендациями Бокса – Дженкинса, сначала определяются структурные параметры, а затем, при выбранных значениях p, d, q , параметры $\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q$. Общее количество параметров (степеней сво-

боды модели) достаточно велико и предоставляет широкие возможности для определения скрытых механизмов, регулирующих поведение временных рядов, и, следовательно, для прогнозирования процессов.

Процедуру идентификации – подбора параметров модели по временному ряду – целесообразно разбить на этапы.

Этап 1. Определение порядка разности d .

Этап 2. Определение при фиксированных структурных параметрах величин α_i, β_j ($i = 1, \dots, p; j = 1, \dots, q$).

Этап 3. Обоснование наиболее подходящих значений структурных параметров p и q методом вычислительного эксперимента.

Охарактеризуем каждый из этих этапов. На **этапе 1** исходный временной ряд приводится к стационарному, если это необходимо. Техника сведения была описана выше (п. 3.12). Фактически обработка временных рядов проводится с помощью программных комплексов. Можно рекомендовать, например, систему STATISTICA, подробное описание которой приведено в монографии [6]. При подборе подходящего порядка разности d поступают следующим образом. Последовательно увеличивая d , сопоставляют получающиеся при этом значения критерия V_r (3.12.2), пока этот критерий не стабилизируется. На практике чаще всего довольствуются разностями первого или второго порядка.

К началу этапа 2 значение d уже выбрано и рассматривается модель

$$Y_t + \alpha_1 Y_{t+1} + \dots + \alpha_p Y_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}, \quad (3.17.2)$$

где $Y_t = \nabla^d X_t$.

Процедуру **этапа 2** представим как многошаговую.

Шаг 1. Соотношение (3.17.2) умножаем последовательно на $Y_{t-(q+1)}, \dots, Y_{t-(q+p)}$ и к произведению применяем операцию математического ожидания. В правой части полученных равенств будет стоять 0, так как $Y_{t-(q+1)}, \dots, Y_{t-(q+p)}$ не зависят от $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$.

Получим систему линейных уравнений

$$\begin{aligned} c_{q+1} + \alpha_1 c_q + \alpha_2 c_{q-1} + \dots + \alpha_p c_{q-p+1} &= 0, \\ c_{q+2} + \alpha_1 c_{q+1} + \alpha_2 c_q + \dots + \alpha_p c_{q-p+2} &= 0, \\ &\dots \\ c_{q+p} + \alpha_1 c_{q+p-1} + \alpha_2 c_{q+p-2} + \dots + \alpha_p c_q &= 0, \end{aligned} \quad (3.17.3)$$

Здесь c_τ — оценка автоковариации порядка τ $c_\tau = \frac{1}{n} \sum_{t=\tau+1}^n Y_t Y_{t-\tau}$.

В случае $q+1 < p$ появляются величины c с отрицательным индексом, тогда $c_{-\tau} = c_\tau$. Решение линейной системы уравнений (3.17.3) даст оценки $\hat{\alpha}_1^{(1)}, \dots, \hat{\alpha}_p^{(1)}$ параметров $\alpha_1, \dots, \alpha_p$, которые будем рассматривать как первое приближение итерационного процесса.

Шаг 2. Определение первого приближения для величин β_j .

Для величин β_j получается нелинейная система уравнений. Введем эту систему для случая $p = q = 2$

$$Y_t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}.$$

Этот избавит нас от громоздких записей, но позволит понять, как надо действовать в общем случае.

Обозначим $Z_t = Y_t + \hat{\alpha}_1^{(1)} Y_{t-1} + \hat{\alpha}_2^{(1)} Y_{t-2}$, где $\hat{\alpha}_1^{(1)}$, $\hat{\alpha}_2^{(1)}$ — оценки параметров, найденные на шаге 1. Получаем $Z_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$. Оценки ковариаций величин Z_t обозначим C_τ

$C_\tau = \frac{1}{n} \sum_{t=\tau+1}^n Z_t Z_{t-\tau}$. Для оценки величин β_1, β_2 имеем

$$\begin{aligned} C_0 &= (1 + \beta_1^2 + \beta_2^2) \mathbf{D}\varepsilon, \\ C_1 &= (\beta_1 + \beta_1 \beta_2) \mathbf{D}\varepsilon, \\ C_2 &= \beta_2 \mathbf{D}\varepsilon. \end{aligned} \quad (3.17.4)$$

Уравнения системы (3.17.4) получены из определения входящих в нее величин, например, второе следует из соотношения $\mathbf{M}(Z_t Z_{t-1}) = \mathbf{M}[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2})(\varepsilon_{t-1} + \beta_1 \varepsilon_{t-2} + \beta_2 \varepsilon_{t-3})] = (\beta_1 + \beta_1 \beta_2) \mathbf{D}\varepsilon$. Система (3.17.4), помимо β_1, β_2 , содержит величину $\mathbf{D}\varepsilon$, которую тоже надо в общем случае считать неизвестной.

Исключая из (3.17.4) $\mathbf{D}\varepsilon$, получаем для оценки $\hat{\beta}_1, \hat{\beta}_2$ систему двух уравнений с двумя неизвестными

$$1 + \hat{\beta}_1^2 + \hat{\beta}_2^2 - \frac{C_0}{C_2} \hat{\beta}_2 = 0, \quad \hat{\beta}_1 + \hat{\beta}_1 \hat{\beta}_2 - \frac{C_1}{C_2} \hat{\beta}_2 = 0. \quad (3.17.5)$$

То, что система нелинейна, вносит в ее решение некоторые трудности.

Шаг 3. Найденные оценки $\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q$ используем как начальное приближение для процедуры минимизации остаточной суммы квадратов.

Идею метода продемонстрируем на примере

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = \varepsilon_t + \beta_1 \varepsilon_{t-1}. \quad (3.17.6)$$

Напомним, что метод наименьших квадратов находит строгое обоснование в случае нормального закона. Положим, что величины $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-\tau+1}$ независимы и имеют распределение Гаусса $\varepsilon_i \in N(0, \sigma^2)$, $i = t - \tau + 1, \dots, t$. Значит, плотность их совместного распределения имеет вид

$$f(z_1, \dots, z_\tau) = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^\tau \exp\left(-\frac{z_1^2 + \dots + z_\tau^2}{2\sigma^2} \right).$$

Метод максимального правдоподобия — наиболее обоснованный и естественный метод получения оценок — сводится к минимизации суммы квадратов (логарифма от функции правдоподобия)

$$z_1^2 + \dots + z_\tau^2 \rightarrow \min. \quad (3.17.7)$$

В модели (3.17.6) оцениваются 3 параметра $\alpha_1, \alpha_2, \beta_1$. Последовательность реализаций z_t величин ε_t получаем из соотношений

$$\begin{aligned} z_3 &= X_3 + \alpha_1 X_2 + \alpha_2 X_1, \\ z_4 &= X_4 + \alpha_1 X_3 + \alpha_2 X_2 - \beta_1 z_3 \text{ и т.д.} \end{aligned} \quad (3.17.8)$$

Подставляя эти значения в (3.17.7), получаем задачу минимизации суммы квадратов

$$\begin{aligned} S(\alpha_1, \alpha_2, \beta_1) &= (X_3 + \alpha_1 X_2 + \alpha_2 X_1)^2 + \\ &+ (X_4 + \alpha_1 X_3 + \alpha_2 X_2 - \beta_1 z_3)^2 + \dots \\ &= \sum_{t=1}^{n-2} (X_{t+2} + \alpha_1 X_{t+1} + \alpha_2 X_t - \beta_1 z_{t+1})^2 \rightarrow \min. \end{aligned} \quad (3.17.9)$$

Общая процедура параметрической оптимизации строится следующим образом. Начальное приближение $\hat{\alpha}_i^{(1)}, \hat{\beta}_j^{(1)}$, найденное на шагах 1 и 2, позволяет вычислить $z_t^{(1)}$ ($t = 3, 4, \dots, n$) по формулам (3.17.8).

Подставляя $z_t^{(l)}$, найденные на итерации l , в (3.17.9), находим $\hat{\alpha}_i^{(l+1)}, \hat{\beta}_j^{(l+1)}$ путем минимизации функции $S(\alpha_1, \alpha_2, \beta_1)$. Если итерационный процесс нельзя считать завершенным по требованиям точности, то эти величины используются для вычисления $z_t^{(l+1)}$.

Этап 3. Структурная оптимизация

Структурные параметры модели p и q находятся неформальными методами. Модель стационарного процесса (3.17.2) применяется для некоторых, выбранных заранее значений p и q . Обычно сопоставляются модели, указанные в табл. 3.5.

Таблица 3.5 — Наиболее применяемые модели стационарных процессов

Модель	p	q	Название	Признак
Модель 1	0	0	Белый шум	Критерии случайности
Модель 2	0	1	СС(1)	Малы значения $\rho(\tau)$ $\tau \geq 2$
Модель 3	0	2	СС(2)	Малы значения $\rho(\tau)$ $\tau \geq 3$
Модель 4	1	0	АР(1)	Малы ч.к.к. *) при $\tau \geq 2$
Модель 5	2	0	АР(2)	Малы ч.к.к. при $\tau \geq 3$
Модель 6	1	1	АРСС(1,1)	

*) ч.к.к. — частные коэффициенты корреляции

Критерием для выбора подходящей модели служит обычно остаточная сумма квадратов $S(\hat{\alpha}_1, \dots, \hat{\alpha}_p; \hat{\beta}_1, \dots, \hat{\beta}_q)$, то есть сумма

в выражении типа (3.17.9), куда подставлены найденные оценки коэффициентов $\hat{\alpha}_i, \hat{\beta}_j$. При этом используются стандартные рассуждения регрессионного анализа, то есть рассматривается отношение двух статистик вида

$$s(p, q) = \frac{1}{n - p - q} S(\hat{\alpha}_1, \dots, \hat{\alpha}_p; \hat{\beta}_1, \dots, \hat{\beta}_q),$$

которое удовлетворяет распределению Фишера. Если при увеличении числа параметров модели путем введения еще одного параметра, остаточная сумма квадратов уменьшается незначительно, то введение этого параметра в модель нецелесообразно.

Так, может оказаться предпочтительной модель 1, если переход к однопараметрическим моделям 2 и 4 не приводит к существенному снижению остаточной суммы квадратов.

3.18 Оценка среднего, корреляционной функции и спектральной плотности

При оценке числовых характеристик, связанных с моделями стационарных случайных процессов, следует помнить замечания о свойстве эргодичности стационарных случайных процессов, приведенные в п. 3.16.

Если установлено, что наблюдаемый временной ряд x_1, \dots, x_n относится к стационарному случайному процессу, то несмещенной оценкой для среднего μ этого процесса служит

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \quad (3.18.1)$$

Эта оценка состоятельная. Доказывается [2, 6], что дисперсия этой оценки имеет порядок n^{-1} $\mathbf{D}\bar{x} \sim \frac{2\pi}{n} w(0)$, где $w(\alpha)$ — спектральная плотность порождающего процесса X_t .

Автоковариационная функция процесса X_t оценивается в зависимости от того, известно ли среднее значение μ процесса X_t или оно оценивается по временному ряду. В первом случае оценку автоковариации $\text{cov}(X_t, X_{t-\tau})$ следует проводить по формуле

$$C_n(\tau) = \frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (x_t - \mu)(x_{t+\tau} - \mu), \quad (3.18.2)$$

а во втором по формуле

$$\tilde{C}_n(\tau) = \frac{1}{n-\tau} \sum_{t=1}^{n-\tau} (x_t - \bar{x})(x_{t-\tau} - \bar{x}), \quad (3.18.3)$$

где \bar{x} определяется соотношением (3.18.1).

В частности, (3.18.3) дает следующую оценку дисперсии

$$S^2 = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2. \quad (3.18.4)$$

Следует заметить, что множитель перед суммой равен n^{-1} , а не $(n-1)^{-1}$. Это значит, что оценка дисперсии получается смещенной.

Иногда и при оценке автокорреляций в формуле (3.18.3) множитель $(n-\tau)^{-1}$ заменяют на n^{-1} (см. формулу для оценки автоковариаций в п. 3.17). Это имеет определенные плюсы. Ковариационная матрица, составленная из таких оценок, будет положительно определенной.

В качестве оценки частной автоковариации $\hat{\varphi}_{\tau\tau}$ (см. п. 3.10) принимается оценка коэффициента $\varphi_{\tau\tau}$ в следующей регрессионной модели

$$x_t - \bar{x} = \varphi_{\tau 1} (x_{t-1} - \bar{x}) + \dots + \varphi_{\tau\tau} (x_{t-\tau} - \bar{x}) + \varepsilon_t \quad (3.18.5)$$

Оценка спектральной плотности строится при известном среднем μ по формуле

$$f_n(\alpha) = \sum_{|\tau| < n} C_n(\tau) \left(1 - \frac{|\tau|}{n}\right) \cos k\alpha, \quad \alpha \in [-\pi, \pi], \quad (3.18.6)$$

а при неизвестном среднем по формуле

$$\tilde{f}_n(\alpha) = \sum_{|\tau| < n} \tilde{C}_n(\tau) \left(1 - \frac{|\tau|}{n}\right) \cos k\alpha, \quad \alpha \in [-\pi, \pi]. \quad (3.18.7)$$

В формуле (3.18.6) $C_n(\tau)$ определяется формулой (3.18.2), а в формуле (3.18.7) $\tilde{C}_n(\tau)$ представляется через (3.18.3). Кроме того, спектральная плотность может оцениваться в соответствии с выведенной ранее формулой (3.13.9) как $\hat{w}(\alpha) = \frac{\pi}{S^2} I(\alpha)$, где $I(\alpha)$ определено в п. 3.13.

Со свойствами приведенных здесь оценок можно познакомиться по работам [1, 8, 12].

3.19 Локальное сглаживание временных рядов (метод скользящих средних)

Одной из естественных и часто используемых процедур первичной обработки временных рядов является локальное сглаживание. В сглаженном ряде устраняется в какой-то мере слу-

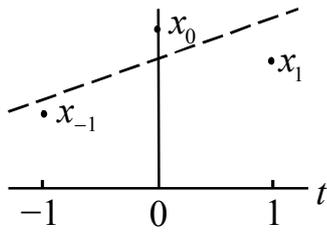


Рисунок 3.12 — Локальное сглаживание по трем точкам

чайность, тенденции локального изменения процесса становятся яснее.

Начнем с простейшего примера. Возьмем три последовательных значения ряда, приписав им номера $-1, 0, 1$ (рис. 3.12). Оценивая поведение ряда при $t = 0$ по трем значениям в этих точках, следует сделать вывод, что имеет место тенденция к возрастанию. Чтобы численно охарактеризовать эту тенденцию, попробуем представить поведение процесса на отрезке $-1, 1$ линейной функцией $y(t) = a_0 + a_1 t$. Будем при этом исходить из требования минимума среднеквадратичного рассогласования

$$\sum_{t=-1}^1 (x_t - a_0 - a_1 t)^2 \rightarrow \min. \quad (3.19.1)$$

Для определения коэффициентов a_0, a_1 , приравняем к нулю производные по a_0 и a_1 от суммы в (3.19.1). Получим

$$\sum_{t=-1}^1 (x_t - a_0 - a_1 t) = 0, \quad \sum_{t=-1}^1 t(x_t - a_0 - a_1 t) = 0,$$

откуда $a_0 = \frac{1}{3} \sum x_t = \bar{x}$, $a_1 = \frac{-x_{-1} + x_1}{2}$.

