

Ю.Н. Орлов

КИНЕТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ

В основе книги лежит курс лекций, читаемый автором студентам магистратуры и аспирантам Московского физико-технического института. Излагается кинетический подход к анализу и прогнозированию нестационарных временных рядов, т.е. подход, основанный на кинетических уравнениях относительно функции распределения случайной величины. Основное внимание уделено методам оценки эмпирической плотности функции распределения и построения кинетических уравнений, моделирующих ее эволюцию. Определяется горизонт квазистационарного прогнозирования функции распределения временного ряда и находится оптимальный объем выборки данных для построения оператора эволюции внутри этого горизонта. Рассматриваются также и стандартные методы, развитые для стационарных случайных процессов, с точки зрения их адаптации к применению в нестационарных случаях. Теория сопровождается примерами анализа и прогнозирования временных рядов, которые встречаются в практической деятельности: динамические системы с хаосом, возникающие при дискретизации обыкновенных дифференциальных уравнений и уравнений с особенностями, последовательности символов в литературных текстах, биржевые ряды.

Книга рассчитана на студентов старших курсов и аспирантов, а также на специалистов в области прикладной математической статистики.

ОГЛАВЛЕНИЕ

Список сокращений.....	5
Предисловие.....	6
Введение.....	9
ЧАСТЬ I. ОСНОВЫ СТАТИСТИЧЕСКОЙ МЕХАНИКИ	17
Глава I. Гладкие динамические системы	17
1.1. Уравнение Лиувилля в классической статистической механике.....	17
1.2. Цепочка Боголюбова.....	27
1.3. Уравнение Больцмана.....	37
1.4. Уравнение Власова.....	53
1.5. Уравнения эволюции моментов.....	57
Глава II. Вырожденные динамические системы	64
2.1. Слаборелятивистские динамические системы.....	64
2.2. Сингулярное множество и хаотизация движения.....	77
2.3. Уравнение Лиувилля для вырожденных ДС.....	81
2.4. Гамильтоновы системы с высшими производными.....	87
2.5. Замечания о дискретизации вырожденных уравнений движения.....	93
Глава III. Хаотические динамические системы	97
3.1. Хаотическое движение гладких гамильтоновых систем.....	97
3.2. Хаотические динамические системы с дискретным временем.....	100
3.3. Вырожденные хаотические динамические системы.....	110
3.4. Уравнение Колмогорова–Фоккера–Планка.....	115
3.5. Модели фрактального блуждания.....	121
ЧАСТЬ II. ВЫБОРОЧНЫЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ	131
Глава IV. Нестационарные временные ряды	131
4.1. Понятие временного ряда.....	131
4.2. Вероятностные распределения и стационарность.....	136
4.3. Выборочные функции распределения и моменты.....	143
4.4. Отличие моментов стационарных рядов от нестационарных.....	150
4.5. Проблема оптимизации объема выборки.....	153
Глава V. Согласованный уровень стационарности	160
5.1. Согласованный доверительный интервал.....	160
5.2. Оценка плотности функции распределения.....	170
5.3. Оптимальное равномерное разбиение гистограммы.....	176
5.4. Распределение расстояний между выборочными плотностями.....	187
5.5. Индекс нестационарности выборочных распределений.....	201
Глава VI. Горизонтный ряд	205
6.1. Определение горизонтного ряда.....	205
6.2. Распределение горизонтного ряда стационарного процесса.....	212
6.3. Горизонтный ряд хаотической динамической системы.....	219
6.4. Распределения горизонтных рядов в нестационарном случае.....	223
6.5. Алгоритмы нахождения оптимального объема выборки.....	228

Глава VII. Индикаторы разладки	233
7.1. Скачки значений горизонтного ряда	233
7.2. Статистическая добротность временного ряда.....	238
7.3. Относительная энтропия и другие расстояния между ВПФР.....	242
7.4. Оптимальный дефлятор коинтеграции.....	247
7.5. Плотность распределения показателя Херста.....	251
ЧАСТЬ III. КИНЕТИЧЕСКИЕ УРАВНЕНИЯ И МОДЕЛИ	
ВРЕМЕННЫХ РЯДОВ	258
Глава VIII. Регрессионные модели	258
8.1. Стационарная линейная регрессия	258
8.2. Линейная регрессия в скользящем окне	270
8.3. Коинтегрированные регрессионные модели.....	275
8.4. Оценка корреляции нестационарных временных рядов	281
8.5. Метод численного распознавания нелинейной динамической системы.....	294
Глава IX. Кинетический метод прогнозирования.....	301
9.1. Уравнение эволюции выборочной плотности распределения.....	301
9.2. Модели замыкания эмпирического уравнения Лиувилля.....	305
9.3. Уравнение Фоккера-Планка для выборочных распределений	316
9.4. Кинетический подход к выводу моделей временных рядов.....	330
9.5. Эволюция внутри горизонта квазистационарности	339
Глава X. Маркированные точечные случайные процессы.....	345
10.1. Нестационарный пуассоновский поток событий.....	345
10.2. Уравнение Лиувилля в терминах времени и номера события	350
10.3. Прогнозные модели интенсивности потока событий.....	352
10.4. Кластеризация паттернов маркированных временных рядов.....	359
10.5. Зависимость выборочных статистик от времени.....	364
Глава XI. Прикладные задачи нестационарной статистики.....	369
11.1. Слабо меняющиеся временные ряды.....	369
11.2. Ряды с высоким уровнем нестационарности	374
11.3. Ряды с сезонной периодичностью	380
11.4. Анализ устойчивости распределения значений индикатора	388
11.5. Метод функций распределения в математической лингвистике.....	391
Глава XII. Вычислительные аспекты нестационарной статистики.....	404
12.1. Генерация нестационарного временного ряда.....	404
12.2. Численное моделирование фрактального блуждания	415
12.3. Возникновение искусственной корреляции	421
12.4. Возникновение отрицательных вероятностей	424
12.5. Численные схемы решения уравнения переноса	427
Заключение	430
Литература.....	431

Список сокращений

АКФ — автокорреляционная функция

АР — авторегрессия

АРИСС — авторегрессия интегрированных скользящих средних

АРСС — авторегрессия скользящих средних

ВПФР — выборочная плотность функции распределения

ВФР — выборочная функция распределения

ДС — динамическая система

ЛРМ — линейная регрессионная модель

МНК — метод наименьших квадратов

ПФР — плотность функции распределения

СДУ — стохастическое дифференциальное уравнение

СУС — согласованный уровень стационарности

СС — скользящие средние

ФР — функция распределения

ХФ — характеристическая функция

ЭЛС — эмпирическая лиувиллева скорость

DFP — distribution function parameters, параметры функций распределения

HDP — hydrodynamic parameters, гидродинамические параметры

Предисловие

В настоящей монографии излагается кинетический метод анализа и прогнозирования нестационарных временных рядов. Это метод, основанный на изучении эволюции выборочной функции распределения случайной величины. В книге кратко описываются также и традиционные методы и модели, применяющиеся для анализа стационарных временных рядов. Однако, хотя для стационарных моделей получаемые результаты являются корректно обоснованными в рамках классической математической статистики, в нестационарном случае получаемые выводы не имеют достоверной оценки точности.

Следует сказать, что существуют и эвристические подходы к анализу нестационарных рядов, когда предлагается некоторая модель, но математически корректного вывода ее не делается. Все такие модели, для которых невозможно указать точно величину ошибки прогноза или сформулировать требования к временному ряду, подчиняющемуся этой модели, здесь не рассматриваются.

Книга представляет собой учебное пособие по применению методов кинетической теории к исследованию нестационарных временных рядов. При этом предполагается, что читатель владеет основными понятиями теории вероятностей и математической статистики в рамках классических университетских курсов. По этой причине известные результаты теории вероятностей приводятся в книге без доказательств.

Достаточно подробно излагается метод вывода кинетических уравнений из цепочки эволюционных уравнений для многомерных функций распределения, описывающих статистическое поведение ансамбля для гамильтоновых систем. Дается вывод и анализ свойств кинетических уравнений Больцмана и Власова. Это необходимо потому, что классические курсы статистической механики и математической статистики обычно разведены по разным специализациям, а данный курс представляет собой объединение подходов обоих направлений. Чтобы использовать уравнение эволюции выборочной функции распределения для получения конкретных ответов о свойствах нестационарного временного ряда, необходимо не только владеть математическими приемами анализа уравнений определенного типа, но и представлять физические основы их вывода и пределы корректной применимости. С этой же целью подробно доказы-

ваются и утверждения о свойствах так называемых индикативных статистик нестационарных временных рядов, которые составляют теоретическую базу данного курса.

По структуре книга разделяется на три части. Первая часть посвящена введению в кинетическую теорию. В ней выводятся основные уравнения классической статистической механики: уравнение Лиувилля, цепочка Боголюбова для многомерных функций распределения гамильтоновых систем, а также некоторые модельные кинетические уравнения, имеющие широкое применение на практике. Эти уравнения выписываются для динамических систем разного типа: регулярных с гладким векторным полем, сингулярных с особенностями в фазовом пространстве и с хаотизацией траекторий на множестве особых точек, а также для так называемых хаотических динамических систем.

Во второй части вводятся понятия нестационарного статистического анализа и решаются такие задачи, как оптимизация объема выборки, построение индикаторов нестационарности временного ряда, оптимальное разбиение гистограммы, определение момента разладки выборочного распределения. Изучаются свойства специальных статистик, важных для решения указанных задач: это горизонтный ряд, согласованный уровень стационарности, статистическая добротность.

Третья часть посвящена собственно анализу нестационарных временных рядов. В ней выводятся различные модельные кинетические уравнения, описывающие эволюцию вероятностных распределений, и решается задача прогнозирования распределения и самого ряда на заданное число шагов вперед с заданной точностью. Строятся различные модели прогнозирования, выводимые из кинетических уравнений, и приводятся примеры практического применения развитой теории для анализа и прогнозирования временных рядов, встречающихся в различных областях деятельности: это ряды цен на финансовых и сырьевых рынках, ряды с сезонной и назначенной периодичностью, последовательности знаков в процессах передачи информации.

В книге принята двойная нумерация формул внутри каждой главы, первым идет номер параграфа; при ссылке на формулу из другой главы используется тройная нумерация, где первым идет номер главы. Такая же нумерация принята и для важных определений и теорем, выделяемых в тексте символами ▼ и ▲. Теоремы, известные из общеинститутских курсов, приводятся без доказательства.

Рисунки и таблицы имеют тройную нумерацию в соответствии с главой и параграфом.

Список литературы составлен в порядке очередности появления ссылок в тексте.

Введение

Создание математических моделей, с помощью которых можно изучать статистические зависимости между величинами, изменение во времени которых обусловлено случайными факторами, скрытыми от наблюдателя, является не только перспективной задачей современного естествознания, но и актуальной в смысле многочисленных практических приложений — от прогнозирования природных катаклизмов до игры на бирже. Эти явления предстают перед исследователем в форме временного ряда, т.е. в виде некоторой последовательности результатов наблюдений, сделанных в упорядоченные (не обязательно равноотстоящие) моменты времени.

Статистические методы позволяют получить количественные оценки зависимостей между величинами, по видимости не находящимися в прямой функциональной связи. Поскольку результаты таких оценок часто являются основой для принятия решений, то корректная обоснованность методов анализа случайных величин имеет большое практическое значение.

В настоящей монографии представлены результаты работы автора в области прикладной математической статистики. В основе развиваемого подхода лежит понятие выборочной плотности функции распределения, меняющейся с течением времени. Представление об эволюции выборочного распределения используется затем для построения прогноза нестационарного временного ряда. Тем самым задача прогнозирования может быть сформулирована как построение модели эволюции функции распределения наблюдаемого процесса. Поскольку наиболее полная информация о процессе содержится в функции распределения значений ряда, то именно эволюция функции распределения, а не тех или иных функционалов от нее (математического ожидания, дисперсии и т.п.), должна быть первичным объектом изучения.

Уравнения эволюции функции распределения традиционно строятся либо для определенных типов случайных процессов, когда задаются вероятности перехода системы из одного состояния в другое в единицу времени, либо для динамических систем с большим числом степеней свободы. В последнем случае эти уравнения называются кинетическими и изучаются в курсе статистической механики как разделе теоретической физики.

Кинетические уравнения обычно ассоциируются с фазовым пространством некоторой детерминированной механической системы, состоящей из большого числа объектов, не обязательно идентичных, каждый из которых подчиняется определенным законам — уравнениям движения. Если оказалось возможным свести одновременное рассмотрение движения большого числа идентичных (или в чем-то похожих) объектов к анализу эволюции некоторой функции, зависящей от меньшего числа аргументов, то говорят о сокращении описания изучаемой системы. В таком случае кинетическое уравнение позволяет изучать влияние большого количества факторов, каждый из которых меняется детерминированным образом, на наблюдаемую величину, являющуюся, как правило, функционалом от функции распределения. Чаще всего в качестве таких функционалов выступают моменты функции распределения. Например, в статистической механике газов уравнения эволюции первых трех моментов функции распределения (нулевого, первого и центрального момента второго порядка) известны как уравнения гидродинамики — неразрывности (Эйлер), вязкости (Навье–Стокс) и теплопроводности. Таким образом, функции распределения в статистической механике существенно нестационарны, но они подчиняются определенным уравнениям эволюции.

В теории вероятностей и математической статистике, напротив, изучаются статистические свойства случайного процесса с некоторой неизменной функцией распределения. Постановке задачи построения эволюции эмпирического распределения в математической статистике отчасти препятствует существующая аксиоматика теории вероятностей, в которой наблюдаемое значение случайной величины трактуется как выборка из некоторой так называемой «генеральной совокупности». Однако следует помнить, что генеральная совокупность является теоретической абстракцией, некоторым бесконечным набором элементарных носителей изучаемого признака, распределение которого и определяется этим набором. В реальности не обязательно должны фактически присутствовать компоненты, составляющие содержание абстрактных конструкций, с помощью которых можно математически корректно доказывать существование решений тех или иных модельных задач.

В некоторых случаях концепция генеральной совокупности является хорошим приближением. Например, в упомянутой выше кинетической теории газов удобно считать, что вместо одного — реального — сосуда с газом экспериментатор имеет дело с набором таких сосудов, находящихся

в одних и тех же внешних условиях, но несколько отличающихся внутренне: положения и скорости частиц в этих сосудах различны. Предположим, что число сосудов таково, что в них реализованы все возможные механические состояния частиц. Такой набор называется статистическим ансамблем. Распределение изучаемого признака в этом ансамбле образует соответствующую генеральную совокупность. Пусть эксперимент с имеющимся в наличии единственным сосудом состоит в том, что в момент времени t_1 наудачу выбирается одна из молекул и определяется, например, ее скорость или положение. Затем эксперимент повторяется в момент t_2 и т.д. В результате получается временной ряд наблюдений за указанным признаком. Здесь не обсуждается, с какой точностью можно провести такой эксперимент, а также то, как измерение повлияет на состояние содержимого сосуда в дальнейшем. Предполагая, что распределение ансамбля не меняется (на физическом языке это означает, что газ находится в равновесии), можно определить эмпирическую вероятность распределения частиц по скоростям. Формально с точки зрения математической статистики для этого надо провести одновременные измерения данного признака по всем системам ансамбля, однако физически можно считать, что большое число частиц газа позволяет реализовать «почти все» возможные состояния одной частицы в одном имеющемся сосуде. Тогда выборка наблюдаемых значений может трактоваться как выборка из генеральной совокупности, параметры распределения которой тогда могут быть найдены с той или иной точностью.

Концепция «сосуда со всевозможными значениями признака», из которого черпаются значения временного ряда, оказалась математически удобной, но она не всегда физически адекватна изучаемой системе. В действительности часто приходится иметь дело единственной реализацией некоторого процесса, причем невозможен мысленный эксперимент размножения сосудов. Если процесс стационарный, то обоснованность статистических выводов из наблюдений велика. Если же нет, надо иметь инструмент исследования, позволяющий и в этой ситуации отвечать на практически важные вопросы.

Следует подчеркнуть, что, строго говоря, в природе вообще нет стационарных процессов, как они определены в соответствующих математических курсах. На практике идеальные условия теорем выполняются лишь приближенно, причем точно они не могут быть выполнены в принципе.

Исследователю же предстоит установить, насколько в условиях проведения конкретного статистического эксперимента можно считать условия теорем выполненными. Так, при анализе разброса размеров вытачиваемых однотипных деталей можно брать на вооружение модели различной точности. Например, а) не принимать во внимание процесс износа рабочей части станка, т.е. пренебречь слабой нестационарностью процесса изготовления; б) принять износ во внимание, измерить изменение формы рабочей поверхности, но пренебречь взаимным влиянием измерительного инструмента и измеряемой конструкции; в) учесть это взаимное влияние, и после каждого измерения измерять сам измерительный инструмент, но пренебречь изменением инструмента в процессе такого измерения, и т.д.

Кроме того, все без исключения тела меняют свои размеры и вес в результате испарения атомов с поверхности (или, наоборот, осаждения на поверхность), ядерных реакций под воздействием фона нейтронов или иных частиц и т.д. Разумеется, чаще всего эти изменения не существенны, т.е. эксперимент имеет обычно гораздо более низкую точность, чем уровень таких изменений, но это не означает, что изменений вообще нет.

Следовательно, необходимо помнить, что определения и утверждения теории вероятностей и математической статистики идеализируют реальность. В этом одновременно состоят их слабость и сила. Сильная сторона теоретических построений в том, что они позволяют выбрать главную часть изучаемого явления, которая, собственно, и наблюдается в эксперименте. Слабость — на самом деле это слабость не утверждений, а исследователя, — в том, что иногда может возникнуть желание подмены явления на его главную часть, которая в определенных условиях может и перестать быть главной.

В математической статистике рассматриваются следующие два основных типа задач: по имеющимся данным оценить параметры распределения, вид которого априори известен, либо проверить статистическую гипотезу относительно вида распределения. В обоих случаях предполагается, что генеральная совокупность существует. Если процесс нестационарный и неизвестной природы, то в рамках существующего формализма указанные задачи не могут быть решены, поскольку значения временного ряда принадлежат разным генеральным совокупностям, и потому для оценки параметров распределения в каждый момент времени у нас имеется только одно значение, чего явно недостаточно. Не желая, однако, закрывать на этом изучение явления, которое объективно существует, сле-

дует понять, какие содержательные утверждения в таком случае можно сделать относительно этого явления.

Допустим, изучается временной ряд биржевых цен на нефть. Классическая теория вероятностей предлагает следующую модель (известную как урна с возвращениями): биржа — это такая своеобразная урна с шарами, на которых написана цена, например, барреля нефти в определенном контракте. Этих шаров там бесконечно много, надписи на них могут и совпадать. Каждый день невидимая «рука рынка» достает из урны один шар, который мы и наблюдаем. На следующий день состав шаров в урне, вообще говоря, совершенно другой. Процесс повторяется ежедневно. Понятно, что на основе таких наблюдений нельзя определить, какой был состав шаров в урне в некоторый (любой) день. Временной ряд наблюдений будет отражать не объективную реальность (вернее, он и ее будет отражать, но очень слабо, на уровне единичного эксперимента), каковой является состав шаров в урне в определенный день, а наше восприятие этой реальности. Но, заметим, что именно это последнее и имеет для нас практическое значение. Предсказательная сила теории тем выше, чем ближе субъективное восприятие к реальному положению дел. Следовательно, в данном случае следует попытаться минимизировать ошибку, которую можно сделать при прогнозе очередного значения цены, но не на основе данных о содержимом урны (эта реальность в рамках поставленного эксперимента остается неизвестной), а только на основе результатов эксперимента. Эти результаты не являются выборкой из генеральной совокупности, но зато они формируют выборку результатов наблюдений, и распределение этой выборки, меняясь с течением времени, может дать ответ об ожидаемой ошибке прогноза, а иногда, если изменение сравнительно мало, и эмпирически определить приближенное правило такого изменения.

Для того, чтобы с некоторой точностью отнести изучаемый случайный процесс к тому или иному классу, разработано большое количество статистических тестов или критериев. Большинство методов формулируются в терминах стационарных в широком или узком смысле процессов. Это связано с тем, что для стационарных случайных процессов доказаны теоремы, позволяющие получать корректные оценки параметров соответствующих распределений по данным наблюдений.

Обычно ряд пытаются представить в виде суммы некоторой детерминированной составляющей и остатка, автокорреляционная функция кото-

рого с достаточной точностью близка к нулю, что позволяет выдвинуть предположение о близости остатка к белому шуму. Строго говоря, именно в такую сторону утверждение и не работает: белый шум имеет нулевую автокорреляцию, но не всякая нулевая автокорреляция порождена белым шумом, это может быть и детерминированный процесс. Тем не менее, на практике как раз исходят из рабочей гипотезы сопоставлению процессу с нулевой автокорреляцией некоторого белого шума. После этого параметрическими или иными методами находят наиболее близкую статистику, моделирующую поведение остатка. Существуют различные модификации такого подхода. В настоящее время основными статистическими методами исследования временных рядов являются: метод выделения полиномиального временного тренда или иной параметрической функциональной зависимости, регрессионный, автокорреляционный, адаптивный (скользящих средних), метод гармонического анализа, сингулярного спектрального анализа, нейросетевой метод.

Подчеркнем, что вышеперечисленные методы корректно обоснованы только для стационарных рядов. Однако, как уже говорилось, существует большой класс актуальных временных рядов, которые не являются стационарными. В этом случае выборочные средние не являются состоятельными оценками моментов распределения, и потому применение асимптотических критериев, гарантирующих в стационарном случае увеличение точности аппроксимации генеральной совокупности или ее моментов с увеличением объема выборки, не увеличивают точности прогноза. В то же время за неимением лучшего эти критерии продолжают применяться ко всем рядам, которые возникают в случайном эксперименте. Тогда перед исследователем встает проблема корректной оценки точности получаемых им результатов, на основе которой адаптивно настраиваются параметры используемых моделей. Основным параметром является длина выборки, отсчитываемая назад от текущего момента времени. При этом актуальным вопросом является следующий: по выборке какого объема следует проводить скользящее усреднение, чтобы получить наименьшую ошибку прогноза на заданном временном горизонте? Учет большого числа членов ряда может привести к тому, что текущий тренд будет перекрыт «вышедшими из моды» тенденциями. Малое же число членов ряда не даст статистической репрезентативности выборки.

Таким образом, основной проблемой прогнозирования временного ряда (в том числе и стационарного) является оценка сверху среднеквадра-

тичной ошибки, которую можно допустить, если применить к нему тот или иной метод анализа. Для ее решения существующие методы для стационарных рядов следует дополнить оценками временных границ их применимости.

В отличие от некоторого (впрочем, достаточно широкого) класса «естественно-физических» проблем, при решении которых гипотеза стационарности кажется разумной, временные ряды в иных областях человеческой деятельности в подавляющем большинстве не являются стационарными. Проблемы, возникающие при их анализе, широко обсуждаются в специальной литературе, но, как правило, на уровне «контрпримеров», когда оказывалось, что та или иная априорная гипотеза несправедлива, или на уровне примеров удачных моделей в отдельных конкретных случаях. Первое систематическое изложение методов анализа нестационарных временных рядов было предпринято в монографии автора [100], где содержалось большое количество примеров из практической области. В настоящей книге основное внимание уделяется теоретической базе — динамическим системам и кинетическим уравнениям для генеральных и для выборочных плотностей функций распределения.

Конкретной целью статистического анализа является построение метода минимизации ошибки прогноза выборочной функции распределения и собственно временного ряда на заданном временном горизонте. Параметром, по которому минимизируется ошибка прогноза, является текущий объем выборки элементов ряда. Оптимизация объема выборки возможна в силу того, что ошибка прогноза состоит из ошибок двух типов. Ошибка первого типа — это погрешность оценки статистических свойств временного ряда за счет конечности объема выборки, т.е. за счет недостаточной репрезентативности. Ошибка второго типа — это погрешность за счет нестационарности статистики. Ошибка первого типа с ростом объема выборки уменьшается, а ошибка второго типа возрастает. Таким образом, корректный метод прогнозирования должен учитывать взаимосвязь между точностью прогноза, объемом выборки и горизонтом прогноза. В стандартных прогнозных моделях стационарных временных рядов эти понятия разобщены.

Для определения оптимального объема выборки и создания адекватной прогнозной модели временного ряда необходимо решить следующие задачи.

Первой задачей является создание системы необходимых понятий для корректного прогнозирования нестационарных временных рядов, когда отсутствует само понятие генеральной совокупности, и разработка подходящего теоретического инструмента для минимизации ошибки прогноза.

Вторая задача связана с созданием математической прогнозной модели для выборочной функции распределения. Рассматривая задачу прогнозирования как задачу о построении оператора эволюции функции распределения, можно сконструировать эмпирическое кинетическое уравнение типа уравнения Лиувилля и определить некоторый кинетический аналог скорости изменения плотности вероятности, т.е. построить квазидинамическую модель временного ряда.

Третья задача — создание численного алгоритма, реализующего предложенную методику.

Обсуждению подходов к решению этих задач и результатов, получаемых в рамках применения кинетических моделей прогнозирования, и посвящена данная монография.

ЧАСТЬ I. ОСНОВЫ СТАТИСТИЧЕСКОЙ МЕХАНИКИ

Глава I. Гладкие динамические системы

1.1. Уравнение Лиувилля в классической статистической механике

В данной монографии изложен метод анализа и прогнозирования эмпирических временных рядов, использующий кинетические уравнения для моделирования эволюции выборочной плотности функции распределения изучаемой случайной величины. В этой связи следует подчеркнуть, что кинетическая теория традиционно понимается как статистическая теория движения молекул газа, причем каждая из молекул подчиняется строго детерминированным уравнениям движения, а статистическая механика как наука занимается изучением вероятностных распределений кинематических характеристик механической системы. Понимаемая в более общем смысле, статистическая механика изучает вероятностные распределения обобщенных координат, характеризующих систему, которые эволюционируют в соответствии с определенными уравнениями. Эти уравнения представляют собой корректно поставленную начальную (или начально-краевую) задачу Коши.

Объект же математической статистики — временной ряд — в общем случае не представляет собой реализацию какой-либо динамической системы. Тем не менее, выборочная функция распределения любого временного ряда существует, и относительно нее можно ставить вопрос о ее изменении с течением времени. Соответствующее уравнение эволюции естественно назвать кинетическим, хотя оно и не будет иметь того изначального смысла, которым обладает, например, уравнение Лиувилля в классической механике. Но интерпретация такого уравнения может быть дана чисто в механических терминах, что и позволяет использовать методы статистической механики для изучения статистических характеристик временных рядов. Этот подход является новым и весьма перспективным для анализа нестационарных случайных процессов.

Существует большое количество литературы по теории динамических систем (далее ДС), рассматриваемых с различных позиций: топологические ДС, гладкие ДС, гамильтоновы ДС и др. Систематическое изложение теории ДС содержится, например, в монографии [1].

▼ **Определение 1.1.** Динамической системой называется однопараметрическая группа (или полугруппа) G_{t,t_0} преобразований $x_t = G_{t,t_0} x_0$ координат, действующая на некотором множестве точек $x \in \Gamma$, называемом фазовым пространством системы. Координата $x_0 \equiv x_{t_0} \in \Gamma$ есть значение x_t при значении параметра (т.е. времени) $t = t_0$. ▲

Обычно множество Γ является дифференцируемым многообразием. Динамическая система в контексте определения 1.1 называется также потоком (или динамическим потоком). Если преобразование $x_t = G_{t,t_0} x_0$ определено для дискретного параметра t , то такая ДС называется каскадом. Удобно считать начальный момент времени нулевым, полагая $t_0 = 0$, и соответствующий поток обозначать для краткости $G_{t,0} \equiv G_t$.

▼ **Определение 1.2.** ДС называется гладкой, если поток G_t таков, что существует гладкое векторное поле

$$v(x) = \frac{d}{dt} G_t x \Big|_{t=0}, \quad (1.1)$$

называемое фазовой скоростью. ▲

Задание векторного поля полностью определяет динамический поток в том смысле, что при заданном начальном условии x_0 решением дифференциального уравнения $\dot{x} = v(x)$ является интегральная кривая $x_t = G_t x_0$. Вследствие этого уравнение $\dot{x} = v(x)$ часто само рассматривается как определение гладкой динамической системы.

▼ **Определение 1.3.** Гладкая ДС называется автономной, если фазовая скорость не зависит явно от времени: $\dot{x} = v(x)$, $x \in M \subset \mathbf{R}^n$. ▲

Если взять любую динамическую величину, т.е. некоторую функцию $B(x)$, аргумент которой зависит от параметра t в соответствии с автономным уравнением $\dot{x} = v(x)$, то ее значение в момент времени t символически записывается в виде

$$B_t(x) \equiv B(x_t) = B(G_t x_0) = G_t(B(x_0)). \quad (1.2)$$

Иными словами, на гладкой траектории динамический поток коммутирует с функцией от обобщенной координаты.

Важный класс ДС составляют лагранжевы системы классической механики. Это гладкие ДС в пространстве размерности $2n$, интегральные кривые которых являются решением экстремальной задачи для функционала, называемого действием:

$$S[L] = \int_{t_0}^{t_1} L(q, v) dt. \quad (1.3)$$

Функция $L(q, v)$ $2n$ переменных — обобщенных координат $q = (q_1, \dots, q_n)$ и скоростей $v = \dot{q}$ — называется лагранжианом. Если действие (1.3) имеет стационарную точку, то эта точка называется экстремалью и является решением уравнений Эйлера–Лагранжа вариационной задачи с закрепленными концами

$$\frac{d}{dt} \frac{\partial L}{\partial v_i} = \frac{\partial L}{\partial q_i}, \quad (1.4)$$

которые представляют собой систему дифференциальных уравнений второго порядка по времени t относительно координат:

$$\frac{\partial^2 L}{\partial v_i \partial v_j} \dot{v}_j = - \frac{\partial}{\partial q_j} \left(\frac{\partial L}{\partial v_i} v_j - L \delta_{ij} \right). \quad (1.5)$$

При условии невырожденности матрицы вторых производных лагранжиана по скоростям система (1.5) может быть записана в виде системы дифференциальных уравнений первого порядка относительно фазовой переменной $\xi = (q, v)$, т.е. приведена к виду определения 1.1. Такое приведение осуществляется преобразованием Лежандра.

▼ **Определение 1.4.** Пусть матрица $J = (J_{ij}) = (\partial^2 L / \partial v_i \partial v_j)$ невырождена: $|J| = \det J \neq 0$. Тогда преобразование координат $\mathfrak{J}: \{q, v\} \rightarrow \{q, p\}$, где $p_i = \partial L / \partial v_i$ — обобщенный импульс, называется преобразованием Лежандра. ▲

Для автономной ДС величина

$$E = \sum_i p_i(v) v_i - L(q, v) \quad (1.6)$$

есть первый интеграл уравнений движения (1.4), т.е. она сохраняется в силу этой системы. Если якобиан $|J|$ невырожден, то уравнения $p_i = \partial L / \partial v_i$ имеют единственные непрерывно дифференцируемые и обратимые решения $v_i = v_i(q, p)$. Тогда преобразование Лежандра позволяет перейти к гамильтонову описанию динамики с функцией Гамильтона $H = \mathfrak{J}E = p v(p) - L(q, v(p))$.

В переменных $\{q, p\}$ уравнения (1.5) запишутся в виде системы двух уравнений первого порядка (уравнений Гамильтона) относительно фазовой скорости $\dot{\eta} = (\dot{q}, \dot{p})$, которая определяется кососимметрическим градиентом (скобкой Пуассона):

$$\dot{\eta} = [\eta, H] = I_\Gamma \frac{\partial H}{\partial \eta}, \quad I_\Gamma = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad (1.7)$$

где I — единичная $n \times n$ -матрица. Соответствующая ДС, задаваемая уравнениями (1.7), называется гамильтоновой.

В более общих случаях функция Лагранжа может зависеть и от высших производных (см., например, [2, 3]): $L_N = L(q, \dot{q}, \ddot{q}, \dots, q^{(N)})$ где $q \in \mathbf{R}^n$, $q^{(s)} = d^s q / dt^s$. Уравнения экстремалей для действия (1.3), порождаемые этим лагранжианом, имеют вид

$$\dot{p}_i = \frac{\partial \mathcal{L}}{\partial q_i}, \quad p_i = \sum_{s=0}^{N-1} \left(-\frac{d}{dt} \right)^s \frac{\partial \mathcal{L}}{\partial q_i^{(s+1)}}. \quad (1.8)$$

Уравнения (1.8) также имеют интеграл энергии, записывающийся в виде

$$E = -L + \sum_{s=0}^N \left(-\frac{d}{dt} \right)^s \sum_{l=1}^N \mathbf{q}_i^{(l)} \frac{\partial \mathcal{L}}{\partial \mathbf{q}_i^{(l+s)}}. \quad (1.9)$$

Переход к гамильтонову описанию осуществляется посредством преобразования Остроградского [2]: вводятся обобщенные координаты $Q_s = q^{(s-1)}$ и сопряженные импульсы P_s так, что

$$\mathfrak{J}E(q, \dot{q}, \dots, q^{(N)}) = H(Q, P) = -\mathfrak{J}L + \sum_{s=1}^N Q_s P_s, \quad (1.10)$$

$$P_s = \sum_{k=s}^N \left(-\frac{d}{dt} \right)^{k-s} \left(\frac{\partial L}{\partial q^{(s)}} \right).$$

Достаточным условием невырожденности преобразования (1.10) является возможность однозначно выразить старшую производную $q^{(N)}$ через обобщенный импульс P_N .

Таким образом, можно сказать, что классическая теория динамических систем есть теория обыкновенных дифференциальных уравнений, разрешенных относительно старшей производной.

Задачей анализа ДС является изучение взаимного расположения интегральных кривых соответствующего векторного поля. При этом удобно представлять себе каждую интегральную кривую как траекторию движения некоторой точечной массы. Тогда естественным представлением об интегральных кривых данной ДС является аналогия с течением идеальной сплошной среды, никакие части которой не взаимодействуют между собой. Распределение плотности этой среды отвечает гипотетическому распределению точечных масс.

Построение статистической механики для ДС, определяемой потоком G_t , начинается с задания плотности функции распределения начальных состояний $\rho_0(x_0)$. Пусть $x \in \Gamma \subset \mathbf{R}^n$. Тогда среднее значение динамической величины (1.2) в момент времени t по этому распределению ρ_0

определяется как интеграл от $B(x_t)$ по фазовому пространству с весом ρ_0 . Имеем следующую цепочку преобразований:

$$\begin{aligned}\langle B_t \rangle_0 &= \int_{\Gamma} B(G_t(x_0)) \rho_0(x_0) dx_0 \equiv \int_{\Gamma} B(G_t(x)) \rho_0(x) dx = \\ &= \int_{\Gamma} \rho_0(x) dx \int_{\Gamma} B(x') \delta(G_t(x) - x') dx' = \int_{\Gamma} B(x') \rho_t(x') dx' \equiv \langle B \rangle_t,\end{aligned}$$

где введено обозначение

$$\rho_t(x') = \int_{\Gamma} \rho_0(x) \delta(G_t(x) - x') dx. \quad (1.11)$$

В вышеприведенной последовательности равенств действие эволюционного оператора G_t переносится с динамической величины на плотность функции распределения (далее ПФР). Из (1.11) следует уравнение эволюции ПФР:

$$\begin{aligned}\frac{\partial}{\partial t} \rho_t(x) &= \int_{\Gamma} \rho_0(x') \left(v(x'), \frac{\partial}{\partial x'} \right) \delta(G_t(x') - x) dx' = \\ &= - \int_{\Gamma} \rho_0(x') \left(v(x'), \frac{\partial}{\partial x} \right) \delta(G_t(x') - x) dx' = \\ &= - \operatorname{div} \int_{\Gamma} \rho_0(x') v(x') \delta(G_t(x') - x) dx' = \\ &= - \operatorname{div} v(x) \rho_t(x) \equiv - \operatorname{div}(\dot{x} \rho_t(x)).\end{aligned} \quad (1.12)$$

Это уравнение, записанное в виде закона неразрывности течения точек с меняющейся во времени фазовой плотностью, называется уравнением Лиувилля. Подчеркнем, что оно имеет место только в случае невырожденности (гладкости) ДС. Нижний индекс (время) далее внесем в аргументы ПФР, так что $\rho_t(x) \equiv \rho(x, t)$.

Если теперь определить меру в Γ как абсолютно непрерывную относительно меры Лебега, т.е. имеющую плотность $\rho(x, t)$, так что $d\Gamma = \rho(x, t) dx$, то уравнение (1.12) будет описывать эволюцию этой меры в силу рассматриваемой динамической системы.

▼ **Определение 1.5.** Динамическая система называется консервативной, если расходимость фазового потока равна нулю: $\operatorname{div} \dot{x} = \partial v(x) / \partial x = 0$. Соответствующая мера называется динамически-инвариантной. ▲

Для консервативных систем уравнение Лиувилля записывается в виде

$$\frac{\partial \rho}{\partial t} + \dot{x} \frac{\partial \rho}{\partial x} = 0. \quad (1.13)$$

В частности, для гамильтоновых систем в силу кососимметричности (1.7) выполняется условие консервативности $\operatorname{div} \dot{\eta} = 0$.

Существование динамически-инвариантной меры для автономной ДС в общем случае дается следующей теоремой [1].

▼ **Теорема 1.1.** (Боголюбов–Крылов, 1937). Пусть в фазовом пространстве Γ (конечномерном евклидовом или метрическом компактном) задана ДС $\dot{x}_i = v_i(x_1, \dots, x_n)$. Тогда, если функции v_i действительные, ограниченные и удовлетворяют условиям Липшица, то в пространстве Γ существует динамически-инвариантная мера. ▲

Следует сделать важное замечание относительно аналогии уравнения (1.13) с уравнением неразрывности для идеальной сплошной среды. Эта аналогия не полная. Дело в том, что идеальная сплошная среда может быть удобной моделью большого, но конечного набора невзаимодействующих частиц, заключенных в некотором сосуда и абсолютно упруго отражающихся от его стенок. Для ПФР этих частиц также может быть выписано уравнение вида (1.13), но его решение имеет несколько иные свойства, чем решение того же уравнения для сплошной среды. Различие между «частицами» и «сплошной средой» заключается в том, что в этих случаях ПФР $\rho(x, t)$ принадлежит разным функциональным классам. Это различие приводит к тому, что для «сплошной среды» существует слабый предел ПФР при $t \rightarrow \infty$ [4], тогда как для точечных «частиц» такой предел отсутствует.