Значит, линейный прогноз по точкам x_{t-1}, x_t, x_{t+1} определяется формулой

$$\begin{aligned} y_t = y(t) &= \bar{x} + \frac{1}{2}(-x_{t-1} + x_{t+1})t = \\ &= \frac{1}{6}[(2 - 3t)x_{t-1} + 2x_t + (2 + 3t)x_{t+1}]. \end{aligned} \quad (3.19.2)$$

Чем больше упреждение прогноза, тем больше должна быть ошибка. Проверим, как это проявится в рассматриваемом примере. Мерой ошибки служит дисперсия. Обозначая дисперсию наблюдения $\mathbf{D}X = \sigma^2$, в соответствии с (3.19.2) получим $\mathbf{D}Y = \frac{1}{36} \left[(2-3t)^2 + 4 + (2+3t)^2 \right] \sigma^2 = \frac{2+3t^2}{6} \sigma^2$. Это значит $\mathbf{D}y(0) = 0,33\sigma^2$; $\mathbf{D}y(1) = 0,87\sigma^2$; $\mathbf{D}y(2) = 2,33\sigma^2$; $\mathbf{D}y(3) = 4,83\sigma^2$; $\mathbf{D}y(4) = 8,33\sigma^2$; $\mathbf{D}y(5) = 12,83\sigma^2$.

В этом примере мы произвели сглаживание по 3 точкам полиномом 1-й степени. Нетрудно получить общие результаты, проводя сглаживание по $2m+1$ точкам полиномом степени p . Припишем точкам номера $-m, -m+1, \dots, m$. Если количество точек четно, рассуждения следует несколько изменить. Случай четного количества точек $2m$ имеет меньшее значение, поэтому его исследование не проводится.

Положим для определенности $2m+1 = 7, p = 3$, то есть сгладим ряд по 7 последовательным значениям $x_{-3}, x_{-2}, \dots, x_3$ полиномом 3-й степени

$$y(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3. \quad (3.19.3)$$

В качестве критерия для определения неизвестных параметров a_0, \dots, a_3 возьмем минимум суммы квадратов отклонений.

$$\sum_{t=-3}^3 (x_t - a_0 - a_1 t - a_2 t^2 - a_3 t^3)^2 \rightarrow \min.$$

Действуя так же, как и раньше, для неизвестных a_0, \dots, a_3 получим 4 уравнения

$$\begin{array}{rcl}
7a_0 & +28a_2 & = \Sigma x_t \\
& 28a_1 & +196a_3 = \Sigma tx_t \\
28a_0 & +196a_2 & = \Sigma t^2 x_t \\
& 196a_1 & +1588a_3 = \Sigma t^3 x_t
\end{array} \quad (3.19.4)$$

Наибольший интерес представляет значение сглаживающего полинома (3.19.3) при $t = 0$, то есть a_0 . Решая систему (3.19.4), получаем

$$y(0) = a_0 = \frac{1}{21}(-2x_{-3} + 3x_{-2} + 6x_{-1} + 7x_0 + 6x_1 + 3x_2 - 2x_3). \quad (3.19.5)$$

Сглаживающую формулу (3.19.5), можно записать в следующем виде

$$y_t = \frac{1}{21}(9\Delta^4 + 9\Delta^5 + 2\Delta^6)x_{t-3}. \quad (3.19.6)$$

где Δ — оператор разности вперед. Формула (3.19.6) проверяется непосредственно.

На практике применяется не только сглаживание по критерию наименьших квадратов, но и другие методы, чаще всего — скользящее среднее арифметическое (простое осреднение). Применяются также комбинированные процедуры. Например, можно взять простое осреднение по трем точкам, а затем в сглаженном ряде простое осреднение по 5 точкам. В результате получим сглаживающую формулу по 7 точкам

$$\begin{array}{cccccc}
 1 & 1 & 1 & & & \\
 & 1 & 1 & 1 & & \\
 & & 1 & 1 & 1 & \\
 & & & 1 & 1 & 1 \\
 & & & & 1 & 1 & 1 \\
 \hline
 1 & 2 & 3 & 3 & 3 & 2 & 1
 \end{array}$$

$$y_t = \frac{1}{15}(x_{t-3} + 2x_{t-2} + 3x_{t-1} + 3x_t + 3x_{t+1} + 2x_{t+2} + x_{t+3}). \quad (3.19.7)$$

Сглаживающие формулы могут использоваться не только для определения сглаженных значений в середине интервала, но и в других, в том числе в крайних точках. Сглаженный ряд применяется для экстраполяции – прогноза.

Эффект Слуцкого-Юла

Применяя процедуры сглаживания, надо помнить, что сглаженный ряд приобретает некоторые свойства, не присущие исходному ряду. Чтобы это продемонстрировать, возьмем в качестве исходного ряда абсолютно случайный ряд ε_t (белый шум) и применим к нему процедуру простого осреднения по трем точкам

$$y_t = \frac{1}{3}(\varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1}).$$

Получим
$$\mathbf{D}y = \frac{1}{9} \cdot 3 \cdot \mathbf{D}\varepsilon = \frac{1}{3}\sigma^2, \quad \text{cov}(y_t, y_{t+1}) =$$

$$\text{cov}\left(\frac{\varepsilon_{t-1} + \varepsilon_t + \varepsilon_{t+1}}{3}, \frac{\varepsilon_t + \varepsilon_{t+1} + \varepsilon_{t+2}}{3}\right) = \frac{2}{9}\sigma^2. \quad \text{Значит, } \rho_1 = \frac{2}{3}.$$

Так же подсчитывается $\rho_2 = \frac{1}{3}$. Другими словами, процедура простого осреднения снизила дисперсию исходного ряда и превратила белый шум в ряд с коррелированными членами.

В общем случае, если $y_{t+m+1} = \sum_{j=1}^{2m+1} a_j \varepsilon_{t+j}$, то

$$Dy = D\varepsilon \sum_{\tau=1}^{2m+1} a_{\tau}^2, \quad \text{cov}(y_t, y_{t+\varepsilon}) = (a_1 a_{1+\tau} + a_2 a_{2+\tau} + \dots + a_{2m+1-\tau} a_{2m+1}) D\varepsilon.$$

В случае $2m+1=7$, $p=3$ получим следующие коэффициенты корреляции $\rho_0 = 1$; $\rho_1 = 0,73$; $\rho_2 = 0,37$; $\rho_3 = 0,05$; $\rho_4 = 0,10$; $\rho_5 = 0,08$; $\rho_6 = 0,03$; $\rho_{\tau} = 0$ при $\tau \geq 7$.

Эффект существенного отличия характеристик сглаженного ряда от характеристик исходного был выявлен независимо Е.Е. Слуцким и Г. Юлом.

4 ЭКСПЕРТНЫЙ ЛОГИЧЕСКИЙ АНАЛИЗ

4.1 Краткая справка

Формальные методы оказываются полезными, когда прогноз строится на базе статистической информации. В таком случае обычно полагают, что условия в будущем на момент прогноза не слишком сильно отличаются от текущих условий и ретроспективы, которой отвечает накопленная информация. Однако в эпоху перемен, турбулентной смены обстоятельств в стране и в мире фактическая экстраполяция текущего положения на будущее становится малоубедительной. Для прогнозирования в сфере экономики, энергетики, техники и технологий остается еще один путь – воспользоваться экспертными знаниями, опереться на интуицию и опыт специалистов.

Но здесь возникает проблема, как лучше раскрыть знания, интуицию и опыт специалистов. Ее решению способствует экспертный логический анализ (ЭЛА). ЭЛА в целом упорядочивает имеющуюся качественную (неформализованную) и количественную (статистическую) информацию, а также мнения и оценки экспертов. ЭЛА имитирует и дополняет реальный процесс подготовки и принятия ответственных решений. ЭЛА является систематической процедурой, позволяющей formalизовать участие экспертов в процессе принятия решений.

Экспертный логический анализ основывается на методе, который разработан Беллманом – Бруком – Бурковым, но получил широкую известность по работам Т.Саати. Публикации Саати [14, 15] более полно раскрыли возможности процедуры, назван-

ной им методом анализа иерархий. Однако это название, на наш взгляд, уводит от сути дела. Термин экспертный логический анализ является значительно более выразительным, и мы им будем пользоваться как эквивалентом термина метод анализа иерархий.

4.2 Техника экспертного логического анализа

Процедура ЭЛА включает итеративную декомпозицию и обработку суждений экспертов по парным сравнениям, выраженным в специальных шкалах. Структурируем процедуру – разобьем ее на несколько этапов.

Этап 1 — построение логической схемы исследования

Экспертный анализ начинается с логической схемы исследования. Логическая схема состоит из блоков (элементов), которые располагаются на разных уровнях – уровнях иерархии. (Этим, кстати, и объясняется название – метод анализа иерархий.) Рассмотрим один из возможных вариантов – пятиуровневую логическую схему (рис. 4.1)

Верхний (нулевой) уровень отражает общую цель исследования. На следующем (первом) уровне в рассматриваемом нами варианте логической схемы располагаются акторы. Этим словом в ЭЛА обозначаются субъекты, оказывающие влияние на принятие решения. Акторами могут быть люди, принимающие решение, субъекты экономической деятельности, административные и общественные организации, от которых зависит принятие решения и т.д. На втором уровне размещены критерии, по которым сопоставляются конкурирующие варианты, на третьем – подкритерии, а на четвертом – сами варианты.



Рисунок 4.1 — Логическая схема исследований для ранжирования вариантов решений с учетом значимости акторов, критериев оценки с детализацией последних (подкритериев)

Все этапы процедуры ЭЛА будем иллюстрировать примерами. В примере 4.1 рассматривается бытовая ситуация, понимание которой не требует специальных знаний в предметной области. Пример 4.2, хотя также имеет иллюстративный характер, относится к актуальной проблеме нефтегазового комплекса.

Пример 4.1. Семья, состоящая из 4-х человек, решает ответственный для себя вопрос о покупке автомашины. Целью исследования (уровень 0) является выбор автомашины. Акторами (уровень 1) выступают все члены семьи: муж, жена, дочь и сын. Они же в данном случае являются экспертами, на основании суждений которых принимается окончательное решение.

Из большого количества предлагаемых к продаже автомашин, новых и с пробегом, акторами предварительно были отобраны 3 варианта. Конечно же, при этом первым ориентиром служили финансовые возможности семьи, и все 3 оставленные для дальнейшего обсуждения варианта лежат в приемлемом стоимостном диапазоне. Не вдаваясь в конкретику – производитель машины и ее марка (читатель может здесь дать волю своей фантазии), – будем называть сопоставляемые решения вариантами № 1, 2 и 3 соответственно (уровень 4).

Критериями для сопоставления (уровень 2) служат: технические характеристики, удобство обслуживания, внешний вид, стоимость. При желании более тщательно и всесторонне исследовать проблему вводят также подкритерии (уровень 3). А вообще-то можно обойтись и без этого уровня, естественно, что логическая схема исследования при этом упростится. Подкритериями критерия *технические характеристики* могут служить: расход топлива, проходимость транспортного средства, удобство управления им, безопасность; критерий *внешний вид* можно конкретизировать, включив в него подкритерии: тип кузова, цвет, качество окраски. Критерий *стоимость* вряд ли нуждается в уточнениях.

Пример 4.2. Один из ключевых вопросов, обсуждение которых должно предшествовать решению об освоении новых нефтегазоносных провинций, состоит в том, каков будет масштаб негативных техногенных последствий промышленной деятельности. Рассмотрим, к примеру, месторождения в малоизученном и труднодоступном регионе – на полуострове Ямал и в акватории

морей Северного Ледовитого океана. Построение моделей и научно обоснованное прогнозирование природных явлений, которые могут сопровождать освоение месторождений углеводородов в этом регионе, является новой, очень трудной задачей, требующей времени для создания научного задела, наблюдений и целенаправленных натуральных экспериментов, разработки сценариев, по которым будет развиваться ситуация и т.д. Предлагаемая логическая схема анализа предназначена для выявления и количественной характеристики сложившихся в кругу специалистов представлений по этой проблеме.

Схема экспертного логического анализа содержит 4 уровня иерархии (рис. 4.2): 0 – цель анализа, 1 – перечень экспертов (и их весовые коэффициенты), 2 – варианты негативных проявлений, которыми может сопровождаться разработка месторождений (во избежание недоразумений эти варианты названы сценариями), 3 — варианты решений по освоению ямальских месторождений (принципиальная возможность освоения и необходимая проектная проработка).

Цель исследования (уровень 0): оценить степень опасности техногенных изменений, инициируемых предстоящей разработкой месторождений полуострова Ямал и шельфа Карского моря, и выбрать направление технической политики.

В схему логического анализа были включены следующие негативные процессы (сценарии), проявление которых возможно при освоении месторождений (уровень 2):

Сценарий 1 — масштабное (в пределах газоносных залежей) проседание земной поверхности, приводящее к трудно устранимым последствиям.

Сценарий 2 — масштабное (в пределах газоносных залежей) проседание земной поверхности, приводящее к гидрологическим и прочим изменениям, влияние которых устранимо.

Сценарий 3 — деградация грунтов на ограниченных территориях.

Сценарий 4 — глобальные деграционные процессы, устранение которых невозможно или требует весьма существенных затрат.

Сценарий 5 — подъем уровня мирового океана как следствие парникового эффекта и затопление сооружений на территории промыслов.

Варианты решений, по разработке месторождений рассматриваемого региона (уровень 3):

Вариант 1 — выполненные ранее проекты разработки месторождений полуострова Ямал: Бованенковского, Крузенштерновского и Харасавэйского – следует реализовывать.

Вариант 2 — в выполненные проекты следует внести коррективы с целью минимизации возможных техногенных проявлений.

Вариант 3 — проекты освоения месторождений следует проводить специальными проектами ликвидации негативных техногенных последствий.

Вариант 4 — разработку месторождений полуострова Ямал отложить из-за тяжелых технологических и экологических последствий.



Рисунок 4.2 — Схема экспертного логического анализа для определения технической политики при подготовке к освоению месторождений полуострова Ямал и Карского моря

Этап 2 – оценка значимости введенных критериев

На этом этапе от эксперта требуется провести оценку относительной значимости критериев. Для оценок все эксперты должны использовать одинаковые шкалы. Весьма распространена девятибалльная шкала (табл. 4.1), которая и рекомендуется к использованию.

Проводя оценку, эксперт квантифицирует интуитивную информацию, сложившиеся у него представления о проблеме. Обратимся к примеру 4.1. В таблице 4.2 приведены возможные оценки одного из акторов. Так может оценить значимость крите-

риев жена. Ее не волнует удобство обслуживания, мало волнуют технические характеристики, а на первом плане находится внешний вид – престижность автомобиля.

Итоговая оценка получается на основании суждений всех экспертов.

Таблица 4.1 — Интерпретация оценок для ранжирования критериев

Интервал оценок, баллы	Группа оценок, степень влияния критерия	Оценка, баллы	Смысл оценки
1–3	1) Слабая	1	Незначимость критерия
		2	Очень низкая значимость критерия
		3	Низкая значимость критерия
4–6	2) Существенная	4	Умеренная значимость критерия
		5	Средняя значимость критерия
		6	Существенная значимость критерия
7–9	3) Доминирующая	7	Высокая значимость критерия
		8	Очень высокая значимость критерия
		9	Абсолютное доминирование значимости критерия

Таблица 4.2 — Ранжирование критериев экспертом

№	Критерии	Оценка в баллах
1	Технические характеристики	4
2	Удобство обслуживания	1
3	Внешний вид	8
4	Стоимость	6

Точно так же (в примере 4.1) проводится оценка значимости подкритериев каждого критерия.

Примечание: В процедуру оценки можно ввести весовые коэффициенты для экспертов, учитывающие их квалификацию или значимость.

Этап 3 – сопоставление вариантов решения по критериям

На этом этапе эксперт должен провести несколько туров сравнений (по числу критериев) каждой пары вариантов. Начнем опять-таки с примера 4.1. Представим для упрощения, что в логической схеме рис. 4.1 опущен уровень подкритериев. В этом случае варианты решения должны сравниваться по каждому из 4-х критериев: сначала по критерию технические характеристики, затем по удобству обслуживания, потом по внешнему виду и наконец по стоимости.

ЭЛА позволяет учитывать как измеряемые факторы, так и качественные. Три первые критерия – качественные, для оценки вариантов по этим критериям необходимо прибегнуть к какому-нибудь способу квантификации, например, к балльным оценкам в соответствии с таблицей 4.3. Таблица является модификацией таблицы 4.1 и приспособлена для случая, когда эксперт сопоставляет пару вариантов. Четвертый критерий является объективно измеряемым, и для сопоставления вариантов по нему не нужно прибегать к шкалам суждений.