▼ **Теорема 1.2.** (Пуанкаре, 1906–Козлов, 2001). Пусть в фазовом пространстве $\Gamma = \mathbf{T}^n \times \mathbf{R}^n$ задана интегрируемая по Лебегу ПФР $\rho(x, v)$, удовлетворяющая уравнению Лиувилля (1.13), и $g(x)$ есть интегрируемая по Риману функция, заданная на n -мерном торе T^n . Введем функционал

$$K(t) = \int_{\Gamma} \rho(x - vt, v) g(x) d^n x d^n v. \quad (1.14)$$

Тогда существует предел этого функционала при $t \rightarrow \infty$, равный

$$\lim_{t \rightarrow \pm\infty} K(t) = \langle \rho \rangle \bar{g}, \quad (1.15)$$

где

$$\langle \rho \rangle = \int_{\Gamma} \rho(x, v) d^n x d^n v, \quad \bar{g} = \frac{1}{(2\pi)^n} \int_{\mathbf{T}^n} g(x) d^n x. \quad \blacktriangle (1.16)$$

В частном случае теоремы 1.2, когда $g(x)$ является характеристической функцией некоторого множества $G \subset \Pi$, измеримого по Жордану, функционал $K(t)$ может быть представлен в виде

$$K(t) = \int_{\Gamma} \rho(x - vt, v) g(x) d^n x d^n v = \int_{G \times \mathbf{R}^n} G_t(f(x, v)) d^n x d^n v.$$

Последний интеграл представляет собой долю частиц, находящихся в момент времени t в объеме G . Следовательно, теорема имеет следующую интерпретацию:

$$\begin{aligned} \lim_{t \rightarrow \pm\infty} K(t) &= \langle \rho \rangle \bar{g} = \\ &= \int_{\Gamma} \rho(x, v) d^n x d^n v \cdot \frac{1}{(2\pi)^n} \int_{\mathbf{T}^n} g(x) d^n x = \frac{mes G}{mes \Pi}. \end{aligned} \quad (1.17)$$

Из (1.17) следует, что частицы со временем стремятся равномерно заполнить сосуд: их асимптотическая плотность в слабом смысле стремится к постоянной. Выравнивание же плотности в слабом смысле означает, что в том же смысле имеет место диффузия, при которой начальные неоднородные условия необратимо размываются. Это показывает, что при определенных условиях необратимость может появиться и в идеальной среде. Можно распространить утверждение теоремы 1.2 и на другие классы функций, которым принадлежит ПФР, например, на ПФР вида $\rho(x, v) = \delta(x) f(v)$.

▼ **Теорема 1.3.** Пусть $\rho(x, \omega) = \delta(x) f(v)$, где $f(v)$ — интегрируемая по Лебегу функция на \mathbf{R}^n , и пусть $g(x)$ есть интегрируемая по Риману функция на \mathbf{T}^n . Тогда существует предел при $t \rightarrow \infty$ функционала

$$K(t) = \int_{\Gamma} f(x - vt, v) g(x) d^n x d^n v,$$

равный

$$\lim_{t \rightarrow \infty} K(t) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} f(v) d^n v \int_{\mathbf{T}^n} g(x) d^n x. \quad (1.18)$$

Доказательство проводится в несколько этапов. Предварительно заметим, что, поскольку любую интегрируемую по Лебегу функцию можно представить как разность двух положительных функций, то доказательство достаточно провести только для положительных функций f . Сначала докажем теорему для случая, когда g является тригонометрическим полиномом. Затем получим то же равенство для всех функций g из пространства интегрируемых по Риману на \mathbf{T}^n .

1) Пусть $g = C$. Тогда сразу же получаем формулу (1.18):

$$\begin{aligned} K(t) &= \int_{\mathbf{R}^n} C \cdot f(v) dv = C \int_{\mathbf{R}^n} f(v) dv = \frac{1}{(2\pi)^n} \int_{\mathbf{T}^n} C d^n x \int_{\mathbf{R}^n} f(v) d^n v = \\ &= \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} f(v) d^n v \int_{\mathbf{T}^n} g(x) d^n x. \end{aligned}$$

2) Пусть $g(x) = \exp(ikx)$. Тогда

$$K(t) = \int_{\Gamma} \exp(ikx) \delta(x - vt) f(v) d^n x d^n v = \int_{\mathbf{R}^n} \exp(ikvt) f(v) d^n v,$$

и по теореме Римана $\lim_{t \rightarrow \infty} K(t) = 0$. В силу линейности функционала

$K(t)$ отсюда следует, что равенство (1.18) выполнено для всех тригонометрических полиномов.

3) Пусть теперь $g(x)$ — функция, интегрируемая по Риману. Тогда для любого ε существуют такие тригонометрические полиномы $g_1(x)$ и $g_2(x)$, что $g_1(x) \leq g(x) \leq g_2(x)$, и $\overline{g_2} - \overline{g_1} < \varepsilon$ где чертой сверху обозначено усреднение функции по тору. Обозначим

$$K_1(t) = \int_{\mathbf{R}^n} g_1(vt) \cdot f(v) dv, \quad K_2(t) = \int_{\mathbf{R}^n} g_2(vt) \cdot f(v) dv.$$

Т.к. $h(v) \geq 0$, то для всех моментов времени имеет место система неравенств $K_1(t) \leq K(t) \leq K_2(t)$ и, кроме того,

$$\lim_{t \rightarrow \infty} (K_2(t) - K_1(t)) = (\overline{g_2} - \overline{g_1}) \int_{\mathbf{R}^n} f(v) dv < \varepsilon \int_{\mathbf{R}^n} f(v) dv.$$

Следовательно, получаем оценку

$$\begin{aligned} & |K(t) - \overline{g} \int_{\mathbf{R}^n} f(v) dv| \leq \\ & \leq |K(t) - K_1(t)| + |K_1(t) - \overline{g_1} \int_{\mathbf{R}^n} f(v) dv| + |\overline{g_2} - \overline{g_1}| \int_{\mathbf{R}^n} f(v) dv \leq \\ & \leq \text{const} \cdot \varepsilon. \end{aligned}$$

Теорема 1.3 доказана. ▲

Таким образом, формулировка теоремы 1.2 в части интегрируемой по Лебегу функции может быть изменена так, чтобы включить случай, когда $\rho(x, v) = \delta(x) f(v)$, где только $f(v)$ является интегрируемой по Лебегу функцией. Заметим, что для ПФР с дельта-распределением по скоростям $\rho(x, v) = f(x) \delta(v)$ утверждение теоремы 1.2 уже не имеет места.

Смысл проводимого анализа слабой сходимости решений уравнения Лиувилля состоит в том, что традиционная трактовка стремления замкнутой физической системы к термодинамическому равновесию обычно связывается с ростом энтропии и с уравнением Больцмана. Важно понимать, что некоторые равновесные свойства могут быть присущи и идеализированным системам без взаимодействия.

1.2. Цепочка Боголюбова

Программа вывода кинетических уравнений, описывающих выделенный комплекс из S частиц среди полного их числа N и уравнений эволюции средних значений динамических величин из так называемых «первых принципов», т.е. из микроскопических уравнений движения в лагранжевой или гамильтоновой формах, была сформулирована в основополагающих работах Н.Н. Боголюбова [5, 6] в 1946–1948 гг. В этих работах для систем классических частиц-молекул с парным потенциалом взаимодействия, зависящим от разности координат, был построен аппарат S -частичных функций распределения, удовлетворяющих эволюционному уравнению, называемому цепочкой (или иерархией) ББГКИ (Боголюбов–Борн–Грин–Кирквуд–Ивон). В зависимости от способа обрыва цепочки на каком-либо уровне (т.е. при расцеплении корреляций высших порядков через низшие), из нее можно получить замкнутые уравнения относительно функций распределения. В частности, расцепление бинарной корреляционной функции позволяет получить кинетические уравнения Больцмана и Власова.

Методы решений конкретных кинетических уравнений имеют индивидуальные особенности, также как и получаемые из этих уравнений макроскопические законы сохранения (гидродинамика). Цепочка же является эффективным инструментом для построения и исследования кинетических и гидродинамических уравнений в целом.

Ниже кратко излагаются основные положения функциональной гипотезы Боголюбова [6], обобщенной на случай систем с многократным взаимодействием, зависящим от координат и импульсов частиц в декартовом фазовом пространстве и описывается метод, позволяющий исследовать общую структуру уравнений гидродинамики. В последующих главах этот метод будет применяться для конструирования моделей временных рядов, удовлетворяющих определенным свойствам.

Одной из центральных задач статистической механики является нахождение коэффициентов переноса, связывающих в линейной теории термодинамические потоки с градиентами параметров, характеризующих конкретную систему. В качестве таких параметров, называемых гидродинамическими, обычно используются масса, полный импульс и полная энергия. Эти связи, будучи подставленными в законы сохранения, приводят к уравнениям вязкой гидродинамики и теплопроводности. Исходным

пунктом при выводе законов сохранения являются динамические уравнения в гамильтоновой или лагранжевой формах и основанное на них уравнение Лиувилля относительно функции распределения частиц системы в фазовом пространстве.

Рассматривается система из N тождественных частиц без внутренней структуры, заключенных в объеме V , гамильтониан которых представляется в виде суммы величин унарного, бинарного и т.д. типов, симметричных относительно перестановки индексов, нумерующих частицы:

$$H_N = \sum_{k=1}^K \sum_{1 \leq i_1 < \dots < i_k \leq N} h_k(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_k}). \quad (2.1)$$

Здесь $\mathbf{x}_i = (\mathbf{q}_i, \mathbf{p}_i)$ — фазовая координата i -ой частицы, $\mathbf{q}_i, \mathbf{p}_i \in \mathbf{R}^n$, n — размерность координатного пространства одной частицы.

Вводится функция распределения всей системы $F_N(x, t)$ как плотность вероятности нахождения системы в окрестности фазовой точки $x = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ в момент времени t . Ее эволюция определяется уравнением Лиувилля (1.12), которое в данном случае имеет вид

$$\frac{\partial F_N(x, t)}{\partial t} = [H_N, F_N], \quad (2.2)$$

где $[\cdot]$ — классическая собка Пуассона.

Вводятся также s -частичные функции распределения комплексов частиц, удовлетворяющих условиям нормировки:

$$F_s(x^s, t) = V^s \int F_N d\mathbf{x}_{s+1} \dots d\mathbf{x}_N, \quad x^s = (\mathbf{x}_1, \dots, \mathbf{x}_s);$$

$$\lim_{V \rightarrow \infty} \frac{1}{V} \int F_1(\mathbf{x}, t) d\mathbf{x} = 1, \quad F_s = \lim_{V \rightarrow \infty} \frac{1}{V} \int F_{s+1} d\mathbf{x}_{s+1}. \quad (2.3)$$

Интегралы берутся по фазовому пространству, отвечающему числу переменных интегрирования. На границе пространства импульсов функция распределения предполагается стремящейся к нулю быстрее любой степенной зависимости. При этих предположениях в так называемом термодинамическом пределе, когда число частиц и объем системы стремятся к бесконечности, но средняя плотность частиц остается конечной

$$N \rightarrow +\infty, \quad V \rightarrow +\infty, \quad \lim_{TD} \left(\frac{V}{N} \right) = \omega = const,$$

из (2.2), (2.3) следует уравнение эволюции s -частичной функции распределения, называемое цепочкой ББГКИ:

$$\frac{\partial F_s(x^s, t)}{\partial t} = [H_s, F_s] + \sum_{k=2}^K \sum_{m=1}^{k-1} \sum_{1 \leq i_1 < \dots < i_{k-1} \leq s} \frac{1}{m! \omega^m} \cdot \int [h_k(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{k-m}}, \mathbf{x}_{s+1}, \dots, \mathbf{x}_{s+m}), F_{s+m}(x^{s+m}, t)] d\mathbf{x}_{s+1} \dots d\mathbf{x}_{s+m}. \quad (2.4)$$

В частности, если взаимодействие между частицами описывается слагаемым бинарного типа $h_2(\mathbf{x}_i, \mathbf{x}_j) \equiv h(\mathbf{x}_i, \mathbf{x}_j)$, цепочка (2.4) принимает более простой вид:

$$\frac{\partial F_s(x^s, t)}{\partial t} = [H_s, F_s] + \frac{1}{\omega} \sum_{i=1}^s \int [h(\mathbf{x}_i, \mathbf{x}_{s+1}), F_{s+1}(x^{s+1}, t)] d\mathbf{x}_{s+1}. \quad (2.5)$$

Следует сделать замечание относительно корректности перехода к пределу $N \rightarrow \infty$ в цепочке (2.4). Дело в том, что формально в основе цепочки лежит уравнение Лиувилля для системы хоть и с большим, но все-таки конечным числом частиц. В пределе $N \rightarrow \infty$ не очевидно, что такое уравнение вообще может быть записано. Следовательно, требуется корректное определение плотности предельной меры и получении уравнения эволюции этой меры. Такое определение было дано в недавней работе В.В. Козлова и О.Г. Смолянова [7], где построена также и общая схема перехода от уравнения Лиувилля относительно мер на бесконечномерных пространствах к равносильной этому уравнению бесконечной системе уравнений относительно плотностей функций распределения на конечномерных пространствах в предположении, что функция Гамильтона является цилиндрической. Последнее означает, что эта функция измерима относительно борелевской σ -алгебры подмножеств симплектического локально-выпуклого фазового пространства системы. Мы не будем углубляться в теорию, развитую в [7], поскольку наша задача формулирую-

ется в терминах конечномерных пространств, но упомянуть об этой проблеме было необходимо для целостности концепции. В [7] показано, что решение бесконечномерного уравнения Лиувилля относительно мер является поточечным слабым пределом мерозначных функций, являющихся решениями уравнений Лиувилля относительно мер на конечномерных подпространствах исходного бесконечномерного пространства.

Таким образом, можно считать, что уравнение Лиувилля для плотности меры $d\Gamma$, выражающее закон сохранения нормировки, корректно, как и представления проекций на конечномерные подпространства. Тем самым определена и цепочка ББККИ в пределе бесконечного числа частиц, но конечной их средней плотности.

Чтобы получить из цепочки (2.4) практически полезную информацию, надо найти ее частные решения для определенного класса изучаемых систем. Один такой класс — это системы, близкие к устойчивому равновесию. При малом отклонении — случайном или вынужденном — от положения равновесия система релаксирует к этому же состоянию. Для описания релаксации вводится разложение по малому параметру, характеризующему величину отклонения функций распределения от равновесных.

Отметим, что если слагаемые в гамильтониане (2.1) инвариантны относительно сдвига по координатам, то тем же свойством обладают и стационарные решения цепочки (2.4), которые будем обозначать F_s^0 . В частности, F_1^0 однородна по $|\mathbf{q}_1|$, F_2^0 зависит только от модуля разности $|\mathbf{q}_{12}| = |\mathbf{q}_1 - \mathbf{q}_2|$, и т.д. Вблизи стационарного состояния функции распределения модифицируются: вводится параметр неоднородности μ , имеющий порядок отношения радиуса действия межмолекулярных сил к характерному размеру макроскопической неоднородности, и дополнительная модифицированная координата $\mathbf{r} = \mu\mathbf{q}_1$. Значение $\mu = 0$ отвечает пространственно-однородному состоянию. Функции распределения переопределяются следующим образом:

$$F_1(x_1, t) \mapsto \tilde{F}_1(\mathbf{r}, \mathbf{p}_1, t; \mu); \quad F_s(x^s, t) \mapsto \tilde{F}_s(\mathbf{r}, x^s, t; \mu), s \geq 2.$$

Далее для краткости тильду опускаем. Вместо (2.4) получаем модифицированную цепочку:

$$\begin{aligned} \frac{\partial F_s}{\partial t} + \mu \frac{\partial H_s}{\partial p_1^\alpha} \frac{\partial F_s}{\partial r^\alpha} = [H_s, F_s] + \sum_{k=2}^K \sum_{m=1}^{k-1} \sum_{1 \leq i_1 < \dots < i_{k-1} \leq s} \frac{1}{m! \omega^m} \cdot \\ \cdot \left\{ \left[h_k(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{k-1}}, \mathbf{x}_{s+1}, \dots, \mathbf{x}_{s+m}), F_{s+m} \right] - \right. \\ \left. - \mu \frac{\partial h_k}{\partial p_1^\alpha} \frac{\partial F_{s+m}}{\partial r^\alpha} \right\} d\mathbf{x}_{s+1} \dots d\mathbf{x}_{s+m}. \end{aligned} \quad (2.6)$$

Здесь обозначение скобки Пуассона сохранено для канонических переменных q, p (модифицированная координата $\mathbf{r} = \mu\mathbf{q}_1$ не является канонической переменной); по дважды повторяющимся индексам α подразумевается суммирование от 1 до n . Слагаемое с $\mu \frac{\partial h_k}{\partial p_1^\alpha} \frac{\partial F_s}{\partial r^\alpha}$ в (2.6)

отлично от нуля в том случае, если энергия взаимодействия зависит не только от координат, но и от импульсов частиц.

Первое уравнение модифицированной цепочки (2.6), т.е. уравнение для $s = 1$ имеет вид

$$\begin{aligned} \frac{\partial F_1}{\partial t} + \mu \frac{\partial}{\partial r^\alpha} (v_{eff}^\alpha F_1) = \\ = [H_1, F_1] + \sum_{k=2}^K \frac{1}{(k-1)! \omega^{k-1}} \int [h_k(x^k), F_k] d\mathbf{x}_2 \dots d\mathbf{x}_k, \quad (2.7) \\ v_{eff}^\alpha = \frac{\partial H_1}{\partial p_1^\alpha} + \sum_{k=2}^K \frac{1}{(k-1)! \omega^{k-1}} \frac{1}{F_1} \int \frac{\partial h_k(x^k)}{\partial p_1^\alpha} F_k d\mathbf{x}_2 \dots d\mathbf{x}_k. \end{aligned}$$

Введенная здесь эффективная скорость \mathbf{v}_{eff} есть аналог унарной динамической величины, среднее значение которой определяет среднюю локальную скорость частиц системы:

$$\begin{aligned} u^\alpha(\mathbf{r}, t) = \sum_{k=1}^K k \left\langle \frac{\partial h_k}{\partial p_1^\alpha} \right\rangle = \frac{1}{\omega \rho} \int v_{eff}^\alpha F_1 d\mathbf{p}_1, \quad (2.8) \\ \rho = \rho(\mathbf{r}, t) = \int F_1(\mathbf{r}, \mathbf{p}_1, t; \mu) d\mathbf{p}_1 \end{aligned}$$

Угловые скобки в (2.8) обозначают усреднение данной физической величины по мере с плотностью F_k в фазовом пространстве соответствующей размерности, причем по фазовому пространству первой частицы интеграл берется только по импульсу, а не по пространственной координате. Если проинтегрировать и по ней, то получится не плотность величины, а ее среднее значение по распределению в текущий момент времени.

Следующее предположение состоит в том, что функции $F_s(\mathbf{r}, x^s, t; \mu)$ считаются асимптотически регулярными по параметру μ в окрестности нуля:

$$F_s(\mathbf{r}, x^s, t; \mu) = F_s^0(\mathbf{r}, x^s, t) + \mu F_s^1(\mathbf{r}, x^s, t) + \dots + \mu^k F_s^k(\mathbf{r}, x^s, t) + \dots, \quad (2.9)$$

а функции распределения нулевого приближения $F_s^0(\mathbf{r}, x^s, t)$ определяются как локально-равновесные решения стационарной цепочки (2.6), куда \mathbf{r} и t входят как параметры, причем только через посредство гидродинамических переменных, т.е. через моменты нулевого, первого и второго порядков по импульсам одночастичной функции распределения. Подстановка разложения (2.9) в цепочку (2.7) и приравнивание членов при одинаковых степенях параметра μ приводит к методу последовательных приближений для нахождения нестационарных функций распределения. Рассмотрим для примера вывод уравнений для определения $F_2^1(\mathbf{r}, x^2, t) = F_2^1(\mathbf{r}, x^2; [F_1])$ в классическом случае парного взаимодействия, зависящего только от модуля разности пространственных координат.

В гидродинамике используются следующие гидродинамические переменные (далее HDP), обозначаемые ρ^v :

$$\begin{aligned} \rho^v, v = 1, \dots, n+2: \\ \rho^1 = \rho \equiv \rho \langle 1 \rangle, \quad \rho^{\alpha+1} = \rho \langle p^\alpha \rangle \equiv \rho p^{*\alpha}, \\ \rho^{n+2} = \rho \langle \tilde{H}_N \rangle \equiv \rho \varepsilon. \end{aligned} \quad (2.10)$$

Здесь n — размерность одночастичного координатного пространства. Тильда у гамильтониана означает, что он записан в системе отсчета,

движущейся со скоростью, отвечающей среднему импульсу \mathbf{p}^* . Согласно функциональной гипотезе, нулевое приближение определяется локально-равновесными функциями распределения, которые зависят от фазовых переменных через аддитивные интегралы движения, а интенсивные параметры, сопряженные этим инвариантам (далее DFP, параметры функций распределения), зависят от \mathbf{r} и t . Обычно DFP — это плотность, импульс, температура, обозначаемые далее π^v , которые затем связываются с HDP (2.10) посредством уравнений

$$\rho^v = \rho^v \Big|_{F_s^0[\pi]}; \quad \rho^v \Big|_{F_s^k} = 0, \quad k \geq 1. \quad (2.11)$$

Для согласования локально-равновесного приближения с макроскопическими средними (т.е. наблюдаемыми параметрами) требуется невырожденность матрицы

$$Q_v^\mu = \frac{\partial \rho^\mu}{\partial \pi^v}. \quad (2.12)$$

Кроме того, т.к. уравнения гидродинамики связывают в локально-равновесном приближении временные производные HDP с градиентами DFP, то уже в первом приближении по параметру неоднородности для получения замкнутой системы надо иметь возможность записывать эти уравнения в единых переменных, что требует невырожденности гидродинамических уравнений.

Локально-равновесные функции распределения классической системы имеют вид

$$F_s^0 = (\omega \rho)^s \left(\frac{\beta}{2\pi} \right)^{ns/2} g_s(q^s) \prod_{k=1}^s \exp \left(-\beta (\mathbf{p}_k - \mathbf{u})^2 / 2 \right), \quad (2.13)$$

где ρ, \mathbf{u}, β — функции от \mathbf{r} и t (это и есть DFP π^v), а g_s — равновесная координатная функция распределения, зависящая от β, ρ как от параметров; $g_1 \equiv 1, g_2 = g(|\mathbf{q}_{12}|)$. Будем бинарную координатную обозначать просто $g(q)$, без нижних индексов. Согласно (2.11), имеем

$$\mathbf{p}^* = \mathbf{u}, \quad \varepsilon = \frac{1}{2\beta} (n + U(\rho, \beta)), \quad (2.14)$$

$$U(\rho, \beta) = \beta \rho \int \Phi(q) g(q) dq.$$

Из цепочки (2.6) следует, что уравнения эволюции величин HDP (2.14) имеют вид:

$$\begin{aligned} \partial_t \rho &= -\nabla_\alpha \rho u^\alpha, \\ \partial_t u^\alpha &= -u^\beta \nabla_\beta u^\alpha - \frac{1}{\rho} \nabla_\beta \rho P^{\alpha\beta}, \\ \partial_t \varepsilon &= -u^\alpha \nabla_\alpha \varepsilon - \frac{1}{\rho} \nabla_\alpha \rho J^\alpha - P^{\alpha\beta} \nabla_\alpha u^\beta. \end{aligned} \quad (2.15)$$

Потоки $P^{\alpha\beta}$ и J^α определяются формулами

$$\begin{aligned} P^{\alpha\beta} &= \langle p^\alpha p^\beta \rangle - u^\alpha u^\beta - \left\langle q^\beta \frac{\partial \Phi(q)}{\partial q^\alpha} \right\rangle, \\ J^\alpha &= \frac{1}{2} \langle p_1^\alpha (\mathbf{p} - \mathbf{u})^2 \rangle + \langle p_1^\alpha \Phi \rangle - u^\alpha \varepsilon - \\ &\quad - \left\langle (p_1^\beta - u^\beta) q^\alpha \frac{\partial \Phi(q)}{\partial q^\beta} \right\rangle. \end{aligned} \quad (2.16)$$

В локально-равновесном приближении (2.13) выражения для потоков (2.16) имеют вид:

$$\begin{aligned} P_0^{\alpha\beta} &= \frac{1}{\beta} \delta^{\alpha\beta} (1 - G(\rho, \beta)), \quad J_0^\alpha = 0; \\ G(\rho, \beta) &= \frac{\beta \rho}{2n} \int q \Phi'(q) g(q) dq. \end{aligned} \quad (2.17)$$

Далее ищутся те решения уравнений цепочки (2.6), которые зависят от времени и параметра неоднородности только через посредство гидродинамических переменных. Поскольку правая часть уравнений (2.15) уже имеет порядок $O(\mu)$, то в иерархии (2.6) производные в левых частях этих уравнений следует выразить через градиенты HDP согласно найден-

ным законам (2.15) и нулевого (т.е. локально-равновесного) приближения (2.13). Например, для одночастичной функции распределения имеем

$$\begin{aligned} \partial_t F_1^0 &= F_1^0 \sigma^\nu \partial_t \pi^\nu = O(\mu), \\ \sigma^\nu &= \frac{\partial \ln F_1^0}{\partial \pi^\nu} = \left(1, \beta \mathbf{w}, \frac{n - \beta w^2}{2\beta} \right), \quad \mathbf{w} = \mathbf{p} - \mathbf{u}. \end{aligned} \quad (2.18)$$

Подставляя разложение (2.9) для F_1 в уравнение цепочки (2.6) и приравнявая в нем члены порядка $O(\mu)$, получаем с учетом (2.18) уравнение для поправки F_2^1 к бинарной локально-равновесной функции распределения:

$$\begin{aligned} \sigma^\nu \partial_t \pi^\nu + \left(\sigma^\nu v_{eff}^\alpha + \frac{\partial v_{eff}^\alpha}{\partial \pi^\nu} \right) \nabla_\alpha \pi^\nu &= \\ &= \frac{1}{\omega F_1^0} \int [\Phi_{12}, F_2^1] d\mathbf{x}_2 \end{aligned} \quad (2.19)$$

Если теперь выразить производные по времени от DFP через градиенты HDP согласно (2.15), то поправку к локально-равновесному распределению F_2^1 первого порядка по неоднородности можно представить в виде разложения по градиентам DFP. В рассматриваемом случае уравнения эволюции HDP (2.15) в терминах DFP имеют вид

$$\partial_t \pi^\nu = \Lambda_\alpha^{\nu\mu} \nabla^\alpha \pi_\mu, \quad (2.20)$$

где коэффициенты $\Lambda_\alpha^{\mu\nu}$ зависят от π^ν и объединяются в блочную матрицу:

$$\Lambda_{cl} = \left\| \begin{array}{ccc} -\mathbf{u} & -\hat{\mathbf{I}} & \mathbf{0} \\ \hat{\mathbf{I}} \frac{1}{\beta} \left(\frac{\partial \rho G}{\partial \rho} - 1 \right) & \mathbf{u} \otimes \hat{\mathbf{I}} & \hat{\mathbf{I}} \left(\frac{1}{\beta^2} + \frac{\partial (G/\beta)}{\partial \beta} \right) \\ \mathbf{0} & \hat{\mathbf{I}} \beta \frac{\rho \partial U / \partial \rho + 2(G-1)}{2\beta \partial U / \partial \beta - U - n} & \mathbf{u} \end{array} \right\|, \quad (2.21)$$

где $\mathbf{0}$ — строка длины n , а $\hat{\mathbf{I}}$ — единичная матрица $n \times n$. После этого ищем F_2^1 в виде

$$F_2^1 = F_2^0 X_\nu^\alpha \nabla_\alpha \pi^\nu. \quad (2.22)$$

Подставляя (2.20)–(2.22) в (2.19), получаем уравнение относительно неизвестных коэффициентов разложения X_ν^α :

$$\left(\Lambda_{\alpha}^{\nu\mu} + v_{eff}^{\alpha} \delta^{\nu\mu} \right) \mathcal{J}_{\nu} + \frac{\partial v_{eff}^{\alpha}}{\partial \pi_{\mu}} \equiv Y_{\mu}^{\alpha} = \frac{1}{\omega F_1^0} \int d\mathbf{x}_2 \left[\Phi_{12}, F_2^0 X_{\mu}^{\alpha} \right]. \quad (2.23)$$

Левые части уравнений (2.23) для рассматриваемого случая имеют следующий вид:

$$Y_1^{\alpha} = w^{\alpha} \frac{\partial \rho G}{\partial \rho}, \quad Y_3^{\alpha} = \frac{w^{\alpha}}{2\beta} \left(n + 2 - \beta w^2 + \beta^2 \frac{\partial}{\partial \beta} \frac{G}{\beta} \right), \quad (2.24)$$

$$Y_2^{\alpha\gamma} = \beta w^{\alpha} w^{\gamma} + \delta^{\alpha\gamma} \left(-1 + \frac{1}{2} (n - \beta w^2) \frac{\rho \partial U / \partial \rho + 2(G-1)}{2\beta \partial U / \partial \beta - U - n} \right).$$

Уравнения (2.23) представляют собой интегральные уравнения первого рода, и в этом случае решение может не быть определено однозначно. Действительно, удастся найти только нечетную по координате \mathbf{q}_{12} компоненту функций X_{μ}^{α} , а четная компонента ортогональна интегральному оператору в (2.23). Характерно, что для определения потоков в первом приближении нужна именно четная часть X_{μ}^{α} , которая не определяется из уравнений цепочки. Нечетная же часть X_{μ}^{α} нужна для определения потоков во втором приближении, а для первого приближения она бесполезна, ибо обращает в ноль интегралы в (2.17). Таким образом, хотя дифференциальная структура уравнений эволюции средних величин может быть определена вышеописанным методом, сами коэффициенты переноса не определяются непосредственно из цепочки Боголюбова. Для их нахождения надо сначала получить из цепочки замкнутое кинетическое уравнение относительно одночастичной функции распределения, а затем в этом

последнем применить разложение потоков по градиентам макроскопических параметров.

Решения уравнений (2.23) относительно нечетной части X_{μ}^{α} ищутся стандартным методом [3, 8] путем разложения неизвестных функций X_{μ}^{α} в ряды по полиномам Сонина–Лагерра, аргументом которых является безразмерная относительная энергия βw^2 . После того, как X_{μ}^{α} найдены (хотя бы в виде рядов), подстановка (2.22) в (2.16) с учетом условий (2.11) дает второе приближение для потоков, и уравнения гидродинамики (2.15) дополняются членами второго порядка по градиентам NDP. К сожалению, коэффициенты переноса — диффузии, вязкости и теплопроводности — не могут быть определены этим методом, поскольку, как уже говорилось, из уравнений цепочки в первом приближении находятся лишь нечетные части функций X_{μ}^{α} .

Таким образом, подробно описанная в этом параграфе функциональная гипотеза и разложение функций распределения в окрестности равновесного состояния не вносят диссипацию в обратимые по времени уравнения — Лиувилля и эволюции моментов. Этим методом можно получить лишь невязкую эйлерову гидродинамику. Уравнения же гидродинамики с необратимыми потоками — уравнения Навье–Стокса, диффузии и теплопроводности — надо получать из других моделей. Эти модели сами могут быть выведены из цепочки путем обрыва (замыкания) ее на каком-либо уравнении. Примеры замыкающих условий рассматриваются в следующих двух разделах. Тем не менее, все методы построения эволюционных уравнений для многомерных плотностей функций распределения опираются на цепочку, которая является теоретическим базисом вывода математических моделей макроскопических процессов из определенных предположений вероятностного характера относительно микроскопической динамики.

1.3. Уравнение Больцмана

Рассмотрим первое уравнение цепочки Боголюбова (2.5) для случая парного потенциала, зависящего только от модуля разности координат положения частиц. Оно имеет вид

$$\frac{\partial F_1(q, p, t)}{\partial t} + \frac{p}{m} \frac{\partial F_1(q, p, t)}{\partial q} = \frac{1}{\omega} \int [h(|\mathbf{q} - \mathbf{q}'|), F_2(x, x', t)] d\mathbf{x}'.$$

Интеграл со скобкой Пуассона преобразуется следующим образом. Поскольку

$$[h(|\mathbf{q} - \mathbf{q}'|), F_2(x, x', t)] = \frac{\partial h}{\partial q} \frac{\partial F_2}{\partial p} + \frac{\partial h}{\partial q'} \frac{\partial F_2}{\partial p'},$$

и потенциал $h(q)$ не зависит от импульсов, то после интегрирования по штрихованному импульсу с учетом нулевого граничного условия для функции распределения получаем

$$\frac{\partial F_1(q, p, t)}{\partial t} + \frac{p}{m} \frac{\partial F_1(q, p, t)}{\partial q} = \frac{1}{\omega} \frac{\partial}{\partial p} \int \frac{\partial h(|\mathbf{q} - \mathbf{q}'|)}{\partial q} F_2(x, x', t) d\mathbf{x}'. \quad (3.1)$$

Исследуем, какие уравнения получаются при втором способе разложения функций распределения по малому параметру — не по величине отклонения от неоднородности, а по удельному объему $1/\omega$, причем разлагаются по нему только функции распределения комплексов частиц больше одного, так что

$$F_s(x, t) = F_s^0(x, t) + \frac{1}{\omega} F_s^1(x, t) + \dots + \frac{1}{\omega^k} F_s^k(x, t) + \dots, \quad s \geq 2. \quad (3.2)$$

Подстановка (3.2) в цепочку (2.5) дает процесс последовательных приближений для нахождения F_s путем приравнивая членов при одинаковых степенях $1/\omega$:

$$\begin{aligned} \frac{\partial F_s^0}{\partial t} &= [H_s, F_s^0]; \\ \frac{\partial F_s^1}{\partial t} &= [H_s, F_s^1] + \sum_{i=1}^s \int [h(|\mathbf{q}_i - \mathbf{q}_{s+1}|), F_{s+1}^0] d\mathbf{x}_{s+1}; \dots \end{aligned} \quad (3.3)$$

Этот подход применяется для получения кинетического уравнения, которым описываются разреженные газы. Кинетическим уравнением называется эволюционное уравнение первого порядка по времени относительно одночастичной плотности функции распределения, т.е. всякое уравнение вида

$$\frac{\partial F_1(q, p, t)}{\partial t} = A(q, p; F_1). \quad (3.4)$$

Его специфика проявляется в том, что зависимость от времени входит в правую часть уравнения только через посредство самой функции F_1 . Поскольку же в силу зацепляющегося характера уравнений цепочки правая часть зависит от бинарных, тернарных и т.д. распределений, то аналогичное функциональное предположение следует сделать и для этих распределений: их зависимость от времени также определяется только функционалом от F_1 .

Рассмотрим метод последовательных приближений для оператора $A(q, p; F_1)$ из правой части (3.4): пусть

$$A(x; F_1) = A_0(x; F_1) + \frac{1}{\omega} A_1(x; F_1) + \frac{1}{\omega^2} A_2(x; F_1) + \dots \quad (3.5)$$

Чтобы определить явный вид операторов $A_k(x; F_1)$, подставим разложение (3.2) в уравнение цепочки (3.1) и приравняем величины, стоящие при одинаковых степенях $1/\omega$. Получаем:

$$\begin{aligned} A_0(x; F_1) &= [H_0(p), F_1] = -\frac{\partial H_0}{\partial p} \frac{\partial F_1}{\partial q}; \\ A_1(x; F_1) &= \int \frac{\partial h(|\mathbf{q} - \mathbf{q}'|)}{\partial q} \frac{\partial F_2^0(x, x'; F_1)}{\partial p} d\mathbf{x}'; \dots; \\ A_k(x; F_1) &= \int \frac{\partial h(|\mathbf{q} - \mathbf{q}'|)}{\partial q} \frac{\partial F_2^{k-1}(x, x'; F_1)}{\partial p} d\mathbf{x}'; \dots \end{aligned} \quad (3.6)$$

Таким образом, в нулевом приближении получаем уравнение эволюции ансамбля из одной частицы, на которую остальные частицы не действуют, т.е. уравнение Лиувилля для газа невзаимодействующих частиц. В следующем приближении эволюция F_1 определяется функцией распределения $F_2^0(x, x'; F_1)$ нулевого приближения для комплекса из двух частиц и т.д. Поскольку $F_2(x, x', t)$ удовлетворяет второму уравнению цепочки Боголюбова, которое для рассматриваемого случая парного потенциала взаимодействия имеет вид

$$\frac{\partial F_2(x^2, t)}{\partial t} = [H_2, F_2] + \frac{1}{\omega} \int \left(\frac{\partial h(|\mathbf{q}_1 - \mathbf{q}_3|)}{\partial q_1} \frac{\partial F_3(x^3, t)}{\partial p_1} + \frac{\partial h(|\mathbf{q}_2 - \mathbf{q}_3|)}{\partial q_2} \frac{\partial F_3(x^3, t)}{\partial p_2} \right) d\mathbf{x}_3, \quad (3.7)$$

то в нулевом приближении по $1/\omega$ левая часть (3.6) выражается через производную F_2^0 по F_1 с последующей заменой производной $\partial F_1 / \partial t$ на оператор $A_0(x; F_1)$ в соответствии с предположением о функциональной зависимости $F_2^0(x, x'; F_1)$. Обозначим оператор такой сложной производной по времени в нулевом приближении по $1/\omega$ через D_0 . Тогда

$$D_0 F_s^0 = [H_s, F_s^0], \quad s \geq 2. \quad (3.8)$$

Условие, которое налагается дополнительно на все частные решения цепочки рассматриваемого функционального вида, состоит в независимости распределений частиц в бесконечно удаленном прошлом. Это так называемое условие ослабления корреляций. Оно и является недостающим по сравнению с цепочкой условием, которое вносит необратимость в кинетическую теорию. Опишем соответствующую математическую конструкцию, следуя [5].