Таблица 4.3 — Девятибалльная шкала для сопоставления двух вариантов решения A и B по одному из критериев

Оценка	Смысл оценки	Пояснения
1	Одинаковая значимость	Два варианта равноценны
3	Слабое преобладание	Некоторое предпочтение отдается варианту A
5	Существенное преобладание	Сильное предпочтение отдается варианту A
7	Очевидное или очень сильное преобладание	Превосходство варианта A с высокой вероятностью
9	Абсолютное доминирование	Несомненное превосходство варианта A
2, 4, 6, 8	Промежуточные значения преобладания	

Примечание. Если эксперт отдает предпочтение варианту B , то он пользуется обратными оценками. Например, если эксперт считает, что вариант B существенно преобладает над A , то его оценка варианта A равна $1/5$.

В таблице 4.4 приведены оценки вариантов одним из экспертов по критерию технические характеристики (пример 4.1). Как видно, эксперт по этому критерию отдает варианту №1 некоторое предпочтение перед вариантом № 2 и безусловное предпочтение перед вариантом № 3. Вторым вариантом, по его мнению, имеет существенное преобладание над третьим.

В таблицах парных сравнений заполняются обычно только клетки над главной диагональю. Предполагается, что в симметричных клетках стоят обратные величины. Каждый эксперт сравнивает варианты по каждому из критериев. Общее количество таблиц парных сравнений типа табл. 4.4 будет равно $n_1 \times n_2$, где n_1 — количество экспертов, а n_2 — количество вариантов.

Таблица 4.4 — Ранжирование вариантов экспертом по критерию технические характеристики

	Вариант № 1	Вариант № 2	Вариант № 3
Вариант № 1	1	3	8
Вариант № 2	1/3	1	5
Вариант № 3	1/8	1/5	1

Этап 4 – подведение результатов

Обработка таблиц производится специальной программой. Математический аппарат обработки изложен в приложении Е. По полученным суждениям экспертов программа вычисляет оценки приоритетов для всех вариантов. Пример приведен в таблице 4.5.

Таблица 4.5 — Результаты обработки экспертных суждений (пример)

Номер варианта	1	2	3	4
Приоритеты	40,6	34,2	15,2	10,0

Итоговая строка, приведенная в таблице 4.5, означает, что вариант № 1 получил как бы 40,6% голосов из 100% возможных, № 2 – 34,2% и т.д. Приоритеты указывают сравнительную ценность вариантов с точки зрения всех использованных критериев и с учетом предпочтений всех экспертов.

4.3 Достоинства экспертного логического анализа.

Некоторые рекомендации по проведению экспертизы

ЭЛА является весьма полезным инструментом для принятия ответственных решений, когда приходится сталкиваться с наличием многих (в том числе противоречащих один другому) крите-

риев и отсутствием объективных измерителей. Обычно цель ЭЛА состоит в выборе одного из вариантов инвестиционных проектов. Критерии должны отражать интересы участников проекта, риски, связанные с его реализацией, факторы неопределенности, в том числе стоимостные показатели, текущую конъюнктуру, тенденции развития политической и экономической ситуации и т.д.

Весьма эффективным применение ЭЛА может оказаться при разработке масштабных проектов и в процессе их реализации. Такие проекты представляют собой сочетание весьма разнородных решений – технологических, строительных, финансовых, экономических, организационных, – которые основываются в большей степени на представлениях экспертов об условиях функционирования объектов в отдаленном будущем, чем на фактической количественной информации.

ЭЛА дает рекомендации по подготовке, проведению и обработке результатов экспертизы, позволяет более глубоко исследовать проблему и получить дополнительную информацию для принятия ответственных решений.

В процессе подготовки ЭЛА требуется проработать и обосновать:

- состав экспертной группы;
- логическую схему анализа; процедуру экспертного опроса.

Экспертную группу следует комплектовать из высококвалифицированных специалистов разных профилей так, чтобы был

охвачен весь спектр факторов, оказывающих существенное влияние на экспертируемую проблему.

Разработка логической схемы является этапом работы, который не может проводиться стандартными методами и должен максимально учитывать специфику проблемы. Целесообразно, чтобы разработкой логической схемы занимались специалисты, профессионально владеющие как проблемой, так и информационными технологиями.

Экспертам должны быть предоставлены возможности глубоко изучить проблему. Рекомендуемый порядок организации экспертного опроса включает следующие этапы:

- первоначальное обсуждение логической схемы анализа;
- проведение предварительной экспертизы и обработка экспертных суждений;
- обсуждение результатов предварительной экспертизы, на основании чего возможна корректировка логической схемы анализа и суждений экспертов;
- повторное проведение экспертизы и обработка экспертных суждений.

Последний этап может быть повторен один или несколько раз. В процессе проведения экспертизы может быть востребована дополнительная информация, возникнуть необходимость привлечения новых специалистов в состав экспертной группы и т.д.

Экспертизу целесообразно повторять по мере появления новой и уточнения имеющейся информации.

ЭЛА имеет следующие достоинства:

- широта охвата проблемы, выражающаяся в сопоставлении альтернативных вариантов на многокритериальной основе, в использовании всей имеющейся информации для принятия решений, включая информацию качественного характера;
- оперативность – собственно экспертный опрос занимает немного времени у специалистов, обработка результатов производится незамедлительно;
- объективность – мнение экспертов при их желании не оглашается, в случае необходимости эксперты, высказавшие маргинальное мнение, имеют возможность подробно его аргументировать в процессе коллективных обсуждений;
- преемственность – при необходимости экспертиза повторяется, что позволяет экспертам скорректировать свои суждения с учетом вновь появившейся информации;
- адаптивность – в процессе проведения экспертизы легко модифицируется как логическая схема анализа, так и состав экспертной группы;
- организационная системность – возможность включения в состав экспертных групп специалистов разных профилей, представителей разных организаций (в том числе, конкурирующих), а также независимых экспертов;
- доступность – для проведения экспертизы требуется минимальная (по применению техники ЭЛА) подготовка экспертов, результаты анализа легко интерпретируются.

Важно, что в процессе анализа эксперт постоянно получает от программы сведения о непротиворечивости введенных в ма-

шину суждений. Это позволяет ему отыскивать и корректировать с помощью программы те его суждения и оценки, которые находятся в логическом противоречии со всей совокупностью введенных им же данных.

Одной из программ, реализующих процедуру ЭЛА, является программа “SHAST”, разработанная Е.Р.Ставровским и И.А.Шабановым. Программа обладает хорошим сервисом и удобна в обращении.

В приложении Е приведены некоторые сведения о математической модели, лежащей в основе экспертного логического анализа. В подробном изложении с математическим аппаратом можно познакомиться по работам Т.Саати [14, 15].

Приложение А — Необходимые сведения из теории вероятности

А.1 Условное математическое ожидание

Рассмотрим пару случайных величин X и Y . Закон распределения пары, когда X и Y являются величинами непрерывного типа^{*)}, определяется совместной плотностью $f_{XY}(x, y)$. Если же X и Y величины дискретного типа, то их распределение задается вероятностями $\mathbf{P}\{X = x, Y = y\}$, где x и y всевозможные значения с ненулевой вероятностью.

Плотность совместного распределения $f_{XY}(x, y)$ позволяет найти плотность условного распределения Y при заданном значении $X = x$.

$$f_{Y|X}(y|x) = f_{XY}(x, y) / f_X(x), \quad (\text{A.1})$$

где

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \quad (\text{A.2})$$

плотность маргинального распределения величины X . В дискретном случае вместо интегралов появятся суммы

$$\mathbf{P}\{Y = y|X = x\} = \mathbf{P}\{X = x, Y = y\} / \sum_{j=1}^{\infty} \mathbf{P}\{X = x, Y = y_j\}. \quad (\text{A.3})$$

^{*)} Строго говоря, следовало бы использовать термин абсолютно непрерывное распределение. Но пособие рассчитано на подготовку инженеров и магистров прикладной направленности, поэтому мы считаем излишним сосредотачиваться на математических тонкостях.

Математическое ожидание этого условного распределения $\mathbf{M}(Y|X = x)$ называется функцией регрессии Y на X и обозначается $m_Y(x)$ или $m_Y(X = x)$

$$\mathbf{M}(Y|X = x) = m_Y(x) = m_Y(X = x). \quad (\text{A.4})$$

Для дисперсии этой условной величины будем использовать обозначение

$$\mathbf{D}(Y|X = x) = \mathbf{M}\left[(Y - m_Y(x))^2 | X = x\right] \quad (\text{A.5})$$

Вычисление введенных функций производится по формулам

$$m_Y(x) = \begin{cases} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \\ \sum_{j=1}^{\infty} y_j P\{Y = y_j | X = x\}, \end{cases} \quad (\text{A.6})$$

$$\mathbf{D}(Y|X = x) = \begin{cases} \int_{-\infty}^{\infty} (y - m_Y(x))^2 f_{Y|X}(y|x) dy \\ \sum_{j=1}^{\infty} (y_j - m_Y(x))^2 \mathbf{P}\{Y = y_j | X = x\}. \end{cases} \quad (\text{A.7})$$

Верхние формулы относятся к величинам непрерывного, а нижние — дискретного типа.

Можно рассматривать случайные величины $m_Y(X)$, $\mathbf{D}(Y|X)$, которые при $X = x$ принимают значения, определяемые формулами (A.4), (A.5).

Пример A.1. Бросаются две игральные кости. Выигрыш 1-го игрока (равный проигрышу 2-го) определяется по следующему правилу:

- если дубль (одинаковое число очков на обеих костях), то число очков, умноженное на 30,
- если не дубль, то -42 , независимо от числа очков.

Рассматриваем следующие величины X — суммарное число очков при одном бросании костей, Y — выигрыш 1-го игрока. Распределение $P\{X=x|Y=y\}$ приведено в таблице А.1. При подсчете вероятностей руководствуются следующими соображениями. Рассмотрим для примера $X=6$. Этому событию благоприятствуют элементарные исходы $1\cup 5$ (на первой кости 1, на второй 5, какую кость считать первой безразлично), $2\cup 4$, $3\cup 3$, $4\cup 2$, $5\cup 1$. Вероятность каждого из элементарных исходов равна $1/36$. При выпадении дубля $3\cup 3$ 1-й игрок выигрывает $(3+3)\times 30 = 180$, в остальных 4-х случаях проигрывает 42. Значит, $\mathbf{P}\{X=6, Y=180\}=1/36$, $\mathbf{P}\{X=6, Y=-42\}=4/36$, остальные вероятности в строке $X=6$ равны нулю.

Таблица А.1 — Совместное распределение $\mathbf{P}\{X=x|Y=y\}$, функции регрессии $m_Y(x)$, $m_X(y)$ и дисперсия $\mathbf{D}\{Y|X=x\}$ в примере А.1

	1	2	3	4	5	6	7	8	9	10
Y		-42	+60	+120	+180	+240	+300	+360	$m_Y(x)$	$\mathbf{D}\{Y X=x\}$
$X=2$		0	1/36	0	0	0	0	0	+60	0
$X=3$		2/36	0	0	0	0	0	0	-42	0
$X=4$		2/36	0	1/36	0	0	0	0	+12	5832
$X=5$		4/36	0	0	0	0	0	0	-42	0
$X=6$		4/36	0	0	1/36	0	0	0	+2,4	7885
$X=7$		6/36	0	0	0	0	0	0	-42	0
$X=8$		4/36	0	0	0	1/36	0	0	+14,4	12724
$X=9$		4/36	0	0	0	0	0	0	-42	0

1	2	3	4	5	6	7	8	9	10
$X=10$	2/36	0	0	0	0	1/36	0	+7,2	25992
$X=11$	2/36	0	0	0	0	0	0	-42	0
$X=12$	0	0	0	0	0	0	1/36	+360	0
$m_X(y)$	7	2	4	6	8	10	12		

Совместное распределение $\mathbf{P}\{X=x, Y=y\}$ позволяет построить маргинальное распределение $\mathbf{P}\{Y=y\}$ величины Y и условные распределения $P\{Y=y|X=x\}$ (таблица А.2) и $P\{X=x|Y=y\}$ (таблица А.4).

Таблица А.2 — Маргинальное распределение $\mathbf{P}\{Y=y\}$ и условные распределения $\mathbf{P}\{Y=y|X=x\}$ в примере А.1

Y	-42	+60	+120	+180	+240	+300	+360
$P\{Y=y\}$	30/36	1/36	1/36	1/36	1/36	1/36	1/36
$P\{Y=y X=2\}$	0	1	0	0	0	0	0
$P\{Y=y X=3\}$	1	0	0	0	0	0	0
$P\{Y=y X=4\}$	2/3	0	1/3	0	0	0	0
$P\{Y=y X=5\}$	1	0	0	0	0	0	0
$P\{Y=y X=6\}$	4/5	0	0	1/5	0	0	0
$P\{Y=y X=7\}$	1	0	0	0	0	0	0
$P\{Y=y X=8\}$	4/5	0	0	0	1/5	0	0
$P\{Y=y X=9\}$	1	0	0	0	0	0	0
$P\{Y=y X=10\}$	2/3	0	0	0	0	1/3	0
$P\{Y=y X=11\}$	1	0	0	0	0	0	0
$P\{Y=y X=12\}$	0	0	0	0	0	0	1

В графе 9 таблицы А.1 представлена функция регрессии $m_Y(x)$. Как видно, эта функция принимает 7 значений. В таблице

А.3 приведены эти значения и их вероятности, то есть распределение случайной величины $m_Y(X)$.

Таблица А.3 — Распределение случайных величин $m_Y(X)$ и $m_X(Y)$ в примере А.1

$m_Y(X)$	значение	-42	+2,4	+12	+14,4	+60	+72	+360
	вероятность	18/36	5/36	3/36	5/36	1/36	3/36	1/36
$m_X(Y)$	значение	2	4	6	7	8	10	12
	вероятность	1/36	1/36	1/36	30/36	1/36	1/36	1/36

Условными распределениями $\mathbf{P}\{X = x|Y = y\}$ и случайной величиной $m_X(Y)$ мы пользоваться не будем, тем не менее, в таблицах А.3 и А.4 они приведены. Читатель может проверить себя на этих примерах.

Таблица А.4 — Маргинальное распределение $\mathbf{P}\{X = x\}$ и условные распределения $\mathbf{P}\{X = x|Y = y\}$ в примере А.1

X	2	3	4	5	6	7	8	9	10	11	12
$P\{X = x\}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$P\{X = x Y = 60\}$	1	0	0	0	0	0	0	0	0	0	0
$P\{X = x Y = 120\}$	0	0	1	0	0	0	0	0	0	0	0
$P\{X = x Y = 180\}$	0	0	0	0	1	0	0	0	0	0	0
$P\{X = x Y = 240\}$	0	0	0	0	0	0	1	0	0	0	0
$P\{X = x Y = 300\}$	0	0	0	0	0	0	0	0	1	0	0
$P\{X = x Y = 360\}$	0	0	0	0	0	0	0	0	0	0	1
$P\{X = x Y = -42\}$	0	$\frac{2}{30}$	$\frac{2}{30}$	$\frac{4}{30}$	$\frac{4}{30}$	$\frac{6}{30}$	$\frac{4}{30}$	$\frac{4}{30}$	$\frac{2}{30}$	$\frac{2}{30}$	0

А.2 Двумерное нормальное распределение

Нормальное распределение двух величин X и Y характеризуется 5-ю параметрами: средними $\mathbf{M}X = m_X$, $\mathbf{M}Y = m_Y$, дисперсиями $\mathbf{D}X = \sigma_X^2$, $\mathbf{D}Y = \sigma_Y^2$ каждой из них и коэффициентом корреляции $\rho = \text{cov}(X, Y) / \sigma_X \sigma_Y$. Плотность распределения имеет вид

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right]\right\}. \quad (\text{A.8})$$

Интегрируя по одной из переменных в соответствии с (A.2), получаем маргинальные плотности

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{(x-m_X)^2}{2\sigma_X^2}\right], \\ f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left[-\frac{(y-m_Y)^2}{2\sigma_Y^2}\right]. \end{aligned} \quad (\text{A.9})$$

Отсюда видно, кстати, что параметры распределения (A.8) m_X , m_Y и σ_X^2 , σ_Y^2 действительно являются средними и дисперсиями величин X, Y . Вычислив $\text{cov}(X, Y) = \mathbf{M}(X-m_X)(Y-m_Y)$, можно убедиться, что $\text{cov}(X, Y) = \sigma_X\sigma_Y\rho$, то есть параметр ρ действительно является коэффициентом корреляции.

Найдем условную плотность $f_{Y|X}(y|x)$. По формуле (A.1) получаем

$$f_{y|x}(y|x) = \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}} \exp \left\{ -\frac{\left[y - m_Y - \frac{\rho\sigma_Y}{\sigma_X}(x - m_X) \right]^2}{2\sigma_Y^2(1-\rho^2)} \right\} \quad (\text{A.10})$$

При этом использовано следующее тождество

$$\begin{aligned} & \frac{1}{2(1-\rho^2)} \left[\left(\frac{x_0}{\sigma_X} \right)^2 - \frac{2\rho x_0 y_0}{\sigma_X \sigma_Y} + \left(\frac{y_0}{\sigma_Y} \right)^2 \right] - \frac{x_0^2}{2\sigma_X^2} = \\ & = \frac{1}{2(1-\rho^2)\sigma_Y^2} \left[y_0^2 - \frac{2\rho\sigma_Y x_0 y_0}{\sigma_X} + \frac{\rho^2 \sigma_Y^2 x_0^2}{\sigma_X^2} \right] = \frac{1}{2(1-\rho^2)\sigma_Y^2} \left(y_0 - \frac{\rho\sigma_Y x_0}{\sigma_X} \right)^2, \end{aligned}$$

где $x_0 = x - m_X$; $y_0 = y - m_Y$.

Из (A.10) следует, что условное распределение $Y|X$ является нормальным со средним (функцией регрессии)

$$m_Y(x) = m_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - m_X) \quad (\text{A.11})$$

и дисперсией

$$\mathbf{D}(Y|X = x) = (1 - \rho^2)\sigma_Y^2. \quad (\text{A.12})$$

A.3 Многомерное нормальное распределение

Для двумерного нормального распределения введем вектор средних \mathbf{m} , матрицу ковариаций $\mathbf{\Sigma}$ и вектор переменных \mathbf{z}

$$\begin{aligned} \mathbf{m} &= \begin{Bmatrix} m_X \\ m_Y \end{Bmatrix}, \quad \mathbf{\Sigma} = \begin{Bmatrix} \sigma_X^2 & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \sigma_Y^2 \end{Bmatrix} = \\ &= \begin{Bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{Bmatrix}, \quad \mathbf{z} = \begin{Bmatrix} x \\ y \end{Bmatrix}. \end{aligned}$$

Тогда совместная плотность (A.8) примет вид

$$f_{XY}(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{m})^T \Sigma^{-1}(\mathbf{z} - \mathbf{m})\right\}. \quad (\text{A.13})$$

Для обратной матрицы удобно использовать обозначения с верхними индексами элементов $\Sigma^{-1} = \|\sigma^{ij}\|$. Перепишем формулы (A.11), (A.12), учитывая,

$$\begin{aligned} \sigma^{11} &= \sigma_Y^2 / (\sigma_X^2 \sigma_Y^2 - \text{cov}^2(X, Y)) = 1/\sigma_X^2 (1 - \rho^2), \\ \sigma^{12} &= \text{cov}(X, Y) / (\sigma_X^2 \sigma_Y^2 - \text{cov}^2(X, Y)) = \rho / \sigma_X \sigma_Y (1 - \rho^2), \\ \sigma^{22} &= 1/\sigma_Y^2 (1 - \rho^2). \end{aligned}$$

Получаем

$$m_Y(x) = m_Y + \frac{\sigma^{12}}{\sigma^{22}}(x - m_X), \quad (\text{A.11}')$$

$$\mathbf{D}(Y|X = x) = 1/\sigma^{22}. \quad (\text{A.12}')$$

В (A.13) и далее использованы стандартные обозначения $|\cdot|$ — определитель матрицы, $\|\cdot\|^{-1}$ — обратная матрица, T — индекс транспонирования.

Формула (A.13) допускает непосредственное обобщение на нормальное распределение совокупности любого конечного числа случайных величин.

Рассмотрим вектор $\tilde{\mathbf{X}} = \|X_1, \dots, X_s, X_{s+1}\|^T$ размерности $s+1$. В качестве первых s компонент можно рассматривать независимые переменные (регрессоры), а в качестве последней компоненты зависимую $X_{s+1} = Y$. Введем векторные обозначения $\tilde{\mathbf{m}} = \|m_1, \dots, m_{s+1}\|^T$ — вектор средних, $\tilde{\Sigma} = \|\sigma_{ij}\|$ — матрица ковариаций

$\sigma_{ij} = \text{cov}(X_i, X_j)$, ($i, j = 1, \dots, s+1$). Совместная плотность распределения запишется в виде

$$f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{X}}) = \frac{1}{\sqrt{(2\pi)^{s+1} |\tilde{\Sigma}|}} \exp\left\{-\frac{1}{2}(\tilde{\mathbf{x}} - \tilde{\mathbf{m}})^T \tilde{\Sigma}^{-1}(\tilde{\mathbf{x}} - \tilde{\mathbf{m}})\right\}. \quad (\text{A.14})$$

Если рассматриваются только регрессоры X_1, \dots, X_s , то в соответствующих обозначениях опустим тильду, то есть будем писать $\mathbf{X} = \|X_1, \dots, X_s\|^T$ и т.д. Рассуждая так же, как в случае двумерного распределения, из (A.14) получим функцию регрессии

$$m_Y(\mathbf{X}) = m_Y - \sum_{i=1}^s (x_i - m_i) \sigma^{i,Y} / \sigma^{YY}. \quad (\text{A.15})$$

Здесь $m_Y = m_{s+1}$, $\sigma^{i,Y} = \sigma^{i,s+1}$, $\sigma^{YY} = \sigma^{s+1,s+1}$ — элементы вектора $\tilde{\mathbf{m}}$ и матрицы $\tilde{\Sigma}^{-1} = \|\sigma^{ij}\|$ соответственно. Таким образом, функция регрессии является линейной функцией переменных x_1, \dots, x_s так же, как и в двумерном случае.

Величина $m_Y(\mathbf{X})$ распределена по нормальному закону со средним (A.15) и дисперсией $1/\sigma^{YY}$. Чтобы лучше понять ситуацию, попробуем обойтись без матричных обозначений, рассмотрим случай $s = 2$. Запишем показатель экспоненты в (A.14), опустив множитель $-(1/2)$

$$\begin{aligned} (\tilde{\mathbf{x}} - \tilde{\mathbf{m}})^T \tilde{\Sigma}^{-1}(\tilde{\mathbf{x}} - \tilde{\mathbf{m}}) &= \sigma^{11} x_{10}^2 + \sigma^{22} x_{20}^2 + \sigma^{YY} y_0^2 + 2\sigma^{12} x_{10} x_{20} \\ &+ 2\sigma^{1Y} x_{10} y_0 + 2\sigma^{2Y} x_{20} y_0 = \left(\sigma^{YY} y_0 + \frac{\sigma^{1Y}}{\sigma^{YY}} x_{10} + \frac{\sigma^{2Y}}{\sigma^{YY}} x_{20} \right)^2 + [\dots]. \end{aligned}$$

Здесь $x_{10} = x_1 - m_1$, $x_{20} = x_2 - m_2$, $y_0 = y - m_Y$. Выражение в квадратной скобке не зависит от y . Отсюда и следует, что

$$Y | (\mathbf{X} = \mathbf{x}) \in N(m_Y(\mathbf{x}), 1/\sigma^{YY}). \quad (\text{A.16})$$

Пример А.2. Рассмотрим матрицу ковариаций трехмерного нормального вектора X_1, X_2, Y

$$\tilde{\Sigma} = \begin{vmatrix} 1 & -0,4 & 1,2 \\ -0,4 & 4 & -1,5 \\ 1,2 & -1,5 & 2,25 \end{vmatrix}.$$

Матрица отвечает дисперсиям

$\mathbf{D}(X_1) = \sigma_1^2 = 1$, $\mathbf{D}(X_2) = \sigma_2^2 = 4$, $\mathbf{D}Y = \sigma_{YY}^2 = 2,25$ и коэффициентам корреляции $\rho_{12} = -0,2$; $\rho_{1Y} = 0,8$; $\rho_{2Y} = -0,5$.

Обратная матрица имеет вид

$$\tilde{\Sigma}^{-1} = \begin{vmatrix} 3,261 & -0,435 & -2,029 \\ -0,435 & 0,391 & 0,493 \\ -2,029 & 0,493 & 1,855 \end{vmatrix}.$$

Таким образом, в центрированных координатах предиктор (А.15) записывается в виде

$$m_{Y_0}(\mathbf{X}_0) = - \left(\frac{-2,029}{1,855} X_{10} + \frac{0,493}{1,855} X_{20} \right) - (-1,094 X_0 + 0,266 X_{20}),$$

а величина σ_{YX}^2 равна

$$\sigma_{YX}^2 = 1/\sigma^{YY} = 1/1,855 = 0,539.$$

Отсюда в соответствии с (2.2.9) получаем корреляционное отношение $\eta_{YX}^2 = 1 - 0,240 = 0,760$.

Таким образом, 76% дисперсии Y обусловлено дисперсией предиктора и 24% дисперсией относительно предиктора, что и является характеристикой ошибки прогноза.

Пример А.3. рассмотрим нормальное распределение, плотность которого с точностью до числового множителя представля-

ется экспоненциальной функцией $\exp\{-\frac{1}{2}Q(\mathbf{x})\}$, где

$Q(\mathbf{x}) = x_1^2 + 2x_2^2 + 4x_3^2 - x_1x_2 - 3x_1 - 2x_2 + 4$. Найдем параметры распределения. Для определения математических ожиданий \mathbf{m} представим $Q(\mathbf{x})$ в виде

$$Q(\mathbf{x}) = (x_1 - m_1)^2 + 2(x_2 - m_2)^2 + 4(x_3 - m_3)^2 - (x_1 - m_1)(x_2 - m_2) + 2x_1m_1 - m_1^2 + 4x_2m_2 - 2m_2^2 + 8x_3m_3 - 4m_3^2 - x_1m_2 - x_2m_1 + m_1m_2 - 3x_1 - 2x_2 + 4.$$

Приравнивая к нулю коэффициенты при x_1, x_2, x_3 , получим линейную систему с неизвестными m_1, m_2, m_3

$$\begin{aligned} 2m_1 - m_2 - 3 &= 0 \\ 4m_2 - m_1 - 2 &= 0 \\ 8m_3 &= 0. \end{aligned}$$

Получаем $m_1 = 2, m_2 = 1, m_3 = 0$, при этом свободный член, как и должно быть, оказывается равным нулю. Получившаяся запись $Q(\mathbf{x})$ позволяет также найти матрицу Σ^{-1}

$$\Sigma^{-1} = \begin{vmatrix} 1 & -0,5 & 0 \\ -0,5 & 2 & 0 \\ 0 & 0 & 4 \end{vmatrix}.$$

Матрицу ковариаций находим как матрицу, обратную к Σ^{-1}

$$\Sigma = \begin{vmatrix} 8/7 & 2/7 & 0 \\ 2/7 & 4/7 & 0 \\ 0 & 0 & 1/4 \end{vmatrix}.$$

Замечание. Поскольку функция должна быть квадратичной формой от компонент вектора $\mathbf{x} - \mathbf{m}$, то $\left. \frac{dQ(\mathbf{x})}{d\mathbf{x}} \right|_{\mathbf{x}=\mathbf{m}} = 0$. Это условие дает систему уравнение для определения \mathbf{m} . Естественно, что

эта система совпадает с той, которая получена ранее методом выделения полных квадратов, но техника составления системы проще.

А.4 Свойства условных математических ожиданий

Условные математические ожидания удовлетворяют (как случайные величины) следующим двум соотношениям

$$\mathbf{M}Y = \mathbf{M}[\mathbf{M}(Y|\mathbf{X})] = \mathbf{M}m_Y(\mathbf{X}), \quad (\text{A.17})$$

$$\mathbf{D}Y = \mathbf{D}m_Y(\mathbf{X}) + \mathbf{M}\mathbf{D}(Y|\mathbf{X}). \quad (\text{A.18})$$

Оба равенства следуют из определения условных математических ожиданий. Докажем их на примере величин непрерывного типа. Равенство (A.17) очевидно

$$\begin{aligned} \mathbf{M}Y &= \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f_{\mathbf{X}Y}(\mathbf{x}, y) d\mathbf{x} \right) y dy = \\ &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \int_{-\infty}^{\infty} y f_{Y|\mathbf{X}}(y|\mathbf{x}) dy = \int_{-\infty}^{\infty} m_Y(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathbf{M}m_Y(\mathbf{X}). \end{aligned}$$

Докажем теперь равенство (A.18)

$$\begin{aligned} \mathbf{D}Y &= \mathbf{M}(Y - \mathbf{M}Y)^2 = \mathbf{M}[(Y - m_Y(\mathbf{X})) + (m_Y(\mathbf{X}) - \mathbf{M}Y)]^2 = \\ &= \mathbf{M}[Y - m_Y(\mathbf{X})]^2 + \mathbf{M}[m_Y(\mathbf{X}) - \mathbf{M}Y]^2 + \\ &\quad + 2\mathbf{M}[(Y - m_Y(\mathbf{X}))(m_Y(\mathbf{X}) - \mathbf{M}Y)]. \end{aligned}$$

Первое из этих слагаемых в соответствии с (A.17) и (A.5) равно

$$\mathbf{M}[Y - m_Y(\mathbf{X})]^2 = \mathbf{M}\{\mathbf{M}[Y - m_Y(\mathbf{X})]^2 | \mathbf{X}\} = \mathbf{M}\mathbf{D}(Y|\mathbf{X}).$$

Второе в соответствии с (A.17)

$$\mathbf{M}[m_Y(\mathbf{X}) - \mathbf{M}Y]^2 = \mathbf{M}[m_Y(\mathbf{X}) - \mathbf{M}m_Y(\mathbf{X})]^2 = \mathbf{D}m_Y(\mathbf{X}).$$

И, наконец,

$$\begin{aligned}
 & \mathbf{M}\left[(Y - m_Y(\mathbf{X}))(m_Y(\mathbf{X}) - \mathbf{M}m_Y(\mathbf{X}))\right] = \\
 & = \mathbf{M}\left\{\mathbf{M}\left[(Y - m_Y(\mathbf{X}))(m_Y(\mathbf{X}) - \mathbf{M}m_Y(\mathbf{X}))\mid\mathbf{X}\right]\right\} = \quad (\text{A.19}) \\
 & = \mathbf{M}\left\{(m_Y(\mathbf{X}) - \mathbf{M}m_Y(\mathbf{X}))\mathbf{M}\left[(Y - m_Y(\mathbf{X}))\mid\mathbf{X}\right]\right\} = 0,
 \end{aligned}$$

так как $\mathbf{M}\left[(Y - m_Y(\mathbf{X}))\mid\mathbf{X}\right] = 0$ по определению.