Если имеется динамический поток G_t (1.1), то для любой дифференцируемой функции F динамических переменных и времени выполняются уравнения Гамильтона

$$\frac{\partial}{\partial t} G_{-t} F(x^s, t) = [H_s, G_{-t} F(x^s, t)],$$

причем $G_t F(x(0)) = F(x(t))$. Сравнивая это выражение с (3.8), получаем

$$D_0 F_s^0(x^s; G_{-t} F_1) = \frac{\partial}{\partial t} F_s^0(x^s; G_{-t} F_1). \quad (3.9)$$

В терминах оператора динамического потока условие ослабления корреляций означает, что функции распределения нулевого приближения факторизуются в бесконечно удаленном прошлом, а функции распределе-

ния в следующих приближениях стремятся к нулю в силу сохранения нормировки. Такая функциональная гипотеза записывается в виде предельных соотношений на F_s при $s \geq 2$:

$$G_{-t} F_s^0(x^s, G_t F_1) - \prod_{i=1}^s (G_t F_1) \rightarrow 0, \quad t \rightarrow +\infty; \quad (3.10)$$

$$G_{-t} F_s^1(x^s, G_t F_1) \rightarrow 0, \quad t \rightarrow +\infty.$$

Поскольку тождественно выполняется равенство

$$F_s^0(x^s; F_1) = G_{-\tau} F_s^0(x^s; G_{\tau} F_1),$$

где левая часть не зависит явно от τ , то, переходя к пределу $\tau \rightarrow +\infty$, получаем с учетом первого из условий (3.10)

$$F_s^0(x^s; F_1) = \lim_{\tau \rightarrow +\infty} G_{-\tau} \prod_{i=1}^s (G_{\tau} F_1(x_i)). \quad (3.11)$$

В терминах оператора потока формальное решение уравнений (3.3) имеет вид

$$F_s^1(x^s, t) = G_{-t} F_s^1(x^s, 0) + \int_0^t G_{-(t-\tau)} \sum_{i=1}^s [h(|\mathbf{q}_i - \mathbf{q}_{s+1}|), F_{s+1}^0(x^s, \tau)] d\mathbf{x}_{s+1} d\tau. \quad (3.12)$$

Как следует из (3.6), в первом приближении кинетического уравнения достаточно знать s -частичные функции распределения в нулевом приближении, т.е. решения вида (3.11). Координаты в бесконечно удаленном прошлом трактуются с физической точки зрения как координаты «до столкновения».

Дальнейшие преобразования уравнения, получающегося в первом приближении по $1/\omega$, связаны с механической задачей разлета двух тел в поле потенциала $h(q)$. Если бы мы рассматривали тройной потенциал, то возникли бы тройные столкновения, и т.д. Вводится дифференциальное сечение рассеяния $\sigma(w, \cos \theta)$, где $w = |\mathbf{p}_1 - \mathbf{p}_2|/m$ есть модуль относительной скорости частиц до столкновения, а θ — угол между относительными скоростями до и после столкновения. Поскольку в уравнении фигурируют далее только унарные функции распределения частиц, то

нижний индекс будем у них опускать. Обозначим импульсы частиц после столкновения штрихами. Тогда получаем уравнение

$$\begin{aligned} & \frac{\partial F(q_1, p_1, t)}{\partial t} + p_1^\alpha \frac{\partial F(q_1, p_1, t)}{\partial q_1^\alpha} = \\ & = \int_{R^n \times S^{n-1}} w \sigma(w, \cos \theta) [F(q_1, p'_1, t) F(q_1, p'_2, t) - \\ & - F(q_1, p_1, t) F(q_1, p_2, t)] d\mathbf{p}_2 d\mathbf{\Omega}. \end{aligned} \quad (3.13)$$

Импульсы после столкновения находятся из законов сохранения импульса и энергии с точностью до произвольного направления, единичный вектор которого обозначен в (3.13) через $\mathbf{\Omega}$. Интегрирование по $d\mathbf{\Omega}$ проводится по $n-1$ -мерной, а не n -мерной сфере, потому что модуль $\mathbf{\Omega}$ фиксирован. Пусть в общем случае массы сталкивающихся частиц различны, что актуально для смесей газов. Тогда вместо одного уравнения (3.13) получаем систему аналогичных уравнений на унарные функции распределения частиц разных сортов. В этих уравнениях импульсы после столкновения выражаются через импульсы до столкновения следующим образом:

$$\begin{aligned} \mathbf{p}'_1 &= \frac{\mathbf{p}_1 + \mathbf{p}_2}{2} (1 + \delta) + w \cos \theta \mathbf{\Omega}, \\ \mathbf{p}'_2 &= \frac{\mathbf{p}_1 + \mathbf{p}_2}{2} (1 - \delta) - w \cos \theta \mathbf{\Omega}, \\ \delta &= \frac{m_1 - m_2}{m_1 + m_2}, \quad \mathbf{\Omega} \in R^n, \quad |\mathbf{\Omega}| = 1. \end{aligned} \quad (3.14)$$

Уравнение (3.13) называется уравнением Больцмана. Опишем кратко его основные свойства.

Равновесное решение. Если рассмотреть стационарный случай уравнения (3.13), причем предположить, что функция распределения однородна по пространственным координатам, то левая часть этого уравнения станет равной нулю. Следовательно, такое однородное стационарное решение, если оно существует, должно занулять правую часть уравнения (3.13), называемую столкновительным членом. Такое решение называется равновесным. Введем обозначение для столкновительного члена

$$I[F, F] = \int_{R^n \times S^{n-1}} w \sigma(w, \cos \theta) [F(q_1, p'_1, t) F(q_1, p'_2, t) - F(q_1, p_1, t) F(q_1, p_2, t)] d\mathbf{p}_2 d\mathbf{\Omega}. \quad (3.15)$$

Равновесное решение $F(q_1, p_1, t) = F_{eq}(p_1)$ должно, таким образом, обладать свойством

$$I[F_{eq}, F_{eq}] = \int_{R^n \times S^{n-1}} w \sigma(w, \cos \theta) [F_{eq}(p'_1) F_{eq}(p'_2) - F_{eq}(p_1) F_{eq}(p_2)] d\mathbf{p}_2 d\mathbf{\Omega} = 0.$$

Следовательно, можно предложить такое частное решение этого уравнения, когда логарифм равновесной функции распределения является функцией, сохраняющейся в процессе механического столкновения двух тел. В таком случае функция распределения является экспонентой от инвариантов столкновения, причем интеграл от нее по импульсам должен сходиться. Этому условию удовлетворяет распределение Максвелла:

$$F_{eq}(p) = \left(\frac{\beta}{2\pi m} \right)^{n/2} \exp \left(-\beta \frac{p^2}{2m} \right), \quad (3.16)$$

где β есть постоянный неотрицательный параметр, называемый фактором Гиббса (обратная температура, выраженная в энергетических единицах).

Законы сохранения. Из (3.13) следует, что если ввести гидродинамические переменные ρ^V , аналогичные (2.10), как средние значения (т.е. HDP) по унарному распределению от следующих динамических величин

$$\begin{aligned} \rho^1 &= \rho \equiv \rho \langle 1 \rangle, \quad \rho^{\alpha+1} = \rho \langle p^\alpha \rangle \equiv \rho p^{*\alpha}, \\ \rho^{n+2} &= \frac{\rho \langle (p - p^*)^2 \rangle}{2m} \equiv \rho \varepsilon, \quad \alpha = 1, \dots, n, \end{aligned} \quad (3.17)$$

то эволюция этих средних значений представляется в дивергентной форме, т.е. имеет вид законов сохранения. Это связано с тем, что суммарно по паре частиц величины $1, \mathbf{p}$ и p^2/m сохраняются в процессе столкновения, так что интеграл от столкновительного члена по импульсам оказывается равным нулю. Рассмотрим величину

$$g(q_1, t) = \int F(q_1, p_1, t) \psi(p_1) d\mathbf{p}_1.$$

В силу (3.13) для этой величины получаем уравнение эволюции

$$\frac{\partial g(q_1, t)}{\partial t} + \frac{\partial}{\partial q_1^\alpha} \rho \left\langle \rho p_1^\alpha \psi(p_1) \right\rangle = \int I[F, F] \psi(p_1) d\mathbf{p}_1. \quad (3.18)$$

С использованием симметрии подынтегрального выражения относительно перенумерации частиц последний интеграл может быть преобразован к виду

$$\frac{1}{4} \int I[F, F] (\psi(p_1) + \psi(p_2) - \psi(p'_1) - \psi(p'_2)) d\mathbf{p}_1.$$

Написанное выражение обращается в ноль, если $\psi(p_1) + \psi(p_2) = \text{const}$. Следовательно, для величин, сохраняющихся в процессе бинарного соударения, интеграл от столкновительного члена равен нулю.

Нетривиальным фактом является справедливость и обратного утверждения: те и только те функционалы, называемые сумматорными инвариантами уравнения Больцмана, обращают в ноль интеграл от столкновительного члена (3.15), которые представляются в виде линейной комбинации HDP (3.17).

H-теорема. Для уравнения Больцмана существует убывающий функционал, называемый *H-функцией* Больцмана:

$$H = \int f \ln f dq dp, \quad \frac{dH}{dt} \leq 0. \quad (3.19)$$

Опуская для краткости у функций распределения все аргументы, кроме импульсов, получаем из (3.13) с учетом симметрии столкновительного члена по номерам частиц:

$$\begin{aligned} \frac{dH}{dt} &= \frac{1}{4} \int I[F, F] \ln \left(\frac{F(q_1, p_1, t) F(q_1, p_2, t)}{F(q_1, p'_1, t) F(q_1, p'_2, t)} \right) d\mathbf{p}_1 d\mathbf{q}_1 = \\ &= \frac{1}{4} \int w \sigma(w, \cos \theta) \ln \left(\frac{F(p_1) F(p_2)}{F(p'_1) F(p'_2)} \right) \times \\ &\times [F(p'_1) F(p'_2) - F(p_1) F(p_2)] d\mathbf{p}_1 d\mathbf{p}_2 d\Omega d\mathbf{q}_1 \leq 0. \end{aligned}$$

Последнее неравенство следует из того алгебраического факта, что для любых двух положительных вещественных чисел x, y выполняется

неравенство $(x - y) \ln \left(\frac{y}{x} \right) \leq 0$. Знак равенства имеет место только в случае $x = y$.

Величина $S = -H$ называется энтропией. Наличие убывающего функционала означает существование устойчивого предельного состояния при $t \rightarrow +\infty$, при котором функция распределения оказывается равновесной, т.к. условие $\frac{dH}{dt} = 0$ означает равенство произведений функ-

ций распределения до и после столкновения, что приводит к выражению (3.16). Следовательно, уравнение Больцмана описывает необратимые по времени процессы, причем эти процессы направлены к равновесному максвелловскому состоянию.

Уравнение переноса. Это уравнение, следующее из уравнения Больцмана для смеси двух газов. Пусть f есть плотность функции распределения (ПФР) первой компоненты, а F — второй. Тогда вместо одного уравнения (3.13) имеем систему из двух уравнений

$$\begin{aligned} \frac{\partial f}{\partial t} + \frac{p_1}{m_1} \frac{\partial f}{\partial q_1} &= I[f, f] + I[f, F]; \\ \frac{\partial F}{\partial t} + \frac{p_2}{m_2} \frac{\partial F}{\partial q_1} &= I[F, F] + I[F, f]. \end{aligned}$$

В приближении, когда распределение второй компоненты известно, а плотность первой компоненты мала, либо мало сечение столкновений, так что столкновениями молекул первого газа между собой можно пренебречь, остается одно уравнение, описывающее движение первого газа в среде, свойства которой известны:

$$\frac{\partial f}{\partial t} + \frac{p_1}{m_1} \frac{\partial f}{\partial q_1} = I[f, F]. \quad (3.20)$$

Это — линейное уравнение по ПФР $f(q, v, t)$. Оно имеет только один линейный закон сохранения — числа частиц. Рассмотрим для простоты пространственно-однородный случай, когда ПФР не зависит от q_1 .

Тогда при известном сечении рассеяния уравнение (3.20) удобно записать в виде

$$\begin{aligned} \frac{\partial f(v, t)}{\partial t} &= \int (K(v, u)f(u, t) - K(u, v)f(v, t)) du = \\ &= \int K(v, u)f(u, t) du - a(v)f(v, t), \end{aligned} \quad (3.21)$$

где положительное ядро рассеяния $K(v, u)$ определяется по сечению σ и по функции F .

Обозначим через $g(v)$ стационарное решение уравнения (3.21) и введем функционал

$$H_{KL}(f, g) = \int f(v) \ln \left[\frac{f(v)}{g(v)} \right] dv, \quad (3.22)$$

называемый расстоянием Кульбака–Лэйблера между положительными функциями f и g или относительной энтропией. Расстояние (3.22) не-симметрично относительно функций f и g , поэтому его следует называть «квазирасстоянием», но оно не отрицательно. Поскольку для всех $x > -1$ выполняется оценка $\ln(1+x) \leq x$ и равенство возможно лишь при $x = 0$, то

$$\ln \left(\frac{g}{f} \right) = \ln \left(1 + \left(\frac{g}{f} - 1 \right) \right) \leq \frac{g}{f} - 1.$$

Тогда

$$\int f(x) \ln \left(\frac{g(x)}{f(x)} \right) dx \leq \int f(x) \left(\frac{g(x)}{f(x)} - 1 \right) dx = \int (g(x) - f(x)) dx = 0.$$

Последнее равенство следует из того, что оба распределения f и g нормированы на единицу:

$$\int f(x) dx = \int g(x) dx = 1.$$

Таким образом, квазирасстояние (3.22) неотрицательно при любых ненулевых плотностях распределений вероятностей f и g .

Для функционала (3.22) существует аналог H -теоремы:

$$\begin{aligned} \frac{dH_{KL}}{dt} &= - \int K(v, u)f(u) (\ln z - 1 + 1/z) dudv \leq 0, \\ z &= \frac{f(u)g(v)}{f(v)g(u)}. \end{aligned} \quad (3.23)$$

Следовательно, в неравновесном состоянии функционал квазирасстояния между распределениями убывает в силу уравнения (3.21), т.е. текущее состояние стремится к равновесному.

Линеаризованное уравнение Больцмана и кинетические коэффициенты. Уравнение Больцмана (3.13) в окрестности равновесного максвелловского распределения может быть линеаризовано по величине отклонения от равновесия. Равновесное распределение выберем в виде (3.16), доопределив его множителем ρ :

$$F_{eq}(p) = f_0 = \rho \left(\frac{\beta}{2\pi m} \right)^{n/2} \exp \left(-\beta \frac{(p-u)^2}{2m} \right). \quad (3.24)$$

Связь между параметрами функции распределения π^v

$$\pi^1 = \rho, \quad \pi^{\alpha+1} = u^\alpha, \quad \pi^{n+2} = \beta, \quad \alpha = 1, \dots, n \quad (3.25)$$

и гидродинамическими переменными ρ^v (3.17) дается равенствами:

$$\begin{aligned} \rho^1 &= \rho = \pi^1, \quad \rho^{\alpha+1} = u^\alpha = \pi^{\alpha+1}, \\ \rho^{n+2} &= \frac{n}{2\beta} = \frac{n}{2\pi^{n+2}}. \end{aligned} \quad (3.26)$$

Матрица производных $Q_{\mu\nu} = \frac{\partial \rho^\mu}{\partial \pi^\nu}$ диагональна и невырождена.

Этот факт будет использован далее при нахождении коэффициентов переноса.

Будем искать решение уравнения (3.13) в виде

$$F(q, p, t) = f_0 \cdot (1 + \varphi(q, p, t)). \quad (3.27)$$

Тогда в линейном приближении по φ получаем уравнение

$$f_0(p_1) \left(\frac{\partial \varphi_1}{\partial t} + p_1^\alpha \frac{\partial \varphi_1}{\partial q_1^\alpha} \right) =$$

$$= \frac{1}{\omega} \int w \sigma(w, \cos \theta) f_0(p_1) f_0(p_2) (\varphi_1' + \varphi_2' - \varphi_1 - \varphi_2) d\mathbf{p}_2 d\mathbf{\Omega} \equiv$$

$$\equiv L[\varphi] . \quad (3.28)$$

Линейный оператор L является эрмитовым, причем все его собственные значения неположительны. Он имеет $n+2$ собственных функции, отвечающие нулевому собственному значению, так как в этом случае функции φ являются линейными комбинациями сумматорных инвариантов. Они имеют вид

$$\varphi^{(0)1} = 1; \quad \varphi^{(0)\alpha} = p^\alpha \sqrt{\beta/m};$$

$$\varphi^{(0)n+2} = \frac{1}{\sqrt{2n}} (\beta p^2 / m - n). \quad (3.29)$$

Линеаризованное уравнение Больцмана (3.28) используется для нахождения связи между термодинамическими потоками и градиентами GDP в линейном приближении. Соответствующие коэффициенты называются кинетическими коэффициентами переноса (вязкости и теплопроводности). Будем искать решение точного уравнения Больцмана (3.13) методом последовательных приближений

$$F = f_0 + f_1[f_0] + f_2[f_1] + \dots, \quad (3.30)$$

где f_0 есть локально-равновесное максвелловское распределение, а f_1 — неравновесная поправка, пропорциональная градиентам DFP, которые (градиенты), в свою очередь, пропорциональны градиентам GDP, как в методе Боголюбова.

Поскольку столкновительный член на локально-равновесном распределении обращается в ноль, то первое приближение в уравнении (3.13), получаемое подстановкой разложения (3.30) в (3.13), оказывается линейно по f_1 . Из соображений нормировки первую поправку (как и последующие) к локально-равновесному распределению удобно записывать в виде $f_1 = f_0 \varphi$. Тогда уравнение на φ имеет вид

$$\frac{\partial f_0}{\partial t} + p^\alpha \frac{\partial f_0}{\partial q^\alpha} = L[\varphi]. \quad (3.31)$$

Локально-равновесное распределение дается формулой (3.24), в которой DFP π^ν (3.25) зависят от координаты q и времени t . Для краткости положим массы равными единице, так что импульс и скорость численно совпадают. Поступая так же, как и при выводе уравнений гидродинамики из цепочки Боголюбова в п. 1.2, в первом приближении получаем GDP (3.26), уравнения эволюции которых имеют вид

$$\frac{\partial \rho}{\partial t} = - \frac{\partial}{\partial q^\alpha} (\rho u^\alpha),$$

$$\frac{\partial u^\alpha}{\partial t} = - u^\beta \frac{\partial u^\alpha}{\partial q^\beta} - \frac{1}{\rho} \frac{\partial}{\partial q^\beta} (\rho P^{\alpha\beta}); \quad (3.32)$$

$$\frac{\partial \varepsilon}{\partial t} = - u^\alpha \frac{\partial \varepsilon}{\partial q^\alpha} - P^{\alpha\beta} \frac{\partial u^\beta}{\partial q^\alpha} - \frac{1}{\rho} \frac{\partial}{\partial q^\alpha} (\rho J^\alpha)$$

Потоки в правых частях системы (3.32) определяются формулами:

$$P^{\alpha\beta} = \langle p^\alpha p^\beta \rangle - u^\alpha u^\beta; \quad J^\alpha = \frac{1}{2} \langle p^\alpha (p-u)^2 \rangle - u^\alpha \varepsilon. \quad (3.33)$$

В локально-равновесном приближении потоки (3.33) равны соответственно

$$P_0^{\alpha\beta} = \langle p^\alpha p^\beta \rangle_0 - u^\alpha u^\beta = \frac{\delta^{\alpha\beta}}{\beta};$$

$$J_0^\alpha = \frac{1}{2} \langle p^\alpha (p-u)^2 \rangle_0 - u^\alpha \varepsilon = 0 \quad (3.34)$$

Чтобы найти поправку φ из уравнения (3.31), выразим производные f_0 в его левой части через DFP:

$$\frac{\partial f_0}{\partial t} = \frac{\partial f_0}{\partial \pi^\nu} \frac{\partial \pi^\nu}{\partial t} = \frac{\partial f_0}{\partial \rho} \frac{\partial \rho}{\partial t} + \frac{\partial f_0}{\partial u^\alpha} \frac{\partial u^\alpha}{\partial t} + \frac{\partial f_0}{\partial \beta} \frac{\partial \beta}{\partial t} =$$

$$= \frac{\partial f_0}{\partial \rho} \frac{\partial \rho}{\partial t} + \frac{\partial f_0}{\partial u^\alpha} \frac{\partial u^\alpha}{\partial t} + \frac{\partial f_0}{\partial \beta} \frac{\partial \beta}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial t}.$$

Аналогично преобразуется и второе слагаемое в (3.31) $p^\alpha \frac{\partial f_0}{\partial q^\alpha}$.

Из (3.24) имеем

$$\begin{aligned} \frac{\partial f_0}{\partial \rho} &= \frac{f_0}{\rho}; \quad \frac{\partial f_0}{\partial u^\alpha} = -\beta(p^\alpha - u^\alpha)f_0; \\ \frac{\partial f_0}{\partial \beta} &= \frac{1}{2} \left(\frac{n}{\beta} - (\mathbf{p} - \mathbf{u})^2 \right) f_0 \end{aligned} \quad (3.35)$$

Поскольку $\varepsilon = n/(2\beta)$, то $\partial\beta/\partial\varepsilon = -n/(2\varepsilon^2)$. В результате левая часть (3.31) может быть выражена только через градиенты GDP в соответствии с выведенными уравнениями гидродинамики первого приближения (3.32). Вводя относительный импульс

$$w^\alpha = p^\alpha - u^\alpha,$$

получаем:

$$\begin{aligned} \frac{\partial f_0}{\partial t} + p^\alpha \frac{\partial f_0}{\partial q^\alpha} &= \\ &= \frac{f_0}{\rho} \left(\frac{\partial \rho}{\partial t} + p^\alpha \frac{\partial \rho}{\partial q^\alpha} \right) - \beta w^\alpha f_0 \left(\frac{\partial u^\alpha}{\partial t} + p^\beta \frac{\partial u^\alpha}{\partial q^\beta} \right) - \\ &- \frac{n}{4\varepsilon^2} \left(\frac{n}{\beta} - w^2 \right) f_0 \left(\frac{\partial \varepsilon}{\partial t} + p^\alpha \frac{\partial \varepsilon}{\partial q^\alpha} \right) = \\ &= f_0 \left(\left(-\frac{\beta w^2}{n} \operatorname{div} \mathbf{u} \right) + \frac{1}{2} \beta w^\alpha w^\beta \left(\frac{\partial u^\alpha}{\partial q^\beta} + \frac{\partial u^\beta}{\partial q^\alpha} \right) + \right. \\ &\quad \left. + \frac{1}{2} w^\alpha (n + 2 - \beta w^2) \frac{\partial \ln \beta}{\partial q^\alpha} \right). \end{aligned} \quad (3.36)$$

Последняя строка в (3.36) дает выражения для термодинамических сил, которые пропорциональны градиентам соответствующих параметров. Отметим, что градиент плотности в локально-равновесном приближении отсутствует.

После этого неизвестная функция φ из (3.31) ищется в виде разложения по градиентам величин в соответствии с представлением (3.36) с коэффициентами, зависящими от относительного импульса w^α . Эти коэф-

фициенты определяются затем в виде рядов по специальным многочленам Сонина–Лагерра [8]. Подстановка функции φ в уравнения для потоков (3.33) позволяет найти коэффициенты пропорциональности между потоками и градиентами GDP. Эти коэффициенты называются коэффициентами переноса. Таким образом, уравнение Больцмана позволяет в явном виде получить выражения для вязкого и теплового потоков. Найдем, например, коэффициент теплопроводности.

Будем искать φ в виде

$$\begin{aligned} \varphi &= X_1^\alpha \frac{\partial \ln \beta}{\partial q^\alpha} + X_2^{\alpha\beta} \left(\frac{\partial u^\alpha}{\partial q^\beta} - \frac{\partial u^\beta}{\partial q^\alpha} \right) + \\ &+ X_3^{\alpha\beta} \left(\frac{\partial u^\alpha}{\partial q^\beta} + \frac{\partial u^\beta}{\partial q^\alpha} - \frac{2}{n} \delta^{\alpha\beta} \operatorname{div} \mathbf{u} \right). \end{aligned} \quad (3.37)$$

Здесь величины X считаются зависящими только от относительного импульса \mathbf{w} . Безразмерной величиной является $z = \beta w^2/2$, так что, например, $X_1^\alpha = w^\alpha g_1(z)$, где g_1 есть искомая функция. Если она найдена, то

$$\begin{aligned} J_1^\alpha &= \frac{1}{2} \left\langle p^\alpha (p - u)^2 \right\rangle_1 = \\ &= \frac{1}{2\omega\rho} \int p^\alpha (p - u)^2 f_0(p) w^\alpha g_1(z) \frac{\partial \ln \beta}{\partial q^\alpha} dp = \kappa \frac{\partial \ln \beta}{\partial q^\alpha}, \\ \kappa &= \frac{1}{2\omega\rho} \int p^\alpha (p - u)^2 f_0(p) w^\alpha g_1(z) \rho \left(\frac{\beta}{2\pi} \right)^{n/2} \times \\ &\times \exp \left(-\beta \frac{(p - u)^2}{2} \right) dp = \\ &= \left(\frac{\beta}{2} \right)^{n/2} \frac{1}{\omega} \int_0^\infty \exp(-\beta w^2/2) w^{n+3} g_1(\beta w^2/2) dw = \\ &= \frac{2}{\beta^2 \omega} \int_0^\infty e^{-z} z^{n/2+1} g_1(z) dz. \end{aligned} \quad (3.38)$$

Поскольку законы сохранения удовлетворяются на локально-равновесном распределении, то первая поправка должна быть ортогональна сумматорным инвариантам:

$$\int X_1^\alpha f_0 \cdot \{1, p^\beta, p^2\} dp = 0. \quad (3.39)$$

Уравнение на g_1 решается с учетом условий (3.39) и имеет вид:

$$f_0 w^\alpha (1 + n/2 - z) = L[w^\alpha g_1(z)]. \quad (3.40)$$

Функция g_1 разлагается в ряд по ортогональным многочленам Сонина–Лагерра $S_\nu^k(z)$, которые определяются формулой

$$S_\nu^k(z) = \frac{1}{k!} e^z z^{-\nu} \frac{d^k}{dz^k} e^{-z} z^{\nu+k}.$$

В частности,

$$S_\nu^0 = 1, \quad S_\nu^1 = \nu + 1 - z, \quad S_\nu^2 = \frac{(\nu+1)(\nu+2)}{2} - (\nu+2)z + \frac{z^2}{2}.$$

Обобщенный порядок ν полиномов, по которым надо разлагать функцию g_1 , определяется видом левой части уравнения (3.40). Поскольку $1 + n/2 - z = S_{n/2}^1(z)$, то для нахождения теплового потока разложение следует вести по полиномам $S_{n/2}^k(z)$:

$$g_1(z) = \sum_{k=0}^{\infty} A_k S_{n/2}^k(z). \quad (3.41)$$

Используем далее условие ортогональности

$$\int_0^{\infty} e^{-z} z^\nu S_\nu^k(z) S_\nu^m(z) dz = \frac{\Gamma(\nu+k+1)}{k!} \delta_{km}. \quad (3.42)$$

Если подставить теперь (3.41) в последнее выражение в (3.38), получим с учетом условий ортогональности (3.39) и (3.42) выражение для коэффициента теплопроводности

$$\kappa = \frac{2A_1 \Gamma(2 + n/2)}{\beta^2 \omega}. \quad (3.43)$$

Сами коэффициенты A_k находятся из алгебраической системы, по-

лучаемой из (3.40) путем домножения его на w^α и интегрированием по импульсам.

Максвелловский потенциал. Отметим, что для степенных потенциалов $h(r) = 1/r^\alpha$ из анализа механики столкновений двух точечных тел следует $\sigma(w, \cos \theta) = w^{-4/\alpha} g(\cos \theta)$, где $g(\cos \theta)$ — функция, не зависящая от w . При $\alpha = 4$ индикатриса рассеяния $B = w\sigma(w, \cos \theta)$ не зависит от относительной скорости сталкивающихся частиц, что упрощает интеграл столкновений и позволяет в явном виде вычислить коэффициенты переноса (вязкость, теплопроводность), не прибегая к линеаризации. Соответствующий потенциал $h(r) = 1/r^4$ называется потенциалом максвелловских молекул. Однако следует заметить, что именно для таких сравнительно слабо убывающих потенциалов как функций расстояния уравнение Больцмана и не применимо, поскольку столкновение в этом случае не короткодействующее.

1.4. Уравнение Власова

Для дальнодействующих потенциалов предлагается другое уравнение, получаемое из цепочки при разложении функций распределения по третьему малому параметру — так называемой силы межчастичной связи или интенсивности взаимодействия, которое при средней плотности предполагается малым по сравнению с кинетической энергией тел. Получаемые таким путем уравнения в первом приближении описывают движение в самосогласованных полях. Эти уравнения применительно к задачам теории плазмы были введены А.А. Власовым [9].

Рассмотрим замыкание на первом уравнении цепочки Боголюбова, но обрыв будем делать не в приближении малой плотности, а по малому взаимодействию. Если между частицами нет взаимодействия или оно проявляется в виде точечного столкновения, то бинарная функция распределения факторизуется через унарные распределения. Если ввести парное взаимодействие со слабой константой связи, то можно разложить бинарную функцию распределения по степеням этой связи. Поскольку потенциал уже входит в правую интегральную часть цепочки, то в первом

приближении под интегралом можно взять бинарную функцию распределения в виде $F_2(1;2) = F_1(1)F_1(2)$. Тогда получается кинетическое уравнение Власова. Поскольку далее в этом параграфе рассматривается только одночастичная функция распределения, индекс «1» у нее опускаем. Уравнение имеет вид:

$$\frac{\partial F}{\partial t} = [H_V, F], \quad (4.1)$$

$$H_V = H_V(q, p; [F]) = H_0(p) + \frac{1}{\omega} \int h(\mathbf{x}_1, \mathbf{x}_2) F(x_2, t) d\mathbf{x}_2$$

Эффективный гамильтониан H_V определяется не только энергией свободного движения H_0 , но и средней энергией взаимодействия частицы со своим окружением, т.е. гамильтониан H_V является функционалом от искомой функции распределения. Такие силовые поля называются самосогласованными. Они характерны для дальнедействующих потенциалов, когда сложно выделить микроскопически малую пространственную локализацию процесса столкновения частиц.

В отличие от цепочки, для этого уравнения не проходит метод разложения функции распределения по малому параметру пространственной неоднородности, т.к. стационарные распределения могут не быть однородными. Действительно, любое стационарное решение уравнения (4.1) имеет вид произвольной дифференцируемой функции от средней энергии самосогласованного поля: $F_{st} = f(H_V)$. Полагая, в частности, функцию f экспонентой, получаем для классического гамильтониана с бинарным взаимодействием $\Phi(|\mathbf{q} - \mathbf{q}'|)$ следующее равновесное распределение:

$$F_{st} = \frac{1}{Z_0 Q_V} \exp(-\beta H_V), \quad H_V = \frac{p^2}{2m} + U_V(q), \quad (4.2)$$

$$Z_0 = \int \exp(-\beta H_0(p)) dp, \quad Q_V = \int \exp(-\beta U_V(q)) dq.$$

Отсюда следует, что самосогласованный потенциал $U_V(q)$ должен удовлетворять уравнению

$$U_V(q) \int \exp(-\beta U_V(q')) dq' = \frac{1}{\omega} \int \Phi(|\mathbf{q} - \mathbf{q}'|) \exp(-\beta U_V(q')) dq'. \quad (4.3)$$

Множество локальных парных потенциалов Φ , при которых существует решение уравнения (4.3), не пусто. В нем во всяком случае содержится осцилляторный потенциал, для которого решение уравнения (4.3) можно получить явно. Например, в \mathbf{R}^1 для $\Phi(|\mathbf{q} - \mathbf{q}'|) = k(q - q')^2$ из (4.3) получаем, что $U_V(q)$ является полиномом второго порядка по q с коэффициентами, зависящими от β и представляющими моменты равновесного самосогласованного распределения:

$$U_V(q) = \frac{k}{\omega} (q^2 - 2qm_1 + m_2),$$

$$m_1 = \int_{-\infty}^{\infty} q \exp(-\beta U_V(q)) dq, \quad m_2 = \int_{-\infty}^{\infty} q^2 \exp(-\beta U_V(q)) dq,$$

откуда следует, что

$$Q_V = \sqrt{\pi\omega/\beta ek}, \quad m_2 = m_1^2 + 1/2\beta k,$$

а коэффициент m_1 произвольный. В итоге получаем равновесное распределение в виде

$$F_{st} = \frac{1}{Z_0} \sqrt{\frac{\beta k}{\pi\omega}} \exp\left(-\beta H_0(p) - \frac{\beta k}{\omega} (q - m_1)^2\right). \quad (4.4)$$

Если потенциал взаимодействия имеет более высокий порядок — тернарный и более, то постулат факторизации применяется ко всем функциям распределения, участвующим в первом уравнении цепочки для одночастичной функции. Тогда для гамильтониана (2.1) следует кинетическое уравнение, описывающее движение системы в самосогласованном поле:

$$\frac{\partial F}{\partial t} = [H_V, F],$$

$$H_V(q, p; [F]) = H_0(p) + \sum_{k=2}^K \sum_{m=1}^{k-1} \frac{1}{(k-l-1)! \omega^{k-l}} \cdot \quad (4.5)$$

$$\cdot \int h_{k-l+1}(\mathbf{x}_1, \dots, \mathbf{x}_{k-l+1}) F(x_2) \dots F(x_{k-l+1}) d\mathbf{x}_2 \dots d\mathbf{x}_{k-l+1}.$$

Свойства уравнения Власова (4.1) отличны от свойств уравнения Больцмана.

Во-первых, уравнение допускает микроскопические решения, т.е. решения вида

$$F(q, p, t) = \delta(q - q(t)) \delta(p - p(t)), \quad (4.6)$$

где $q(t)$ и $p(t)$ удовлетворяют уравнениям Гамильтона с потенциалом H_V , что проверяется непосредственной подстановкой (4.6) в (4.1).

Во-вторых, уравнение Власова сохраняет энтропию:

$$\begin{aligned} \frac{dS}{dt} &= \frac{d}{dt} \int f \ln f dp dq = \\ &= \int \frac{\partial f}{\partial t} (1 + \ln f) dp dq = \int [H_V, f] (1 + \ln f) dp dq = \\ &= \int \left(\frac{\partial H_V}{\partial q} \frac{\partial(f \ln f)}{\partial p} - \frac{\partial H_V}{\partial p} \frac{\partial(f \ln f)}{\partial q} \right) dp dq = \\ &= \int [H_V, f \ln f] dp dq = 0. \end{aligned}$$

Если в уравнении Власова (4.1) сделать так называемую гидродинамическую подстановку

$$F(q, p, t) = \rho(q, t) \delta(p - u(q, t)), \quad (4.7)$$

то для функций ρ и u получаются уравнения гидродинамического типа:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{1}{m} \frac{\partial(\rho u^\alpha)}{\partial q^\alpha} &= 0; \\ \frac{\partial u^\alpha}{\partial t} + \frac{u^\beta}{m} \frac{\partial u^\alpha}{\partial q^\beta} &= \int \frac{\partial h(q, q')}{\partial q} \rho(q', t) dq'. \end{aligned} \quad (4.8)$$

Уравнения гидродинамики, т.е. уравнения эволюции моментов по пространству импульсов для функции распределения, удовлетворяющей уравнению (4.1), отличаются, естественно, от аналогичных уравнений, следующих из цепочки Боголюбова или из уравнения Больцмана. Каждый подход, использующий то или иное кинетическое уравнение, определяет свою гидродинамику. При этом важно понимать, что между уравнениями гидродинамики нет той генетической связи, какая есть между кинетическими уравнениями: если те или иные кинетические уравнения можно вывести из более общих уравнений, например, из цепочки Боголюбова, то уравнения гидродинамики, следующие из более общих уравнений, не позволяют получить уравнения гидродинамики, следующие из частных кинетических уравнений.

1.5. Уравнения эволюции моментов

Пусть имеется динамическая величина s -кратного типа, симметричная относительно перестановок индексов частиц

$$B^{(s)}(x) = \sum_{1 \leq i_1 < \dots < i_s \leq N} b_s(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_s}). \quad (5.1)$$

Ее среднее значение в момент времени t в термодинамическом пределе есть

$$\beta_s(t) = \frac{1}{s! \omega^s} \int b_s(x^s) F_s dx^s. \quad (5.2)$$

Локальная одночастичная координатная плотность определяется как момент нулевого порядка унарной функции распределения:

$$\rho(\mathbf{q}, t) = \frac{1}{\omega} \int F_1(\mathbf{q}, \mathbf{p}, t) d\mathbf{p}. \quad (5.3)$$

Средней локальной координатной плотностью величины $B^{(s)}$ называется величина

$$\langle B_s \rangle = \frac{1}{\rho \omega^s s!} \int b_s(x^s) F_s d\mathbf{p}_1 d\mathbf{x}_2 \dots d\mathbf{x}_s. \quad (5.4)$$

Непосредственно из цепочки Боголюбова (2.4) интегрированием по фазовому пространству комплекса частиц, отвечающих порядку взаимодействия в изучаемой функции от динамических переменных (унарная,

бинарная и т.д.), получаем уравнения эволюции координатной плотности этой функции. Из (5.4) следует, что изменение локальной плотности динамической величины $B(q, p)$ определяется не только уравнением эволюции функции распределения, но и изменением плотности ρ . Эволюция локальной плотности ρ (5.3) определяется из первого уравнения цепочки (2.4) путем интегрирования этого уравнения по импульсам первой частицы. Имеем

$$\begin{aligned} \frac{\partial F_1(\mathbf{x}_1, t)}{\partial t} &= [H_1, F_1] + \sum_{k=2}^K \frac{1}{(k-1)! \omega^{k-1}} \cdot \\ &\quad \cdot \int [h_k(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k), F_k(x^k, t)] d\mathbf{x}_2 \dots d\mathbf{x}_k, \\ \frac{\partial \rho(\mathbf{q}_1, t)}{\partial t} &= \frac{1}{\omega} \int \frac{\partial F_1(\mathbf{q}_1, \mathbf{p}_1, t)}{\partial t} d\mathbf{p}_1 = \\ &= \frac{1}{\omega} \int \left(\frac{\partial H_1}{\partial q_1^\alpha} \frac{\partial F_1}{\partial p_1^\alpha} - \frac{\partial H_1}{\partial p_1^\alpha} \frac{\partial F_1}{\partial q_1^\alpha} \right) d\mathbf{p}_1 + \\ &+ \sum_{k=2}^K \frac{1}{(k-1)! \omega^k} \int [h_k(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k), F_k(x^k, t)] d\mathbf{p}_1 d\mathbf{x}_2 \dots d\mathbf{x}_k. \end{aligned}$$

Выписанные выше интегралы от скобки Пуассона в правой части уравнения эволюции плотности преобразуются путем интегрирования по частям с учетом нулевых пределов у функции распределения на границе фазового пространства, а также выносе производной по первой пространственной координате из-под интеграла. Предполагая, что гамильтониан является дважды непрерывно дифференцируемой функцией своих аргументов, так что его смешанные производные при перестановке порядка дифференцирования совпадают, для первого слагаемого получаем

$$\begin{aligned} &\int \left(\frac{\partial H_1}{\partial q_1^\alpha} \frac{\partial F_1}{\partial p_1^\alpha} - \frac{\partial H_1}{\partial p_1^\alpha} \frac{\partial F_1}{\partial q_1^\alpha} \right) d\mathbf{p}_1 = \\ &= \frac{\partial H_1}{\partial q_1^\alpha} F_1 \Big|_{-\infty}^{+\infty} - \int \frac{\partial^2 H_1}{\partial q_1^\alpha \partial p_1^\alpha} F_1 d\mathbf{p}_1 - \frac{\partial}{\partial q_1^\alpha} \int \frac{\partial H_1}{\partial p_1^\alpha} F_1 d\mathbf{p}_1 + \\ &+ \int \frac{\partial^2 H_1}{\partial p_1^\alpha \partial q_1^\alpha} F_1 d\mathbf{p}_1 = - \frac{\partial}{\partial q_1^\alpha} \int \frac{\partial H_1}{\partial p_1^\alpha} F_1 d\mathbf{p}_1. \end{aligned}$$

Аналогично интеграл от суммы для k -кратного взаимодействия преобразуется к виду

$$\begin{aligned} &\sum_{k=2}^K \frac{1}{(k-1)! \omega^k} \int [h_k(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k), F_k(x^k, t)] d\mathbf{p}_1 d\mathbf{x}_2 \dots d\mathbf{x}_k = \\ &= - \sum_{k=2}^K \frac{1}{(k-1)! \omega^k} \frac{\partial}{\partial q_1^\alpha} \int \frac{\partial h_k(x^k)}{\partial p_1^\alpha} F_k(x^k, t) d\mathbf{p}_1 d\mathbf{x}_2 \dots d\mathbf{x}_k. \end{aligned}$$

Вводя теперь эффективную одночастичную скорость, как в (2.7)

$$v_{eff}^\alpha(\mathbf{x}_1, t) = \frac{\partial H_1}{\partial p_1^\alpha} + \sum_{k=2}^K \frac{1}{(k-1)! \omega^{k-1}} \frac{1}{F_1} \int d\mathbf{x}_2 \dots d\mathbf{x}_k \frac{\partial h_k(x^k)}{\partial p_1^\alpha} F_k,$$

получаем уравнение эволюции локальной плотности в виде

$$\begin{aligned} \frac{\partial \rho(\mathbf{q}, t)}{\partial t} + \frac{\partial}{\partial q^\alpha} (\rho(\mathbf{q}, t) u^\alpha(\mathbf{q}, t)) &= 0, \\ u^\alpha(\mathbf{q}, t) &= \frac{1}{\omega \rho(\mathbf{q}, t)} \int v_{eff}^\alpha(\mathbf{x}, t) F_1(\mathbf{x}, t) d\mathbf{p} \end{aligned} \quad (5.5)$$

Аналогично получаем уравнение эволюции средней локальной плотности (5.4) динамической величины S -кратного типа:

$$\frac{\partial \langle \rho(B_s) \rangle}{\partial t} = - \frac{\partial}{\partial q^\alpha} (\rho \Pi_s^\alpha) + \rho T_s, \quad (5.6)$$

где введены обозначения

$$\begin{aligned} \Pi_s^\alpha &= \frac{1}{\rho \omega^s s!} \int d\mathbf{p}_1 d\mathbf{x}_2 \dots d\mathbf{x}_s b_s(x^s) \times \\ &\times \left\{ \frac{\partial H_s}{\partial p_1^\alpha} F_s + \sum_{k=2}^K \sum_{l=1}^{k-1} \sum_{2 \leq i_2 < \dots < i_{k-l} \leq s} \frac{1}{l! \omega^l} \times \right. \\ &\times \left. \int d\mathbf{x}_{s+1} \dots d\mathbf{x}_{s+l} \frac{\partial h_k(x^{k-l}, \mathbf{x}_{s+1}, \dots, \mathbf{x}_{s+l})}{\partial p_1^\alpha} F_{s+l} \right\}; \\ T_s &= \langle [b_s, H_s^{eff}] \rangle; \\ H_s^{eff} &= H_s + \sum_{k=2}^K \sum_{l=1}^{k-1} \sum_{2 \leq i_2 < \dots < i_{k-l} \leq s} \frac{F_s^{-1}}{l! \omega^l} \times \\ &\times \int d\mathbf{x}_{s+1} \dots d\mathbf{x}_{s+l} h_k(x^{k-l}, \mathbf{x}_{s+1}, \dots, \mathbf{x}_{s+l}) F_{s+l}. \end{aligned} \quad (5.7)$$

Таким образом, получаем, что эволюция средней локальной плотности S -кратной динамической величины определяется дивергенцией потока этой плотности и средним значением скобки Пуассона этой динамической величины и эффективного гамильтониана S -частичного взаимодействия.

Это общая структура уравнений гидродинамического типа. Она представляет собой незамкнутую бесконечную систему моментных уравнений независимо от того, какое кинетическое уравнение положено в основу. Обрыв моментной системы состоит в выборе некоторого правила замыкания, когда высшие моменты выражены через низшие. Важно понимать, что замыкание — это не математическое следствие кинетической теории, а модельное требование.

Уравнения гидродинамики, следующие из уравнения Больцмана, имеют более простой вид и более ясное физическое содержание. Гидродинамическими переменными для этого кинетического уравнения являются средние значения (3.17) сумматорных инвариантов. Уравнения эволюции для них следующие:

$$\begin{aligned}\frac{\partial \rho}{\partial t} + \frac{1}{m} \frac{\partial}{\partial q^\alpha} (\rho p^{*\alpha}) &= 0; \\ \frac{\partial}{\partial t} (\rho p^{*\alpha}) + \frac{1}{m} \frac{\partial}{\partial q^\beta} (\rho P^{\alpha\beta}) &= 0; \\ \frac{\partial (\rho \varepsilon)}{\partial t} + \frac{1}{m} \frac{\partial}{\partial q^\alpha} (\rho J^\alpha) &= 0.\end{aligned}\quad (5.8)$$

Введенные здесь средние величины определяются по унарному распределению:

$$\begin{aligned}\rho &= \langle 1 \rangle, \quad p^{*\alpha} = \langle p^\alpha \rangle, \quad \varepsilon = \frac{1}{2m} \langle (p - p^*)^2 \rangle, \\ P^{\alpha\beta} &= \langle p^\alpha p^\beta \rangle, \quad J^\alpha = \frac{1}{2m} \langle (p - p^*)^2 p^\alpha \rangle.\end{aligned}\quad (5.9)$$

Сравним полученные уравнения с гидродинамикой, следующей из уравнения Власова (4.1). Поскольку это уравнение описывает движение тела в эффективном потенциале H_V , то и гидродинамические параметры

в части средней энергии следует выбрать самосогласованными. Введем «власовские средние»

$$\begin{aligned}\rho_V &= \frac{1}{\omega} \int F(q, p, t) dp, \quad \mathbf{p}_V^* = \frac{1}{\omega \rho_V} \int \mathbf{p} F(q, p, t) dp, \\ \varepsilon_V &= \frac{1}{\omega \rho_V} \int H_V F(q, p, t) dp.\end{aligned}\quad (5.10)$$

Дифференцируя равенства (5.10) в силу уравнения эволюции функции распределения (4.1), получаем для первых двух GDP

$$\begin{aligned}\frac{\partial \rho_V}{\partial t} &= - \frac{\partial}{\partial q^\alpha} \rho_V u_V^\alpha, \quad \rho_V u_V^\alpha = \frac{1}{\omega} \int \frac{\partial H_V}{\partial p^\alpha} F(q, p, t) dp; \\ \frac{\partial \rho_V p_V^{*\alpha}}{\partial t} &= - \frac{1}{\omega} \int \frac{\partial H_V}{\partial q^\alpha} F(q, p, t) dp - \frac{1}{\omega} \frac{\partial}{\partial q^\beta} \int p^\alpha \frac{\partial H_V}{\partial p^\beta} F(q, p, t) dp.\end{aligned}\quad (5.11)$$

Заметим, что в уравнении эволюции плотности фигурирует средняя скорость u_V^α как скорость, построенная по эффективному гамильтониану самосогласованного поля, которая не обязана, вообще говоря, явно выражаться через средний импульс $p_V^{*\alpha}$. Это означает, что для скорости u_V^α также следует получить уравнение эволюции, которое будет иметь значительно более сложный вид, поскольку эта скорость определяется через гамильтониан, который сам зависит от времени в соответствии с (4.1). Получим сначала уравнение эволюции для средней власовской энергии. Поскольку

$$\begin{aligned}H_V &= H_V(q, p; [F]) = H_0(p) + \frac{1}{\omega} \int h(\mathbf{x}_1, \mathbf{x}_2) F(x_2, t) d\mathbf{x}_2, \\ \frac{\partial F}{\partial t} &= [H_V, F]\end{aligned}$$

то

$$\begin{aligned}\frac{\partial H_V}{\partial t} &= \frac{1}{\omega} \int h(\mathbf{x}_1, \mathbf{x}_2) \frac{\partial F(x_2, t)}{\partial t} d\mathbf{x}_2 = \\ &= \frac{1}{\omega} \int h(\mathbf{x}_1, \mathbf{x}_2) [H_V(q_2, p_2; [F]), F(q_2, p_2, t)] d\mathbf{x}_2.\end{aligned}$$

Тогда

$$\begin{aligned}\frac{\partial(\rho_V \varepsilon_V)}{\partial t} &= \frac{1}{\omega} \int \frac{\partial H_V}{\partial t} F(q, p, t) dp + \frac{1}{\omega} \int H_V \frac{\partial F(q, p, t)}{\partial t} dp = \\ &= \frac{1}{\omega} \int h(\mathbf{x}_1, \mathbf{x}_2) [H_V(x_2; [F]), F(x_2, t)] F(x_1, t) dp_1 d\mathbf{x}_2 + \\ &+ \frac{1}{\omega} \int H_V(x_1; [F]) [H_V(x_1; [F]), F(x_1, t)] dp_1.\end{aligned}$$

Для краткости в промежуточных вычислениях опустим индекс «V» у средних величин, а также аргументы у функций распределения, отмечая зависимость от «штрихованных» переменных штрихом над функцией. Тогда первый интеграл в правой части уравнения эволюции средней энергии имеет вид

$$\begin{aligned}\int \frac{\partial H}{\partial t} F dp &= \frac{1}{\omega} \int F dp \int \Phi(|q - q'|) \frac{\partial F'}{\partial t} dq' dp' = \\ &= \frac{1}{\omega} \int F dp \int \Phi(|q - q'|) [H', F'] dq' dp',\end{aligned}$$

который интегрированием по частям с учетом нулевых граничных условий, а также того, что $\partial \Phi(|q - q'|) / \partial q' = -\partial \Phi(|q - q'|) / \partial q$, преобразуется к выражению

$$\begin{aligned}\int \frac{\partial H}{\partial t} F dp &= \frac{1}{\omega} \int F dp \int \frac{\partial \Phi(|q - q'|)}{\partial q'} \frac{\partial H}{\partial p'} F' dq' dp' = \\ &= -\rho \frac{\partial}{\partial q} \int \Phi(|q - q'|) \rho' u' dq'.\end{aligned}$$

Второй интеграл в правой части уравнения эволюции средней энергии преобразуется интегрированием по частям к следующему виду:

$$\begin{aligned}\int H \left(\frac{\partial H}{\partial q} \frac{\partial F}{\partial p} - \frac{\partial H}{\partial p} \frac{\partial F}{\partial q} \right) dp &= \\ &= -\int F \frac{\partial}{\partial p} \left(H \frac{\partial H}{\partial q} \right) dp - \frac{\partial}{\partial q} \int H F \frac{\partial H}{\partial p} dp + \\ &+ \int H F \frac{\partial^2 H}{\partial p \partial q} dp + \int F \frac{\partial H}{\partial q} \frac{\partial H}{\partial q} dp = -\frac{\partial}{\partial q} \int H F \frac{\partial H}{\partial p} dp.\end{aligned}$$

В итоге для эволюции средней vlasовской энергии получаем уравнение

$$\begin{aligned}\frac{\partial(\rho_V \varepsilon_V)}{\partial t} &= -\rho_V \frac{\partial}{\partial q^\alpha} \int \Phi(|q - q'|) \rho'_V u'_V{}^\alpha dq' - \\ &- \frac{1}{\omega} \frac{\partial}{\partial q^\alpha} \int H_V F \frac{\partial H_V}{\partial p^\alpha} dp.\end{aligned}\quad (5.12)$$

Аналогично получаем уравнение эволюции средней vlasовской скорости:

$$\frac{\partial(\rho_V u_V^\alpha)}{\partial t} = -\frac{1}{\omega} \frac{\partial}{\partial q^\beta} \int \frac{\partial H_V}{\partial p^\alpha} \frac{\partial H_V}{\partial p^\beta} F dp - \frac{1}{m\omega} \int \frac{\partial H_V}{\partial q^\alpha} F dp. \quad (5.13)$$

Таким образом, каждое кинетическое уравнение порождает собственную гидродинамику. Важность этих уравнений в том, что именно средние величины и являются наблюдаемыми, а уравнения их эволюции определяют некоторую макроскопическую динамическую систему. Тем самым уравнения эволюции моментов могут быть положены в основу модели временного ряда, для которого удалось написать уравнение эволюции его выборочной функции распределения. Эта связь будет реализована далее в части III.

Глава II. Вырожденные динамические системы

2.1. Слаборелятивистские динамические системы

В первой главе были построены кинетические уравнения для гладких динамических систем, т.е. для систем, у которых к интегральным кривым решений уравнений движения в фазовом пространстве в каждой точке можно провести единственную касательную. В лагранжевом формализме классической механики это означает, что матрица вторых производных лагранжиана по скоростям невырождена.

В ряде физических приложений возникают модели, которые допускают неоднозначность решений. Такие примеры хорошо известны из курса обыкновенных дифференциальных уравнений: если, например, уравнение второго порядка имеет особое решение (дискриминантную кривую), то на таком решении теряется однозначность динамики. Если нет указаний на то, как следует продолжать траекторию из точки особого решения, то вырожденные динамические системы естественным образом порождают вероятностную динамику. Таким образом, точки множества, на котором вырождается лагранжиан динамической системы, могут рассматриваться как источники хаоса. Это — дополнительный элемент неопределенности к случайному выбору начальных условий при описании эволюции статистического ансамбля. Если у системы нет особых точек, то эволюция начального распределения дается уравнением Лиувилля (1.1.12) или, для консервативных систем, (1.1.13). Но, как было показано в параграфе 1.1, динамическим основанием для уравнения Лиувилля служат уравнения Эйлера–Лагранжа (1.1.4), которые определены только для регулярных траекторий ДС. Поэтому следует сначала определить правила прохождения через особые точки.

Существует ряд физически важных примеров ДС с особенностями в фазовом пространстве [3]. В частности, особые траектории порождаются лагранжианами Дарвина и Фока–Фихтенгольца–Инфельда в слаборелятивистской теории взаимодействующих частиц — электрических зарядов и тяготеющих масс. Практически важным является лагранжиан системы слаборелятивистских частиц с учетом запаздывания взаимодействия в первом порядке по отношению энергии движения к энергии покоя, описывающий горячую электронную плазму или быстрые пучки заряженных

частиц. Модель такой системы приводит к наличию сингулярных точек в фазовом пространстве, что требует разработки специальной методики продолжения траектории через сингулярное множество.

Чтобы обосновать идею о продолжении траектории, рассмотрим сначала простой пример одномерной динамики слаборелятивистского осциллятора. Эта модельная система имеет нетривиальную точку поворота, в которой скорость отлична от нуля, а ускорение стремится к бесконечности.

Рассмотрим лагранжиан

$$L = -mc^2 \sqrt{1 - (v/c)^2} - \frac{1}{2} kq^2 \left(1 + (a-1) \frac{v^2}{2c^2} \right), \quad (1.1)$$

где m — масса покоя частицы, k — жесткость осцилляторной связи, a — некоторая неотрицательная целочисленная постоянная (так называемый ранг поля взаимодействия). При $a = 1$ модель (1.1) невырождена, так что будем считать, что $a \neq 1$.

Согласно (1.1.4), динамическая масса этой системы равна

$$M = \frac{\partial^2 L}{\partial v^2} = \frac{m}{\left(1 - v^2/c^2\right)^{3/2}} - \frac{k(a-1)q^2}{2c^2}, \quad (1.2)$$

а энергия определяется выражением

$$E = v \frac{\partial L}{\partial v} - L = \frac{mc^2}{\sqrt{1 - (v/c)^2}} + \frac{1}{2} kq^2 \left(1 - (a-1) \frac{v^2}{2c^2} \right). \quad (1.3)$$

Заметим, что энергия как функция, определяемая по лагранжиану по правилу (1.3), определена для всех точек фазового пространства, если, как в данном примере, лагранжиан является дифференцируемой функцией скорости. Однако смысл инварианта уравнений движения эта функция приобретает только в регулярных точках траектории, т.е. там, где определены уравнения Эйлера–Лагранжа.

В точках кривой S_D , определяемой уравнением $M = 0$, т.е. на кривой

$$S_D : \quad q^2 = \frac{2mc^2 / [k(a-1)]}{\left(1 - v^2/c^2\right)^{3/2}}, \quad (1.4)$$

векторное поле динамической системы не определено. Уравнение $E = \text{const}$ определяет траекторию, которая может пересекаться с кривой S_D при некоторых значениях координат и скоростей. Введем для сокращения записи безразмерные координаты $u = v/c$ и $x = (q/c)\sqrt{k/m}$, а энергию покоя обозначим как $E_0 = mc^2$. В этих координатах энергия имеет вид

$$E = E_0 \left(\frac{1}{\sqrt{1-u^2}} + \frac{1}{2} x^2 \left(1 - \frac{a-1}{2} u^2 \right) \right),$$

а на особой траектории выполняется условие

$$x^2 = \frac{2/(a-1)}{(1-u^2)^{3/2}}.$$

Подставляя его в уравнение энергии, получаем, что условие достижимости особого решения при заданном значении величины $\varepsilon = E/E_0$ состоит в существовании решений относительно величины $\gamma = 1/\sqrt{1-u^2}$ на множестве $[1; +\infty]$ следующего уравнения:

$$\begin{cases} \gamma^3 + \frac{3}{2\alpha} \gamma - \frac{\varepsilon}{\alpha} = 0, & \alpha = \frac{1}{a-1} - \frac{1}{2} \neq 0; \\ \gamma = \frac{2\varepsilon}{3}, & a = 3. \end{cases} \quad (1.5)$$

Кубическое уравнение относительно γ при $a > 1$ имеет единственное действительное решение. Например, при $a = 2$ решение имеет вид

$$\gamma = \left(\sqrt{1+\varepsilon^2} + \varepsilon \right)^{1/3} - \left(\sqrt{1+\varepsilon^2} - \varepsilon \right)^{1/3}.$$

Необходимое условие $\gamma \geq 1$ выполняется только при $\varepsilon \geq 2$. В этом случае особая траектория достижима. При стремлении к ней регулярной траектории ускорение тела стремится к бесконечности и направлено либо против движения, либо по движению, в зависимости от того, в какой области фазового пространства происходит приближение к особой точке — там, где $M > 0$ или где $M < 0$. Следовательно, можно предложить та-

кую интерпретацию происходящего на особой траектории: частица либо отражается от нее как от абсолютно жесткой стенки, либо проскакивает в ту часть фазового пространства, где $M > 0$. Хотя на самой S_D траектория не определена, она может быть продолжена на нее по непрерывности функции энергии. При таком подходе отражение или проскакивание не сопровождаются изменением модуля скорости в бесконечно малой окрестности особой траектории.

Рассмотрим теперь более общий пример — одну частицу в R^n , лагранжиан которой зависит от модулей координаты и скорости: $L = L(q, v)$, $q = |\mathbf{q}|$, $v = |\mathbf{v}|$. Тогда тензор массы такой частицы равен

$$M_{ij} = \frac{1}{v} \frac{\partial L}{\partial v} \Delta_{ij}^\perp + \frac{1}{v} \frac{\partial E}{\partial v} \Delta_{ij}, \quad (1.6)$$

где $E = v \frac{\partial L}{\partial v} - L$ — энергия системы, а Δ_{ij} и Δ_{ij}^\perp — взаимно-ортогональные проекторы на направление вектора скорости:

$$\Delta_{ij} = e_i e_j, \quad \Delta_{ij}^\perp = \delta_{ij} - \Delta_{ij}, \quad \mathbf{e} = \mathbf{v}/v.$$

Определитель матрицы M_{ij} легко вычисляется, поскольку $\frac{1}{v} \frac{\partial E}{\partial v}$ — ее собственное значение кратности 1, а $\frac{1}{v} \frac{\partial L}{\partial v}$ — собственное значение кратности $n-1$, так что

$$J = \left(\frac{1}{v} \frac{\partial L}{\partial v} \right)^{n-1} \frac{1}{v} \frac{\partial E}{\partial v}, \quad (M^{-1})_{ij} = \left(\frac{1}{v} \frac{\partial L}{\partial v} \right)^{-1} \Delta_{ij}^\perp + \left(\frac{1}{v} \frac{\partial E}{\partial v} \right)^{-1} \Delta_{ij}. \quad (1.7)$$

Отсюда следует, что особыми точками являются нетривиальные (с ненулевым значением скорости) стационарные точки энергии и лагранжиана, причем последние существуют только в пространстве размерности больше 1. Стационарные точки энергии отвечают разрыву компоненты ускорения, касательной к траектории, а стационарные точки лагранжиана — разрыву его нормальной составляющей. Разрешая формально уравнения Эйлера–Лагранжа (1.1.4) относительно ускорений, получаем

$$\dot{v}_i = \left(\frac{\partial L / \partial q}{\partial L / \partial v} \Delta_{ij}^\perp - \frac{\partial E / \partial q}{\partial E / \partial v} \Delta_{ij} \right) v n_j, \quad \mathbf{n} = \frac{\mathbf{q}}{q}. \quad (1.8)$$

Отсюда следует, что в окрестности сингулярного множества ускорение стремится в бесконечность, если в этих точках $\partial L / \partial \dot{q} \neq 0$, $\partial E / \partial \dot{q} \neq 0$.

Назовем сингулярным множеством S_D динамической системы множество точек, в которых $J = 0$. В нашем примере S_D разбивается в прямую сумму двух множеств: сингулярного множества S_D^1 первого типа, отвечающего разрыву компоненты ускорения, направленной вдоль движения, и сингулярного множества S_D^2 второго типа, отвечающего разрыву нормальной компоненты ускорения.

В той области фазового пространства, где $\frac{\partial E / \partial \dot{q}}{\partial E / \partial v} > 0$, в окрестности S_D^1 ускорение имеет направление, противоположное движению (знак «минус» при проекторе Δ_{ij}), а в окрестности S_D^2 — ортогонально скорости. Поэтому можно предложить следующую интерпретацию поведения системы в окрестности S_D : траектория отражается от S_D^1 с сохранением модуля нормальной к S_D^1 составляющей скорости, и спирально накручивается на S_D^2 . На самом множестве S_D векторное поле динамической системы не определено, но первые интегралы — в частности, энергия системы — могут быть продолжены на него по непрерывности. Тогда достижимость системой своего сингулярного множества зависит от того, имеет ли S_D общие точки с изоэнергетической поверхностью $E = \text{const}$. Поэтому может возникнуть такая ситуация, что, хотя динамическая система и имеет сингулярное множество S_D , оно недостижимо в силу имеющихся законов сохранения. Пусть, в частности, движение тела в R^n задается лагранжианом

$$L = -mc^2 \sqrt{1 - v^2 / c^2} - \Phi(q) \left(1 + \frac{(a-1)v^2}{2c^2} \right), \quad (1.9)$$

Сингулярные поверхности определяются уравнением

$$J = m^n \left(\frac{1}{\sqrt{1 - v^2 / c^2}} - \frac{(a-1)\Phi(q)}{mc^2} \right)^{n-1} \times \left(\frac{1}{(1 - v^2 / c^2)^{3/2}} - \frac{(a-1)\Phi(q)}{mc^2} \right) = 0 \quad (1.10)$$

Для случая притяжения ($\Phi < 0$) при $a \geq 1$ лагранжиан (1.9) невырожден. Это важное замечание, поскольку из него следует невырожденность слабoreлятивистской задачи Кеплера о движении в поле притягивающего центра.

Динамическая система (1.9) имеет интеграл энергии того же вида, что и (1.3)

$$E = \frac{mc^2}{\sqrt{1 - v^2 / c^2}} + \Phi(q) \left(1 - \frac{(a-1)v^2}{2c^2} \right), \quad (1.11)$$

а при $n > 1$ существует также еще один первый интеграл — кинетический момент:

$$\mathbf{K} = \left[\mathbf{q}, \frac{\partial L}{\partial \dot{\mathbf{v}}} \right] = m \left(\frac{1}{\sqrt{1 - v^2 / c^2}} - \frac{(a-1)\Phi(q)}{mc^2} \right) [\mathbf{q}, \mathbf{v}]. \quad (1.12)$$

Поэтому, если $\mathbf{K} \neq 0$, то сингулярное множество второго типа недостижимо, т.к. на нем должно быть $\frac{\partial L}{\partial v} \Big|_{S_D^2} = 0$. Что касается S_D^1 , то из

(1.10) следует, что на этом множестве выполняется равенство

$$\Phi(q) = \frac{mc^2}{(a-1)} \gamma^3, \text{ и потому законы сохранения (1.11), (1.12), непрерывно продолженные на } S_D^1, \text{ должны быть совместны.}$$

Как и в рассмотренном выше одномерном случае, достижение S_D^1

определяется решением уравнения (1.5). Дальнейшее исследование зависит от конкретного вида потенциала взаимодействия. Исследуем ситуацию для случая релятивистского осциллятора при $a = 2$. Условия совместности интегралов движения на S_D^1 дают систему

$$\gamma^3 + 3\gamma - 2\varepsilon = 0, \quad (1.13)$$

$$(\gamma^2 - 1)^3 \sin^2 \phi = k^2,$$

где ϕ — угол между векторами \mathbf{q} и \mathbf{v} , а $k = K/(mc)$.

Если $\gamma < \sqrt{2}$ (т.е. $\varepsilon < 5/\sqrt{2}$) и $k^2 > 1$, то ни при каких углах ϕ множество S_D^1 недостижимо, т.к. должно быть $|\sin \phi| \leq 1$. Таким образом, условие достижимости сингулярного множества имеет вид

$$\frac{1}{k^2} \left[\left(\sqrt{1+\varepsilon^2} + \varepsilon \right)^{2/3} + \left(\sqrt{1+\varepsilon^2} - \varepsilon \right)^{2/3} - 3 \right] \geq 1. \quad (1.14)$$

Обратимся теперь к моделям конкретных физических систем. Лагранжианы, задающие динамику системы с учетом конечности скорости распространения взаимодействия, инвариантные относительно группы преобразований Пуанкаре, построены в работе [10]. Слаборелятивистские лагранжианы с бинарным взаимодействием имеют вид

$$\begin{aligned} L_{PG} = & - \sum_{i=1}^N m_i c^2 \sqrt{1 - (\mathbf{v}_i / c)^2} - \\ & - \sum_{1 \leq i < j \leq N} \Phi_{ij} \left[1 + \frac{a-1}{2c^2} \mathbf{v}_{ij}^2 - \frac{\mathbf{v}_i \mathbf{v}_j}{2c^2} + \frac{\theta_{ij}}{2c^2} (\mathbf{n}_{ij} \mathbf{v}_i) (\mathbf{n}_{ij} \mathbf{v}_j) \right] - \\ & - \frac{b}{2c^2} \sum_{1 \leq i < j \leq N} \Phi_{ij}^2 - \frac{l}{6c^2} \sum_{i < j < k} (\Phi_{ij} \Phi_{ik} + \Phi_{ij} \Phi_{jk} + \Phi_{ik} \Phi_{jk}), \end{aligned} \quad (1.15)$$

где через m_i обозначены массы покоя частиц, $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$, $\mathbf{n}_{ij} = \mathbf{q}_{ij} / q_{ij}$ — единичные векторы в направлении $\mathbf{q}_{ij} = \mathbf{q}_i - \mathbf{q}_j$, $\Phi_{ij} = \Phi(q_{ij})$ — классический потенциал парного взаимодействия, а $\theta_{ij} = q_{ij} \Phi'_{ij} / \Phi_{ij}$ — логарифмическая производная потенциала. При $l = 0$, $a = 1$, $b = 0$ из (1.15) получается лагранжиан Дарвина [11], а при $l = 3$, $a = 4$, $b = 1$ — лагранжиан Фока–Фихтенгольца–Инфельда [12]. Параметр a определяется тензорным рангом поля взаимодействия [10]: для скалярного поля $a = 0$, для системы электрических зарядов $a = 1$

(векторное поле), для гравитационного поля $a = 4$ (ранг тензора Римана).

Найдем сингулярное множество двухчастичной системы, задаваемой лагранжианом (1.15). Тензор массы $M_{ij}^{\alpha\beta} = \partial^2 L / \partial v_i^\alpha \partial v_j^\beta$ имеет вид

$$M = \begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix}, \quad (1.16)$$

$$\begin{aligned} M_{ii}^{\alpha\beta} &= \left(m_i \gamma_i - \frac{(a-1)\Phi}{c^2} \right) \delta^{\alpha\beta} + m_i \frac{v_i^2}{c^2} \gamma_i^\alpha \gamma_i^\beta e_i^\alpha e_i^\beta, \\ M_{ij}^{\alpha\beta} \Big|_{i \neq j} &= \frac{\Phi}{2c^2} \left[(2a-1) \delta^{\alpha\beta} - \theta n^\alpha n^\beta \right]. \end{aligned} \quad (1.17)$$

Для сокращения записей введем обозначения

$$z = \Phi / c^2, \quad \mu_i = m_i \gamma_i - \frac{(a-1)\Phi}{c^2}, \quad \tilde{\mu}_i = m_i \gamma_i^3 - \frac{(a-1)\Phi}{c^2}.$$

По формуле Шура [13] для определителя блочной матрицы имеем

$$J = \det M = \det M_{11} \det T, \quad T = M_{22} - M_{21} M_{11}^{-1} M_{12},$$

где $\det M_{11}$ вычисляется по формуле (1.10), а матрица T имеет более сложный вид:

$$\begin{aligned} T^{\alpha\beta} &= A \delta^{\alpha\beta} + B_1 e_1^\alpha e_1^\beta + B_2 e_2^\alpha e_2^\beta + B_3 n^\alpha n^\beta + B_{13} (e_1^\alpha n^\beta + e_1^\beta n^\alpha), \\ A &= \mu_2 - \frac{z^2}{4\mu_1} (2a-1)^2, \quad B_1 = \frac{z^2}{4} (2a-1)^2 \left(\frac{1}{\mu_1} - \frac{1}{\tilde{\mu}_1} \right), \\ B_2 &= \tilde{\mu}_2 - \mu_2, \quad B_3 = \frac{z^2}{4\mu_1} \theta \left[2(2a-1) - \theta \left(1 - \left(1 - \frac{\mu_1}{\tilde{\mu}_1} \right) (\mathbf{n} \mathbf{e}_1)^2 \right) \right], \\ B_{13} &= -\frac{z^2}{4} \theta (2a-1) \left(\frac{1}{\mu_1} - \frac{1}{\tilde{\mu}_1} \right) (\mathbf{n} \mathbf{e}_1). \end{aligned} \quad (1.18)$$

Для вычисления определителя матрицы вида

$$T^{\alpha\beta} = A \delta^{\alpha\beta} + \sum_{i,j=1}^N B_{ij} e_i^\alpha e_j^\beta, \quad \mathbf{e} \in R^n, |\mathbf{e}| = 1,$$

воспользуемся формулой, выводимой непосредственно путем разложения определителя по строкам:

$$\det T = A^n + \sum_{k=1}^n \frac{A^{n-k}}{k!} \sum_{i_1, j_1=1}^N B_{i_1 j_1} \dots \sum_{i_k, j_k=1}^N B_{i_k j_k} V_{i_1 \dots i_k}^{(k)} V_{j_1 \dots j_k}^{(k)}, \quad (1.19)$$

где $V_{i_1 \dots i_k}^{(k)}$ — объем k -мерного направленного параллелепипеда, построенного на векторах $\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k}$. В частности, в трехмерном случае

$$V_i^{(1)} = \mathbf{e}_i, \quad V_{ij}^{(2)} = [\mathbf{e}_i, \mathbf{e}_j], \quad V_{ijk}^{(3)} = (\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k). \quad (1.20)$$

Для матрицы (1.18) по формуле (1.20) получаем

$$\begin{aligned} \det T = A^{n-3} \{ & A^3 + A^2 (B_1 + B_2 + B_3 + 2B_{13} \mathbf{n} \mathbf{e}_1) + \\ & + A [(B_1 B_3 - B_{13}^2) [\mathbf{n} \mathbf{e}_1]^2 + B_1 B_2 [\mathbf{e}_1 \mathbf{e}_2]^2 + B_2 B_3 [\mathbf{n} \mathbf{e}_2]^2 + \\ & + 2B_2 B_{13} [\mathbf{e}_1 \mathbf{e}_2] [\mathbf{n} \mathbf{e}_2]] + B_2 (B_1 B_3 - B_{13}^2) (\mathbf{e}_1 \mathbf{e}_2 \mathbf{n})^2 \}. \end{aligned} \quad (1.21)$$

Эта формула определяет общий вид поверхности вырождения для пары частиц в фазовом пространстве. Однако практическую значимость имеет та часть сингулярного множества, на которой происходит отражение траектории. Для этого надо учесть, что условие существования нетривиальных решений (относительно скоростей) системы

$$\begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix} \begin{vmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{vmatrix} = 0$$

уже, чем условие $\det M = 0$. Исключая из диагональных блоков матрицы M проекторы, ортогональные скорости, получаем гораздо более простое выражение для соответствующего сингулярного множества S_D^1 :

$$\begin{aligned} J = & [m_1 m_2 \gamma_1^3 \gamma_2^3 - (a-1)z(m_1 \gamma_1^3 + m_2 \gamma_2^3) - \\ & - \frac{z^2}{4} (\theta^2 - 2\theta(2a-1) + 4a-3)]^{n-1} \times \\ & \times [m_1 m_2 \gamma_1^3 \gamma_2^3 - (a-1)z(m_1 \gamma_1^3 + m_2 \gamma_2^3) - az^2] = 0. \end{aligned} \quad (1.22)$$

Для лагранжиана Дарвина обе особые поверхности в (1.22) совпадают, и условие вырождения принимает вид

$$\gamma_1^3 \gamma_2^3 = \frac{1}{x^2}, \quad x = \frac{q}{q_0^{el}}, \quad q_0^{el} = \sqrt{q_1^{el} q_2^{el}}, \quad q_i^{el} = \frac{e_i^2}{m_i c^2}. \quad (1.23)$$

Отметим, что вырождение может происходить при любых скоростях (в том числе и нулевых — на расстоянии q_0^{el} между зарядами) и при любых знаках зарядов, причем, если $q_1^{el} \neq q_2^{el}$, то сближение происходит за счет «сдавливания» частицы с большим классическим радиусом.

Для гравитационного взаимодействия $\Phi = Gm_1 m_2 / q$ условие вырождения имеет вид

$$\begin{aligned} & \left[\gamma_1^3 \gamma_2^3 + \frac{3}{x} \left(\sqrt{\eta} \gamma_1^3 + \frac{1}{\sqrt{\eta}} \gamma_2^3 \right) - \frac{4}{x^2} \right] \times \\ & \times \left[\gamma_1^3 \gamma_2^3 + \frac{3}{x} \left(\sqrt{\eta} \gamma_1^3 + \frac{1}{\sqrt{\eta}} \gamma_2^3 \right) - \frac{7}{x^2} \right] = 0, \end{aligned} \quad (1.24)$$

$$\eta = \sqrt{m_1 / m_2}, \quad x = \frac{q}{q_0^g}, \quad q_0^g = \sqrt{q_1^g q_2^g}, \quad q_i^g = \frac{Gm_i}{c^2}.$$

Приведем также уравнение сингулярного множества первого типа для системы двух электрических зарядов с учетом гравитации. Для удобства сравнения с (1.22)–(1.24) введем следующие безразмерные величины: $\eta = m_1 / m_2$ — отношение масс покоя двух тел, $x = q / q_0$ — безразмерное расстояние между частицами, нормированное на среднее геометрическое

от двух классических радиусов зарядов, $\lambda = \sqrt{\frac{q_1^{el} q_2^{el}}{q_1^g q_2^g}} \sigma$ — относи-

тельная сила электромагнитного и гравитационного взаимодействий, $\sigma = \pm 1 = \text{sgn}(e_1 e_2)$ — знак произведения электрических зарядов. В этих обозначениях для определителя матрицы Якоби преобразования Лежандра имеем

$$J = \left[\gamma_1^3 \gamma_2^3 + \frac{3\sigma}{\lambda x} \left(\sqrt{\eta} \gamma_1^3 + \frac{1}{\sqrt{\eta}} \gamma_2^3 \right) - \frac{1}{x^2} \left(1 + \frac{14\sigma}{\lambda} + \frac{13}{\lambda^2} \right) \right]^{n-1} \times \left[\gamma_1^3 \gamma_2^3 + \frac{3\sigma}{\lambda x} \left(\sqrt{\eta} \gamma_1^3 + \frac{1}{\sqrt{\eta}} \gamma_2^3 \right) - \frac{1}{x^2} \left(1 + \frac{8\sigma}{\lambda} - \frac{20}{\lambda^2} \right) \right]. \quad (1.25)$$

При $\lambda \rightarrow +\infty$ (т.е. когда силы гравитации много меньше электростатических), из (1.25) получаем (1.22). Важной особенностью (1.25) является то, что существует область значений параметра λ , когда якобиан невырожден для случая разноименных зарядов. Именно, если отношение масс удовлетворяет условию $\eta \in (7 - 4\sqrt{3}, 7 + 4\sqrt{3})$, а отношение сил лежит в пределах

$$\lambda \in \left(7 - 6\sqrt{1 - \left((\sqrt{\eta} + 1/\sqrt{\eta})/4 \right)^2}, 4 + 6\sqrt{1 - \left((\sqrt{\eta} + 1/\sqrt{\eta})/4 \right)^2} \right). \quad (1.26)$$

то определитель (1.25) отличен от нуля. Таким образом, действие «в одном направлении» электромагнитных и гравитационных сил приблизительно одинаковой величины (слабо разноименно заряженные большие массы) снимает вырождение преобразования Лежандра во всем фазовом пространстве.