Равенство (A.18) может быть записано в других обозначениях как

$$\sigma_Y^2 = \sigma_{m_Y(\mathbf{X})}^2 + \sigma_{Y\mathbf{X}}^2. \quad (\text{A.18}')$$

A.5 Свойства оптимального предиктора

Теорема A.1. Оптимальный предиктор совпадает с функцией регрессии Y на \mathbf{X} , то есть

$$\hat{Y}^* = m_Y(\mathbf{X}).$$

Пусть $\hat{Y}(\mathbf{X})$ – некоторый предиктор. Повторяя выкладки, проведенные при доказательстве формулы (A.18), получаем

$$\begin{aligned}
 \Delta(\hat{Y}) &= \mathbf{M}\left(\hat{Y}(\mathbf{X}) - Y\right)^2 = \mathbf{M}\left[\left(\hat{Y}(\mathbf{X}) - m_Y(\mathbf{X})\right) + (m_Y(\mathbf{X}) - Y)\right]^2 = \\
 &= \mathbf{M}\left[\hat{Y}(\mathbf{X}) - m_Y(\mathbf{X})\right]^2 + \mathbf{M}\left[m_Y(\mathbf{X}) - Y\right]^2 \geq \mathbf{M}\left(m_Y(\mathbf{X}) - Y\right)^2.
 \end{aligned}$$

Знак равенства имеет место при $\hat{Y}(\mathbf{X}) = m_Y(\mathbf{X})$, откуда и следует утверждение.

Минимальная ошибка прогноза

$$\Delta^* = \mathbf{M}\left[m_Y(\mathbf{X}) - Y\right]^2 = \mathbf{M}\mathbf{D}(Y|\mathbf{X}) = \sigma_{Y\mathbf{X}}^2.$$

Пример A.4. Проверим выполнение соотношений (A.17), (A.18) в условиях примера A.1 (таблицы A.1 – A.4).

Соотношение (А.17). Все три величины в двойном равенстве (А.17) равны 0. Действительно, из таблицы А.2

$$\begin{aligned} \mathbf{M}Y &= -42 \cdot \frac{30}{36} + 60 \cdot \frac{1}{30} + 120 \cdot \frac{1}{30} + 180 \cdot \frac{1}{36} + 240 \cdot \frac{1}{36} + \\ &+ 300 \cdot \frac{1}{36} + 360 \cdot \frac{1}{36} = 0 \end{aligned}$$

В среднем выражении (А.17) условное математическое ожидание $\mathbf{M}(Y|X)$ надо рассматривать как функцию с.в. X . Из таблицы А.1 имеем

$$\mathbf{M}[\mathbf{M}(Y|X)] = 60 \cdot \frac{1}{36} - 42 \cdot \frac{2}{36} + 12 \cdot \frac{3}{36} + \dots + 360 \cdot \frac{1}{36} = 0.$$

Последнее выражение (А.17) предполагает вычисление математического ожидания с.в. $\mathbf{M}m_Y(X)$. Согласно таблице А.3

$$\begin{aligned} \mathbf{M}m_Y(X) &= -42 \cdot \frac{18}{36} + 2,4 \cdot \frac{5}{36} + 12 \cdot \frac{3}{36} + 14,4 \cdot \frac{5}{36} + 60 \cdot \frac{1}{36} + \\ &+ 72 \cdot \frac{3}{36} + 360 \cdot \frac{1}{36} = 0. \end{aligned}$$

Соотношение (А.18). Согласно таблице А.2

$$\sigma_Y^2 = \mathbf{D}Y = \mathbf{M}Y^2 = (-42)^2 \cdot \frac{30}{36} + 60^2 \cdot \frac{1}{36} + \dots + 360^2 \cdot \frac{1}{36} = 10570.$$

Согласно таблице А.3

$$\begin{aligned} \sigma_{m_Y(X)}^2 &= \mathbf{D}m_Y(X) = \mathbf{M}[m_Y(X)]^2 = (-42)^2 \cdot \frac{18}{36} + \\ &+ (2,4)^2 \cdot \frac{5}{36} + \dots + 360^2 \cdot \frac{1}{36} = 5035,6. \end{aligned}$$

И, наконец, в последней графе таблицы А.1 представлены значения условной дисперсии $\mathbf{D}(Y|X=x)$. Откуда получаем

$$\begin{aligned}\sigma_{YX}^2 = \mathbf{M}[\mathbf{D}(Y|X)] &= 5832 \cdot \frac{3}{36} + 7885,44 \cdot \frac{5}{36} + 12723,84 \cdot \frac{5}{36} + \\ &+ 25992 \cdot \frac{3}{36} = 5514,4.\end{aligned}$$

Таким образом, σ_Y^2 — общая дисперсия (разброс) с.в. Y представляется как сумма двух составляющих: первая — дисперсия (разброс) функции регрессии $\sigma_{m_Y(X)}^2$ и вторая — дисперсия (разброс) σ_{YX}^2 условного распределения $Y|X$. Эти составляющие в примере оказались примерно одинаковыми (48% и 52% соответственно).

Пример А.5. Двумерное нормальное распределение

Соотношение (А.17). Согласно второму равенству из (А.9)

$$\begin{aligned}\mathbf{M}Y = m_Y. \text{ Из (А.11) получаем } \mathbf{M}m_Y(X) &= \mathbf{M}\left[m_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - m_X)\right] = \\ &= m_Y.\end{aligned}$$

Соотношение (А.18). Согласно второму равенству из (А.9)

$\mathbf{D}Y = \sigma_Y^2$. Используя (А.11), имеем

$$\sigma_{m_Y(X)}^2 = \mathbf{D}[m_Y(X)] = \mathbf{D}\left[m_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - m_X)\right] = \frac{\rho^2\sigma_Y^2}{\sigma_X^2}\mathbf{D}X = \rho^2\sigma_Y^2.$$

И, наконец, в силу (А.12)

$$\sigma_{YX}^2 = \mathbf{M}[\mathbf{D}(Y|X)] = \mathbf{M}[(1-\rho^2)\sigma_Y^2] = (1-\rho^2)\sigma_Y^2. \quad (\text{А.20})$$

Таким образом, соотношение (А.18) подтверждается в данном случае.

А.6 Функции регрессии (пример обработки данных)

В таблице А.5 приведены данные (пример А.6), полученные по замерам двух величин в технологическом процессе. Те же данные нанесены 25 точками на рис. А.1. Полученное облако точек называют обычно корреляционным полем. Употребляются также термины поле рассеяния и диаграмма рассеяния.

Расчетом получаем оценки параметров распределения $\hat{m}_X = \bar{x} = 52,6$; $\hat{m}_Y = \bar{y} = 9,424$; $\hat{\sigma}_X = 17,265$; $\hat{\sigma}_Y = 1,631$; $\text{cov}(X, Y) = -22,8$; $\hat{\rho} = \frac{\text{cov}(X, Y)}{\hat{\sigma}_X \hat{\sigma}_Y} = -0,811$; $\rho \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} = -0,0766$; $\rho \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} = -8,587$;
 $\mathbf{D}(Y|X=x) = 0,910$.

Таким образом, оценка функций регрессии $Y|X$ и $X|Y$ имеет вид

$$m_Y(x) = 9,424 - 0,0766 \cdot (x - 52,6)$$
$$m_X(Y) = 52,6 - 8,587 \cdot (y - 9,424)$$

На рис. А.1 представлены обе функции регрессии и линия уровня нормальной двумерной плотности (пунктиром).

Найдем составляющие равенства (А.18') $\sigma_Y^2 = 2,66$, $\sigma_{m_Y(X)}^2 = 1,75$. Дисперсия функции регрессии $\sigma_{m_Y(X)}^2$ составляет 66%, а средняя дисперсия условного распределения σ_{YX}^2 — 34% от суммарной дисперсии Y .

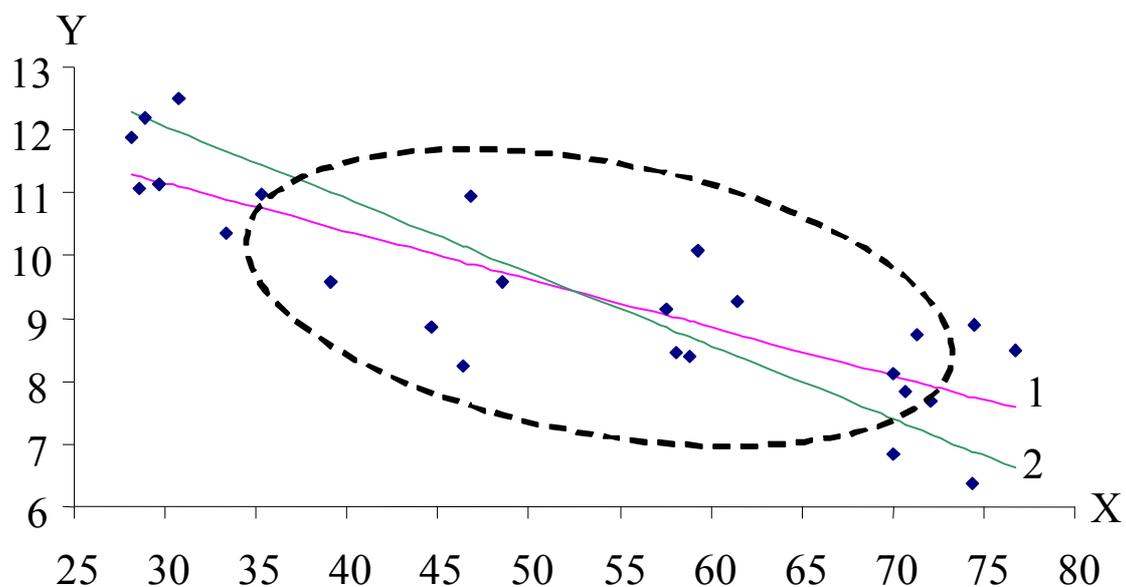


Рисунок А.1 — Корреляционное поле и функции регрессии.
1 — Y на X, 2 — X на Y

Таблица А.5 — Исходные данные для обработки (пример А.6)

i	x_i	y_i	i	x_i	y_i
1	35,3	10,98	14	39,1	9,57
2	29,7	11,13	15	46,8	10,94
3	30,8	12,51	16	48,5	9,58
4	58,8	8,40	17	59,3	10,09
5	61,4	9,27	18	70,0	8,11
6	71,3	8,73	19	70,0	6,83
7	74,4	6,36	20	74,5	8,88
8	76,7	8,50	21	72,1	7,68
9	70,7	7,82	22	58,1	8,47
10	57,5	9,14	23	44,6	8,86
11	46,4	8,24	24	33,4	10,36
12	28,9	12,19	25	28,6	11,08
13	28,1	11,88			

Приложение Б — Некоторые распределения, связанные с нормальным

В настоящем приложении рассматриваются распределения, которым подчиняются выборочные функции (статистики) для выборок из нормального распределения (нормальных выборок). Предположения о нормальном распределении членов выборки часто бывают оправданными благодаря центральной предельной теореме. Можно утверждать, что на практике с нормальными выборками приходится оперировать чаще, чем с выборками из других генеральных совокупностей.

Б.1 Распределение χ^2

Другие названия: распределение хи-квадрат, распределение Пирсона.

Определение Б.1. Распределение с плотностью

$$f_{\chi^2}(x) = \begin{cases} C_m e^{-\frac{x}{2}} x^{\frac{m}{2}-1} & \text{при } x > 0 \\ 0 & \text{при } x \leq 0 \end{cases} \quad (\text{Б.1})$$

называется распределением χ^2 с m степенями свободы.

Константа C_m определяется через Γ -функцию

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx, \quad (\text{Б.2})$$

а именно $C_m^{-1} = 2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)$.

Целочисленный параметр m называется числом степеней свободы.

Среднее и дисперсия распределения χ^2 равны соответственно

$$\mathbf{M}\chi^2 = m, \mathbf{D}\chi^2 = 2m. \quad (\text{Б.3})$$

Форма графика плотности зависит от m . При $m=1$ плотность имеет особенность в точке $x=0$. При $m=2$ распределение χ^2 совпадает с показательным распределением, параметр которого $\lambda=0,5$.

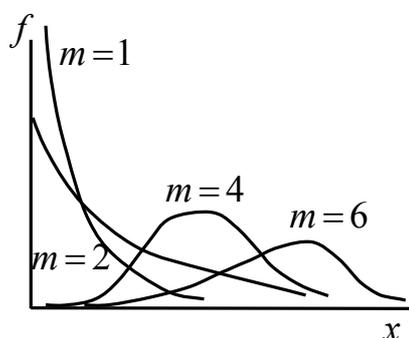


Рисунок Б.1 — Графики плотностей распределения χ^2 при различных числах степеней свободы

При $m > 2$ график плотности представляет собой одногорбую кривую, максимум которой достигается в точке $x = m - 2$. Эскизы графиков плотности представлены на рис. Б.1.

Чаще всего распределение χ^2 появляется в связи со следующим его свойством. Если с.в. Y_1^0, \dots, Y_m^0 независимы и имеют нормальное распределение со стандартными параметрами $Y_i^0 \in N(0,1) \quad i=1, \dots, m$, то сумма их квадратов имеет распределение χ^2 с m степенями свободы

$$(Y_1^0)^2 + \dots + (Y_m^0)^2 \in \chi^2(m). \quad (\text{Б.4})$$

Теорема Б.1 (распределение суммы)

Если с.в. χ_1^2 и χ_2^2 независимы и имеют распределение хи-квадрат с числами степеней свободы m_1 и m_2 , то с.в. $\chi^2 = \chi_1^2 + \chi_2^2$

имеет распределение хи-квадрат с числом степеней свободы равным $m = m_1 + m_2$.

Утверждение легко доказывается с привлечением аппарата характеристических функций случайных величин.

Следствие Б.1. Если $\chi_i^2 \in \chi^2(m_i)$ ($i = 1, 2$) и $\chi_1^2 - \chi_2^2 > 0$, то $\chi^2 = \chi_1^2 - \chi_2^2 \in \chi^2(m)$, где $m = m_1 - m_2$.

Теорема Б.2. Если $Y_i \in N(0, \sigma^2)$, то выборочная функция

$$\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \in \chi^2(n-1). \quad (\text{Б.5})$$

Следует обратить внимание, что число степеней свободы распределения равно не n , а $(n-1)$. Это можно сопроводить следующим нестрогим рассуждением. В левой части равенства стоит сумма квадратов нормальных с.в., которые зависимы, так как связаны соотношением $\sum_{i=1}^n (Y_i - \bar{Y}) = 0$. Эта зависимость обуславливает снижение на 1 числа степеней свободы.

Связь с Γ -распределением. Распределение χ^2 связано с Γ -распределением. Если $F_\Gamma(u; p)$ — функция распределения Γ с параметром p , то

$$F_\Gamma(u; p) = F_{\chi^2}(2u; 2p), \quad (\text{Б.6})$$

где F_{χ^2} — функция распределения χ^2 .

Б.2 Распределение Стьюдента (t -распределение)

Определение Б.2. Распределение с плотностью

$$f_t(x; m) = D_m \left(1 + \frac{x^2}{m} \right)^{-\frac{m+1}{2}} \quad (\text{Б.7})$$

называется t -распределением Стьюдента с m степенями свободы. Константа D_m выражается через Γ -функцию (Б.2)

$$D_m = \Gamma\left(\frac{m+1}{2}\right) / \sqrt{\pi m} \Gamma\left(\frac{m}{2}\right).$$

Распределение Стьюдента имеет следующие среднее и дисперсию $\mathbf{M}t = 0$, $\mathbf{D}t = \frac{m}{m-2}$ ($m > 2$).

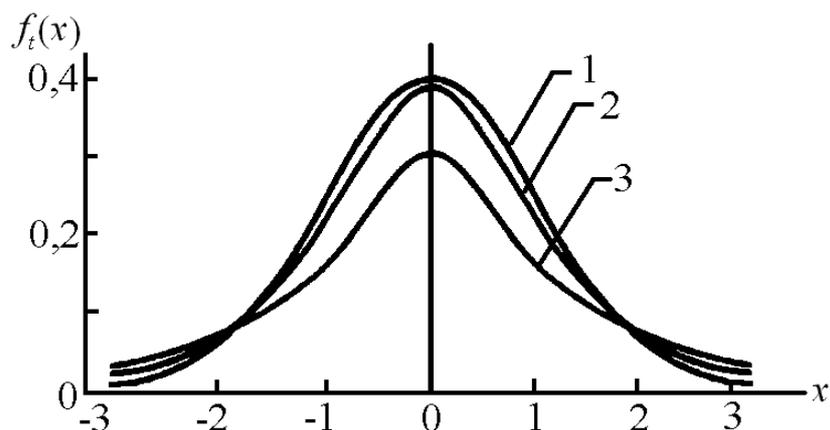


Рисунок Б.2 — Графики плотности: 1 — нормального распределения $N(0,1)$, 2,3 — t -распределения (2 — $m = 5$, 3 — $m = 1$)

График плотности (рис. Б.2) представляется колоколообразной кривой, максимум которой $f_t(0, m)$ снижается с уменьшением m . При достаточно большом числе степеней свободы ($m \geq 30$) график $f_t(x; m)$ близок к плотности стандартного гауссовского распределения.

Теорема Б.3. Пусть ξ и η независимы $\xi \in N(0,1)$, $\eta \in \chi^2(m)$.

Тогда

$$\xi / \sqrt{\frac{\eta}{m}} \in t(m). \quad (\text{Б.8})$$

Следствие Б.2. Если $Y_i \in N(m_Y, \sigma^2)$, то

$$\frac{\bar{Y} - m_Y}{\sigma} \sqrt{n} : \sqrt{\frac{(n-1)s^2}{\sigma^2(n-1)}} = \frac{(\bar{Y} - m_Y)\sqrt{n}}{s} \in t(n-1). \quad (\text{Б.9})$$

Б.3 Распределение Фишера

Определение Б.3. Распределение с плотностью

$$f_F(x; m_1, m_2) = \begin{cases} E_{m_1 m_2} x^{\frac{m_1}{2}-1} (m_2 + m_1 x)^{-\frac{m_1 m_2}{2}} & \text{при } x > 0 \\ 0 & \text{при } x \leq 0 \end{cases} \quad (\text{Б.10})$$

называется распределением Фишера с m_1 и m_2 степенями свободы.

Константа $E_{m_1 m_2}$ вычисляется через Γ -функцию (Б.2)

$$E_{m_1 m_2} = \Gamma\left(\frac{m_1 + m_2}{2}\right) m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}} / \Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right).$$

Моменты распределения – среднее и дисперсия – определяются

формулами $\mathbf{MF} = \frac{m_2}{m_2 - 2}$, при $m_2 > 2$, $\mathbf{DF} =$

$$= \frac{2m_2^2(m_1 + m_2 - 2)}{m_1(m_2 - 2)^2(m_2 - 4)}, \text{ при } m_2 > 4.$$

Теорема Б.4. Пусть с.в. ξ_1, ξ_2 независимы, $\xi_1 \in \chi_1(m_1)$,

$\xi_2 \in \chi_2(m_2)$. Тогда $\frac{\xi_1}{m_1} : \frac{\xi_2}{m_2} \in F(m_1, m_2)$.

Следствие Б.3. Пусть s_1^2 и s_2^2 — выборочные дисперсии, вычисленные по двум выборкам из n_1 и n_2 членов. Тогда

$$\frac{s_1^2}{s_2^2} \in F(n_1 - 1, n_2 - 1).$$

На рис. Б.3 представлен эскиз графика плотности распределения Фишера.

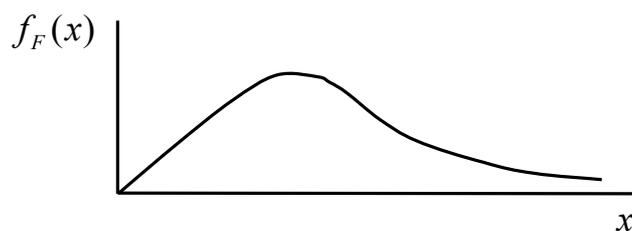


Рисунок Б.3 — Типичный вид плотности F -распределения

Приложение В — Геометрический подход к регрессионному анализу

В.1 Геометрическая интерпретация метода наименьших квадратов

Основную задачу регрессионного анализа, пользуясь геометрическими представлениями, можно сформулировать следующим образом. Вектор наблюдений \mathbf{Y} в соответствии с формулой (2.3.3) надо разложить на 2 составляющие $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ так, чтобы норма вектора $\|\boldsymbol{\varepsilon}\| = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|$ была минимальной. В качестве нормы выбирается длина вектора в n -мерном евклидовом пространстве. Вектор $\mathbf{X}\boldsymbol{\beta}$ принадлежит линейному подпространству Ω — множеству значений (образу) линейного оператора \mathbf{X} . Меняя вектор $\boldsymbol{\beta}$, получим минимум $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|$ при таком $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, когда

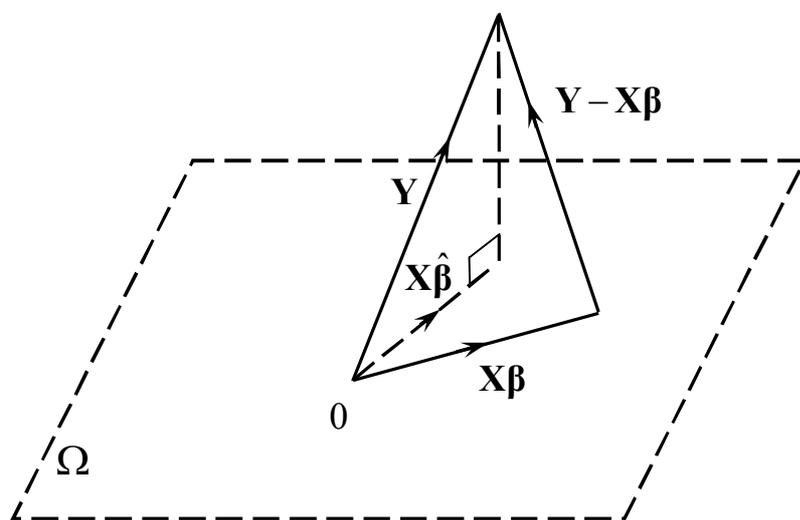


Рисунок В.1 — Разложение вектора \mathbf{Y} по методу наименьших квадратов

$\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ принадлежит ортогональному дополнению Ω (рис. В.1).

Условием ортогональности вектора $\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ подпространству Ω является ортогональность

каждому вектору, образующему Ω , то есть каждому столбцу матрицы \mathbf{X}

$$\mathbf{X}^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}. \quad (\text{B.1})$$

Таким образом, для определения $\hat{\boldsymbol{\beta}}$ получена система (2.5.1), выведенная в п. 2.5 другим способом.

В.2 Операторы проектирования и их свойства.

Пользуясь равенством (2.5.2), представим проекцию $\mathbf{X}\hat{\boldsymbol{\beta}}$ в виде

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \mathbf{P}\mathbf{Y}. \quad (\text{B.2})$$

Отсюда видно, что $\mathbf{P} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ есть оператор проектирования на подпространство Ω , а $\mathbf{I}_n - \mathbf{P}$ — оператор проектирования на ортогональное подпространство. (Здесь операторы отождествляются с их матричными представлениями.)

Из линейной алгебры известно, что

- представление $\mathbf{Y} = \mathbf{P}\mathbf{Y} + (\mathbf{I}_n - \mathbf{P})\mathbf{Y}$ единственно,
- матрицы \mathbf{P} и $\mathbf{I}_n - \mathbf{P}$ идемпотентны (идемпотентной называется матрица, удовлетворяющая соотношению $\mathbf{P}^2 = \mathbf{P}$),
- матрицы \mathbf{P} и $\mathbf{I}_n - \mathbf{P}$ симметричны,
- $\text{rank}(\mathbf{P}) = \text{rank}(\mathbf{X}) = s + 1$,
- $\text{rank}(\mathbf{I}_n - \mathbf{P}) = n - \text{rank}(\mathbf{P})$, (B.3)
- $(\mathbf{I}_n - \mathbf{P})\mathbf{X} = \mathbf{0}$. (B.4)

Все эти соотношения доказываются без особого труда. Докажем, например, что $\mathbf{P}^2 = \mathbf{P}$

$$\mathbf{P}^2 = \left(\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \mathbf{X} \mathbf{I}_{s+1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \mathbf{P}.$$

В основе доказательств соотношения (В.3) и соотношения, ему предшествующего, лежат следующие рассуждения.

Собственные числа идемпотентной матрицы равны либо 1, либо 0. Действительно, пусть λ собственное число матрицы \mathbf{P} . Поскольку $\mathbf{P}^2 = \mathbf{P}$, то из соотношения $\mathbf{P}\mathbf{X} = \lambda\mathbf{X}$ следует $\lambda\mathbf{X}^T\mathbf{X} = \mathbf{X}^T\mathbf{P}\mathbf{X} = \mathbf{X}^T\mathbf{P}^2\mathbf{X} = (\mathbf{P}\mathbf{X})^T(\mathbf{P}\mathbf{X}) = \lambda^2\mathbf{X}^T\mathbf{X}$, то есть $\lambda(\lambda - 1) = 0$.

Далее, известно, что симметричная матрица невырожденным ортогональным преобразованием может быть приведена к сумме квадратов. Значит, для матрицы $\mathbf{I}_n - \mathbf{P}$ найдется ортогональная матрица \mathbf{A} такая, что $\mathbf{A}^T(\mathbf{I}_n - \mathbf{P})\mathbf{A} = \mathbf{\Lambda}$, где $\mathbf{\Lambda}$ — диагональная матрица, составленная из собственных чисел матрицы $\mathbf{I}_n - \mathbf{P}$. В силу невырожденности матрицы \mathbf{A} $\text{rank}[\mathbf{A}^T(\mathbf{I}_n - \mathbf{P})\mathbf{A}] = \text{rank}(\mathbf{\Lambda}) = \text{tr}(\mathbf{\Lambda})$, так как $\text{tr}\mathbf{\Lambda}$ равен количеству собственных значений равных 1.

Приведенные выше свойства матриц \mathbf{P} и $\mathbf{I}_n - \mathbf{P}$ позволяют получить важные выводы.

В.3 Теорема о распределении квадратичной формы

Теорема. Пусть выборка из n членов взята из нормального распределения $\mathbf{Y} \in N(\mathbf{m}, \sigma^2\mathbf{I}_n)$. Тогда, если \mathbf{P} — идемпотентная матрица $\mathbf{P}^2 = \mathbf{P}$ ранга $s+1$, то квадратичная форма $Q = (\mathbf{Y} - \mathbf{m})^T \mathbf{P}(\mathbf{Y} - \mathbf{m})/\sigma^2$ имеет распределение $\chi^2(s+1)$.

Доказательство. Обозначим через \mathbf{B} ортогональную матрицу, приводящую \mathbf{P} к диагональному виду $\mathbf{B}^T \mathbf{P} \mathbf{B} = \mathbf{\Lambda}'$. При этом, естественно, $\mathbf{P} = \mathbf{B} \mathbf{\Lambda}' \mathbf{B}^T$. Первые $s+1$ диагональных элементов матрицы $\mathbf{\Lambda}'$ равны 1, остальные элементы равны нулю. Подвергнем вектор $\mathbf{Y} - \mathbf{m}$ ортогональному преобразованию с матрицей \mathbf{B}^T . Полученная система случайных величин (вектор) $\mathbf{Z} = \mathbf{B}^T (\mathbf{Y} - \mathbf{m})$ независима и распределена по нормальному закону

$$\mathbf{Z} \in N(\mathbf{0}, \sigma^2 \mathbf{I}_n). \quad (\text{B.5})$$

Нормальность распределения вектора \mathbf{Z} следует из того, что \mathbf{Z} линейно зависит от \mathbf{Y} . Матрица ковариаций вектора \mathbf{Z} находится непосредственно

$$\mathbf{DZ} = \mathbf{B}^T \mathbf{D}(\mathbf{Y} - \mathbf{m}) \mathbf{B} = \sigma^2 \mathbf{B}^T \mathbf{B} = \sigma^2 \mathbf{I}_n.$$

Квадратичная форма Q представляется в виде

$$\begin{aligned} \sigma^2 Q &= (\mathbf{Y} - \mathbf{m})^T \mathbf{P} (\mathbf{Y} - \mathbf{m}) = \\ &= (\mathbf{Y} - \mathbf{m})^T \mathbf{B} \mathbf{\Lambda}' \mathbf{B}^T (\mathbf{Y} - \mathbf{m}) = \mathbf{Z}^T \mathbf{\Lambda}' \mathbf{Z} = Z_1^2 + \dots + Z_{s+1}^2. \end{aligned} \quad (\text{B.6})$$

Из этого в соответствии с (Б.4) следует, что Q распределена по χ^2 с $s+1$ степенями свободы $Q \in \chi^2(s+1)$. Теорема доказана.

В.4 Остаточная сумма квадратов

Остаточная сумма квадратов $\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|$ представляется в виде

$$\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P})^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y} = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y}. \quad (\text{B.7})$$

Точно так же, как и при доказательстве теоремы В.3, получаем

$$\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}) \mathbf{Y} = \mathbf{Y}^T \mathbf{A}^T \boldsymbol{\Lambda} \mathbf{A} \mathbf{Y} = Z_1^2 + \cdots + Z_{n-s-1}^2. \quad (\text{B.8})$$

Причем, $\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}$ имеет с точностью множителя распределение χ^2 с $n - s - 1$ степенями свободы: $\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} / \sigma^2 \in \chi^2(n - s - 1)$.

Приложение Г — Теоремы о распределениях выборочных сумм

Г.1 Теорема о распределении выборочной функции

$$\Sigma(Y_i - \bar{Y})^2$$

Теорема. Пусть Y_1, \dots, Y_n — независимые случайные величины из нормального распределения $Y_i \in N(m, \sigma^2)$. Тогда $(n-1)s^2/\sigma^2 = \Sigma(Y_i - \bar{Y})^2/\sigma^2$ не зависит от \bar{Y} и имеет распределение $\chi^2(n-1)$.

Доказательство. Центрируем и нормируем с.в. Y_i . Получим стандартные нормальные величины $Y_i^0 = \frac{Y_i - m}{\sigma}$, $Y_i^0 \in N(0,1)$. Поло-

жим

$$\begin{aligned} Z_1 &= \frac{Y_1^0 + \dots + Y_n^0}{\sqrt{n}}, Z_2 = \frac{Y_1^0 - Y_2^0}{\sqrt{2}}, Z_3 = \frac{Y_1^0 + Y_2^0 - 2Y_3^0}{\sqrt{6}}, \dots, \\ Z_n &= \frac{Y_1^0 + Y_2^0 + \dots + Y_{n-1}^0 - (n-1)Y_n^0}{\sqrt{n(n-1)}}. \end{aligned} \tag{Г.1}$$

В матричном виде преобразование (Г.1) можно записать как

$$\mathbf{Z} = \mathbf{A}\mathbf{Y}^0$$

$$\mathbf{A} = \left\| \begin{array}{cccccc} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \dots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \dots & -\frac{n-1}{\sqrt{n(n-1)}} \end{array} \right\|.$$

Коэффициенты в формулах (Г.1) подобраны так, чтобы матрица \mathbf{A} была ортогональной. Это легко проверяется. Строки матрицы \mathbf{A} , векторы \mathbf{a}_i^T — ортогональны, $\mathbf{a}_i^T \mathbf{a}_j = 0$ (при $i \neq j$) и по модулю равны 1 $\|\mathbf{a}_i^T\| = \sum_{j=1}^n a_{ij}^2 = 1$.

Из ортогональности \mathbf{A} следует, что с.в. Z_i независимы. Так как преобразование (Г.1) линейно, то Z_i распределены по нормальному закону.

Далее имеем $\sum_{i=1}^n Z_i^2 = \mathbf{Z}^T \mathbf{Z} = \mathbf{Y}^{0T} \mathbf{A}^T \mathbf{A} \mathbf{Y}^0 = \mathbf{Y}^{0T} \mathbf{Y}^0 = \sum_{i=1}^n (Y_i^0)^2 = n(\bar{Y}^0)^2 + \sum_{i=1}^n (Y_i^0 - \bar{Y}^0)^2 = Z_1^2 + \sum_{i=1}^n (Y_i^0 - \bar{Y}^0)^2$. Z_1 не зависит от Z_2, Z_3, \dots, Z_n , значит, не зависит от $\sum_{i=1}^n (Y_i^0 - \bar{Y}^0)^2$. Так как $Z_1^2 \in \chi^2(1)$, то в силу следствия Б.1 $\sum_{i=1}^n (Y_i^0 - \bar{Y}^0)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2$ и, значит, $(n-1)s^2/\sigma^2$ имеет распределение χ^2 с $n-1$ степенями свободы $(n-1)s^2/\sigma^2 \in \chi^2(n-1)$.