Следует подчеркнуть, что рассматриваемая ситуация принципиально отличается от классической динамики. Во-первых, лагранжиан системы не содержит «стенок» и иных «отражателей», за которые частица не может пройти в силу наложенных внешних условий; он является непрерывно дифференцируемой функцией своих аргументов бесконечное число раз везде, кроме сингулярных точек потенциала (если они есть) и предельных значений скоростей $u = \pm 1$; тем не менее, в фазовом пространстве существуют области, в которые точка никогда не сможет попасть ни при каких значениях своей энергии. Более того, из курсов статистической физики известно, что равновесное распределение, выводимое из принципа максимума энтропии, зависит только от полной энергии системы, а зависимость от других аддитивных инвариантов микроскопического движения (им-

пульса и момента импульса) может быть исключена выбором подходящей системы отсчета. В нашем же случае доступная область фазового пространства определяется всеми инвариантами, которыми обладает система.

В качестве примера, иллюстрирующего важность учета ограничения доступной области фазового пространства невырожденностью якобиана преобразования Лежандра, рассмотрим точно решаемый случай релятивистского газа в линейном внешнем поле:

$$L = \sum_i \left[-mc^2 \sqrt{1 - v_i^2/c^2} - \lambda |q_i| \left(1 + \frac{v_i^2}{c^2} \right) \right]. \quad (1.27)$$

Определитель тензора масс в этом случае факторизуется:

$$J = \prod_i J_i, \quad J_i = \frac{m}{(1 - v_i^2/c^2)^{3/2}} - \frac{2\lambda |q_i|}{c^2},$$

$$Z = (Z_1)^N, \quad Z_1 = \int_{J_1 > 0} \exp(-\beta E_1) J_1 dq_1 dv_1, \quad (1.28)$$

$$E_1 = \frac{mc^2}{\sqrt{1 - v_1^2/c^2}} + \lambda |q_1| \left(1 - \frac{v_1^2}{c^2} \right).$$

Насколько такая ситуация отличается от классического случая? Введем классический радиус взаимодействия $q_0 = mc^2/2\lambda$ и параметр $b = \beta mc^2$. Тогда получаем для одночастичной статистической суммы Z_1 выражение

$$Z_1 = 2mcq_0 \int_0^1 du \exp(-b/\sqrt{1-u^2}) \times \int_0^{(1-u^2)^{-3/2}} \left[\frac{1}{(1-u^2)^{3/2}} - 2x \right] \exp\left(-\frac{bx}{2}(1-u^2)\right) dx. \quad (1.29)$$

Используя представление функции Мак-Дональда [14]

$$K_\nu(x) = \frac{\sqrt{\pi}}{\Gamma\left(\nu - \frac{1}{2}\right)} \left(\frac{x}{2}\right)^{\nu-1} \int_0^\infty \exp\left(-x\sqrt{1+p^2}\right) p^{2(\nu-1)} dp,$$

получаем точное выражение для статистической суммы:

$$Z_1 = \frac{2mc}{\beta\lambda} \left[\Psi_1(b) + \Psi_1(3b/2) + \frac{4}{b} (\Psi_2(3b/2) - \Psi_2(b)) \right], \quad (1.30)$$

$$\Psi_1(b) = K_1(b) + \frac{1}{b} K_2(b), \quad \Psi_2(b) = K_2(b) - \frac{1}{b} K_1(b).$$

Член $\Psi_1\left(\frac{3b}{2}\right)$ дает принципиальное отличие (1.30) от случая идеального релятивистского газа, которое проявилось из-за ограничения $J > 0$, т.е. из-за условия $\frac{|q|}{q_0} < \frac{1}{(1-v^2/c^2)^{3/2}}$.

Средняя энергия системы определяется формулой

$$\varepsilon = -\frac{\partial \ln Z_1}{\partial \beta}. \quad (1.31)$$

В случае низких температур ($b \gg 1$) используем асимптотическое представление K -функции

$$K_\nu(b) \approx \sqrt{\frac{\pi}{2b}} e^{-b} \left(1 + \frac{4\nu^2 - 1}{8b} + o(1/b) \right).$$

Тогда

$$\varepsilon = -\frac{\partial \ln Z_1}{\partial \beta} = mc^2 + \frac{1}{2\beta} + O(1/b), \quad (1.32)$$

что совпадает с квазиклассическим приближением для средней энергии. В другом предельном случае высоких температур ($b \ll 1$) используем формулу [14]

$$K_n(b) = (-1)^{n+1} I_n(b) \ln\left(\frac{b}{2}\right) + \frac{1}{2} \sum_{k=0}^{n-1} (-1)^k \left(\frac{b}{2}\right)^{2k-n} \frac{(n-k-1)!}{k!} +$$

$$+ \frac{(-1)^n}{2} \sum_{k=0}^{\infty} \left(\frac{b}{2}\right)^{2k+n} \frac{\psi(n+k+1) + \psi(k+1)}{k!(n+k)!},$$

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x), \quad I_n(x) = \sum_{k=0}^{\infty} \left(\frac{b}{2}\right)^{2k+n} \frac{1}{k!(n+k)!}.$$

Тогда получаем вместо (4.10)

$$\varepsilon = \frac{3}{\beta} (1 + o(b)), \quad (1.33)$$

что совпадает с ультрарелятивистским пределом. В случае промежуточных температур надо использовать общую формулу (1.31), из которой нельзя вычленить отдельные вклады свободного движения и эффектов, связанных с сингулярностью преобразования Лежандра.

2.2. Сингулярное множество и хаотизация движения

Принципы продолжения траектории через сингулярное множество, для которого якобиан преобразования Лежандра вырожден, основаны на следующем утверждении.

▼ **Теорема 2.1.** Пусть лагранжиан $L(q, v)$ является непрерывно дифференцируемой функцией координат и дважды непрерывно дифференцируемой функцией скоростей. Тогда множество локальных экстремумов траектории (т.е. точек, в которых $\partial q_i / \partial v_j = 0$) является подмножеством сингулярного множества $S : \det J = 0$.

Доказательство. Поскольку энергия $E = -L + \sum v_i \partial L / \partial v_i$ является непрерывно дифференцируемой функцией своих аргументов, то $\partial E / \partial v_i = \sum_j v_j \partial^2 L / \partial v_i \partial v_j = \sum_j J_{ij} v_j$. Тогда на траектории, т.е. на линии уровня $E = E_0$ справедливы соотношения

$$\frac{\partial q_i}{\partial v_j} = -\frac{\partial E / \partial v_j}{\partial E / \partial q_i} = -\frac{1}{\partial E / \partial q_i} \sum_k J_{jk} v_k. \quad (2.1)$$

Следовательно, нетривиальные (т.е. с не равными нулю скоростями) точки локального экстремума $\partial q_i / \partial v_j = 0$ существуют только в том случае, если $\det J = 0$. ▲

Пусть система состоит из N частиц и размерность одночастичного координатного пространства равна n , то совокупность $(nN)^2$ величин

$$M_{\alpha\beta}^{ij} = \frac{\partial^2 L(q, v)}{\partial v_i^\alpha \partial v_j^\beta} \quad (2.2)$$

образуют тензор массы динамической системы.

Обозначим через $J = \det(M_{\alpha\beta}^{ij})$ определитель соответствующей матрицы. Согласно определениям [2], лагранжиан L называется невырожденным, если этот определитель не обращается в ноль ни в одной точке фазового пространства: $\forall (q, v) \in \gamma : J \neq 0$. Лагранжиан называется сильно невырожденным, если он невырожденный, и уравнения $p_{i\alpha} = \frac{\partial L}{\partial v_i^\alpha}$ имеют единственные непрерывно дифференцируемые и об-

ратимые решения $v_i^\alpha = v_i^\alpha(q, p)$. Отображение $\mathfrak{Z} : \{q, v\} \rightarrow \{q, p\}$ называется преобразованием Лежандра, если L сильно невырожден. Подчеркнем, что классическая статистическая механика построена для сильно невырожденных лагранжианов. Если в некоторой области фазового пространства γ определитель вырожден ($J = 0$), то там ускорения не могут быть представлены как однозначные функции фазовых переменных.

▼ **Определение 2.1.** Лагранжиан будем называть невырожденным почти всюду в γ , если множество, на котором $J = 0$, имеет Лебегову меру нуль. В этом случае сохраним обозначение \mathfrak{Z} за преобразованием Лежандра, которое определено почти всюду в γ . Будем писать $\mathfrak{Z} : \gamma \rightarrow \Gamma; H = \mathfrak{Z}E$. ▲

Классификация особых точек лагранжевых ДС дается следующей теоремой.

▼ **Теорема 2.2.** Особыми точками лагранжевых ДС являются нетривиальные стационарные точки энергии и лагранжиана. Стационарные точки энергии отвечают разрыву компоненты ускорения, касательной к траектории, а стационарные точки лагранжиана — разрыву его нормальной компоненты. ▲

Сингулярным множеством S_D динамической системы называется множество точек, в которых $J = 0$. При этом S_D разбивается в прямую сумму двух множеств: сингулярного множества S_D^1 первого типа, отвечающего разрыву компоненты ускорения, направленной вдоль движения, и сингулярного множества S_D^2 второго типа, отвечающего разрыву нормальной компоненты ускорения.

Подчеркнем, что формально на сингулярном множестве не определены никакие инварианты движения, поскольку они задаются только для регулярных точек. Однако, если инварианты (энергия, импульс, момент импульса и т.п.) непрерывны в фазовом пространстве, их можно по непрерывности продолжить и на множество S . Поскольку движение происходит на пересечении множеств, определяемых инвариантами, появляется следующее условие достижимости сингулярного множества.

▼ **Теорема 2.3.** Пусть динамическая система с лагранжианом $L(q, v)$ имеет сингулярное множество S_D , на котором вырождается преобразование Лежандра. Если она обладает независимыми непрерывно дифференцируемыми в γ инвариантами движения $I_j, j = 1, 2, \dots, j_m \leq nN$, и $S_{I_j} = S_D \cap I_j$ — множество общих точек S_D и I_j , то S_D достижимо тогда и только тогда, когда $\bigcap_j S_{I_j} \neq \emptyset$. Доступная область фазового

пространства системы ограничена поверхностями вырождения якобиана $J = 0$, в окрестности которых нормальная составляющая ускорения направлена против движения. Точками отражения от S_D являются нетривиальные решения системы уравнений $\partial q_i / \partial v_j = 0$, где траектория, зада-

ваемая системой уравнений $I_j(q, v) = I_j^0$, имеет локальный максимум.

▲

Таким образом, сингулярное множество разделяет области фазового пространства с разными знаками $\det J$. Проекция вектора ускорения на направление вектора скорости вблизи S в одном случае положительна (локальный минимум), а в другом — отрицательна (локальный максимум). Следовательно, в первом случае частица скачком, т.е. с разрывом определенной компоненты ускорения, проходит через сингулярность, а во втором отражается от нее. Характерно, что по другую сторону сингулярности траектория меняет тип экстремума, т.е., например, локальный минимум становится локальным максимумом и наоборот, поэтому такое множество особых точек является полупроницаемым.

В случае, если матрица $M_{\alpha\beta}^{ij}$ положительно определена на S_D , траектория имеет на особой поверхности локальный минимум, а если отрицательно — то локальный максимум. В терминах уравнений движения это означает, что если сингулярное множество разделяет области фазового пространства с разными знаками J , то проекция вектора ускорения на направление вектора скорости вблизи S_D в одном случае сонаправлена движению, а в другом — противонаправлена. Поэтому если с одной стороны S_D (т.е. при $t \rightarrow t_s - 0$, где t_s — момент достижения S_D) траектория имеет локальный максимум, то она не может пройти «сквозь» S_D и отражается от особой поверхности, но тогда с другой стороны этого множества траектория имеет локальный минимум (в силу другого знака J) и «проскакивает» на другую сторону S_D . Это означает, что особое множество полупроницаемо. Именно, те области фазового пространства, где ускорение $\mathbf{a}_v = (\mathbf{a}, \mathbf{v})\mathbf{v}/|\mathbf{v}|^2$ направлено противоположно скорости, имеют отражающую границу, а если \mathbf{a}_v направлено вдоль движения, то соответствующая область односторонне проницаема. Аналогично рассматриваются линии уровня и для других инвариантов.

Заметим теперь, что на множестве S^1 поведение ДС детерминировано в силу теоремы 2.1, а на множестве S^2 рассеивание может происходить на любой угол. Это означает, что на S^2 можно ввести априорную

вероятность возникновения движения в направлении, касательном к S^2 . Соответствующие условия войдут как граничные в уравнение Лиувилля, которое будет описывать эволюцию системы в регулярной области с учетом отражения от (или прохождения сквозь) S^1 и хаотизации направления движения при достижении S^2 .

Таким образом, хаотическое поведение динамической системы может возникнуть по причине неоднозначности продолжения траектории через сингулярное множество.

2.3. Уравнение Лиувилля для вырожденных ДС

Запишем общее уравнение, моделирующее вероятностную эволюцию ПФР, где вероятность перехода зависит от точек фазового пространства и момента времени. Рассмотрим только локальные по времени вероятности перехода.

Пусть $f(x, t)$ есть ПФР системы, рассматриваемой в фазовом пространстве Γ точек x в момент времени t . Пусть также для каждой пары точек $x, x' \in \Gamma$ и для любых двух моментов времени t и t' определена условная вероятность $P(x', t' | x, t)$ того, что система находится в момент t' в точке x' при условии, что она находилась в точке x в момент t . Если считать, что в модели отсутствует действие из будущего, то $P(x', t' | x, t) = 0$ при $t' < t$. Тогда вероятность перехода $w(x, x'; t)$ в единицу времени из точки x в точку x' в момент времени t определяется формулой (см. [15]):

$$w(x, x'; t) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} (P(x', t + \tau | x, t) - \delta(x' - x)). \quad (3.1)$$

Поскольку фазовое пространство считается полным, то переход в какую-либо точку обязателен, откуда следует условие нормировки $\int P(x', t' | x, t) dx' = 1$. Тогда из (3.1) для вероятности перехода получаем условие консервативности $\int w(x, x'; t) dx' = 0$.

Поскольку

$$f(x, t + \tau) = \int F(x', t; x, t + \tau) dx' = \int P(x', t | x, t + \tau) f(x', t) dx',$$

то

$$\begin{aligned}\frac{\partial f(x, t)}{\partial t} &= \lim_{\tau \rightarrow 0} \frac{f(x, t + \tau) - f(x, t)}{\tau} = \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left(\int P(x', t \mid x, t + \tau) f(x', t) dx' - f(x, t) \right) = \\ &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \int \left(P(x', t \mid x, t + \tau) - \delta(x - x') \right) f(x', t) dx' = \\ &= \int w(x', x; t) f(x', t) dx' .\end{aligned}$$

Итак, в терминах вероятности перехода уравнение эволюции $f(x, t)$ имеет вид

$$\frac{\partial f(x, t)}{\partial t} = \int f(x', t) w(x', x; t) dx' . \quad (3.2)$$

В случае детерминированной эволюции переход с вероятностью единица осуществляется между точками, лежащими на одной интегральной кривой системы $\dot{x} = v(x)$, и с вероятностью ноль между другими точками. Тогда легко проверить, что уравнение Лиувилля (1.1.12) следует из (3.2), если

$$w(x, x'; t) = \frac{\partial}{\partial t} \delta(x'(t) - x) . \quad (3.3)$$

Действительно, подставляя (3.3) в (3.2) и выполняя дифференцирование, получаем

$$\begin{aligned}\frac{\partial f(x, t)}{\partial t} &= \int \frac{dx'}{dt} \frac{\partial \delta(x' - x)}{\partial x'} f(x', t) dx' = \\ &= - \int \frac{dx'}{dt} \frac{\partial \delta(x' - x)}{\partial x} f(x', t) dx' = - \frac{\partial}{\partial x} (v(x) f(x, t)) .\end{aligned}$$

Предположим теперь, что динамика задается неоднозначным отображением, причем в самой системе отсутствует механизм выбора той или иной траектории. Тогда в каждой точке неоднозначности можно «руками» задать некоторую вероятность выбора определенной траектории. В результате эволюция ПФР сама будет иметь вероятностный характер. Системы могут иметь такое поведение, например, тогда, когда нарушается

гладкость отображения, что и имеет место в случае вырождающейся динамики.

В соответствии с проведенным анализом в п. 2.2 следует дополнить уравнение Лиувилля (1.1.12) граничными условиями для функции распределения после попадания системы на сингулярную фазовую поверхность. Это условие запишем в интегральном виде по аналогии с (3.2). Тогда

$$\rho(x, t_s + 0) = \int P(x, x') \rho(x', t_s - 0) dx' , \quad (3.4)$$

где ядро $P(x, x')$ описывает процесс прохождения траектории через поверхность $S(x)$. Особую поверхность будем считать гладкой, так что в каждой ее точке определен вектор нормали $\mathbf{n} = \nabla S / |\nabla S|$, где градиент берется именно в фазовом пространстве размерности $2n$. В ядре $P(x, x')$ должны быть отражены правила прохождения особой поверхности: ядро отлично от нуля только на особой поверхности, т.е. в нем должен находиться фактор $\delta(S(x'))$; оно должно сохранять пространственные положения частиц после прохождения, что означает наличие фактора $\delta(\mathbf{q} - \mathbf{q}')$; сохранение первых интегралов системы обеспечивается через $\delta(I(x) - I(x'))$, где под $I(x)$ здесь подразумевается набор имеющихся независимых первых интегралов системы. Для простоты рассмотрим только случай с отражающей поверхностью. Тогда в ядре появляется множитель $\delta(\mathbf{x} - 2\mathbf{n}(\mathbf{x}'\mathbf{n}))$. Т.к. в данном случае поверхность не поглощает систему, то $\int P(x, x') dx' = 1$. В итоге получаем

$$P(x, x') = \delta(S(x')) \delta(I(x) - I(x')) \delta(\mathbf{q} - \mathbf{q}') \delta(\mathbf{x} - 2\mathbf{n}(\mathbf{x}'\mathbf{n})) . \quad (3.5)$$

В качестве простого примера применения правила (3.4) рассмотрим систему, порождаемую дифференциальным уравнением первого порядка, не разрешенным относительно производной по времени:

$$y(x, p, t) = 0, \quad p = \dot{x} . \quad (3.6)$$

Достаточным условием разрешимости уравнения (3.6) относительно p является $\partial y / \partial p \neq 0$. В этом случае через каждую такую точку $\{x, t\}$ в расширенном фазовом пространстве, называемую регулярной, проходит ровно одна интегральная кривая. В точке $\{x, t\}$ определено единственное

векторное поле $\{\dot{x}, 1\} \equiv \{p(x, t), 1\}$, где $p(x, t)$ является локальным решением уравнения (3.6) относительно p . Исследуем эволюцию ансамбля таких систем, различающихся начальными условиями, и введем плотность распределения $f(x, t)$ координаты x , учитывая, что эта координата является решением уравнения (3.6). Тогда эволюция распределения $f(x, t)$ в силу системы (3.6) описывается уравнением Лиувилля

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x}(fp) = 0. \quad (3.7)$$

Однако мера dx в общем случае не инвариантна в R^1 : если $p(x, t)$ не равна постоянной, то $\partial p / \partial x \neq 0$. С другой стороны, так как полная вероятность для системы пребывать в фазовом пространстве γ равна единице и сохраняется во времени, то должна существовать плотность динамически-инвариантной меры, такая, что

$$\int f dy = \int f(x, t) \sqrt{g(x, t)} dx = 1,$$

где \sqrt{g} и есть эта плотность. Легко убедиться, что $\sqrt{g} = 1/p$ есть требуемая плотность, т.к. тогда

$$\operatorname{div} fx = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x} (f \sqrt{g} \dot{x}) = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x} (fp \sqrt{g}) = p \frac{\partial f}{\partial x},$$

и уравнение (3.7) можно записать в дивергентной форме закона сохранения.

В случае же, если $\partial y / \partial p = 0$, такой меры в общем случае может и не быть, так как скорость p может быть не определена. Если $\partial^2 y / \partial p^2 \neq 0$, то уравнение $\partial y / \partial p = 0$ можно локально разрешить относительно p , после чего исключить p и из исходного уравнения (3.6). Получившаяся в результате дискриминантная кривая $S(x, t) = F(x, p(x, t), t) = 0$ в расширенном фазовом пространстве $\{x, t\}$ является особым решением, имеющим общую точку с каждой регулярной интегральной кривой. Движение ансамбля систем в особых точках не определено, так как в них нет единственности решения уравнения (3.6). Следовательно, на дискриминантной кривой следует поставить граничные условия для плотности вероятности $f(x, t)$, которые могут быть

сформулированы в виде некоторого правила рассеяния вероятности на особой траектории.

Для иллюстрации обобщения статистического описания на ДС с особенностями рассмотрим конкретный пример. Пусть динамика системы задана обыкновенным дифференциальным уравнением:

$$y(x, \dot{x}, t) = 5x + \dot{x}^2 - t(t + \dot{x}) = 0. \quad (3.8)$$

Вводя $p(t) = \dot{x}$ и решая совместно с (3.8) уравнение $\partial y / \partial p = 0$, которое имеет вид $2p - t = 0$, находим множество S особых точек уравнения (3.8):

$$S(x, t) = \frac{t^2}{4} - x = 0. \quad (3.9)$$

Регулярные траектории уравнения (3.8) имеют вид

$$x = -\frac{1}{5}C^2 + Ct - t^2, \quad (3.10)$$

где C — произвольный параметр. В зависимости от значений C картина интегральных кривых уравнения (3.8) показана на рис. 2.3.1.

Выберем определенное направление движения по интегральным кривым, например, в сторону возрастания параметра t . Задача статистической механики, порождаемой этой системой, состоит в следующем. Пусть в некоторый момент t_0 заданы вероятности реализации особого решения f_s и регулярных решений f_r , а также соответствующие вероятности перехода. Требуется определить плотность вероятности в заданной точке фазового пространства, т.е. решить уравнение эволюции для ПФР.

На особой кривой функция распределения f_s , как и скорость, зависит только от одного параметра, например, от x . Пусть $\alpha(x)$ есть вероятность ухода в единицу времени системы с особого решения на регулярное, а $\beta(x)$ — с регулярного на особое. Обозначим $f_r^\pm(x)$ пределы $f_r(x, t)$ при $t \rightarrow t_s \pm 0$, где t_s есть момент попадания выбранной регулярной траектории на особое решение.

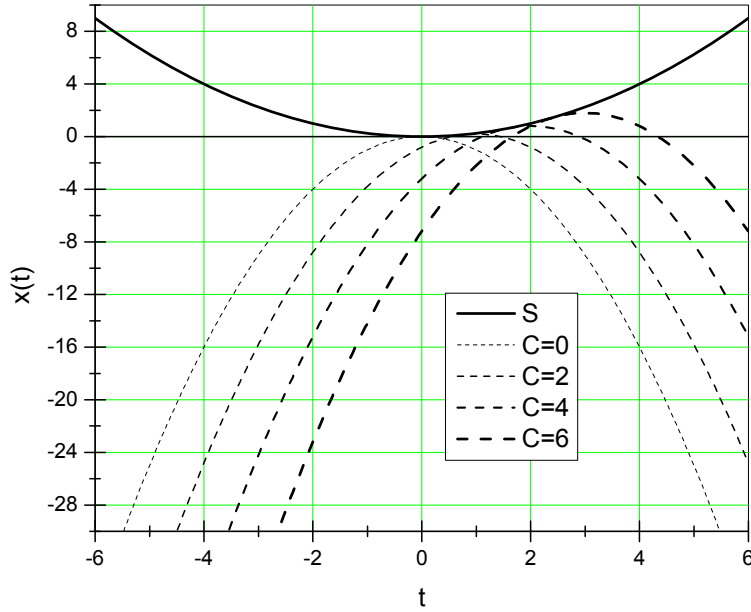


Рис. 2.3.1 – Интегральные кривые уравнения (3.8)

Из (3.9), (3.10) следует, что скорости $p_r(x, t)$, $p_s(x, t)$ на регулярной и особой кривых имеют вид

$$p_r(x, t) = \frac{1}{2} \left(t + \sqrt{5t^2 - 20x} \right), \quad p_s(x, t) = \frac{t}{2} = \pm \sqrt{x} \equiv p_s(x). \quad (3.11)$$

Уравнение Лиувилля для ансамбля таких динамических систем запишется в следующем виде:

$$\begin{cases} 2\sqrt{x} \frac{\partial f_s(x)}{\partial x} = -\frac{1}{2\sqrt{x}} f_s(x) - \alpha(x) f_s(x, t) + \beta(x) f_r^-(x), \\ \frac{\partial f_r(x, t)}{\partial t} + \frac{\partial}{\partial x} (p_r(x, t) f_r(x, t)) = \\ = (-\beta(x) f_r^-(x, t) + \alpha(x) f_s(x)) \delta(S(x, t)). \end{cases} \quad (3.12)$$

В результате движение в регулярной области фазового пространства содержит случайные источники в правой части уравнения (3.12).

2.4. Гамильтоновы системы с высшими производными

Вариационные задачи с высшими производными возникают в различных моделях: это релятивистские системы с ограниченным, подобно скорости, ускорением, некоторые модели квантовой теории поля, включая случай Дирака глобального вырождения лагранжиана, а также лагранжианы, зависящие от главных кривизн траектории.

Рассмотрим лагранжиан $L(q_1, \dots, q^{(N)})$, зависящий от конечного набора высших производных $q^{(s)} = d^s q / dt^s$ как независимых переменных. Уравнения движения порожаемой им ДС имеют вид

$$\dot{p}_i = \frac{\partial L}{\partial q_i}, \quad p_i = \sum_{s=0}^{N-1} \left(-\frac{d}{dt} \right)^s \frac{\partial L}{\partial q_i^{(s+1)}}. \quad (4.1)$$

Обычно в силу физических представлений о системе члены в уравнении Эйлера–Лагранжа, отвечающие высшим производным, предполагаются малыми. Это означает, что рассматривается сингулярно-возмущенная классическая задача, и возникает вопрос о невырожденности преобразования Лежандра.

В качестве иллюстративного примера возьмем модель «свободного движения» со степенной зависимостью лагранжиана от первого ускорения:

$$L = \frac{mv^2}{2} + \frac{|a|^l}{2A}, \quad \mathbf{a} = d\mathbf{v} / dt \in R^n, \quad A = \text{const} \in R. \quad (4.2)$$

Далее будем обозначать $\mathbf{w} = \dot{\mathbf{a}}$. Согласно (4.1), уравнения движения в этом частном случае имеют вид

$$\begin{aligned} m\mathbf{a}^\alpha &= \frac{l}{2A} |a|^{l-2} \left(\delta^{\alpha\beta} + (l-2) \frac{a^\alpha a^\beta}{|a|^2} \right) \dot{w}^\beta + \\ &+ \frac{l(l-2)}{2A} \left[(l-4) |a|^{l-6} (\mathbf{a}\mathbf{w})^2 a^\alpha + |a|^{l-4} \left(|w|^2 a^\alpha + 2(\mathbf{a}\mathbf{w}) w^\alpha \right) \right]. \end{aligned} \quad (4.3)$$

Уравнения (4.3) имеют инварианты: импульс $\mathbf{P} = \mathbf{P}_0$ и энергию $E = E_0$. Энергию можно записать в виде

$$E = \mathbf{P}_0 \mathbf{v} + \mathbf{a} \frac{\partial \mathcal{L}}{\partial \mathbf{a}} - L = \mathbf{P}_0 \mathbf{v} + \frac{l-1}{2A} |a|^l - \frac{mv^2}{2}. \quad (4.4)$$

Из (4.3) следует, что при всех конечных значениях A и при $l \neq 1$ лагранжиан невырожден. Если же $l = 1$, то при $n > 1$ лагранжиан вырожден во всем фазовом пространстве. Из (4.4) следует, что тогда

$$(m\mathbf{v} - \mathbf{P}_0)^2 = P_0^2 - 2mE = \text{const} \geq 0,$$

и потому допустимые скорости определяются выражением

$$\mathbf{v}(t) = \mathbf{P}_0 / m + \mathbf{n}(t) \sqrt{P_0^2 - 2mE}, \quad (4.5)$$

где $\mathbf{n}(t)$ — произвольно направленный единичный вектор.

Сингулярное свойство лагранжиана (4.2) проявилось здесь в том, что движение не определено однозначно. Это находится в согласии с вышеприведенной трактовкой сингулярных поверхностей в фазовом пространстве как множеств, могущих порождать хаотизацию движения.

При $l = 2$ лагранжиан (4.2) невырожден и уравнения (4.3) легко интегрируются:

$$\mathbf{q}(t) = \mathbf{q}_0 + \mathbf{v}_0 t - \frac{\mathbf{a}_0}{\omega_s^2} (\cos \omega_s t - 1) - \frac{\mathbf{w}_0}{\omega_s^3} (\sin \omega_s t - \omega_s t), \quad (4.6)$$

где $\omega_s = \sqrt{mA}$ — частота свободных вращений, совершаемых частицей.

Индексом «0» обозначены начальные значения фазовых переменных. Если $A < 0$, то в (4.6) получаем экспоненциально растущие члены, которые могут быть исключены только нулевыми значениями начальных ускорений, что, в таком случае, будет эквивалентно классическому движению. При $A \rightarrow +\infty$ решение (4.6) переходит в классическое свободное движение материальной точки массы m . Последнее означает, что решения, представляющие собственные вращения частицы, имеют амплитуду, стремящуюся к нулю с увеличением параметра A , и потому не являются сингулярными. Это позволяет рассматривать такие решения как некоторую классическую интерпретацию спина частиц.

Рассмотрим теперь некоторые сингулярные траектории, порождаемые квадратичным лагранжианом с взаимодействием:

$$L_N = -\frac{1}{2} \sum_{s=1}^N (-1)^s m_s |\mathbf{q}^{(s)}|^2 - \Pi(q), \quad m_s \geq 0. \quad (4.7)$$

Этот лагранжиан является сильно невырожденным при $m_N > 0$, и единственная особенность решений уравнений Эйлера–Лагранжа

$$\sum_{s=1}^N m_s q^{\alpha^{(2s)}} + \frac{q^\alpha}{q} \Pi'(q) = 0 \quad (4.8)$$

возникает в случае предельного перехода $m_N \rightarrow +0$.

▼ **Определение 4.1.** Решение $\mathbf{q}(t)$ уравнений (5.34) будем называть регулярным по параметру $m_s, s = 1, \dots, N$, если предельная траектория $\mathbf{q}^*(t) = \mathbf{q}(t; m_s \rightarrow +0)$ существует и является экстремалью лагранжиана $L^* = L_N(m_s \rightarrow +0)$. ▲

Построим класс сингулярных решений уравнений движения (4.8) на сфере в R^n . Полагая $\mathbf{q} = R\mathbf{r}$, $R = \text{const}$, $|\mathbf{r}| = 1$, и обозначая $m_0 = \frac{1}{q} \Pi'(q) \Big|_{q=R}$, получаем из (4.8) линейное уравнение с постоянными коэффициентами:

$$\sum_{s=0}^N m_s r^{\alpha^{(2s)}} = 0, \quad \alpha = 1, \dots, n.$$

Для исследования принадлежности решения этого уравнения сфере в R^n удобно перейти к безразмерным величинам:

$$c_s = \sqrt{m_s / m_N}, \quad \lambda_s = \left(c_s c_{N-1}^{s-N} \right)^2, \quad \tau = c_{N-1} t. \quad (4.9)$$

Сохраняя для производной по безразмерному времени τ прежнее обозначение $r^{(s)}$, записываем уравнение Эйлера–Лагранжа в виде

$$r^{\alpha^{(2N)}} + r^{\alpha^{(2N-2)}} + \sum_{s=0}^{N-2} \lambda_s r^{\alpha^{(2s)}} = 0. \quad (4.10)$$

Если движение происходит на сфере, то соответствующее характеристическое уравнение

$$X_N(\omega^2) = \omega^{2N} - \omega^{2N-2} + \sum_{s=0}^{N-2} (-1)^s \lambda_s \omega^{2s} = 0 \quad (4.11)$$

относительно ω может иметь только простые действительные решения. Обозначим положительные решения уравнения (4.11), занумерованные в порядке возрастания, через ω_k , $k=1, \dots, \tilde{N} \leq N$. Тогда общее решение уравнения (4.10) имеет вид

$$\mathbf{r} = \sum_{k=1}^{\tilde{N}} (\mathbf{a}_k \cos \omega_k \tau + \mathbf{b}_k \sin \omega_k \tau). \quad (4.12)$$

Это решение будем называть полным, если постоянные векторы \mathbf{a}_k , \mathbf{b}_k выбраны так, что все допустимые частоты представлены в решении. Условие $|\mathbf{r}|^2 = 1$ накладывает ограничения на значения частот, с которыми может реализоваться такое движение, т.к. из (4.12) следует

$$\begin{aligned} & \sum_{k=1}^{\tilde{N}} \left[\frac{a_k^2 + b_k^2}{2} + \frac{a_k^2 - b_k^2}{2} \cos 2\omega_k \tau + \mathbf{a}_k \mathbf{b}_k \sin 2\omega_k \tau \right] + \\ & + \sum_{k < j} \left[(\mathbf{a}_k \mathbf{b}_j + \mathbf{a}_j \mathbf{b}_k) \sin(\omega_k + \omega_j) \tau - \right. \\ & \quad \left. - (\mathbf{a}_k \mathbf{b}_j - \mathbf{a}_j \mathbf{b}_k) \sin(\omega_k - \omega_j) \tau \right] + \\ & + \sum_{k < j} \left[(\mathbf{a}_k \mathbf{a}_j + \mathbf{b}_j \mathbf{b}_k) \cos(\omega_k - \omega_j) \tau + \right. \\ & \quad \left. + (\mathbf{a}_k \mathbf{a}_j - \mathbf{b}_j \mathbf{b}_k) \sin(\omega_k + \omega_j) \tau \right] \equiv 1. \end{aligned} \quad (4.13)$$

Поэтому, если все \tilde{N}^2 комбинаций частот в (4.13) различны, то все $2\tilde{N}$ векторов \mathbf{a}_k , \mathbf{b}_k должны быть взаимно ортогональны, поэтому решение может быть полным только при $2\tilde{N} \leq n$. Из (4.13) следует, что набор осцилляций $\{\omega_1, \dots, \omega_N\}$ можно реализовать на сфере в R^3 тогда и только тогда, когда для любой пары частот среди четырех комбинаций $\{2\omega_k, 2\omega_s, |\omega_k \pm \omega_s|\}$ найдется хотя бы одна, совпадающая с какой-

нибудь аналогичной комбинацией для другой пары частот, либо $\omega_k = 3\omega_s$.

Например, для $N=2$ в R^3 существует единственная возможность сферического движения, когда корни полинома (4.11), равные $\omega_{1,2}^2 = \frac{1}{2}(1 \mp \sqrt{1-4\lambda_0})$, удовлетворяют условию $\omega_2 = 3\omega_1$. Это при-

водит к тому, что $\omega_1^2 = 1/10$, $\omega_2^2 = 9/10$, $\lambda_0 = 9/100$, т.е. фиксированы как частоты вращения, так и потенциал взаимодействия (т.е. радиус сферы). В поле с периодическим потенциалом соответствующих радиусов сферы счетное множество. В поле кулоновского потенциала $\Pi = -\alpha/r$ такой радиус единственный, который выражается формулой:

$$R = \left(\frac{\alpha m_2}{\lambda_0 m_1^2} \right)^{1/3} = \left(\frac{100\alpha}{9Am^2} \right)^{1/3}.$$

Общее решение в этом случае определяется (с точностью до ориентации системы координат $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$) двумя параметрами χ, ς : $\chi^2 + \varsigma^2 \equiv \rho^2 \leq 1$, так что

$$\begin{aligned} \mathbf{r}(\tau) = & \left(\frac{\chi^2 - \varsigma^2}{4h} \cos \omega \tau - \frac{\chi \varsigma}{2h} \sin \omega \tau - h \cos 3\omega \tau \right) \mathbf{e}_1 + \\ & + \left(-\frac{\chi^2 - \varsigma^2}{4h} \sin \omega \tau - \frac{\chi \varsigma}{2h} \cos \omega \tau + h \sin 3\omega \tau \right) \mathbf{e}_2 + \\ & + (\chi \cos \omega \tau + \varsigma \sin \omega \tau) \mathbf{e}_3, \end{aligned} \quad (4.14)$$

где $\omega = 1/10$, а h есть корень уравнения $h^2 - |h| + \rho^2/4 = 0$. Длина замкнутой траектории решения (4.14) выражается эллиптическим интегралом:

$$\begin{aligned} l = & 4\sqrt{1+8h^2+2\rho^2} [E(\varphi_1, k) - E(\varphi_2, k)], \\ & k = 2\rho / \sqrt{1+8h^2+2\rho^2}, \\ & \varphi_{1,2} = \arcsin \left(\sqrt{\rho^2 \pm (\chi^2 - \varsigma^2)} / \rho\sqrt{2} \right). \end{aligned} \quad (4.15)$$

Построенное решение (4.14) не имеет предела при $\omega \rightarrow \infty$, т.е. является примером сингулярной траектории.