Г.2 Теорема о распределении суммы квадратов величин из $N(\mathbf{0}, \mathbf{I}_n)$

Пусть $\mathbf{Y}^0 \in N(\mathbf{0}, \mathbf{I}_n)$ и $\mathbf{Y}^{0T} \mathbf{Y}^0 = \mathbf{Y}^{0T} \mathbf{P} \mathbf{Y}^0 + \mathbf{Y}^{0T} \mathbf{B} \mathbf{Y}^0$, где \mathbf{P} — идемпотентная матрица ранга r . Тогда матрица $\mathbf{B} = \mathbf{I}_n - \mathbf{P}$ также идемпотентна и имеет ранг $n - r$.

Матрица \mathbf{B} симметрична, так как симметрична \mathbf{P} . Идем-
тентность \mathbf{B} вытекает из следующего:

$\mathbf{B}^2 = (\mathbf{I}_n - \mathbf{P})^2 = \mathbf{I}_n - 2\mathbf{P} + \mathbf{P}^2 = \mathbf{I}_n - \mathbf{P} = \mathbf{B}$. Так как $\mathbf{Y}^{0T}\mathbf{Y}^0 \in \chi^2(n)$,

$\mathbf{Y}^{0T}\mathbf{P}\mathbf{Y}^0 \in \chi^2(r)$, то в силу предыдущей теоремы

$$\mathbf{Y}^{0T}(\mathbf{I}_n - \mathbf{P})\mathbf{Y}^0 \in \chi^2(n - r).$$

Приложение Д — Изменение количества регрессоров

Представим себе, что нашей целью является сопоставление двух моделей. Одна из них включает все регрессоры другой и плюс к тому еще некоторые регрессоры. Формализуем эту ситуацию, сохранив для модели с большим числом параметров прежние обозначения

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (\text{Д.1})$$

Здесь векторы \mathbf{Y} и $\boldsymbol{\varepsilon}$ имеют размер $n \times 1$, матрица \mathbf{X} размер $n \times p$, вектор $\boldsymbol{\beta}$ размер $p \times 1$.

Пусть модель (Д.1) сопоставляется с моделью, в которой из p регрессоров модели (Д.1) оставлены $p^{(1)}$. Матрицу \mathbf{X} можно записать тогда в блочном виде, поместив сначала столбцы, отвечающие этим регрессорам $\mathbf{X} = \|\mathbf{X}^{(1)} : \mathbf{X}^{(2)}\|$. Матрица $\mathbf{X}^{(2)}$ составлена из оставшихся $p^{(2)} = p - p^{(1)}$ столбцов матрицы \mathbf{X} . Компоненты вектора $\boldsymbol{\beta}$ разобьем соответствующим образом $\boldsymbol{\beta}^T = \|\boldsymbol{\beta}_0^{(1)T}, \boldsymbol{\beta}_0^{(2)T}\|$. С учетом новых обозначений модель (Д.1) примет вид

$$\mathbf{Y} = \|\mathbf{X}^{(1)}, \mathbf{X}^{(2)}\| \cdot \begin{vmatrix} \|\boldsymbol{\beta}_0^{(1)}\| \\ \|\boldsymbol{\beta}_0^{(2)}\| \end{vmatrix} + \boldsymbol{\varepsilon} = \mathbf{X}^{(1)}\boldsymbol{\beta}_0^{(1)} + \mathbf{X}^{(2)}\boldsymbol{\beta}_0^{(2)} + \boldsymbol{\varepsilon}. \quad (\text{Д.2})$$

Будем, как и раньше, считать, что число регрессоров (факторов) p выбрано с учетом объема имеющейся информации: $p < n$ — и регрессоры подобраны удачно, так что ранги матриц $\mathbf{X}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ равны соответственно $p, p^{(1)}, p^{(2)}$.

Модель с $p^{(1)}$ регрессорами, вошедшими в матрицу $X^{(1)}$, имеет вид

$$Y = X^{(1)}\beta^{(1)} + \varepsilon^{(1)}. \quad (\text{Д.3})$$

Очевидно, что оценки первых $p^{(1)}$ компонент вектора β в общей модели (Д.2) будут отличаться от оценок вектора $\beta^{(1)}$ в модели (Д.3). Поэтому в общей модели составляющие вектора β снабжены нижним индексом 0. Наряду с моделью (Д.3) (модель 1) можно было бы рассматривать еще одну регрессионную модель (модель 2) с заменой в соотношении (Д.3) индекса (1) на индекс (2). Дальнейшие рассуждения не изменились бы с заменой модели 1 на модель 2. Свободный коэффициент β_0 и, соответственно, состоящий из единиц первый столбец матрицы X обычно приходится вводить в модель 1, хотя это не обязательно.

Как показано в приложении В, применение метода наименьших квадратов в общей модели (Д.1) приводит к тому, что вектор Y представляется в виде суммы

$$Y = U + V = P_{\Omega}Y + (I_n - P_{\Omega})Y, \quad (\text{Д.4})$$

где первая компонента $U = P_{\Omega}Y$ является проекцией Y на линейное пространство Ω , натянутое на векторы $X\beta$. Вторая компонента $V = (I_n - P_{\Omega})Y$ принадлежит ортогональному дополнению Ω^{\perp} пространства Ω

$$V = (I_n - P_{\Omega})Y \in \Omega^{\perp}. \quad (\text{Д.5})$$

Размерности пространств Ω и Ω^{\perp} равны соответственно p и $n - p$. Матрица P_{Ω} оператора P_{Ω} выражается через X

$$\mathbf{P}_\Omega = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T. \quad (\text{Д.6})$$

Матрицу \mathbf{P}_Ω можно представить в виде

$$\mathbf{P}_\Omega = \mathbf{A}\mathbf{A}^T, \quad (\text{Д.7})$$

где \mathbf{A} — ортогональная матрица, столбцы которой составляют ортонормированный базис пространства Ω . Ранг \mathbf{P}_Ω равен p .

Оценка $\mathbf{X}\hat{\boldsymbol{\beta}}$ вектора $\mathbf{X}\boldsymbol{\beta}$ представляется как

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}_\Omega \mathbf{Y} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (\text{Д.8})$$

Из линейной алгебры известно, что разложение (Д.4) единственно.

Для модели (Д.3) имеют место аналогичные результаты с очевидной заменой обозначений.

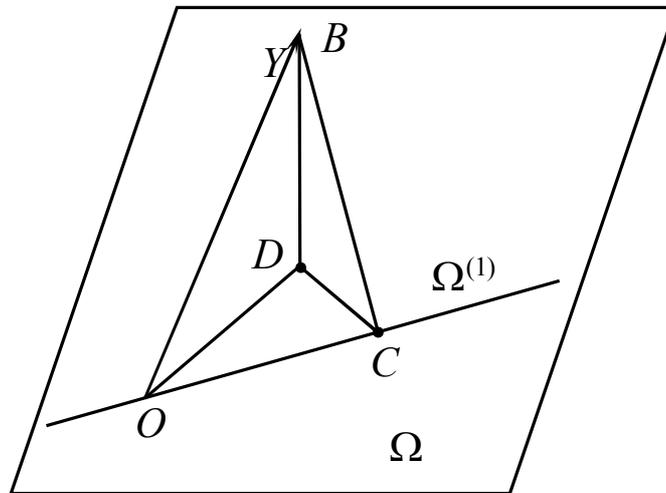


Рисунок Д.1— Геометрическая иллюстрация: применение метода наименьших квадратов в моделях (Д.3), (Д.1)

Оптимальным предиктором в модели (Д.3) будет проекция \mathbf{Y} на пространство $\Omega^{(1)}$, натянутое на векторы $\mathbf{X}^{(1)}\boldsymbol{\beta}^{(1)}$. Рис. Д.1 иллюстрирует ситуацию. Пространство Ω представляется плос-

костью ODC , пространство Ω^\perp — прямой BD , вектор $\mathbf{P}_\Omega \mathbf{Y}$ лежит на прямой OD .

Пространство $\Omega^{(1)}$ поневоле (трехмерное пространство не позволяет изобразить на чертеже $\Omega^{(1)}$ более, чем одномерным) изображено прямой OC . Разложение вектора \mathbf{Y} , аналогичное (Д.4), в модели 1 (формула (Д.3)) имеет вид

$$\mathbf{Y} = \mathbf{P}_{\Omega^{(1)}} \mathbf{Y} + (\mathbf{I}_n - \mathbf{P}_{\Omega^{(1)}}) \mathbf{Y}, \quad (\text{Д.9})$$

Проекция \mathbf{Y} на $\Omega^{(1)}$ изображается на рис. Д.1 вектором \vec{OC} , ортогональное дополнение $\Omega^{(1)\perp}$ пространства $\Omega^{(1)}$ — прямой BC .

Матрицы \mathbf{P}_Ω и $\mathbf{P}_{\Omega^{(1)}}$ связаны равенствами

$$\mathbf{P}_\Omega \mathbf{P}_{\Omega^{(1)}} = \mathbf{P}_{\Omega^{(1)}} \mathbf{P}_\Omega = \mathbf{P}_{\Omega^{(1)}}. \quad (\text{Д.10})$$

Соотношение (Д.10) следующим образом иллюстрируется рисунком Д.1. Проекцию \mathbf{Y} на $\Omega^{(1)}$ $\mathbf{P}_{\Omega^{(1)}} \mathbf{Y}$ (вектор \vec{OC}) можно получить, проектируя \mathbf{Y} на Ω (вектор \vec{OD}), и затем этот вектор на $\Omega^{(1)}$. Соотношение (Д.10) доказывается следующим рассуждением. Применяя оператор \mathbf{P}_Ω к вектору $\mathbf{P}_{\Omega^{(1)}} \mathbf{Y}$, лежащему в плоскости Ω , мы, очевидно, получаем тот же вектор $\mathbf{P}_{\Omega^{(1)}} \mathbf{Y}$.

Из тождества

$$\mathbf{P}_\Omega \mathbf{Y} = \mathbf{P}_{\Omega^{(1)}} \mathbf{Y} + (\mathbf{P}_\Omega - \mathbf{P}_{\Omega^{(1)}}) \mathbf{Y} \quad (\text{Д.11})$$

вытекает, что $(\mathbf{P}_\Omega - \mathbf{P}_{\Omega^{(1)}}) \mathbf{Y} \in \Omega$. В силу (Д.10) векторы $\mathbf{P}_{\Omega^{(1)}} \mathbf{Y}$ и $(\mathbf{P}_\Omega - \mathbf{P}_{\Omega^{(1)}}) \mathbf{Y}$ ортогональны. Значит, (Д.11) является ортогональ-

ным разложением вектора $\mathbf{P}_\Omega \mathbf{Y}$ по подпространствам $\Omega^{(1)}$ и $\Omega^{(1)\perp}$.

На рис. Д.1 это соответствует разложению \vec{OD} в сумму векторов \vec{OC} и \vec{CD} .

Остаточная сумма квадратов в модели (Д.1) вычисляется по формуле

$$R_p^2 = \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}_\Omega) \mathbf{Y}. \quad (\text{Д.12})$$

С точностью до множителя σ^2 эта величина распределена по χ^2 с $n - p$ степенями свободы

$$R_p^2 / \sigma^2 \in \chi^2(n - p). \quad (\text{Д.13})$$

Аналогичные результаты имеют место для модели (Д.3). Остаточная сумма квадратов

$$R_{p^{(1)}}^2 = \hat{\boldsymbol{\varepsilon}}^{(1)T} \hat{\boldsymbol{\varepsilon}}^{(1)} = \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}_{\Omega^{(1)}}) \mathbf{Y}; \quad R_{p^{(1)}}^2 / \sigma^2 \in \chi^2(n - p^{(1)}). \quad (\text{Д.14})$$

В силу следствия Б1 разность $R_{p^{(1)}}^2 - R_p^2$ имеет распределение χ^2

$$\left(R_{p^{(1)}}^2 - R_p^2 \right) / \sigma^2 \in \chi^2(p - p^{(1)}). \quad (\text{Д.15})$$

Величина $R_{p^{(1)}}^2 - R_p^2$ отвечает остаточной сумме квадратов при операторе проектирования $P_\Omega - P_{\Omega^{(1)}}$. Это следует из соотношений

$$\begin{aligned} R_{p^{(1)}}^2 - R_p^2 &= \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}_{\Omega^{(1)}}) \mathbf{Y} - \mathbf{Y}^T (\mathbf{I}_n - \mathbf{P}_\Omega) \mathbf{Y} = \\ &= \mathbf{Y}^T (\mathbf{P}_\Omega - \mathbf{P}_{\Omega^{(1)}}) \mathbf{Y}. \end{aligned} \quad (\text{Д.16})$$

То же самое равенство представляется в виде

$$\|\mathbf{Y} - \mathbf{X}^{(1)}\hat{\boldsymbol{\beta}}^{(1)}\|^2 - \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}^{(1)}\hat{\boldsymbol{\beta}}^{(1)}\|^2. \quad (\text{Д.17})$$

Соотношение (Д.17) иллюстрируется рисунком Д.1 как равенство $\|BC\|^2 - \|BD\|^2 = \|DC\|^2$.

Приложение Е — Краткие сведения о математической модели, лежащей в основе экспертного логического анализа

Е.1 Свойства матрицы суждений и следствия из них

Математическая основа ЭЛА подробно изложена в работах Т.Саати [14, 15]. Ниже представлены только краткие сведения о математическом аппарате. Рассмотрим элементы $C_{1l}, C_{2l}, \dots, C_{nl}$ некоторого уровня l . Это может быть уровень критериев, факторов, действующих лиц (акторов), решений, альтернатив и т.п. В рассматриваемой в ЭЛА иерархии, вообще говоря, все элементы l -го уровня оказывают влияние на любой элемент уровня $l-1$ и через них на элементы высших уровней иерархии. Мету этого влияния показывают приоритеты элементов, которые оцениваются по суждениям экспертов. Эксперты сопоставляют попарно элементы одного уровня и вводят свои оценки в баллах, заполняя квадратные таблицы (матрицы суждений).

Поставим себе целью определить веса w_1, w_2, \dots, w_n влияния элементов $C_{1l}, C_{2l}, \dots, C_{nl}$ на некоторый элемент предыдущего уровня. Анализ существенно опирается на тот факт, что матрицы суждений \mathbf{A} являются обратно-симметрическими $\mathbf{A} = \|\|a_{ij}\|\|$, $a_{ij} = 1/a_{ji}$ при $a_{ij} > 0$.

Введем следующее определение: суждения, помещенные в матрице \mathbf{A} , назовем совершенными, а саму матрицу согласованной, если для всех i, j, k выполняется соотношение $a_{ik} = a_{ij}a_{jk}$.

Смысл этого определения состоит в том, что для любых троек совершенных суждений действует закон логического следствия (транзитивности): если из E следует F , а из F следует G , то из E должно следовать G .

Если элементы a_{ij} могут быть представлены в виде $a_{ij} = w_i / w_j > 0$ для всех i, j , то все суждения являются совершенными и матрица \mathbf{A} согласованной. При этом условии, то есть при существовании n -мерного вектора весов \mathbf{w} .

$$\mathbf{A}\mathbf{w} = \left\| \sum_j a_{ij} \mathbf{w}_j \right\| = n\mathbf{w}. \quad (\text{E.1})$$

Соотношение (E.1) показывает, что n является собственным числом, а \mathbf{w} – отвечающим ему собственным вектором матрицы \mathbf{A} , который, как известно, определяется с точностью до постоянного множителя. Это собственное число является наибольшим по модулю $\lambda_{\max} = n$ и называется числом Фробениуса.