Следует отметить, что в отдельных случаях переход к более высокой размерности в уравнении движения ДС позволяет снять эффект вырождения. Так, слаборелятивистский лагранжиан Дарвина (1.15) является разложением в ряд Тейлора запаздывающих потенциалов [12]. Для оценки влияния ускорений надо учесть соответствующую зависимость в запаздывании взаимодействия. Рассмотрим для простоты две заряженных частицы. Функция Лагранжа прямого взаимодействия, зависящая от ускорений всех порядков, получена в [16], и имеет вид

$$L = e_1 e_2 \sum_{s=0}^{\infty} (-1)^s \frac{(D_1 D_2)^s}{(2s)! c^{2s}} \left(1 - \frac{\mathbf{v}_1 \mathbf{v}_2}{c^2} \right) \Phi_s(|\mathbf{q}_{12}|);$$

$$D_i = \sum_{s=0}^{\infty} q_i^{(s+1)} \frac{\partial}{\partial q_i^{(s)}}, \quad (4.16)$$

$$\frac{\Phi'_s(r)}{r} = (2s-1)\Phi_{s-1}(r), \quad \Phi_0(r) = 1/r.$$

Ограничиваясь только ускорением первого порядка, т.е. членами порядка $O(c^{-4})$, и добавляя в (4.16) «ускорение собственных вращений» квадратичного вида, как в (4.7), получаем из (4.16) функцию Лагранжа в виде

$$L = - \sum_{i=1}^2 \left(m_i c^2 \sqrt{1 - \frac{v_i^2}{c^2}} - \frac{a_i^2}{2A_i} \right) -$$

$$- \frac{e_1 e_2}{r} \left[1 - \frac{\mathbf{v}_1 \mathbf{v}_2 + (\mathbf{n} \mathbf{v}_1)(\mathbf{n} \mathbf{v}_2)}{2c^2} + \frac{\mathbf{a}_1 \mathbf{a}_2 + 3(\mathbf{a}_1 \mathbf{n})(\mathbf{a}_2 \mathbf{n})}{4c^4} r^2 \right] +$$

$$+ \frac{e_1 e_2}{4c^4} \left[(\mathbf{a}_1 \mathbf{v}_2)(\mathbf{n} \mathbf{v}_2) + (\mathbf{a}_2 \mathbf{v}_1)(\mathbf{n} \mathbf{v}_1) + 3 \sum_{i=1}^2 (v_i^2 - (\mathbf{n} \mathbf{v}_i)^2)(\mathbf{a}_i \mathbf{n}) \right] +$$

$$+ O\left(\frac{v^4}{c^4}\right), \quad (4.17)$$

где $\mathbf{n} = \mathbf{q}_{12}/q_{12}$, $q_{12} \equiv r$, а члены $O(v^4/c^4)$ обозначают вклады четвертого порядка только по скоростям, которые будут несущественны при исследовании (4.17) на невырожденность.

Лагранжиан (4.17) невырожден, если

$$J = \det \left(\frac{\partial^2 L}{\partial a_i^\alpha \partial a_j^\beta} \right) = \left(\frac{1}{A_1 A_2} - \frac{e_1^2 e_2^2 r^2}{16c^8} \right)^{n-1} \left(\frac{1}{A_1 A_2} - \frac{5e_1^2 e_2^2 r^2}{8c^8} \right) \neq 0$$

(4.18)

Видно, что без членов «ускорений собственных вращений» ($A_1 \rightarrow +\infty, A_2 \rightarrow +\infty$) определитель невырожден, т.е. постпостгалилеево приближение невырождено. Это — важное обстоятельство, поскольку оно могло бы позволить перейти к гамильтонову описанию во всем фазовом пространстве. Однако, соответствующий гамильтониан не является положительно определенным по импульсам, и потому для него нельзя будет построить равновесную термодинамику. Ее, впрочем, по той же причине нельзя построить и в лагранжевых переменных, т.к. для сходимости статистической суммы члены, зависящие от ускорений, следует полагать распределенными с отрицательной температурой, для чего в настоящий момент нет достаточных оснований.

В каждом из примеров ДС, рассмотренных в этом параграфе, уравнение Лиувилля дополняется, если необходимо, граничными условиями на сингулярном множестве, причем эти условия могут быть заданы в виде некоторых вероятностных распределений. Тем самым в исходную динамику вносится элемент случайности.

2.5. Замечания о дискретизации вырожденных уравнений движения

При численном моделировании динамических систем с вырождением возникает вопрос о том, как моделировать сингулярность, которая реализуется по нулевой мере в фазовом пространстве. Существует несколько путей решения этой задачи, приводящих к принципиально различным результатам.

Например, если исходить из представления о попадании системы на сингулярное решение, то приближенная численная реализация такого движения должна выбирать такую последовательность фазовых точек, которая точно попадает на сингулярное множество. Если дискретная схе-

ма будет выбрана без учета этого требования, то может оказаться, что особые точки ДС игнорируются.

Представляется, что дискретная схема должна адекватно передавать приближенный характер любого реального движения, но в то же время сохранять и специфику теоретической модели. Применительно к сингулярному множеству это означает, что вместо множества нулевой меры следует выбрать определенную ширину дискретной ячейки, отвечающую фактическому вырождению. Поскольку такой выбор должен исходить из представлений о точности реализации модели, полезно провести классификацию ошибок, возникающих на пути численного моделирования ДС. Фактически возникают ошибки пяти типов, каждый из которых существенен как с качественной, так и с количественной точек зрения.

Первый тип ошибок порождается самим вычислительным инструментом, имеющим конечную точность (компьютером с конечной длиной машинного слова). Поэтому теоретическая формула численно будет воспроизведена заведомо не точно. Во многих случаях, связанных с вырождением, эта неточность фатальна для модели, поскольку сингулярность обычно локализована в пространстве меньшей размерности, чем фазовое, и, в частности, может находиться в точке с иррациональной координатой.

Следовательно, чтобы сохранить за численной реализацией модели важные, с точки зрения исследователя, особенности, их следует принудительно размазать в некоторой области, т.е. «офизичить» теоретическую модель. Это приводит к ошибке второго типа, которая несколько отличается по содержанию от обычных ошибок дискретизации, стремящихся к нулю при уменьшении шага. Именно, в модель вводится понятие «характерного расстояния» или «размера нуля», когда всякое число, меньшее определенной величины по модулю, считается не отличимым от этой величины. В частности, вырождение в точке следует заменить на вырождение в интервале характерного размера $(-\varepsilon, \varepsilon)$. Этот характерный размер задается в рамках модели совершенно произвольно, требуется лишь, чтобы он был больше минимального числа, воспринимаемого шкалой компьютера.

Далее, если модель призвана объяснить реально получаемые данные, то надо учесть, что эти данные сами получены с ошибкой, характерной для инструмента наблюдения, и потому к ним не стоит относиться как к «абсолютно точным». Третий тип ошибок состоит, таким образом, в том,

что значения наблюдаемых величин оснащаются параметром ε' , который определяет величину неточности в определении исходных данных.

Следует также учесть, что сама исследуемая физическая система имеет свой собственный порог чувствительности ε'' , ниже которого она не реагирует на изменение своих параметров. Величину ε следует в таком случае соотнести с реальной физической системой, т.е. понять, насколько точно сама система распознает, в какой точке своего фазового пространства она находится. Этот четвертый тип ошибки не зависит от наблюдателя, если предположить, что его присутствие не искажает динамику системы, в модели которой изначально не было места для наблюдателя.

Пятый тип ошибок не связан собственно с возможным сингулярным характером динамических траекторий, а присущ всякой вообще численной аппроксимации производной. Проблема состоит в том, что задача вычисления производной от функции, значения которой известны не везде, а лишь в узлах сетки, является некорректной с той точки зрения, что малым погрешностям в задании функции могут отвечать большие ошибки в вычислении производной. Это означает, что численные методы, вообще говоря, не позволяют реализовать непрерывную зависимость значений производной от значений функции. Поэтому шаг дискретизации не может быть выбран сколь угодно малым не по философским соображениям, апеллирующим к конечным атомарным размерам и к учету разного рода квантовых эффектов, а в силу того, что этот шаг должен быть согласован с точностью вычисления значений самой функции. Пусть, например, задано некоторое разбиение временной оси на моменты t_n , в которые вычисляются значения координаты $x(t)$ с шагом τ как решения некоторого дифференциального уравнения. Значения же этой координаты известны с точностью ε' , т.е. вместо точных значений $x(t_n)$ используются приближенные $\tilde{x}(t_n)$, причем из предыдущего анализа следует, что точные значения нам не известны в принципе. Таким образом, в каждый момент времени выполняется оценка $|x(t_n) - \tilde{x}(t_n)| \leq \varepsilon'$, которая, строго говоря, может не быть равномерной по времени. Но даже при условии равномерной ограниченности ошибки измерения аппроксимация производной dx/dt в точке t_n по формуле

$$\frac{dx(t_n)}{dt} = \frac{\tilde{x}(t_{n+1}) - \tilde{x}(t_n)}{\tau}$$

имеет порядок не $O(\tau)$ в соответствии с представлением $x(t + \tau) = x(t) + \tau x'(t) + o(\tau)$, а $O(\tau) + 2\varepsilon' / \tau$. Погрешность в вычислении производной неточно заданной функции зависит от погрешности аппроксимации производной $O(\tau)$ и от погрешности вычисления значений самой функции. Вклад от последней неограниченно возрастает при $\tau \rightarrow 0$, если шаг по времени не согласован с точностью ε' . Очевидно, чтобы обе погрешности имели один и тот же порядок малости, требуется выполнение условия согласования, которое в данном случае имеет вид

$$\tau = O(\sqrt{\varepsilon'}). \quad (5.1)$$

Все эти типы ошибок следует учитывать при формировании требований к точности воспроизведения временного ряда с помощью компьютерной модели. Невозможно, да в этом и нет необходимости идеально воспроизводить неточно измеряемые данные. Самым удобным для исследователя является вариант, когда $\varepsilon \leq \varepsilon' \leq \varepsilon''$. Если же $\varepsilon'' < \varepsilon'$, то неточность измерений может свести на нет существенные для модели аспекты поведения, такие, как чувствительность к начальным условиям.

Глава III. Хаотические динамические системы

3.1. Хаотическое движение гладких гамильтоновых систем

В этой главе будут рассмотрены различные варианты возникновения хаотического, т.е. по видимости нерегулярного, поведения траекторий динамических систем. Изложение носит ознакомительный характер, поскольку по теории динамических систем существует большое количество специальной литературы, в том числе и по хаотическим аспектам динамики. Цель этой главы состоит в том, чтобы дать представление о возможных способах интерпретации моделей нестационарных временных рядов. Поскольку модели эти будут строиться с помощью кинетических уравнений, то естественно возникает аналогия с динамической системой, которой отвечает, например, эмпирическое уравнение Лиувилля или другое модельное уравнение, приближенно описывающее наблюдаемую эволюцию выборочных средних значений временного ряда или эволюцию выборочной функции распределения его значений. Следовательно, надо представлять себе, какого типа ДС могут использоваться для моделирования хаотического поведения временных рядов. Это могут быть гладкие, но неинтегрируемые гамильтоновы системы, нелинейные дискретные отображения, вырожденные динамические системы, а также системы, заданные на фрактальных множествах. В настоящем параграфе рассматриваются гладкие гамильтоновы системы.

Напомним некоторые факты из курса дифференциальных уравнений.

▼ **Теорема 1.1.** Если в уравнении $dx/dt = g(x)$, $x \in \mathbb{R}^n$, векторное поле $g(x)$ является дифференцируемым (или, в более слабой форме, удовлетворяет условию Липшица), то через каждую точку расширенного фазового пространства проходит одна и только одна интегральная кривая решения. ▲

▼ **Теорема 1.2 (Лиувилль).** Пусть ДС гамильтонова с гамильтонианом $H(q, p)$, $(q, p) \in \mathbb{R}^{2n}$. Если эта система имеет n независимых первых интегралов $F_j(q, p)$, $j = 1, 2, \dots, n$, которые находятся в инволю-

ции, т.е. $[F_i, F_j] = 0$, то траектории системы лежат на n -мерном инвариантном торе T^n , движение по которому является условно-периодическим с n частотами $\omega_k(F_1, \dots, F_n)$, и угловые переменные θ_k , характеризующие координаты на торе, удовлетворяют уравнениям $d\theta_k / dt = \omega_k$. ▲

В случае выполнимости теоремы Лиувилля существует переход к переменным «действие-угол», причем преобразованный гамильтониан зависит только от переменных действия I_k . От них же зависят и первые интегралы. Такая система называется полностью интегрируемой по Лиувиллю. Неинтегрируемость возникает тогда, когда указанные преобразования выполнить невозможно. В качестве примера часто приводят ДС Энона–Эйлеса, задаваемую гамильтонианом двумерного осциллятора с кубической нелинейностью:

$$H = \frac{1}{2}(p_1^2 + q_1^2 + p_2^2 + q_2^2) + q_1^2 q_2 - q_2^3. \quad (1.1)$$

Известно [1], что у этой системы есть только один первый интеграл (энергия). При $H = 1/6$ на изоэнергетической поверхности система (1.1) демонстрирует хаотическое поведение.

Если представить неинтегрируемую ДС в виде

$$H = H_0(I_1, \dots, I_n) + V(I, \theta), \quad (1.2)$$

где H_0 отвечает интегрируемой системе, а добавка $V(I, \theta)$ трактуется как возмущение, то инвариантные торы «идеальной» системы с гамильтонианом H_0 при добавлении неинтегрируемого возмущения разрушаются, и движение внешне приобретает хаотические черты, хотя по-прежнему выполнены условия локальной интегрируемости в смысле существования интегральной кривой по теореме 1.1.

▼ **Опр. 1.1.** Инвариантный тор T^n называется резонансным, если существуют некоторые не все равные нулю целые числа k_i такие, что

$$\sum_{i=1}^n k_i \omega_i = 0. \quad \blacktriangle \quad (1.3)$$

Если ДС интегрируема, то ее фазовое пространство состоит из резонансных и нерезонансных торов, причем ни один из них не является изолированным либо предельным. Возмущение разрушает резонансные торы и, возможно, некоторые нерезонансные. Движение в этих разрушенных областях не является квазипериодическим и представляет пример хаотической динамики, называемой диффузией Арнольда.

Другой пример хаотичности движения дают негамильтоновы ДС, обладающие так называемым стохастическим аттрактором.

▼ **Опр. 1.2.** Автономная система обыкновенных дифференциальных уравнений называется перемешивающей на инвариантном компактном множестве A , если для любых двух областей $\Omega_0 \subset A$ и $M \subset A$ существует конечный предел

$$\lim_{t \rightarrow \infty} \frac{\text{mes}(\Omega_t \cap M)}{\text{mes} M} = \frac{\text{mes}(\Omega_0)}{\text{mes} A}. \quad (1.4)$$

Аттрактор, движение по которому обладает свойством перемешивания, называется стохастическим. ▲

В случае, если аттракторы обладают масштабной инвариантностью, они называются странными или обладающими фрактальной структурой. Такие аттракторы не являются конечным объединением подмногообразий фазового пространства. Они характерны для так называемых диссипативных систем и имеют дробную размерность Хаусдорфа.

Гладкие ДС обычно не рассматриваются как модельные прототипы временных рядов. Краткое упоминание о них вызвано тем, что в отдельных случаях к таким системам можно свести эмпирическую эволюцию выборочных моментов, трактуемую как непрерывный случайный процесс. Тем не менее, больший интерес представляют ДС с дискретным временем, поскольку к ним в явном виде сводятся регрессионные и авторегрессионные модели временных рядов.

3.2. Хаотические динамические системы с дискретным временем

Автономные ДС в моделях с дискретным временем задаются рекуррентными соотношениями (так называемым каскадом g)

$$x(t+1) = g(x(t), x(t-1), \dots, x(t-k)) \quad (2.1)$$

и соответствующим количеством начальных условий. Переходя к многомерному представлению, рекурсию (2.1) можно записать как рекурсию первого порядка относительно $k+1$ -мерного вектора. Для удобства записи параметр t будем иногда заменять на индекс последовательности n , так что рекурсия запишется в виде

$$x_{n+1} = g(x_n) \equiv g^n(x_1). \quad (2.2)$$

Итеративная процедура, задаваемая отображением (2.2), определяет траекторию как последовательность значений $\{x_1, x_2, \dots, x_n, \dots\}$.

Траектория ДС $x_{n+1} = g(x_n)$ называется периодической (циклом), если существует такое натуральное K , что для некоторого n оказалось $x_{n+K} = x_n$. Число K называется периодом. Интерес представляют незамкнутые траектории, отличные от циклов. Совокупность траекторий, получаемых при реализации различных начальных условий, образует статистический ансамбль.

Пусть $f_0(x)$ представляет начальную плотность распределения величины x , заданную на конечном измеримом множестве. Тогда, действуя так же, как при выводе уравнения Лиувилля для непрерывных систем (1.1.12), получаем его аналог в дискретном случае, называемый уравнением Перрона–Фробениуса [17]:

$$f_{n+1}(x) = \int f_n(y) \delta(x - g(y)) dy. \quad (2.3)$$

Если функция $g(x)$ имеет производную, то (2.3) можно переписать в виде

$$f_{n+1}(x) = \sum_k |g'(y_k)|^{-1} f_n(y_k), \quad (2.4)$$

где суммирование ведется по всем корням уравнения $x = g(y)$.

Стационарные точки отображения (2.2) определяются решениями уравнения (если таковые есть) $x = g(x)$. В свою очередь динамически-инвариантные меры (не обязательно единственные) определяются стационарными точками отображения (2.3):

$$f(x) = \int f(y) \delta(x - g(y)) dy. \quad (2.5)$$

Правая часть отображения, задаваемого формулой (2.5), называется оператором Перрона–Фробениуса \hat{P} . Таким образом, инвариантная плотность является собственной функцией оператора \hat{P} , отвечающей собственному значению $\lambda = 1$. Оператор, сопряженный к оператору \hat{P} , называется оператором Купмана \hat{Q} . Он определяется равенством

$$\hat{Q}f(x) = f(g(x)). \quad (2.6)$$

Пусть $f_{inv}(x)$ — инвариантная плотность одномерного отображения $x_{n+1} = g(x_n)$. Его автокорреляционной функцией (АКФ) $A(n)$, $n > 1$ называется величина

$$A(n) = \int x g^n(x) f_{inv}(x) dx - \int x f_{inv}(x) dx \int g^n(y) f_{inv}(y) dy. \quad (2.7)$$

Естественно предположить, что если дискретная АКФ в чем-то подобна АКФ случайных процессов, то свойства траекторий таких динамических систем будут внешне походить на траектории случайных процессов.

Динамическим белым шумом называется такое отображение $x_{n+1} = g_0(x_n)$, для которого $A(n) = \delta_{n0}$. Примеры таких ДС существуют. Это, например, рассматриваемое ниже логистическое отображение. Тем самым к рассмотрению привлекаются системы, эволюция которых вполне детерминирована, но в ней присутствуют и некоторые хаотические черты. Дискретные системы с динамическим хаосом являются естественными кандидатами на тестирование моделей временных рядов, поэтому им стоит уделить особое внимание. Ниже рассматриваются некоторые «популярные» модели динамического хаоса и изучаются их статистические свойства.

Системы, задаваемые уравнением вида (2.1), полностью детерминированы. Однако на практике детерминированную систему не всегда бывает легко отличить от хаотической по виду ее траекторий. Это связано с

тем, что реальные измерения всегда имеют некоторую погрешность ε . Если различие в начальных данных меньше ε , то с этой точностью данные воспринимаются как совпадающие. Тогда наблюдатель естественно ожидает, что если из каждой точки, взятой из малой области в ε -окрестности некоторого начального условия, выпустить траектории, то в дальнейшем все эти траектории в любой момент времени также будут находиться если и не в ε -окрестности, то достаточно близко одна к другой. Такие системы визуально воспринимаются как детерминированные. Если же траектории, выпущенные из любой сколь угодно малой первоначальной области, разойдутся между собой на расстояние порядка характерного размера всего множества за конечное время, то они воспринимаются как случайные.

Такое восприятие оправданно: пусть, например, несколько раз подряд ставится «эксперимент» по генерации некоторой траектории вида (2.2) для одних и тех же, с точки зрения наблюдателя, начальных условий. Однако условия эти на самом деле различаются, и, будь измерения более точными, это можно было бы обнаружить. Считая начальные данные одинаковыми, экспериментатор тем не менее получает набор различных траекторий. С его точки зрения это означает, что система проявляет случайные свойства. Следовательно, хаотическим поведением обладают ДС с высокой чувствительностью к начальным данным. Строго это можно определить следующим образом.

▼ **Опр. 2.1.** ДС $x_{n+1} = g(x_n)$, заданная на множестве $x \in \Delta = [0; 1]$, называется хаотической, если для любой точки x_0 , принадлежащей любому открытому множеству $O \subset \Delta$, и для любого ε , такого, что $0 < \varepsilon < 1$, найдутся такая точка $y_0 \in O$ и такое число $n \in N$, что $|x_n - y_n| > \varepsilon$. ▲

Интерес к хаотическим ДС вызван тем, что, как оказалось, многие явления, воспринимаемые как случайные, могут быть на самом деле детерминированными, и тогда модели типа (2.2) являются средством для изучения и познания хаоса. С одной стороны, какова бы ни была точность измерений, она принципиально не может быть абсолютной, поэтому окончательно решить вопрос о причинах, приводящих к тем или иным следствиям, невозможно. С другой стороны, наблюдая за выборкой из

некоторого распределения, можно попытаться с некоторой точностью построить ДС с похожим поведением, т.е. с инвариантной плотностью меры, к которой асимптотически сходится выборочная плотность функции распределения. Хотя одна и та же инвариантная мера может отвечать различным ДС, все же ее знание сужает круг возможных модельных функций, порождающих тот или иной временной ряд. Подчеркнем, что наблюдение за конечным числом точек некоторого отображения не позволяет точно определить соответствующий каскад, поэтому в задачах прогнозирования следует моделировать не саму предполагаемую динамическую систему, а эволюцию выборочной меры, и то на небольшое число шагов вперед. Поэтому приобретают большую практическую важность те измеримые величины, которые характеризуют ДС в целом. Определения некоторых таких величин даются ниже в соответствии с [1, 2, 17].

Количественной мерой чувствительности ДС к вариации начальных условий в терминах скорости экспоненциального разбегания близких траекторий служат показатели Ляпунова. Обозначим через $d(0)$ расстояние между начальными точками двух траекторий ДС $x_{n+1} = g(x_n)$, и пусть $d(n)$ есть расстояние между ними на n -ом шаге. Показателем Ляпунова называется предел

$$\lambda = \lim_{\substack{n \rightarrow \infty \\ d(0) \rightarrow 0}} \frac{1}{n} \ln \left(\frac{d(n)}{d(0)} \right). \quad (2.8)$$

В частности, если отображение задается дифференцируемой функцией, то показатель Ляпунова вычисляется как предел вдоль траектории с началом в x_0 :

$$\lambda(x_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln |g'(x_k)|. \quad (2.9)$$

Если существует единственная инвариантная плотность $f_{inv}(x)$, то процесс строго эргодический и предел в (2.9) не зависит от x_0 :

$$\lambda = \int f_{inv}(x) \ln |g'(x)| dx. \quad (2.10)$$

Для k -мерной ДС существует k таких показателей. Если оказалось, что хотя бы один из показателей Ляпунова положителен, то расстояние

между траекториями экспоненциально возрастает с ростом n , что и означает хаотическое поведение ДС.

Пусть на множестве значений x задана некоторая начальная плотность $f_0(x)$. Если система (2.2) строго эргодическая, то итерационный процесс, задаваемый уравнением Перрона–Фробениуса (2.3), сходится к инвариантной мере. При этом, однако, может оказаться так, что носитель инвариантной меры не совпадает со всем фазовым пространством, а сосредоточен на некотором его подмножестве, называемом притягивающим множеством или аттрактором.

▼ **Опр. 3.2.** Множество A называется инвариантным относительно динамического потока G_t (или каскада g^t) если для всех t выполнено условие $G_t(A) = A$. ▲

▼ **Опр. 3.3.** Областью притяжения ДС для множества A называется множество точек U_A фазового пространства, таких, что $\forall x \in U_A \exists t(x) : \forall t \geq t(x) \ g^t(x) \in A$. ▲

▼ **Опр. 3.4.** Замкнутое инвариантное множество A называется аттрактором ДС, если мера его области притяжения U_A строго положительна и не существует строго меньшего замкнутого множества $A' \subset A$ такого, что $U_{A'}$ совпадает с U_A с точностью до множества меры нуль. ▲

Первое условие в определении 3.4 соответствует понятию устойчивости, т.к. небольшое шевеление траектории, выводящее ее за пределы притягивающего множества, приводит к тому, что впоследствии та вновь оказывается захваченной аттрактором. Второе условие в определении 3.4 называется свойством неразложимости множества, и означает отсутствие в нем инвариантных относительно G_t подмножеств.

Если аттрактор неустойчив, т.е. при любом малом шевелении траектория не возвращается к нему, то он называется репеллером. Заметим, что при обращении времени (для диффеоморфизмов) репеллер становится аттрактором и наоборот.

Одной из важнейших характеристик притягивающего множества является его размерность, которая определяется следующим образом. Пусть множество A покрывается некоторыми «эталонными» множествами с ха-

рактерным размером δ и мерой $C(d)\delta^d$, где $C(d)$ — геометрический коэффициент: например, для куба $C(d) = 1$, для шара

$C(d) = \frac{\pi^{d/2}}{\Gamma(1 + d/2)}$. Пусть для покрытия рассматриваемого множества потребовалось $N(\delta)$ эталонных множеств. Введем величину (верхнюю меру) $mesA = C(d) \lim_{\delta \rightarrow 0} \delta^d \inf N(\delta)$, где нижняя грань берется по

всем возможным полным покрытиям множества A .

Размерностью Хаусдорфа–Безиковича множества A называется такое число D , что при $d > D$ $mesA = 0$, а при $d < D$ $mesA = \infty$.

▼ **Опр. 3.5.** Аттрактор называется регулярным, если он является точкой, или периодической траекторией, или поверхностью, размерность Хаусдорфа–Безиковича которой равна ее топологической размерности. В противном случае аттрактор называется странным. Множество, размерность Хаусдорфа–Безиковича которого строго больше его топологической размерности, называется фракталом, а соответствующая размерность — фрактальной. ▲

Другой информативной характеристикой отображения $x_{n+1} = g(x_n)$ на некотором многообразии M является его топологическая энтропия или энтропия Колмогорова, которая интерпретируется как мера хаоса данной системы. Введем расстояние на M , положив

$$\forall n > 0 \quad \forall x, y \in M \quad \rho_n(x, y) = \max_{0 \leq k \leq n} \rho(g^k(x), g^k(y)).$$

Обозначим через $N_M(n, \delta)$ наименьшее количество шаров радиуса δ в метрике ρ_n , достаточных для полного покрытия множества M .

Топологической энтропией $h_T[g]$ отображения g называется предел

$$h_T[g] = \lim_{\substack{n \rightarrow \infty \\ \delta \rightarrow 0}} \frac{\ln N_M(n, \delta)}{n}. \quad (2.11)$$

Имеет место следующее утверждение, связывающее топологическую энтропию и показатели Ляпунова.

▼ **Теорема 2.1.** Пусть гладкое отображение g строго эргодично. Тогда топологическая энтропия равна сумме положительных показателей Ляпунова. ▲

Кроме топологической энтропии, для оценки меры информации, необходимой для определения положения системы, используется информационная энтропия S_{inf} , определяемая формулой Шеннона [17]

$$S_{\text{inf}} = - \sum_i p_i \ln p_i, \quad (2.12)$$

где p_i есть вероятность системе находиться в i -ом состоянии. Для оценки выборочной информационной энтропии используется выборочная плотность. Если данных достаточно много, по ним можно оценить размерность («информационную») предполагаемого аттрактора, разбивая область изменения наблюдаемой величины, т.е. фазовое пространство ДС, на меньшие ячейки с характерным размером δ и одновременно увеличивая объем выборки, оценивая вероятность существования предела

$$d_{\text{inf}} = - \lim_{\delta \rightarrow 0} S_{\text{inf}}(\delta) / \ln \delta. \quad (2.13)$$

Практическое применение введенных понятий к задачам анализа случайных процессов состоит в определении (разумеется, приближенном) по имеющейся выборке характеристики возможной динамической системы, порождающей данный временной ряд.

Исследуем теперь некоторые конкретные хаотические дискретные отображения.

Рассмотрим задачу Коши для следующего дифференциального уравнения:

$$\dot{x} = ax - bx^2, \quad x(0) = x_0, \quad a, b > 0. \quad (2.14)$$

Его решение дается формулой

$$x(t) = \frac{ax_0}{bx_0 + (a - bx_0)e^{-at}}, \quad (2.15)$$

из которой следует, что при $t \rightarrow +\infty$ $x \rightarrow a/b$ по монотонной S-образной кривой, называемой логистической, независимо от начального значения x_0 , за исключением случая $x_0 = 0$, при котором $x(t) \equiv 0$.

Заметим, что указанные точки $x_1^* = 0$ и $x_2^* = a/b$ являются стационарными точками ДС (2.14). Линейный анализ устойчивости показывает, что вторая из них устойчива, а первая — нет.

Будем решать ту же задачу (2.14) численно. Не обсуждая здесь разностные схемы, которые можно применить для этой цели, возьмем следующую явную схему первого порядка с шагом дискретизации h , обозначая через x_n значения $x(t)$ в узлах сетки:

$$\frac{x_{n+1} - x_n}{h} = ax_n - bx_n^2.$$

Пусть шаг h выбран так, что $ha \neq 1$. Тогда линейной заменой переменной

$$x_n = \frac{a^2 h^2 - 1}{4bh} y_n + \frac{ah + 1}{2bh}$$

разностное уравнение относительно x_n преобразуется к виду

$$y_{n+1} = 1 - ry_n^2 \equiv g(y_n; r), \quad r = \frac{1}{4}(a^2 h^2 - 1). \quad (2.16)$$

Если существует предел y^* последовательности y_n (2.16) при $n \rightarrow \infty$, то он равен одному из двух значений: $y_{1,2}^* = \frac{-1 \mp \sqrt{1+4r}}{2r}$.

При выборе знака «минус» $y_1^* = -2/(ah - 1)$, а при выборе знака «плюс» $y_2^* = 2/(ah + 1)$. Соответствующие предельные значения для исходной переменной x равны указанным выше стационарным точкам $x_{1,2}^*$. Однако последовательность (2.16) является сходящейся не при всех

r и y_0 : например, если $r > 0$ и $y_0 < y_1^*$, то она расходится. Следовательно, чтобы правильно решить уравнение (2.14) численно, практически важно изучить динамику, задаваемую отображением (2.16), при всевозможных значениях параметра r .

Заметим, однако, что обычно логистическая модель применяется для оценки численности популяции в условиях ограниченного ресурса, поэтому физически содержательные решения системы (2.16) получаются

при $-1/4 \leq r < 0$ и $y_0 < y_1^*$ (последнее есть условие положительности переменной x), либо при $r > 0$ и $y_0 > y_1^*$.

Сходимость последовательности определяется ее поведением в окрестности предельной точки. Если показатель Ляпунова положителен, что означает расширение фазового объема и, следовательно, расходимость, то при дифференцируемости функции $g(y; r)$ по y в предельной точке в этом случае должно выполняться условие

$$\left| \frac{\partial g(y; r)}{\partial y} \right|_{y=y_{1,2}^*} > 1. \quad (2.17)$$

Легко проверяется, что точка $y_2^* = \frac{-1 + \sqrt{1+4r}}{2r}$, отвечающая устойчивому положению равновесия, теряет устойчивость при $r > r_1 = 3/4$. При этом появляется новый тип стационарного решения — цикл периода 2 (существовавшая при $r < 3/4$ единственная устойчивая предельная точка представляла собой цикл периода 1). Соответствующие циклические точки удовлетворяют системе уравнений $y_{2c} = g(y_{1c}; r)$, $y_{1c} = g(y_{2c}; r)$ и имеют вид $y_{1,2c} = \frac{(1 \pm \sqrt{4r-3})}{2r}$.

С увеличением r этот цикл, в свою очередь, теряет устойчивость при значении $r = r_2$ таком, что мультипликатор (2.17) для пары точек, связанных между собой, определяемый по правилу производной сложной функции, также становится больше единицы: $|g'(y_{1c}; r)g'(y_{2c}; r)| > 1$. Из этого условия можно найти, что $r_2 = 5/4$. При $r > r_2$ появляется цикл из четырех точек и т.д. Описанное явление называется бифуркацией удвоения периода. Последующие значения r_k будут уже иррациональными. Последовательность $\{r_k\}$ сходится к некоторому конечному пределу r_c при $k \rightarrow \infty$. Этот предел называется критической точкой. При $r > r_c$ динамическая система демонстрирует хаотическое поведение.

В качестве точно решаемого примера хаотической динамики удобно рассмотреть систему (2.16) с параметром $r = 2$:

$$x_{n+1} = 1 - 2x_n^2, \quad x_0 \in [-1; 1]. \quad (2.18)$$

ДС (2.18) имеет динамически-инвариантную меру

$$f_{inv}(x) = \frac{1}{\pi \sqrt{1-x^2}}, \quad (2.19)$$

позволяющую вычислить показатель Ляпунова (2.10): он равен

$$\lambda = \left(\frac{4}{\pi} - 1 \right) \ln 2.$$

Математическое ожидание μ распределения (2.19) равно, очевидно, нулю, а дисперсия $\sigma^2 = 0,5$. Последнее означает, что спрогнозировать значения ряда точнее, чем с ошибкой, равной $\sigma \approx 0,7$ невозможно, если, разумеется, не известна порождающая его динамика. Расчет показывает, что выборочная АКФ этого процесса хаотически осциллирует с амплитудой 0,05. Точное значение АКФ для меры (3.19) равно нулю, т.е. эта ДС может считаться динамическим генератором белого шума.

Примером двумерного отображения, обладающего хаотической динамикой и обобщающего логистическое (2.16), является система Энона:

$$x_{n+2} = 1 - rx_{n+1}^2 + bx_n. \quad (2.20)$$

При $0 < b < 1$ и $r > r_c$ у этой системы имеются хаотические режимы, аналогичные случаю (2.18). Приращения ряда Энона практически независимы, АКФ непериодически колеблется около нулевого значения с амплитудой 0,05.

Популярной трехмерной моделью с квадратичной нелинейностью, имеющей хаотические режимы, является система Лоренца. Этот пример часто появляется в различных физических приложениях. В дифференциальной форме модель имеет вид

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = \lambda x - y - xz, \\ \dot{z} = -\mu z + xy. \end{cases} \quad (2.21)$$

Она зависит от трех начальных условий и трех параметров. Хаотическое поведение решения этой системы наблюдается при значениях параметров $\sigma = 10$, $\lambda = 28$, $\mu = 8/3$.

Приведенные в этом параграфе примеры показывают, что прогнозировать поведение динамической системы с хаосом в отсутствие знаний о ее динамике так же трудно, как и случайного процесса иной природы. Линейные статистические модели регрессионного и авторегрессионного типов приводят в этих случаях к большим ошибкам, поскольку исходные системы нелинейны. В то же время прогнозирование функции распределения является универсальным подходом к исследованию случайных процессов, в том числе и динамического хаоса, и может быть проведено с определенной точностью, оцениваемой по объему выборки и горизонту прогноза. Основы этого метода описываются далее в части II, а примеры применения эволюции выборочной плотности функции распределения на временной горизонт, согласованный с длиной выборки и требуемой точностью прогноза, даются в части III.

3.3. Вырожденные хаотические динамические системы

Выше был рассмотрен пример логистической ДС, непрерывная версия которой имеет вполне предсказуемое поведение, тогда как дискретная при некоторых значениях параметров, например, при определенном шаге дискретизации проявляет хаотические свойства. Следовательно, при моделировании реальных физических процессов надо внимательно следить за адекватностью используемых моделей и за корректностью выводов, которые делаются из их анализа.

Исходя из современных представлений о мире, следует считать, что реально происходящие в нем процессы по существу дискретны. Например, для того, чтобы движение физического тела конечной и отличной от нуля массы во всех случаях адекватно описывалось бы дифференциальным уравнением, требуется, чтобы у этого тела был «инструмент», способный «сказать» ему, что внешняя сила изменилась на бесконечно малую величину, тогда как мелкость разбиения ограничена естественными атомарными размерами. Хотя это является и общим местом, следует все же подчеркнуть, что непрерывность является хорошим приближением, когда рассматриваются такие размеры тел и промежутки времени, которые «много больше» соответствующих минимальных величин для данной

системы. «Много больше» взято в кавычки потому, что это понятие на практике определяется не в соответствии с курсом математического анализа, а из соображений достаточной точности измерения. Если характерное время, определяющее реакцию системы на внешнее воздействие или определяющее ее внутреннюю эволюцию, «много меньше» промежутка времени, на котором изучается система, непрерывное описание является адекватным. В противном случае применяются дискретные модели.

Но, решая «непрерывную модель» численно, исследователь вводит свой шаг дискретизации и свою численную схему, отличную, вообще говоря, от физически адекватной модели процесса. Кроме того, величина шага дискретизации определяется в основном не физикой, а вычислительными возможностями. Отклонения параметров численной схемы от характерных минимальных величин, присущих исходной системе, может происходить как в большую, так и в меньшую стороны. Если шаг дискретизации меньше минимального размера, то надо отдавать себе отчет, что на каждом шаге вычисляются величины, не имеющие правильной интерпретации в терминах исходной модели. Если же он больше, то надо следить, во-первых, за тем, чтобы шаг все еще оставался малым, в соответствии с требованиями по точности, и, во-вторых, чтобы и рассчитываемые величины определялись достаточно точно.

В частности, говоря о логистической системе (2.14) применительно к описанию эволюции численности популяции, представляется неверным делать вывод о ее хаотичности просто на основании знака показателя Ляпунова для некоторой дискретной схемы. Разумеется, для математического объекта (2.16) существуют различные типы решений. Но физически странно выглядит эволюция, которая на одном шаге уменьшает численность особей почти до нуля, а на следующем — увеличивает до переполнения, так что на следующем шаге все опять «гибнут» и т.д. Это означает, что шаг по времени оказался кратен характерному времени жизни поколения, а должен быть равен характерному времени, необходимому для увеличения численности популяции на одну биологическую единицу. Если эти времена совпадают, то, действительно, система будет вести себя хаотически. Если же, как это чаще всего и бывает, естественный шаг по времени мал по сравнению с промежутком времени, на котором численность меняется существенно, то дискретная (2.16) и непрерывная (2.14) модели будут приводить качественно к одним и тем же результатам. На-

конец, если целью было описание именно эволюции поколений, то логистическая модель не является для этого адекватной.

Если модель изначально является дискретной, то, разумеется, проблемы дискретизации снимаются, и остаются только вопросы устойчивости траектории по отношению к точности вычисления величин на каждом шаге. Однако не все уравнения, которые требуют практического решения, имеют ясное физическое содержание, и могут быть записаны в естественной дискретной форме. В ряде случаев уравнение, имеющее непрерывную форму записи, является в теоретическом плане исходным, и его свойства должны быть по возможности сохранены при дискретизации.

В параграфе 2.1 были приведены примеры вырожденных лагранжевых систем. Возможно, что при некоторых параметрах дискретной схемы численное решение не попадет на особую кривую, но будет хаотическим. Что произойдет, если в области изменения параметров, отвечающей динамическому хаосу, существуют значения, при которых достижимо также и сингулярное множество, на котором, как было показано в параграфе 2.2, можно задать случайное продолжение траектории? Такую динамику, в которой сочетаются два случайных процесса различной природы, будем называть двухуровневой хаотичной динамикой.

Таким образом, нетривиальное поведение решений разностных уравнений может представлять самостоятельную ценность и практическую важность. С другой стороны, хаотическое поведение решения, получаемого численно, может являться и артефактом численной схемы. В обоих случаях статистическое описание таких систем имеет значение для понимания сложного характера динамики.

Приведем пример лагранжевой системы, имеющей хаотическое поведение и, кроме того, вырожденной. За основу возьмем известную модель Энона–Эйлеса двумерной неинтегрируемой системы осцилляторов с кубической нелинейностью. Лагранжиан исходной модели имеет вид

$$L_{HH} = \frac{v_1^2}{2} + \frac{v_2^2}{2} - \frac{q_1^2}{2} - \frac{q_2^2}{2} - q_1^2 q_2 + \frac{q_2^3}{3}, \quad v_i = \dot{q}_i, \quad i = 1, 2. \quad (3.1)$$

Система (3.1) имеет хаотическое поведение при энергиях, приблизительно больших (в безразмерных переменных) числа 1/6. Простейшей модификацией этой системы является модель слабoreлятивистских нелинейных осцилляторов [8] с лагранжианом

$$L = \frac{v_1^2}{2} \left(1 - \frac{q_1^2}{2c^2} \right) + \frac{v_2^2}{2} \left(1 - \frac{q_2^2}{2c^2} \right) - \frac{q_1^2}{2} - \frac{q_2^2}{2} - q_1^2 q_2 + \frac{q_2^3}{3}, \quad (3.2)$$

где c есть некоторый большой положительный параметр (в размерных величинах — скорость света). В регулярных областях фазового пространства энергия системы (3.2) дается выражением

$$E = \frac{v_1^2}{2} \left(1 - \frac{q_1^2}{2c^2} \right) + \frac{v_2^2}{2} \left(1 - \frac{q_2^2}{2c^2} \right) + \frac{q_1^2}{2} + \frac{q_2^2}{2} + q_1^2 q_2 - \frac{q_2^3}{3}. \quad (3.3)$$

Будучи непрерывной функцией координат и скоростей, энергия (3.3) может быть продолжена по непрерывности и на сингулярное множество системы (3.2), определяемое условием

$$\det J = \left| \frac{\partial^2 L}{\partial v_i \partial v_j} \right| = \left(1 - \frac{q_1^2}{2c^2} \right) \left(1 - \frac{q_2^2}{2c^2} \right) = 0. \quad (3.4)$$

Таким образом, сингулярное множество представляет собой четыре плоскости $q_{1,2} = \pm c\sqrt{2}$. Оно достижимо первой частицей, если при заданном значении энергии E_{s1} существуют координата и скорость второй частицы, удовлетворяющие уравнению

$$E = E_{s1} = \frac{v_2^2}{2} \left(1 - \frac{q_2^2}{2c^2} \right) + c^2(1 + q_2) + \frac{q_2^2}{2} - \frac{q_2^3}{3}.$$

Условие достижимости сингулярного множества второй частицей представляется двумя уравнениями:

$$E = E_{s2} = \frac{v_1^2}{2} \left(1 - \frac{q_1^2}{2c^2} \right) + c^2 + \frac{q_1^2}{2} \pm c\sqrt{2} \left(q_1^2 - \frac{2c^2}{3} \right).$$

Обе частицы могут достичь сингулярности при значениях энергии $E_{s3} = c^2 \pm c^3\sqrt{2}/3$. Следуя правилам продолжения траектории через особое множество (теоремы 2.2.1 и 2.2.2), приходим к выводу, что в данном случае будет происходить упругое отражение частиц от точек, определяемых в (3.4): координаты не меняются, а скорости меняют знак. Уравнение Лиувилля (1.1.12) следует дополнить в этом случае граничными условиями для плотности распределения после попадания системы на

сингулярную фазовую поверхность. Это условие запишем в интегральном виде по аналогии с (2.2):

$$\rho(x, t_s + 0) = \int P(x, x') \rho(x', t_s - 0) dx', \quad (3.5)$$

где ядро $P(x, x')$ описывает процесс прохождения траектории через поверхность $S(x)$, момент достижения которой обозначен в (3.5) как t_s . В ядре $P(x, x')$ должны быть отражены правила прохождения особой поверхности: ядро отлично от нуля только на особой поверхности, т.е. в нем должен содержаться член вида $\delta(S(x'))$; оно должно сохранять пространственные положения частиц после прохождения, что означает наличие члена $\delta(\mathbf{q} - \mathbf{q}')$; сохранение энергии может быть обеспечено множителем $\delta(E(x) - E(x'))$. Отражение от сингулярной поверхности в данном конкретном случае описывается множителем $\delta(\mathbf{v} + \mathbf{v}')$. В итоге получаем

$$P(x, x') = \delta(q_1^2 - 2c^2) \delta(q_2^2 - 2c^2) \delta(E(x) - E(x')) \delta(\mathbf{q} - \mathbf{q}') \delta(\mathbf{v} + \mathbf{v}'). \quad (3.6)$$

В качестве примера дискретной вырожденной системы с хаосом возьмем сингулярно-возмущенную дискретную систему Энона (2.20), в которую добавлены малые члены. Например, вместо (2.20) рассмотрим систему

$$\begin{cases} x_{n+1} + \varphi_n y_{n+1} = 1 - rx_n^2 + y_n \\ \varphi_n x_{n+1} + y_{n+1} = bx_n, \end{cases} \quad (3.7)$$

в которой $\varphi_n = \varphi(x_n, y_n)$ — некоторая функция, пренебрежимо малая при почти всех значениях своих аргументов, но принимающая значение, равное единице, если окажется, что $y_n = rx_n^2 + bx_n - 1$. Тогда при достижении соответствующих значений $\{x_n, y_n\}$ дальнейшая эволюция не определена, и, при желании, ее можно продолжить вероятностным путем, задав плотность вероятности точки x_{n+1} в зависимости, например, от x_n .

Приведенные примеры показывают возможность моделирования случайных процессов с помощью более сложных ДС, в которых заложена хаотизация различной природы. Однако на практике использовать некоторую готовую модель для объяснения наблюдаемых данных удастся, как правило, лишь на очень небольшом (по сравнению с имеющимся объемом этих данных) отрезке времени. Это связано с тем, что временной ряд, будучи нестационарным, на больших промежутках времени характеризуется гораздо большей дисперсией, чем на сравнительно малых, но статистически достаточных промежутках. Поэтому интерес представляет моделирование эволюции функции распределения ряда, чтобы оценить характерный промежуток времени, в течение которого наблюдаемую систему можно было бы считать квазистационарной.

3.4. Уравнение Колмогорова–Фоккера–Планка

Получим кинетическое уравнение относительно плотности распределения случайного процесса, если переходная вероятность зависит только от начального и конечного положений системы. Как и в п. 2.3, вывод кинетического уравнения для ПФР опирается на уравнение Колмогорова для вероятности перехода системы из одного состояния в другое в единицу времени. Напомним введенные в п. 2.3 определения. Пусть $f(x, t)$ есть плотность вероятности, рассматриваемая в фазовом пространстве Γ точек x . Для каждой пары точек $x, x' \in \Gamma$ и для любых двух моментов времени t и t' определена условная вероятность $P(x', t' | x, t)$ того, что система находится в точке x' в момент t' при условии, что она находилась в точке x в момент t . Вводится также двумерная плотность вероятности $F(x', t'; x, t)$ того, что система находится в состоянии (x', t') и находилась в состоянии (x, t) . Условие согласованности распределений состоит в том, что

$$\int F(x', t'; x, t) dx' = f(x, t)$$

и, кроме того, по определению условной вероятности

$$P(x', t' | x, t) \cdot f(x, t) = F(x', t'; x, t).$$

Если вероятность перехода в точку x' в момент t' определяется только вероятностью находиться в точке x в момент времени t (марков-

ские процессы), условные вероятности $P(x', t' | x, t)$ удовлетворяют уравнению Колмогорова–Смолуховского–Чепмена

$$\begin{aligned} \forall t'', t \leq t'' \leq t': P(x', t' | x, t) = \\ = \int P(x', t' | x'', t'') P(x'', t'' | x, t) dx'' . \end{aligned} \quad (4.1)$$

Поскольку многие модели случайных процессов выводятся из уравнения (4.1), полезно проследить связь между получающимися уравнениями, призванными описать определенные свойства процесса, и исходными предположениями. Например, уравнение Фоккера–Планка (так называемое прямое уравнение Колмогорова) часто применяют для описания динамики приростов цен на те или иные финансовые инструменты, считая приросты случайным блужданием. При этом надо помнить, что модели, явно или неявно использующие марковское свойство, имеют весьма ограниченную область применения для финансового рынка, так как рынок обладает некоторой конечной памятью, а модель предполагает отсутствие таковой уже через один шаг по времени. И если для физических процессов такое приближение корректно, то на рынке ценных бумаг оно чаще всего не выполняется. Тем не менее, в отдельных случаях такие упрощенные представления могут дать полезные результаты.

Можно построить два уравнения эволюции условной вероятности $P(x, t | x_0, t_0)$ для процесса перехода из точки (x_0, t_0) в точку (x, t) : по конечному времени t и по начальному времени t_0 . В первом случае уравнение называется прямым уравнением Колмогорова или уравнением Фоккера–Планка, а во втором случае — обратным уравнением Колмогорова [15]. Если задано начальное распределение вероятностей $f(x, t)|_{t=t_0} = f_0(x)$ и требуется найти вероятность в будущий момент времени t , то решается прямое уравнение. Если же надо вычислить распределение первого момента времени достижения фиксированного уровня $x = x^*$, то решается обратное уравнение.

Для вывода прямого уравнения рассмотрим в уравнении Колмогорова (4.1) изменение условной вероятности $P(x, t | x_0, t_0)$ при увеличении на τ времени перехода из точки x_0 в точку x . Формально из (4.1) имеем

$$P(x, t + \tau | x_0, t_0) = \int P(x, t + \tau | x', t') P(x', t' | x_0, t_0) dx' .$$

Поскольку время t' любое, большее t_0 и меньшее $t + \tau$, можно взять $t' = t$. Тогда

$$P(x, t + \tau | x_0, t_0) = \int P(x, t + \tau | x', t) P(x', t | x_0, t_0) dx' . \quad (4.2)$$

Для дальнейших преобразований перейдем к представлению характеристической функции (ХФ), т.е. к фурье-образу условной вероятности $P(x, t + \tau | x', t)$ случайного приращения $x - x'$ за время τ при фиксированном значении x' . Введем условную ХФ

$$X(\omega, t + \tau | x', t) = \int_{-\infty}^{\infty} P(x, t + \tau | x', t) \exp(i\omega(x - x')) dx . \quad (4.3)$$

Разлагая экспоненту под интегралом в (4.3) в ряд Тэйлора, получаем эквивалентное представление ХФ через условные моменты:

$$\begin{aligned} X(\omega, t + \tau | x', t) = \sum_{n=0}^{\infty} \frac{(i\omega)^n}{n!} \int_{-\infty}^{\infty} P(x, t + \tau | x', t) (x - x')^n dx \equiv \\ \equiv \sum_{n=0}^{\infty} \frac{(i\omega)^n}{n!} m_n(x', t; \tau) \end{aligned} \quad (4.4)$$

Обратное к (4.3) преобразование имеет вид

$$P(x, t + \tau | x', t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega, t + \tau | x', t) \exp(-i\omega(x - x')) d\omega ,$$

которое с учетом (4.4) можно представить в виде ряда по производным дельта-функции:

$$\begin{aligned} P(x, t + \tau | x', t) = \\ = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{(i\omega)^n}{n!} m_n(x', t; \tau) \exp(-i\omega(x - x')) d\omega = \\ = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} m_n(x', t; \tau) \frac{\partial^n}{\partial x^n} \delta(x' - x) . \end{aligned} \quad (4.5)$$

Подставляя итоговое выражение (4.5) в интеграл в правой части уравнения (4.2) и выделяя в нем первое слагаемое в бесконечной сумме (при $n = 0$), получаем после интегрирования дельта-функций следующее разложение:

$$P(x, t + \tau \mid x_0, t_0) = P(x, t \mid x_0, t_0) + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} (m_n(x, t; \tau) P(x, t \mid x_0, t_0)). \quad (4.6)$$

Разложение (4.6) можно использовать для вычисления производной условной вероятности по конечному моменту времени (т.е. по времени t):

$$\begin{aligned} \frac{\partial P(x, t \mid x_0, t_0)}{\partial t} &= \lim_{\tau \rightarrow +0} \frac{P(x, t + \tau \mid x_0, t_0) - P(x, t \mid x_0, t_0)}{\tau} = \\ &= \lim_{\tau \rightarrow +0} \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} \left(\frac{m_n(x, t; \tau)}{\tau} P(x, t \mid x_0, t_0) \right). \end{aligned} \quad (4.7)$$

Из (4.7) видно, что для существования производной в левой части (4.7) должны существовать пределы условных моментов в правой части.

Итак, пусть существуют величины

$$\begin{aligned} M_n(x, t) &= \lim_{\tau \rightarrow +0} \frac{m_n(x, t; \tau)}{\tau} = \\ &= \lim_{\tau \rightarrow +0} \frac{1}{\tau} \int_{-\infty}^{\infty} (x' - x)^n P(x', t + \tau \mid x, t) dx'. \end{aligned} \quad (4.8)$$

Тогда (4.7) представляется в виде

$$\frac{\partial P(x, t \mid x_0, t_0)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n} (M_n(x, t) P(x, t \mid x_0, t_0)). \quad (4.9)$$

По определению, процесс $x(t)$ называется диффузионным, если в правой части (4.9) отличны от нуля только первые два момента, которые мы будем обозначать как $u(x, t)$ (коэффициент сноса) и $\lambda^2(x, t)$ (по построению неотрицательный коэффициент диффузии). Если $\lambda^2(x, t)$ строго больше нуля, непрерывный марковский процесс называется регулярным.

Аналогичным способом получается и обратное эволюционное уравнение (по начальному времени), которое для диффузионного процесса записывается в виде

$$\begin{aligned} \frac{\partial P(x, t \mid x_0, t_0)}{\partial t_0} &= \\ &= -u(x_0, t_0) \frac{\partial P(x, t \mid x_0, t_0)}{\partial x_0} - \frac{1}{2} \lambda^2(x_0, t_0) \frac{\partial^2 P(x, t \mid x_0, t_0)}{\partial x_0^2}. \end{aligned} \quad (4.10)$$

Для одномерной плотности вероятности $f(x, t)$ регулярного диффузионного процесса в соответствии с представлением $f(x, t) = \int F(x', t; x, t) dx' = \int P(x', t \mid x, t) f(x', t) dx'$ получаем из (4.9) одномерное уравнение Фоккера–Планка:

$$\frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x} (u(x, t) f(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (\lambda^2(x, t) f(x, t)). \quad (4.11)$$

Как следует из вывода уравнения Фоккера–Планка, оно представляет собой обрыв разложения обратного преобразования Фурье для условной вероятности на втором порядке по производной дельта-функции.

Между стохастическим дифференциальным уравнением (СДУ) и уравнением Фоккера–Планка (4.11) существует тесная связь, которую полезно напомнить для того, чтобы представлять, какие именно требования к эмпирическому распределению должны быть выполнены, чтобы о наблюдаемом случайном процессе можно было говорить как о приближенном решении определенного СДУ. Кроме того, следует представлять отличия конкретно производимых вычислений с конечной дискретной выборкой от теоретической схемы, в рамках которой выведена модель процесса.

Для строгого определения стохастического дифференциального уравнения вводится следующая вероятностная схема. Пусть задано вероятностное пространство $\{\Omega, \sigma, P\}$ и измеримый при всех $t \in [0; T]$ винеровский процесс $\xi(t, \omega), \omega \in \Omega$. Параметризация временем «превращает» случайную величину $\xi(\omega)$ в случайный процесс $\xi(t, \omega)$, но обычно для краткости указание на пространство Ω допустимых исходов опускается.

Случайный процесс, определенный при $t \in [0; T]$, называется винеровским, если он:

- 1) является процессом с независимыми приращениями;
- 2) его среднее значение равно нулю $M\xi(t) = 0$;
- 3) дисперсия $\sigma_\xi^2 = const$;
- 4) среднеквадратичное приращение пропорционально разности времен

$$M(\xi(t+\tau) - \xi(t))^2 = \sigma_\xi^2 |\tau|; \quad (4.12)$$

- 5) распределения процессов $\xi(t)$ и $\xi(t+\tau) - \xi(t)$ нормальны.

Функция распределения приращений винеровского процесса, таким образом, имеет вид

$$P\{\xi(t+\tau) - \xi(t) < x\} = \frac{1}{\sqrt{2\pi\tau\sigma_\xi^2}} \int_{-\infty}^x \exp\left(-\frac{z^2}{2\sigma_\xi^2|\tau|}\right) dz.$$

На вероятностном пространстве $\{\Omega, \sigma, P\}$ рассматривается совокупность σ -алгебр $\{\Psi_t\}$ при $t \in [0; T]$, связанная с винеровским процессом $\xi(t)$ условиями:

- 1) $\Psi_s \subset \Psi_t \subset \sigma$ при $s < t$;
- 2) при всех $t \in [0; T]$ процесс $\xi(t)$ измерим относительно Ψ_t ;
- 3) процесс приращений $\xi(t+\tau) - \xi(t)$ при всех $t > 0$, $\tau \geq 0$ не зависит от событий σ -алгебры Ψ_τ .

Пусть x_t есть случайный процесс, являющийся решением СДУ:

$$dx_t = u(x_t, t)dt + \lambda(x_t, t)d\xi_t, \quad x_0 = x(0, \omega), \omega \in \Omega. \quad (4.13)$$

В уравнении (4.13) $\xi_t = \xi(t, \omega)$ — винеровский процесс, $u(x, t)$ и $\lambda(x, t)$ — измеримые функции при каждом $t \in [0, T]$. Тогда [15], если выполнены условия

- 1) $\exists K < \infty: \forall x, y \in R, \forall t \in [0; T]$
 $|u(x, t) - u(y, t)| + |\lambda(x, t) - \lambda(y, t)| \leq K|x - y|;$
- 2) $|u(x, t)|^2 + |\lambda(x, t)|^2 \leq K^2(1 + |x|^2);$
- 3) $M|x(0, \omega)|^2 < \infty,$

то с вероятностью 1 (т.е. почти наверное) существует непрерывное решение СДУ (4.13), и плотность $f(x, t)$ распределения процесса ξ_t удовлетворяет уравнению Фоккера–Планка (4.11). Тем самым обеспечивается перевод с языка СДУ на язык кинетических уравнений. Подчеркнем, что теория оперирует с непрерывными генеральными распределениями случайных процессов, тогда как на практике приходится иметь дело с некоторой конечной реализацией, может быть, даже и не случайного (в смысле данных выше определений) процесса. Поэтому представляется важным описать ту кинетическую структуру, которая появляется в задачах математической статистики нестационарных временных рядов, чтобы теория могла быть применена в конкретном случае. Нарушение вышеперечисленных условий приводит к совершенно другим кинетическим уравнениям. Так, если отказаться от винеровского процесса или от конечности дисперсии, то будут получаться нелокальные интегро-дифференциальные уравнения, обсуждаемые в следующем разделе.

3.5. Модели фрактального блуждания

Поставим теперь вопрос о том, каким уравнением следует описывать эволюцию плотности функции распределения случайного процесса, который удовлетворяет СДУ (4.13), но мера $d\xi_t$ не связана с винеровским процессом. Будем рассматривать такие процессы, для которых распределения случайной величины и ее приращений описываются одним и тем же типом распределения, как и в случае винеровского процесса. В случае винеровского процесса эти распределения нормальны. Такие распределения называются устойчивыми. Уравнение Колмогорова–Чепмена для устойчивой плотности вероятности $f(x, t)$ имеет вид

$$f(x, t) = \int_{-\infty}^{\infty} f(y, t') f(x - y, t - t') dy. \quad (5.1)$$

Аналогичное уравнение записывается и в случае n -мерного пространства.

В характеристических функциях вместо (5.1) получаем

$$X(k, t) = X(k, t')X(k, t - t'). \quad (5.2)$$

Решением этого уравнения в классе автомодельных функций, зависящих только от одного аргумента $z = |k|t^{1/\alpha}$ и произвольного параметра $\alpha > 0$, является функция

$$\hat{g}(z) = \exp(-Dz^\alpha), \quad D = \text{const} > 0. \quad (5.3)$$

Уравнение эволюции для такой ХФ имеет вид

$$\frac{\partial \hat{g}}{\partial t} = -D|k|^\alpha \hat{g}, \quad (5.4)$$

которое называется уравнением аномальной диффузии. В исходных переменных оно символически записывается в виде так называемого оператора Рисса [18]

$$\frac{\partial g}{\partial t} = -D(-\Delta)^{\alpha/2} g, \quad (-\Delta)^{\alpha/2} g = \hat{F}^{-1}|k|^\alpha \hat{F}g, \quad (5.5)$$

где символом \hat{F} обозначено преобразование Фурье. Процесс, удовлетворяющий уравнению (5.5) при $0 < \alpha < 2$, называется процессом Леви. При $\alpha = 2$ он переходит в стандартный винеровский процесс. При $\alpha > 2$ функция $g(x, t)$ теряет положительность, и потому не может рассматриваться как плотность вероятности.

Устойчивые распределения определяются интегралом

$$\begin{aligned} g_n(x) &= \frac{1}{(2\pi)^n} \int \exp(-i\mathbf{k}\mathbf{x} - D|\mathbf{k}|^\alpha) d\mathbf{k} = \\ &= \frac{1}{(2\pi)^{n/2}} \int_0^\infty \exp(-k^\alpha) \frac{J_{n/2-1}(k|x|)}{(k|x|)^{n/2-1}} k^{n-1} dk. \end{aligned} \quad (5.6)$$

Физическая интерпретация СДУ, в котором случайное блуждание описывается не винеровским процессом, а процессом Леви, состоит в том, что такие процессы имеют аномально широкое распределение — дисперсия бесконечно велика. Этот факт трактуется в том смысле, что встреча-

ются очень большие приросты значений случайной величины, называемые «полетами Леви». А так как в обычном пространстве ширина начального распределения в виде дельта-функции пропорциональна квадратному корню из времени, то аномальная диффузия означает, что частицы «как бы пролетают» мимо элементов среды, имеющей, таким образом, в отношении этих частиц пустоты. Такое движение могло бы быть реализовано, например, на фрактальном множестве.

В качестве конкретного множества будем рассматривать Канторово множество C_λ , которое образовано итерационной процедурой по выбрасыванию из средней части предыдущего отрезка открытого интервала длины λ . Исходным множеством является отрезок $[0; 1]$. Длина каждого из 2^{n-1} интервалов, удаляемых на n -ой итерации, равна β^{n-1} , где $\beta = (1 - \lambda)/2$, так что на n -ом шаге удаляются интервалы совокупной длины l_n , равной

$$l_n = \lambda(2\beta)^{n-1}, \quad \beta = \frac{1 - \lambda}{2}. \quad (5.7)$$

Суммарная длина этих интервалов, как легко видеть, равна единице, поэтому мера Лебега множества C_λ равна нулю. Размерность α Хаусдорфа такого множества определяется выражением $\alpha = -\log_\beta 2$, поскольку C_λ является объединением двух подобных ему множеств с масштабным коэффициентом β и может принимать любые значения между 0 и 1. Точки обобщенного множества Кантора могут быть представлены в виде

$$x = (1 - \beta) \sum_{k=1}^{\infty} b_k \beta^k, \quad b_k \in \{0, 1\}. \quad (5.8)$$

Одним из наиболее известных примеров множеств вида C_λ является тернарное множество Кантора с параметрами $\lambda = \beta = 1/3$, которое может быть описано в явном виде как

$$C_{1/3} = [0; 1] \setminus \bigcup_{m=1}^{\infty} \bigcup_{k=0}^{3^{m-1}-1} \left[\frac{3k+1}{3^m}; \frac{3k+2}{3^m} \right]. \quad (5.9)$$

Пусть $f(x, t)$, $x \in C_\lambda$, есть ПФР процесса, определенного на фрактальном множестве, так что

$$\int_{C_\lambda} f(x, t) \mu_\alpha(dx) = 1. \quad (5.10)$$

Интерпретация дробного интеграла как интеграла по фрактальному множеству C_λ , который вычисляется с мерой $\mu_\alpha(dx)$, отвечающей размерности Хаусдорфа множества, предложена в работе Нигматуллина [19]. Интеграл (5.10) вычисляется в соответствии с интегральной формулой Римана–Лувиля. Именно, интеграл с переменным верхним пределом по множеству с ненатуральной размерностью $0 < \alpha \leq 1$ определяется формулой (см., напр., [18])

$$(I_{a+}^\alpha y)(x) = \frac{1}{\Gamma(\alpha)} \int_a^x \frac{y(\xi) d\xi}{(x - \xi)^{1-\alpha}}. \quad (5.11)$$

В терминах интеграла I_{a+}^α условие нормировки может (5.10) имеет вид $(I_{0+}^\alpha f)(1, t) = 1$. В частности, если рассматривается отрезок $[0, 1]$, на котором «живет» множество C_λ , то фрактальная мера μ_α этого множества есть

$$\mu_\alpha(C_\lambda) = (I_{0+}^\alpha 1)(1) = \frac{1}{\Gamma(\alpha)} \int_0^1 \frac{d\xi}{(1 - \xi)^{1-\alpha}} = \frac{1}{\alpha \Gamma(\alpha)}. \quad (5.12)$$

Величина, обратная мере (5.12), есть плотность равномерного распределения на множестве Кантора.

Для вывода уравнения Лиувилля, выражающего закон сохранения нормировки, надо ввести оператор дробного дифференцирования. Оператор, обратный дробному интегралу Римана–Лиувилля (5.11) слева, является дробной производной Римана–Лиувилля [18] того же порядка:

$$(D_{a+}^\alpha y)(x) = \frac{d}{dx} \frac{1}{\Gamma(1-\alpha)} \int_a^x \frac{y(\xi) d\xi}{(x - \xi)^\alpha}, \quad (5.13)$$

$$(D_{a+}^\alpha I_{a+}^\alpha y)(x) = y(x).$$

В частности, когда $0 < \alpha \leq 1$, результат дробного дифференцирования представляется в виде

$$(D_{a+}^\alpha y)(x) = \frac{1}{\Gamma(1-\alpha)} \left(\frac{y(a)}{(x-a)^\alpha} + \int_a^x \frac{y'(\xi) d\xi}{(x-\xi)^\alpha} \right). \quad (5.14)$$

Заметим теперь, что для вывода уравнения эволюции ПФР требуется вычислять интеграл от дивергенции, т.е. применять оператор интегрирования слева от оператора дифференцирования. Но интеграл (5.11) является правым обратным, а не левым обратным оператором для производной (5.13). Поэтому нужно определить коммутатор этих двух операторов. Соответствующая формула имеет вид [18]

$$(I_{a+}^\alpha D_{a+}^\alpha y)(x) = y(x) - \sum_{j=1}^n \frac{y_{n-\alpha}^{(n-j)}(a)}{\Gamma(\alpha - j + 1)} (x-a)^{\alpha-j}, \quad (5.15)$$

$$n = [\alpha] + 1.$$

где $y_{n-\alpha}(x) = (I_{a+}^\alpha y)(x)$, т.е. в сумму (5.15) входят производные дробного интеграла в начальной точке. В частности, при $0 < \alpha \leq 1$ имеем

$$(I_{a+}^\alpha D_{a+}^\alpha y)(x) = y(x) - \frac{(x-a)^{\alpha-1}}{\Gamma(\alpha)} (I_{a+}^{1-\alpha} y)(a). \quad (5.16)$$

В результате оказывается, что уравнение Лиувилля для ПФР на множестве C_λ формально содержит функцию источника, определенную с точностью до произвольной функции $\varphi(x, t)$ такой, что для нее выполняется условие (5.12) нормировки на единицу

$$\int_{C_\lambda} \varphi(x, t) \mu_\alpha(dx) = 1.$$

Это уравнение имеет вид

$$\frac{\partial f(x, t)}{\partial t} + (D_{0+}^\alpha u f)(x, t) = J(t) \varphi(x, t), \quad (5.17)$$

$$J(t) = u(1, t) f(1, t) - \frac{1}{\Gamma(\alpha)} (I_{0+}^{1-\alpha} u f)(0, t).$$

Здесь $u(x, t)$ — вообще говоря, произвольная функция на C_λ , к которой можно применить операцию дробного дифференцирования. По сво-

ему смыслу эта функция представляет среднюю скорость изменения координаты x случайной траектории.

Именно такое уравнение отвечает условию сохранения нормировки (5.10). В обычном случае натуральной размерности для отсутствия источника в правой части уравнения Лиувилля достаточно потребовать естественного «условия непротекания» $u(0,t)f(0,t) = u(1,t)f(1,t) = 0$. Теперь же в правую часть добавляются неочевидные члены от производной (обычной) дробного интеграла от потока $u(x,t)f(x,t)$ в начальной точке. При нулевом граничном условии $u(1,t)f(1,t) = 0$ и аналогичном условии на интеграл дробного порядка в нуле правая часть (5.17) обращается в ноль. Только при этом условии снимается неопределенность правой части в (5.17). Будем считать его далее выполненным.

Поскольку для корректного использования уравнения Лиувилля нужна динамическая система, а на практике случайный процесс порождается динамической системой лишь приближенно, то более естественным для описания случайных траекторий на фрактальных множествах является уравнение Колмогорова и следующее из него уравнение Фоккера–Планка. Однако формальное обобщение на фракталы классического вывода уравнения Фоккера–Планка из уравнения Колмогорова встречает определенные трудности. Дело в том, что при его выводе используется характеристическая функция переходной вероятности, т.е. делается преобразование Фурье плотности вероятности, заданной на фрактале. Но с формальной точки зрения преобразование $f(x) \rightarrow \tilde{f}(k) = \int \exp(ikx)f(x)d^n x$ содержит под знаком экспоненты скалярное произведение в n -мерном пространстве, а на фрактале его определить затруднительно. Кажется естественным, что случайное блуждание на множестве нецелой размерности будет описываться уравнением вида

$$\frac{\partial f(x,t)}{\partial t} + (D_{0+}^\alpha u f)(x,t) = \frac{1}{2} (D_{0+}^\alpha D_{0+}^\alpha \sigma^2 f)(x,t), \quad (5.18)$$

с учетом граничных условий «непротекания». Это во всяком случае так, если случайный процесс изначально задан на «обычном» многообразии, и лишь «волею судеб» его конкретные значения попадают на фрактальное множество. Вывод уравнения (5.18) аналогичен выводу обычного уравнения Фоккера–Планка из уравнения Колмогорова–Чепмена с учетом того, что интеграл по промежуточным состояниям, связывающим плотности

вероятностей в разные моменты времени, берется в смысле Римана–Лиувилля. Соответствующее уравнение называют уравнением Фоккера–Планка–Заславского [20].

С другой стороны, существует довольно большое количество математических моделей фрактальных блужданий, выведенных из эвристических соображений и не имеющих строгого обоснования. Они опираются на уже упоминавшееся выше представление о том, что если распределение, например, приростов некоторой случайной величины (допустим, цен на какой-либо финансовый инструмент) имеет степенные «хвосты», то модельным уравнением должно быть уравнение с дробными производными, что, в свою очередь, ассоциируется с движением по фрактальному множеству. По поводу таких моделей следует сделать замечание, касающееся вероятности не просто больших, а очень больших скачков: вероятность эта не близка к нулю. Но тогда модель становится не адекватна действительности, ибо ни в каких фантазиях, например, суточная цена на нефть не изменится на 1000 долларов за баррель за один шаг (сутки), т.е. такой «полет Леви» совершенно невозможен. Между тем по плотности распределения, спадающей, например, как $1/x^2$, вероятность такого события равна $1/1000$, и потому «должна» наблюдаться в среднем раз в три-четыре года. Такие эффекты связаны с неоправданной интерполяцией степенного «хвоста» в область несуществующих значений. На самом деле все распределения наблюдаемых временных рядов финитны в совершенно практическом смысле, так как отношение суточного изменения цены к тиковому приросту имеет порядок 100 или меньше, что не может считаться «бесконечностью». Тем не менее, уравнения эволюции, трактуемые как «фрактальная динамика», используются рядом авторов для моделирования некоторых реальных ситуаций. Не обсуждая здесь практическую полезность таких моделей, обратимся к математической стороне вопроса о корректности вывода модельных уравнений, приводимых в различных работах по аномальным случайным блужданиям.

Так, часто используется уравнение Фоккера–Планка в виде

$$\frac{\partial f(x,t)}{\partial t} + \frac{\partial}{\partial x} (u f)(x,t) = \frac{1}{2} \frac{\partial^2}{\partial x^{2\alpha}} (\sigma^2 f)(x,t),$$

где вводится производная $\frac{\partial}{\partial x^\alpha} = \frac{|x|^{\alpha-1}}{\alpha} \frac{\partial}{\partial x}$. Но так записанное уравнение эволюции не обладает свойством закона сохранения нормировки при дробном интегрировании по множеству C_λ .

Встречаются модели стационарного фрактального блуждания с использованием уравнения Фоккера–Планка с обычным линейным сносом и дробной диффузионной производной, понимаемой в смысле Римана–Лиувилля:

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}(u(x,t)f(x,t)) + D \frac{\partial^\alpha f(x,t)}{\partial x^\alpha}.$$

Производная первого порядка для описания сноса в такой модели неверна, так как при интегрировании по C_λ опять-таки не будет выполнен закон сохранения нормировки.

Часто используются модели диффузионного типа без сноса вида

$$\frac{\partial f(x,t)}{\partial t} = -D(-\Delta)^{\alpha/2} f(x,t),$$

что было рассмотрено выше в (5.5).

Иногда при выводе уравнения Колмогорова рассматривается эволюция переходной вероятности в виде $\frac{\partial^\beta f(x,t)}{\partial t^\beta}$, где производная по времени понимается в смысле Римана–Лиувилля порядка β , $0 < \beta \leq 1$. Но заметим, что «дробная эволюция» не является естественным обобщением для динамической системы, если бы пришлось задавать таковую на фрактале, так как скорость по-прежнему определяется обычной, а не дробной производной по времени.

Применяются также модели случайного блуждания с непрерывным временем, но дискретными координатами. Исходным является эволюционное уравнение типа Лиувилля с фрактальной эволюцией по времени и с линейным сносом

$$\frac{\partial^\beta f(k,t)}{\partial t^\beta} = -\kappa(f(k+1,t) - f(k,t)),$$

где производная по времени понимается в смысле Капуто (несколько измененный вид производной Римана–Лиувилля):

$$\frac{d^\beta f}{dt^\beta} = \begin{cases} \frac{1}{\Gamma(n-\beta)} \int_0^t \frac{1}{(t-s)^{1+\beta-n}} \frac{d^n f}{ds^n} ds, \\ \frac{d^n f}{dt^n}, \quad \beta = n \in N. \end{cases}$$

Таким образом, существует довольно много именно моделей фрактальной случайной динамики, в которых из тех или иных физических соображений постулируется некоторое удобное уравнение. Существенно, что во всех описанных моделях частица блуждает по обычному множеству (действительной прямой), и лишь дробная производная ассоциируется с фрактальной динамикой на уровне интерпретации результатов.

В заключение сделаем важное замечание относительно вероятностных распределений, заданных на фрактальных множествах. Как видно из проведенного выше краткого обзора результатов работ, плотности этих распределений удовлетворяют некоторым интегро-дифференциальным уравнениям, к которым сводятся выражения с дробными производными. Однако далеко не каждое решение этих уравнений может считаться плотностью вероятности. Дело в том, что любая неотрицательная интегрируемая функция может считаться плотностью функции распределения в обычном пространстве, поскольку интеграл от нее определяет неубывающую неотрицательную непрерывную слева функцию распределения. На фрактальном же множестве это не так.

Рассмотрим интеграл вида (5.11) с переменным верхним пределом, который имеет трактовку интегральной функции распределения на $[0;1]$, причем $0 < \alpha < 1$:

$$F(x) = \frac{1}{\Gamma(\alpha)} \int_0^x \frac{f(t)dt}{(x-t)^{1-\alpha}}. \quad (5.19)$$

Поскольку подынтегральное выражение само зависит от x , то неотрицательности плотности, т.е. функции $f(t)$, недостаточно для монотонности $F(x)$. А она и в данном случае должна быть таковой, поскольку

вероятность того, что случайная величина заключена в пределах $[u; v]$, равна

$$P\{u < x < v\} = (I_{a+}^{\alpha} f)(v) - (I_{a+}^{\alpha} f)(u) = F(v) - F(u).$$

Например, пусть ПФР имеет кусочно-постоянный вид:

$$f(t) = \begin{cases} a, & 0 < x \leq x_0 \\ b, & x_0 < x \leq 1 \end{cases} \quad (5.20)$$

и $a, b > 0$. Тогда из (5.19) получаем, что функция распределения, отвечающая этой плотности, непрерывна, но не монотонна:

$$F(x) = \begin{cases} \frac{a}{\Gamma(\alpha+1)} x^{\alpha}, & 0 < x \leq x_0, \\ \frac{b-a}{\Gamma(\alpha+1)} (x-x_0)^{\alpha} + \frac{a}{\Gamma(\alpha+1)} x^{\alpha}, & x_0 < x \leq 1. \end{cases} \quad (5.21)$$

Легко проверяется, что если $0 < x_0 < 1 - (1 - b/a)^{1/(1-\alpha)}$, то $F(x)$ при $b < a$ имеет экстремум в точке $x_0 < x_{extr} < 1$, и эта точка определяется равенством

$$x_{extr} = \frac{x_0}{1 - (1 - b/a)^{1/(1-\alpha)}}. \quad (5.22)$$

Тем самым гистограммная оценка плотности вероятности на множестве дробной размерности может не быть требуемой функцией, если соответствующая функция распределения окажется немонотонной.

Достаточным (но не необходимым) условием того, что $F(x)$ из (5.19) может считаться функцией распределения, является нестрогое возрастание плотности $f(t)$. То, что это условие не необходимо, следует из примера (5.20): если в (5.21) выражение для x_0 удовлетворяет условию $1 - (1 - b/a)^{1/(1-\alpha)} < x_0 < 1$, то функция распределения монотонна.