Центральная идея метода обработки суждений состоит в том, чтобы оценивать приоритеты (меры значимости, веса) элементов с помощью нормализованного собственного вектора матрицы суждений ("главного правого собственного вектора"), отвечающего главному собственному значению λ_{\max} .

Означенную идею (высказанную К.Бержем) иллюстрирует следующий пример. Пусть в урне имеется N шаров n разных цветов, ровно m_i шаров i -го цвета ($i=1,2,\dots,n$), $\sum_i m_i = N$. Ясно, что вероятность вынуть наудачу шар i -го цвета равна $w_i = m_i / N$. Составим матрицу \mathbf{A} попарных сравнений числа шаров в урне, по-

ложив $a_{ij} = w_i/w_j = m_i/m_j$. Матрица \mathbf{A} оказывается обратно-симметрической и согласованной, а ее нормализованный собственный вектор, отвечающий наибольшему собственному значению ($n = \lambda_{\max}$), оказывается в точности равным вектору $\mathbf{w} = \|w_1, \dots, w_n\|^T$ — распределению вероятностей в этой урновой схеме.

По аналогии, компоненты главного собственного вектора обратно-симметрической матрицы суждений (не обязательно согласованной), составленных в относительных шкалах, должны иметь смысл весов (показателей значимости) соответствующих элементов с точки зрения всей совокупности введенных в матрицу суждений эксперта о сравнительной важности сопоставляемых элементов.

В ЭЛА элементарный шаг вычислений состоит в вычислении относительных приоритетов элементов данного уровня по отношению к элементу предыдущего уровня. Для этого по ответам эксперта строится матрица \mathbf{A} суждений и отыскивается ее число Фробениуса λ_{\max} . В качестве вектора приоритетов определяется нормированный собственный вектор матрицы \mathbf{A} , отвечающий λ_{\max}

$$\mathbf{A}\mathbf{w} = \lambda_{\max} \mathbf{w}; \quad \|\mathbf{w}\| = \sum_i w_i = 1$$

По известной теореме (Фробениуса-Перрона) любая положительная матрица имеет положительное действительное собственное значение λ_{\max} , которому отвечает единственный (с точностью до множителя) собственный вектор с положительными

компонентами. Тем самым существование вектора приоритетов (весов элементов) обеспечивается во всех случаях, когда в матрице суждений имеются лишь положительные элементы. Для вычисления приоритетов достаточно найти любой собственный вектор \mathbf{w} , отвечающий λ_{\max} , и разделить на сумму его компонент. Полученное решение всегда является единственным.

Е.2 Вычислительная процедура метода

Охарактеризуем основные моменты вычислительной процедуры метода. Вычислим вектор \mathbf{w}^1 приоритетов элементов первого уровня $11, 12, \dots, 1m_1$ и матрицу \mathbf{W}^1 , состоящую из собственных векторов-столбцов $\mathbf{w}^{11}, \mathbf{w}^{12}, \dots, \mathbf{w}^{1m_1}$ матриц суждений об элементах второго уровня относительно критериев $11, 12, \dots, 1m_1$. Тогда элементы второго уровня будут обладать приоритетами, являющимися компонентами вектора $\mathbf{w}^2 = \mathbf{W}^1 \mathbf{w}^1 = \mathbf{W}^{11} w_1^1 + \dots + \mathbf{W}^{1m_1} w_{m_1}^1$.

Таким образом, вектор приоритетов \mathbf{w}^2 представляет собой взвешенную сумму приоритетов элементов второго уровня, вычисленных в соответствии с критериями первого уровня, причем весами служат приоритеты (оценки значимости) критериев (акторов, целей, стратегий и т.п.), которые помещены на первом уровне иерархии исследования решений.

Продолжая последовательно ту же процедуру, получим для последнего k -го уровня искомый вектор приоритетов решений $\mathbf{w}^k = \mathbf{W}^{k-1} \mathbf{W}^{k-2} \dots \mathbf{W}^1 \mathbf{w}^1$, то есть для построения k -го финального вектора приоритетов нужно вычислить вектор приоритетов эле-

ментов первого уровня и умножить его слева на произведение матриц локальных приоритетов элементов всех последующих уровней до $k-1$ -го.

Остановимся на свойствах матрицы суждений. Если в балльной шкале суждений отсутствуют отрицательные числа и нуль, то матрица суждений положительна, неразложима на простые блоки (неприводима), примитивна (не сводится к циклу суждений) и устойчива, то есть существует предел $\mathbf{A}^k > \mathbf{0}$ при $k \rightarrow \infty$.

Перечислим другие полезные для анализа свойства матриц суждений как положительных матриц:

➤ число Фробениуса удовлетворяет соотношениям $\lambda_{\max} = \max_{\mathbf{x} \geq \mathbf{0}} \min_{1 \leq i \leq n} [(\mathbf{A}\mathbf{x})_i / x_i] = \min_{\mathbf{x} \geq \mathbf{0}} \max_{1 \leq i \leq n} [(\mathbf{A}\mathbf{x})_i / x_i]$, $\min_{1 \leq i \leq n} [(\mathbf{A}\mathbf{x})_i / x_i] \leq \lambda_{\max} \leq \max_{1 \leq i \leq n} [(\mathbf{A}\mathbf{x})_i / x_i]$,

➤ величина λ_{\max} ограничена максимальной и минимальной суммами элементов в строках или столбцах

$$\min_i [(\mathbf{A}\mathbf{e})_i] \leq \lambda_{\max} \leq \max_i [(\mathbf{A}\mathbf{e})_i],$$

$$\min_j [(\mathbf{e}^T \mathbf{A})_j] \leq \lambda_{\max} \leq \max_j [(\mathbf{e}^T \mathbf{A})_j],$$

где $\mathbf{e} = \|\mathbf{1}, \mathbf{1}, \dots, \mathbf{1}\|^T$, при несовпадении границ неравенство выполняется строго,

➤ если \mathbf{A} — стохастическая матрица (неотрицательная с суммой элементов в каждой строке, равной единице), то $\lambda_{\max} = 1$ и

$\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{e}\mathbf{v}$, $\mathbf{v} = \|v_1, \dots, v_n\| > \mathbf{0}$, $\sum_i v_i = 1$, вектор-строка \mathbf{v} задает фи-

нальное (эргодическое) распределение марковской цепи, для которой матрица \mathbf{A} служит матрицей переходных вероятностей;

➤ существуют положительные правый \mathbf{w} и левый \mathbf{v} собственные векторы матрицы \mathbf{A} , отвечающие λ_{\max} , и такие, что

$$\lim_{k \rightarrow \infty} (\mathbf{A}^k / \lambda_{\max}^k) = \mathbf{w}\mathbf{v},$$

➤ главные собственные векторы ортогональны неглавным: левый – левым, правый – правым;

➤ если \mathbf{A} – неотрицательная неразложимая матрица, то значение λ_{\max} возрастает с увеличением любого элемента a_{ij} , то есть λ_{\max} монотонная функция элементов матрицы \mathbf{A} .

Если все собственные числа простые, то правый собственный вектор \mathbf{w} , соответствующий λ_{\max} , удовлетворяет соотношению $\lim_{k \rightarrow \infty} [(\mathbf{A}^k \mathbf{e}) / (\mathbf{e}^T \mathbf{A}^k \mathbf{e})] = c\mathbf{w}$, где c – постоянный множитель; $\mathbf{A}^k \mathbf{e}$ – вектор-столбец из строчных сумм матрицы \mathbf{A}^k , $\mathbf{e}^T \mathbf{A}^k \mathbf{e}$ – сумма всех элементов матрицы $\mathbf{A}^k \mathbf{e}$.

Чтобы вычислить вектор приоритетов нужно вычислять степени матрицы \mathbf{A} до тех пор, пока максимальная разность элементов в двух последовательных векторах $(\mathbf{A}^{k-1} \mathbf{e}) / (\mathbf{e}^T \mathbf{A}^{k-1} \mathbf{e})$ и $(\mathbf{A}^k \mathbf{e}) / (\mathbf{e}^T \mathbf{A}^k \mathbf{e})$ не станет меньше наперед заданной точности ε .

Для вычисления λ_{\max} лучше всего воспользоваться соотношением $\lambda_{\max} = \lim_{k \rightarrow \infty} [tr(\mathbf{A}^k)]^{1/k}$. Оно показывает, что одновременно с вычислением матрицы \mathbf{A}^k нужно находить ее след (сумму диагональных элементов) и, извлекая из него корень k -й степени,

получать последовательные приближения к λ_{\max} до практического совпадения двух соседних значений.

Е.3 Согласованность суждений.

Индекс согласованности

Для любой матрицы ее след равен сумме собственных значений $\sum_i \lambda_i = tr \mathbf{A} = \sum_i a_{ii}$ (в сумме слева каждое λ_i берется с учетом его кратности). Поэтому для обратнo-симметрической матрицы $\sum_i \lambda_i = n$, а для согласованной матрицы все ее собственные значения, кроме числа Фробениуса, равны нулю (поскольку ее ранг равен единице).

Если суждения несовершенны, то указанные свойства могут нарушаться. При небольших вариациях элементов a_{ii} спектр матрицы изменится незначительно, так как спектр непрерывно зависит от элементов матрицы, и величина λ_{\max} останется близкой к n . Отличие старшего собственного значения от размерности матрицы служит показателем несогласованности суждений.

Если матрица оказывается несогласованной, то выполняется неравенство $\lambda_{\max} > n$, а остальные собственные значения могут быть отрицательными и комплексными (попарно сопряженными). В качестве показателя согласованности матрицы \mathbf{A} следует выбрать так называемый индекс согласованности μ :

$$\mu = (\lambda_{\max} - n)/(n - 1).$$

Индекс согласованности дает количественную оценку меры близости к согласованности всей совокупности суждений, соб-

ранных в матрице A : чем больше μ , тем меньше согласованность суждений.

Проверку согласованности матрицы суждений в случае $\lambda_{\max} > n$ проще всего осуществить, сопоставляя величину μ со «случайным индексом» (СИ), то есть с индексом согласованности обратно-симметрической матрицы той же размерности, выбранной случайным образом (с элементами в шкале $1 \div 9$). Значения СИ, табулированные по данным статистических экспериментов с объемом случайной выборки 500, приведены в таблице Е.1.

Таблица Е.1 — Значения «случайных индексов» (по данным вычислительного эксперимента)

n	1	2	3	4	5	6	7
СИ	0,00	0,00	0,58	0,90	1,12	1,24	1,32
n	8	9	10	11	12	13	14
СИ	1,41	1,45	1,49	1,51	1,54	1,56	1,57

Если величина $\mu/\text{СИ}(n)$ превосходит 0,1, суждения считаются несогласованными и подлежат пересмотру. В противном случае, суждения и результаты вычислений считаются в достаточной мере непротиворечивыми.

Несогласованность следует рассматривать как органическую черту оценочных построений. В таких построениях E может доминировать над F , F над G , а G над E . Такого рода нетранзитивность возникает, например, в таблицах результатов различного рода турниров и нередко встречается в опросах.

Исследование согласованности матриц суждений (то есть всей совокупности суждений об элементах данного уровня отно-

сительно одного элемента предыдущего уровня иерархии) основывается на следующих свойствах обратнo-симметрических положительных матриц.

Собственные значения λ_i обратнo-симметрической матрицы (с учетом их кратности) удовлетворяют соотношению $\sum_{j < k} \lambda_i \lambda_k = 0$. Матрица \mathbf{A} согласованна тогда и только тогда, когда $\lambda_{\max} = n$.

Е.4 Способы выявления и пересмотра несогласованных суждений

Для выявления и пересмотра несогласованных суждений рекомендуется следующий алгоритм.

1. Сформировать матрицу $\| |a_{ij} - w_i / w_j| \|$ и выявить наибольшие ее элементы – отклонения от согласованных суждений.

2. В строках найти средний квадрат отклонений (СКО). Вместо СКО можно пользоваться (или $\max \sum_j |a_{ij} - w_i / w_j|$) и пересмотреть строку с максимальным СКО. Этот способ основан на том наблюдении, что человек имеет склонность ошибаться при оценке одного фактора по отношению ко всем остальным, а не в одном конкретном суждении.

3. Выделить и пересмотреть элементы с наибольшим отношением a_{ij} к w_i / w_j .

Коррекция может состоять в удалении некоторых элементов из сформированной иерархии принятия решений. Если \mathbf{A} – положительная согласованная матрица и \mathbf{A}' получена из нее вычер-

киванием i -й строки и i -го столбца, то \mathbf{A}' согласованна и ее собственный вектор получится из \mathbf{w} — собственного вектора матрицы \mathbf{A} , если положить $w_i = 0$ и нормировать компоненты. В общем случае (когда матрица является несогласованной) так поступать нельзя и все вычисления приходится проводить заново.

Отметим еще, что если сравнения производят несколько человек, то лучше брать среднее геометрическое, а не среднее арифметическое их оценок. Если один человек берет оценку a , а другой противоположную $1/a$, то среднее геометрическое равно 1, а среднее арифметическое $(a + 1/a)/2$, что хуже, ибо дает систематическое смещение результатов.

Литература

1. Андерсен Т. Статистический анализ временных рядов. – М.: Мир, 1976.
2. Арженовский С.В., Федосова О.Н. Эконометрика. Учебное пособие / Рос. гос. экон. унив. – Ростов н/Д., – 2002. – 102с.
3. Беклемишев Д. В. Курс аналитической геометрии и линейной алгебры. – М.: ФИЗМАТЛИТ, 2005. – 304 с.
4. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управления. Вып. 1. – М.: Мир. Вып. 2. – М.: Мир. 1974.
5. Большев Л.Н, Смирнов Н.В. Таблицы математической статистики. – М.: Наука. 1983.
6. Боровиков В.П. Ивченко Г.И. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере. Учебное пособие. 2-е издание. – М.: Финансы и статистика. 2006. – 368 с.
7. Гельфанд А.О. Исчисление конечных разностей. – М.: Наука. 1967.
8. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения. – М.: Мир. Вып. 1, 1971. Вып. 2, 1972.
9. Драйпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Финансы и статистика, в 2-х книгах, 1987. – 392 с.
10. Канторович Г.Г. Анализ временных рядов. Лекционные и методические материалы. Экономический журнал ВШЭ. – 2002. – № 1. С. 85–116, № 2. С. 251–273, № 3. – С. 379–401, 2003. – № 1. С. 79–103.

11. Кендэл М. Временные ряды. – М.: Финансы и статистика. 1981.
12. Миллер Б.М. Панков А.Р. Теория случайных процессов в примерах и задачах. – М.: ФИЗМАТЛИТ. 2002. – 320 с.
13. Носко В.П. Эконометрика. Введение в регрессионный анализ временных рядов. www.iet.ru/mipt/2/text/curs_econometrics.htm
14. Саати Т. Принятие решений. Метод анализа иерархий. – М.: «Радио и связь». 1993. – 320 с.
15. Саати Т., Кернс К. Аналитическое планирование. Организация систем. – М.: «Радио и связь». 1991. – 224 с.
16. Себер Дж. Линейный регрессионный анализ. – М.: Мир, 1980. – 456 с.
17. Сухарев М.Г., Ставровский Е.Р. Оптимизация систем транспорта газа. – М.: Недра. 1975. – 277 с.
18. Тихонов А.Н., Уфимцев М.В. Статистическая обработка результатов эксперимента. – М.: МГУ, 1988.
19. Чураков Е.П. Математические методы обработки экспериментальных данных в экономике. Учебное пособие. – М.: Финансы и статистика, 2004. – 240 с.
20. Эконометрика / под ред. Елисейевой. – М.: Финансы и статистика. 2001. 344 с.

УДК 62-192 – 519.2

М.Г.Сухарев. Методы прогнозирования. Учебное пособие — М.: РГУ нефти и газа, 2009г., 208 с.

Компьютерный набор и верстка *Славоной Т.И.*

Подписано в печать

Формат 60×90/16

Бумага офсетная. Печать офсетная. Усл. п.л.

Тираж 100 экз. Заказ №



Отдел оперативной полиграфии РГУ нефти и газа им. И.М.Губкина

119991, Москва, Ленинский пр., 65.

Тел. 930-93-49