Приведенные примеры показывают, что интерпретация «аномально-го» поведения случайной величины как движения по фракталу имеет определенные практические неудобства. Более естественным представляется описание эволюции именно выборочной функции распределения без апелляции к моделям генеральной совокупности.

ЧАСТЬ II. ВЫБОРОЧНЫЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

Глава IV. Нестационарные временные ряды

4.1. Понятие временного ряда

Пусть на подмножестве $\mathbf{T} \subseteq \mathbf{R}$ множества действительных чисел \mathbf{R} задана некоторая функция $x(t)$, $t \in \mathbf{T}$. Временным рядом называется любое счетное множество X упорядоченных по времени значений величины x . Конечное упорядоченное по времени подмножество Q этого множества, состоящее из всех тех элементов множества X , которые отвечают выбранным значениям времени из промежутка $\Delta(t, \tau) = (t; t + \tau]$, называется выборкой из временного ряда, а промежуток $\Delta(t, \tau)$ — окном выборки.

На практике временной ряд получается как результат наблюдения за некоторой меняющейся во времени величиной $x(t)$. Если обозначить через n порядковый номер момента наблюдения (измерения) величины x , то моменты времени $t_1, t_2, \dots, t_n, \dots$ называются моментами наступления наблюдаемого события, а соответствующие значения $x_1, x_2, \dots, x_n, \dots$ образуют временной ряд в том смысле, что на множестве натуральных чисел величина x определена по результатам наблюдений как функция времени: $x = x(t)$. В результате наблюдений, естественно, всегда получается конечная совокупность величин $x(t)$. Таким образом, временной ряд — это некоторая абстракция, показывающая потенциальную возможность бесконечно долгого наблюдения за изучаемой величиной.

Здесь следует заметить, что для экспериментально получаемого временного ряда время t измеряется по часам наблюдателя, которые имеют, во-первых, конечную шкалу мелкости измерения времени, и, во-вторых, само наблюдение может осуществляться с погрешностью измерения. Поэтому все измеряемые величины являются случайными в том смысле, что абсолютно точно значения пары «момент времени — результат измере-

ния», строго говоря, не известны. Эти погрешности в дальнейшем нас интересовать не будут, и мы будем считать, что моменты времени и значения наблюдаемой величины определяются «абсолютно точно». Анализу подлежит другой аспект случайности — а именно, статистические свойства последовательности абсолютно точно известных величин. Это означает, что возможная нерегулярность поведения последовательности величин $\{x_n\}$ имеет внутреннюю природу, и не всегда допускает интерпретацию «детерминированное движение плюс шум». Один из центральных вопросов состоит в том, с какой точностью «траектория» x_n все-таки может быть приближена динамической системой, и как распознать наилучшую в смысле аппроксимативных свойств динамическую систему.

Изучаемые нами наблюдаемые величины будут относиться к одному из двух типов.

Первый тип — это величины, которые существуют независимо от того, наблюдает за ними кто-нибудь или нет. Это, например, температура воздуха в комнате, количество биомассы на некоторой территории и т.п. Если не иметь в виду тонких физических теорий о квантовом строении материи, в которых пространство и время по существу дискретны, то в остальных случаях можно считать, что время непрерывно. Изучаемая же система предполагается объективно существующей в любой момент времени, т.е. наблюдатель может в любой момент обнаружить такую систему в некотором состоянии посредством проведения определенного измерения. При этом удобно проводить измерения через одинаковые промежутки времени, которые составляют одинаковое количество минимальных промежутков на шкале часов наблюдателя. Положив в таком случае шаг по времени единичным, можно отождествить момент наблюдения с номером члена ряда. Тогда на промежутке времени длительности T измерения проводятся в моменты времени $t_0 + 1, \dots, t_0 + T$. Указанный набор моментов времени будем, как и выше, обозначать $\Delta(t_0, t_0 + T) \equiv \Delta_T(t_0)$, отличая его от самого временного промежутка $(t_0, t_0 + T]$.

Второй тип наблюдаемых — это величины, реализующиеся по субъективным причинам. Например, если изучаемой величиной является длительность телефонного разговора по некоторому телефонному аппарату, то совершенно естественно, что наблюдать такое явление (телефонный разговор) можно не в любой момент времени, а лишь тогда, когда поступил вызов. Точно так же биржевой ряд цен образован актами купли-

продажи, которые совершаются не непрерывно, а в те моменты времени, когда спрос и предложение совпали по цене и объему. Таким образом, в этих случаях моменты времени, в которые измеряется наблюдаемая величина, не зависят от наблюдателя и сами представляют реализацию некоторого случайного процесса.

Для величин первого типа выборкой объема T на промежутке $\Delta(t_0, t_0 + T)$ называется набор из T значений ряда, полученных в моменты времени $t_0 + 1, \dots, t_0 + T$. В этом случае количество элементов выборки численно совпадает с величиной временного промежутка, измеренной в «единичных шагах».

Для величин второго типа существуют две различных выборки, подлежащие изучению: количественная выборка из N элементов и «временная» выборка из элементов, содержащихся в промежутке $(t_0, t_0 + T]$.

В соответствии с этими двумя типами выборок («штучной» и «временной») возможны и два варианта прогнозных задач. Пусть известны значения ряда в моменты времени t_1, t_2, \dots, t_n . Требуется найти значения ряда в моменты t_{n+1}, \dots, t_{n+k} , т.е. дать прогноз на k наблюдений вперед, либо же дать прогноз значений, которые будут наблюдаемы в промежутке $\Delta(t_n, t_n + \tau)$. Во втором случае надо также спрогнозировать и количество событий, которое будет наблюдаемо в этом промежутке.

Для математической корректности описания событий в терминах теории вероятностей вводится вполне определенная аксиоматика. Кратко напомним ее, отсылая за подробностями к фундаментальным курсам [21, 22].

При аксиоматическом определении вероятности первоначально вводимыми понятиями являются событие и вероятностное пространство. Событием A называется некоторое подмножество конечного или бесконечного множества \mathfrak{R} элементов $\{E\}$, называемых элементарными исходами статистического эксперимента. Элементарное событие E элементарно в том смысле, что оно представляет собой так называемый неразложимый исход эксперимента, т.е. такой исход, который представляется одним и только одним событием. Определение единственности исхода требует введения своего рода причинности в вероятностную схему. Именно, говорят, что событие A влечет за собой событие B , если B наступает всегда, когда наступает событие A : $A \subset B$. В этих терминах событие E

называется элементарным, если для любого события A оно влечет либо A , либо \bar{A} .

Элементарные события E_1, E_2, \dots, E_n образуют полный набор событий, если они попарно несовместны (т.е. произведение любых двух из них — невозможное событие, никогда не наблюдаемое в эксперименте) и сумма всех элементарных событий — достоверное событие, т.е. событие, всегда наблюдаемое в эксперименте. Обозначим полное множество событий через S . Вероятностью называется определенная на полном множестве событий вещественная функция P , удовлетворяющая трем условиям:

- 1) $\forall A \in S \quad P(A) \geq 0$;
- 2) $P(S) = 1$;
- 3) для любого конечного или бесконечного набора попарно несовместных событий $E_k \in S$ справедлива формула

$$P(E_1 \cup E_2 \cup \dots) = P(E_1) + P(E_2) + \dots$$

Нормируемость $P(S)$ на единицу связана с геометрическим смыслом вероятности: именно, вероятностью $P(A)$ наступления события A в этой трактовке называется отношение количества благоприятных для A исходов к общему числу исходов.

Непустое множество Ω событий называется алгеброй, если оно включает в себя достоверное и невозможное события и удовлетворяет двум условиям:

- 1) если $A \in \Omega$, то и $\bar{A} \in \Omega$;
- 2) если $A, B \in \Omega$, то и $A \cup B \in \Omega$, а также и $A \cap B \in \Omega$.

Если алгебра событий такова, что вместе с каждой бесконечной последовательностью событий A_k , $k = 1, 2, \dots$ она содержит и события $\bigcup_k A_k$ и $\bigcap_k A_k$, то такая алгебра называется полем событий Ω или σ -алгеброй.

Описанная абстрактная вероятностная модель является обобщением следующей практической ситуации. Ставится эксперимент, состоящий в измерении некоторой величины, значения которой являются действительными числами. Поскольку измерение не может проводиться с любой точностью, то в эксперименте фиксируется принадлежность измеряемой величины одному из промежутков вида $[a, b)$, на которые разбита число-

вая прямая. Пусть промежутки попарно не пересекаются, а их объединение образует всю числовую ось. Тогда указанные промежутки являются элементарными событиями, вся числовая прямая — полным множеством событий, а всевозможные конечные суммы промежутков — алгеброй событий.

Вероятностным пространством или полем вероятностей называется совокупность $\{S, \Omega, P\}$ трех объектов: полного пространства элементарных событий S , σ -алгебры Ω подмножеств пространства S и вероятностной меры $P(A)$, определенной для любого $A \in \Omega$.

Случайной величиной ξ называется любое вещественное отображение $g: S \rightarrow X$ σ -алгебры подмножеств из пространства S в σ -алгебру подмножеств измеримого пространства X , в котором принимает значение результат эксперимента, так что $g(S) = X$, и значения этого отображения $\xi = g(E) = x$, т.е. результаты эксперимента, определены на множестве элементарных событий.

Распределением случайной величины ξ называется вероятностная мера $P_\xi(\omega) = P(\xi(\omega) \in \Omega)$, заданная на σ -алгебре подмножеств вероятностного пространства.

Случайным процессом называется семейство $\{x(t), t \in T\}$ случайных величин $x(t)$, заданных для всех значений t из множества T на некотором измеримом вероятностном пространстве, т.е. это семейство измеримых отображений пространства элементарных событий в фазовое пространство процесса. Если переменная t принимает дискретные значения, процесс называется временным рядом.

В предположении выполнения условий так называемой эргодической теоремы Биргхофа–Хинчина для стационарного процесса $x(t)$, ПФР которого $f(x)$ имеет конечное математическое ожидание и дисперсию, а автокорреляционная функция стремится к нулю с увеличением лага, среднее значение по всей реализации процесса совпадает со средним по генеральной совокупности:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt = \int x f(x) dx. \quad (1.1)$$

Эта теорема позволяет заменить вычисление математических ожиданий случайных величин по функции распределения на вычисление средних арифметических по выборке значений, которая представляет единственную реализацию случайного процесса. И наоборот, во многих прикладных задачах теории надежности вероятность, например, отказа прибора в течение заданного времени оценивается по распределению интервалов времени до первого отказа для достаточно большой совокупности идентичных приборов. Далее нас будет интересовать ситуация, когда процесс нестационарный, а генеральное распределение в текущий момент времени не известно. Тогда анализу доступна только левая часть из (1.1) при конечном промежутке времени T , а в правой части в этом случае интеграл берется от выборочной плотности функции распределения.

4.2. Вероятностные распределения и стационарность

В теории случайных процессов вводятся два уровня стационарности временного ряда — в широком и узком смыслах. Оба они используют понятие генеральной совокупности, т.е. гипотетической плотности функции распределения, определяющей в каждый момент времени вероятность наблюдения того или иного значения случайной величины.

Даже будучи непрерывной, на практике случайная величина ξ принимает дискретное множество значений или принадлежит конечному множеству промежутков значений x_1, x_2, \dots, x_n , которыми исчерпываются все возможные варианты. Таким образом, предполагается, что пространство событий, образованное множеством x_1, x_2, \dots, x_n , является полным.

Распределением дискретной случайной величины ξ называется совокупность $P(\xi = x_i) = p_i$ значений вероятностей в полном пространстве событий, так что $\sum_{i=1}^n p_i = 1$.

Функцией распределения (ФР) $F_\xi(x)$ величины ξ называется вероятность того, что значение ξ не превосходит x : $F_\xi(x) = P(\xi \leq x)$. Если из контекста ясно, о какой именно случайной величине идет речь, нижний индекс ξ у функции F будем опускать и писать просто $F(x)$. Пусть

элементы x_1, x_2, \dots, x_n пространства событий упорядочены по возрастанию. Тогда, если $x < x_1$, то $F(x) = 0$; если же $x_k \leq x < x_{k+1}$, $k \leq n-1$, то

$$F(x) \equiv F(k) = \sum_{i=1}^k p_i. \quad (2.1)$$

Если $x \geq x_n$, то, очевидно, $F(x) = 1$.

Многомерные функции распределения вводятся аналогично: если $\xi, x \in R^S$, то $F(x) = P(\xi_1 \leq x_1, \dots, \xi_S \leq x_S)$ есть совместная вероятность того, что $\xi_1 \leq x_1, \dots, \xi_S \leq x_S$.

Квантилем порядка q для одномерной ФР $F(x)$ называется такое число x_q , что

$$F(x_q) = q. \quad (2.2)$$

В частности, медианой распределения называется квантиль порядка $q = 1/2$.

Генеральной совокупностью при проведении статистического эксперимента является гипотетическая «урна» с бесконечным числом шаров, на каждом из которых указано ровно одно из чисел x_1, x_2, \dots, x_N в одномерном случае или k чисел в k -мерном, причем доля шаров x_i равна p_i . Наблюдением в смысле вероятностной конструкции является выбор из этой урны произвольного шара. В стационарном случае выбор шара не влияет на состав оставшихся шаров, т.е. с технической точки зрения это означает, что шар после наблюдения возвращается обратно. В нестационарном случае после каждого наблюдения состав шаров в урне произвольно меняется.

Концепция «урны с шарами» чисто умозрительная, потому что из реальной урны с реальными шарами конечного радиуса нельзя выбрать шар «на удачу», ибо одни шары перекрывают доступ к другим шарам, т.е. такая модель отвечает коррелированному процессу, и лишь для точечных шаров она имеет смысл как реализация случайного процесса.

Распределение называется стационарным в узком смысле, если набор вероятностей $\{p_i\}$ не меняется во времени. Распределение называется

стационарным в широком смысле, если среднее значение и дисперсия случайной величины не меняются во времени. Эти характеристики распределения вводятся по формуле

$$\mu = E[\xi] = \sum_{i=1}^n x_i p_i, \quad \sigma^2 = E[(\xi - \mu)^2] = \sum_{i=1}^n (x_i - \mu)^2 p_i, \quad (2.3)$$

где символ E обозначает взятие указанной справа суммы и называется математическим ожиданием случайной величины, содержащейся в квадратных скобках.

Аналогичные определения записываются и для непрерывных случайных величин. Если случайная величина ξ непрерывна и принимает значения из множества действительных чисел, то ее ФР не отрицательна, не убывает и непрерывна слева, причем

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1. \text{ Такая функция определяет меру}$$

Лебега–Стилтьеса на R :

$$P(\xi \leq x) = F(x) = \int_{-\infty}^x dF. \quad (2.4)$$

Дадим определения часто встречающихся функционалов от ФР для дискретной или непрерывной случайной величины, понимая под интегралом в дискретном случае сумму вида (2.3).

Моментами порядка r распределения $F(x)$, если они существуют, называются величины

$$m_r = E\xi^r = \int_{-\infty}^{+\infty} x^r dF. \quad (2.5)$$

Центральными моментами порядка r называются величины

$$m_r^0 = E(\xi - m_1)^r = \int_{-\infty}^{+\infty} (x - m_1)^r dF. \quad (2.6)$$

Пусть абсолютные моменты распределения существуют и конечны: $E|\xi|^k < \infty$. Тогда характеристическая функция (ХФ) $\chi(s)$ распределения $F(x)$, которая определяется как математическое ожидание величины $\exp(is\xi)$, может быть представлена сходящимся рядом по моментам ФР:

$$\chi(s) = E \exp(is\xi) = \int_{-\infty}^{\infty} \exp(isx) dF = \sum_{k=0}^{\infty} m_k (is)^k / k!. \quad (2.7)$$

Обратно, если известна характеристическая функция, то моменты распределения выражаются через нее по формуле

$$m_k = (-i)^k \chi^{(k)}(0). \quad (2.8)$$

При достаточно общих предположениях о распределении оно может быть восстановлено по своим моментам. Имеет место следующая теорема [22].

▼ **Теорема 2.1.** (Карлеман, 1926). Одномерное распределение вероятностей с точностью до множества лебеговой меры нуль определяется своими

моментами $\{m_k\}$, если они существуют и ряд $\sum_{k=0}^{\infty} |m_k| s^k / k!$ сходится при некотором $s > 0$. Распределение $F(x)$ определяется своими мо-

ментами однозначно, если ряд $\sum_{k=0}^{\infty} (m_{2k})^{-1/(2k)}$ расходится. ▲

Таким образом, знание всех моментов распределения во многих случаях может быть эквивалентным знанию самого распределения. В то же время существуют примеры распределений, которые не определяются своими моментами однозначно. Таковым является распределение случайной величины $\eta = \exp(\xi)$, где случайная величина ξ распределена нормально. Другим примером является распределение с плотностью

$$f(x) = F'(x) = \exp(-c|x|^p) \cdot (1 + \beta \cos(|x|^p)),$$

$$c > 0, \quad 0 < p < 1, \quad |\beta| < 1, \quad x \in R$$

Для этого распределения при всех значениях β из указанного промежутка моменты одинакового порядка совпадают между собой.

Дадим также определения моментов и ХФ для многомерных распределений.

Смешанным моментом порядка r n -мерной случайной величины ξ называется величина

$$M_{k_1, k_2, \dots, k_n} = E \xi_1^{k_1} \dots \xi_n^{k_n} = \int_{R^n} x_1^{k_1} \dots x_n^{k_n} dF(\mathbf{x}), \quad (2.9)$$

$$k_1 + \dots + k_n = r.$$

Центральный смешанный момент порядка r определяется формулой

$$\alpha_{k_1, k_2, \dots, k_n} = E \left(\xi_1 - M_{1,0,\dots,0} \right)^{k_1} \left(\xi_2 - M_{0,1,0,\dots,0} \right)^{k_2} \dots \left(\xi_n - M_{0,\dots,0,1} \right)^{k_n}. \quad (2.10)$$

Ковариацией компонент ξ_1 и ξ_2 двумерной случайной величины ξ называется величина

$$\text{cov}(\xi_1, \xi_2) = E(\xi_1 - M_{1,0})(\xi_2 - M_{0,1}) = E\xi_1\xi_2 - M_{1,0}M_{0,1}. \quad (2.11)$$

Коэффициент корреляции определяется формулой

$$\rho_{12} = \rho_{21} = \text{cov}(\xi_1, \xi_2) / \sqrt{\alpha_{2,0}\alpha_{0,2}}. \quad (2.12)$$

Следует подчеркнуть, что $\rho_{12} = \pm 1$ тогда и только тогда, когда случайные величины ξ_1 и ξ_2 связаны линейной зависимостью вида $y = ax + b$ с постоянными значениями параметров a и b (см., напр., [22, стр. 243]).

Если случайные величины ξ_1 и ξ_2 независимы, то их коэффициент корреляции равен нулю. Как известно, обратное утверждение неверно. Например, пусть ξ и η — одинаково распределенные случайные величины (в частности, значения занумерованных граней двух игральных костей), и пусть $\xi_{1,2} = \xi \pm \eta$. Тогда $M\xi_1\xi_2 = M\xi^2 - M\eta^2 = 0$, причем $M\xi_2 = M\xi - M\eta = 0$. Следовательно, $\text{cov}(\xi_1, \xi_2) = 0$, но величины ξ_1 и ξ_2 зависимы: сумма и разность двух целых чисел либо обе четны, либо обе нечетны. Поэтому статистический метод не позволяет однозначно отобрать независимые величины.

Пусть существует измеримая по Лебегу функция $f(x_1, \dots, x_n)$ такая, что элемент вероятностной меры $dF(\mathbf{x})$ представляется в виде

$$dF(\mathbf{x}) = \frac{\partial^n F(\mathbf{x})}{\partial x_1 \dots \partial x_n} d\mathbf{x} = f(\mathbf{x}) d\mathbf{x}. \quad (2.13)$$

Тогда функция $f(x_1, \dots, x_n)$ называется плотностью функции распределения (ПФР). Изучение свойств случайных величин удобно проводить в терминах ПФР, поскольку эта функция является естественным обобщением дискретного распределения вероятностей.

Обычно при анализе временных рядов исходят из того, что существует стационарное распределение, из которого и берутся наблюдаемые значения. Тем самым цель эксперимента состоит в определении среднего значения распределения, т.е. его первого момента, а отклонения от среднего трактуются как шум. Если же распределение нестационарно, и закон его изменения во времени не известен, то вышеописанная вероятностная схема теряет обоснованность. Для краткости ряды, распределение которых стационарно, будем называть стационарными, а если их ФР нестационарна, то нестационарными.

Заметим теперь, что фактически в распоряжении исследователя имеется не вся генеральная совокупность, а лишь конечная выборка из нее. При этом не всегда бывает оправданно представление о том, что существует некое теоретическое распределение вероятностей появления того или иного события, реализованное в виде объективно существующего вышеупомянутого «ящика с шарами». Для такого представления и эксперимент должен быть организован соответствующим образом, чтобы экспериментатор «извлекал» значение измеряемой величины из подходящего ящика.

Например, для макроскопического сосуда с газом такое представление естественно, поскольку большое число молекул можно рассматривать как физически существующий генератор бесконечного набора исходов. Однако во многих практических случаях исходы наблюдений не генерируются из некоторой совокупности. Более того, если «ящики» для каждого наблюдения разные, то по единственной реализации опыта невозможно составить представление о распределении «шаров». Гипотеза же о том, что «ящики» содержательно одинаковые, не может быть проверена. Тогда применение математического инструмента, который удобен и полезен в силу того, что позволяет корректно доказывать те или иные утверждения о последовательности наблюдаемых величин, но только если те являются реализациями стационарного процесса, может привести к большим неточностям.

На практике анализ проводится не для бесконечного, а для конечного объема данных, заключенных в определенном временном или количественном окне. При этом «на взгляд» бывает довольно сложно определить, является ли изучаемый ряд стационарным или нет. Сразу оговоримся, что ряды с явно выраженным трендом, который может быть удален посредст-

вом перехода к первым разностям или с помощью подходящей функциональной аппроксимации, мы не рассматриваем. Нашей целью является анализ рядов, которые не могут быть сведены к стационарным такими преобразованиями.

Типичным представителем рядов, которые будут анализироваться в этой книге, является ценовой ряд на финансовом или сырьевом рынках. Обычно при анализе таких рядов сразу переходят к первым разностям, изучая вместо ряда, образованного значениями цен x_n , ряд приростов цен, который определяется как $y_n = x_{n+1} - x_n$.

На рис. 4.2.1 показана выборка, составленная из фрагментов трех рядов с различными статистическими свойствами: нестационарного ряда приростов суточных данных курса евро/доллар, стационарного белого шума с равномерным распределением и ряда, образованного стохастической динамической системой $x_{n+1} = 4x_n(1 - x_n)$. Из каждого ряда были взяты по 1000 данных, нормированных затем для удобства сравнения на отрезок $[0; 1]$.

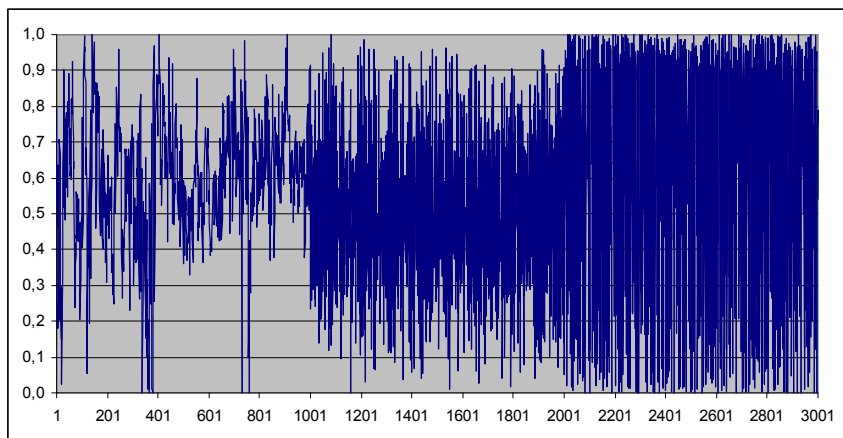


Рис. 4.2.1 — Фрагменты временных рядов

Наша цель состоит в построении статистических критериев, которые позволили бы различать ряды по их уровню нестационарности с точностью, достаточной для практических нужд.

4.3. Выборочные функции распределения и моменты

Как вытекает из предыдущего рассмотрения, основным объектом исследования является не генеральная совокупность, а выборочная функция распределения, построенная по конечному множеству наблюдений.

В действительности множество принимаемых значений $x(t)$ наблюдаемых величин всегда ограничено в силу ограниченности самого промежутка времени наблюдений. Поэтому, чтобы без сопутствующих оговорок рассматривать ряд на любых конечных промежутках, в дальнейшем считаем, что ряд $x(t)$ является равномерно ограниченным по времени. Тогда без ограничения общности множество принимаемых им значений можно считать единичным отрезком $[0; 1]$.

Пусть в выборке объема T значение x_i случайной величины ξ встретилось $n_i(T)$ раз. Относительной частотой этого события называется отношение

$$v_i(T) = n_i(T)/T. \quad (3.1)$$

Совокупность величин $v_i(T)$ называется эмпирическим распределением или выборочной плотностью функции распределения (ВПФР) $f_T(x)$. Эмпирической или выборочной функцией распределения (ВФР) называется ступенчатая неубывающая функция $F_T(x)$, определяемая по эмпирическому распределению частот по формуле, аналогичной (2.1), с той лишь разницей, что теоретические вероятности p_i заменяются на эмпирические $v_i(T)$, зависящие от объема выборки.

Практическая важность введенной ВФР $F_T(x)$ стационарной случайной величины состоит в том (теорема Гливенко, 1933), что она равномерно по x сходится по вероятности к соответствующему распределению $F(x)$ генеральной совокупности при $T \rightarrow \infty$, т.е.

$$P \left\{ \lim_{T \rightarrow \infty} \sup_x |F_T(x) - F(x)| = 0 \right\} = 1. \quad (3.2)$$

Смысл этого утверждения состоит в том, что с увеличением объема выборки вероятность того, что наибольшее отклонение ВФР от генеральной совокупности равно нулю, стремится к единице, т.е. становится достоверным событием. Это, в свою очередь, означает, что эмпирические

вероятности (3.1) сходятся к теоретическим. Именно, относительная частота $V_i(T)$ есть несмещенная состоятельная оценка вероятности p_i .

Напомним, что оценка $\tilde{\theta}_T$ величины θ по выборке длины T называется несмещенной, если ее математическое ожидание равно оцениваемой величине:

$$E[\tilde{\theta}_T] = \theta. \quad (3.3)$$

Равенство (3.3) означает, что если бы нам было известно распределение вероятностей, то, усреднив по этому распределению наблюдаемую эмпирическую статистику, мы получили бы значение параметра θ .

Оценка $\tilde{\theta}_T$ называется состоятельной, если она сходится по вероятности к оцениваемой величине:

$$P\left\{\lim_{T \rightarrow \infty} |\tilde{\theta}_T - \theta| = 0\right\} = 1, \quad (3.4)$$

т.е.

$$\forall \varepsilon > 0 \quad P\left\{|\tilde{\theta}_T - \theta| \leq \varepsilon\right\} \rightarrow 1, T \rightarrow \infty.$$

Если оценка состоятельна, то увеличение объема выборки приводит к уточнению эмпирического значения изучаемой величины.

Мера уклонения ВФР $F_T(x)$ от теоретического распределения $F(x)$ дается следующим утверждением. Если ряд стационарный, то при $T \rightarrow \infty$ разность $F_T(x) - F(x)$ распределена асимптотически нормально с параметрами $\mu = 0$, $\sigma^2 = F(x)(1 - F(x))/T$. Асимптотическая нормальность означает, что при больших объемах выборки плотность вероятности величины $z = F_T(x) - F(x)$ является гауссовой, определяемой выражением

$$P(z \in (a, a + \delta)) = \int_a^{a+\delta} f_G(z; \mu, \sigma) dz, \quad (3.5)$$

$$f_G(z; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z - \mu)^2}{2\sigma^2}\right).$$

Как следует из (3.5), величина уклонения $F_T(x) - F(x)$ неравномерна по x . Для получения равномерных оценок уклонения рассматрива-

ется статистика Колмогорова $D_T = \sup_x |F_T(x) - F(x)|$. Распределение

этой статистики дается теоремой Колмогорова (1933). Если теоретическое распределение $F(x)$ генеральной совокупности непрерывно, то ВФР статистики $\sqrt{T}D_T$ сходится при $T \rightarrow \infty$ к функции Колмогорова $K(z)$:

$$\lim_{T \rightarrow \infty} P\left\{0 < \sqrt{T} \sup_x |F_T(x) - F(x)| < z\right\} = K(z) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2 z^2). \quad (3.6)$$

Обычно распределение генеральной совокупности не бывает известно. Тогда, в предположении, что оно существует, для изучения вопроса принадлежности двух ВФР одному и тому же распределению генеральной совокупности применяется статистика Смирнова

$D_{m,n} = \sup_x |F_{1,m}(x) - F_{2,n}(x)|$, построенная по двум выборкам объ-

мов m и n . Относительно этой статистики справедливо следующее утверждение. Пусть проводятся две независимых серии испытаний по составлению выборок объемов m и n из некоторой генеральной совокупности. Тогда ВФР статистики $\sqrt{mn/(m+n)}D_{m,n}$ сходится при $n \rightarrow \infty$ к функции Колмогорова в смысле формулы (3.6).

При $m = n$ ВФР статистики Смирнова $D_{n,n} \equiv S_n$ определяется в явном виде через конечную сумму отношений биномиальных коэффициентов:

$$P\left\{\sqrt{\frac{n}{2}}S_n < z\right\} = \begin{cases} \sum_{k=-q(n,z)}^{q(n,z)} (-1)^k \frac{C_{2n}^{2n-r(n,z)k}}{C_{2n}^n}, & \frac{1}{\sqrt{2n}} < z \leq \sqrt{\frac{n}{2}}; \\ 1, & z > \sqrt{\frac{n}{2}}; \end{cases} \quad (3.7)$$

$$q(n,z) = \left\lceil \frac{n}{r(n,z)} \right\rceil, r(n,z) = \left\lceil z\sqrt{2n} \right\rceil.$$

Что касается выборочных моментов, т.е. средних значений степеней элементов выборки, то при увеличении объема выборки в стационарном случае они также сходятся к моментам генеральной совокупности. В частности, последовательность средних арифметических сходится к математическому ожиданию не только по вероятности, но и почти наверное, т.е. с вероятностью единица выполняется предельное равенство

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{k=1}^n x_k - \mu \right) = 0. \quad (3.8)$$

Скорость сходимости в (3.8) дается теоремой Леви–Линдеберга, состоящей в том, что если $\{\xi_n\}$ — последовательность взаимно независимых и одинаково распределенных случайных величин, имеющих конечные математическое ожидание μ и дисперсию σ^2 , то последователь-

ность ВФР $\{F_n(x)\}$ нормированных сумм $x_n = \frac{1}{\sigma\sqrt{n}} \left(\sum_{k=1}^n \xi_k - n\mu \right)$

сходится равномерно на \mathbb{R} к нормальному стандартизованному распределению, т.е. распределению Гаусса (3.5) с нулевым средним и единичной дисперсией.

Это утверждение справедливо не только для оценки среднего значения, но и моментов любого порядка. Пусть существуют моменты m_n порядка n генеральной совокупности, называемые генеральными моментами распределения вероятностей $\{p_i\}$:

$$m_n = \sum_{i=1}^N (x_i)^n p_i,$$

где, напомним, x_1, \dots, x_N представляют полный набор возможных исходов для рассматриваемой случайной величины. Тогда выборочные моменты, определяемые по выборке объема T

$$m_n(T) = \frac{1}{T} \sum_{k=1}^T (x_k)^n, \quad (3.9)$$

имеют среднее значение по генеральной совокупности, равное

$$E(m_n(T)) = m_n, \quad (3.10)$$

и дисперсию, равную

$$D(m_n(T)) = \frac{\mu_{2n} - \mu_n^2}{T}. \quad (3.11)$$

В частности, если μ и σ^2 есть генеральные среднее и дисперсия, то

$$Em_1(n) = \mu, \quad Dm_1(n) = \sigma^2 / T. \quad (3.12)$$

Аналогично выборочным центральным моментом $c_n(T)$ порядка n по выборке $\{x_1, \dots, x_T\}$ объема T называется статистика

$$c_n(T) = \frac{1}{T} \sum_{k=1}^T (x_k - m_1(T))^n. \quad (3.13)$$

Обозначим через g_n генеральные центральные моменты порядка n :

$$g_n = \sum_{k=1}^N (x_k - \mu_1)^n. \quad (3.14)$$

Тогда для второго центрального выборочного момента выражения среднего и дисперсии имеют вид:

$$E(c_2(T)) = \left(1 - \frac{1}{T}\right) g_2; \quad (3.15)$$

$$D(c_2(T)) = \frac{g_4 - g_2^2}{T} - 2 \frac{g_4 - 2g_2^2}{T^2} + \frac{g_4 - 3g_2^2}{T^3}.$$

Таким образом, в отличие от нецентральных выборочных моментов, центральные выборочные моменты являются хотя и состоятельными, но смещенными оценками соответствующих теоретических величин. Статистики $s_n(T)$, дающие несмещенные и состоятельные выборочные оценки первых генеральных центральных моментов, имеют вид:

$$s_2(T) = \frac{T}{T-1} c_2(T); \quad s_3(T) = \frac{T^2}{(T-1)(T-2)} c_3(T); \quad (3.16)$$

$$s_4(T) = \frac{T(T^2 - 2T + 3)c_4(T) - 3T(2T-3)c_2^2(T)}{(T-1)(T-2)(T-3)}.$$

В частности, если у генерального распределения существуют конечные второй и четвертый центральные моменты g_2 и g_4 , то

$$M(s_2(T)) = g_2, \quad D(s_2(T)) = \frac{1}{T} \left(g_4 - \frac{T-3}{T-1} g_2^2 \right). \quad (3.17)$$

Выборочным коэффициентом корреляции $R_{xy}(n)$ выборки $\{x_1, \dots, x_n\}$ объема n на выборку $\{y_1, \dots, y_n\}$ называется статистика

$$R_{xy}(n) = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}}. \quad (3.18)$$

Выборочным коэффициентом автокорреляции (или выборочной автокорреляционной функцией, АКФ) $A_T(n)$ выборки $\{x_1, \dots, x_T\}$ объема T с лагом n называется статистика следующего вида:

$$A_N(n) = \frac{(N-n) \sum_{k=1}^{N-n} x_k x_{k+n} - \sum_{k=1}^{N-n} x_k \sum_{j=1}^{N-n} x_{j+n}}{\sqrt{(N-n) \sum_{k=1}^{N-n} (x_k)^2 - \left(\sum_{k=1}^{N-n} x_k \right)^2} \sqrt{(N-n) \sum_{k=1}^{N-n} (x_{k+n})^2 - \left(\sum_{k=1}^{N-n} x_{k+n} \right)^2}}. \quad (3.19)$$

Выборочным квантилем $x_q(N)$ порядка q называется соответствующая квантиль выборочной функции распределения. Если ВФР имеет плотность $f_N(x)$, то x_q определяется по формуле

$$x_q(N) = \int_{-\infty}^{x_q(N)} f_N(x) dx = q = P(\xi \leq x_q). \quad (3.20)$$

Следует отметить, что несмещенные оценки тех или иных статистик нельзя использовать как аргументы для получения несмещенных оценок функций от них (за исключением линейных). Например, если интерес представляет среднеквадратичное отклонение $\sigma(T) = \sqrt{s_2(T)}$, то несмещенной оценкой для σ является не $\sqrt{s_2(T)}$, а величина

$$\sqrt{\frac{2}{T-1}} s_2(T) \frac{\Gamma(T/2)}{\Gamma((T-1)/2)}. \quad (3.21)$$

Подчеркнем, что все формулы этого параграфа, дающие выборочные оценки моментов генеральной совокупности, справедливы только для стационарных рядов.

Если оказалось, что распределение на рассматриваемом промежутке стационарно и близко к генеральному, можно попытаться аппроксимировать распределение его средних значений с помощью нормального распределения. Приведем для справки несколько первых членов такого разложения. Пусть у распределения $F(x)$ существуют конечные центральные моменты g_k вплоть до пятого порядка включительно (g_2 по традиции обозначаем как σ^2). Тогда для ФР $F_n(x)$ нормированных сумм

$z_n = \frac{1}{\sigma \sqrt{n}} \left(\sum_{k=1}^n x_k - n\mu \right)$ равномерно по $x \in \mathbb{R}$ имеет место асимптотическое разложение

$$\begin{aligned} F_n(x) = & F_{G,0}(x) - \frac{g_3}{3! \sigma^3} H_2(x) \frac{f_{G,0}(x)}{\sqrt{n}} - \\ & - \left(\left(\frac{g_4}{\sigma^4} - 3 \right) \frac{H_3(x)}{4!} + \frac{10g_3^2}{\sigma^6} \frac{H_5(x)}{6!} \right) \frac{f_{G,0}(x)}{n} - \\ & - \left(\left(\frac{g_5}{\sigma^5} - \frac{10g_3}{\sigma^3} \right) \frac{H_4(x)}{5!} + \frac{35g_3}{\sigma^3} \left(\frac{g_4}{\sigma^4} - 3 \right) \frac{H_6(x)}{7!} + \frac{280g_3^3}{\sigma^9} \frac{H_8(x)}{9!} \right) \times \\ & \times \frac{f_{G,0}(x)}{n^{3/2}} + o\left(\frac{1}{n^{3/2}}\right), \end{aligned} \quad (3.22)$$

где $H_n(x)$ — полиномы Эрмита, определяемые равенствами

$$\begin{aligned} H_n(x) = & (-1)^n \exp(x^2/2) \frac{d^n}{dx^n} \exp(-x^2/2) = \\ = & n! \sum_{k=0}^{[n/2]} \frac{(-1)^k x^{n-2k}}{k! (n-2k)! 2^k}, \end{aligned} \quad (3.23)$$

$f_{G,0}(x)$ есть стандартизованная плотность распределения Гаусса (3.5) с нулевым средним и единичной дисперсией, а $F_{G,0}(x)$ есть соответствующая интегральная функция распределения

$$F_G(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) dz = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right). \quad (3.24)$$

Функция $\operatorname{erf}(x)$ называется функцией ошибок:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-z^2) dz = 2F_{G,0}(x\sqrt{2}) - 1. \quad (3.25)$$

Величины $\gamma_1 = \frac{g_3}{\sigma^3}$, $\gamma_2 = \frac{g_4}{\sigma^4} - 3$, составленные из центральных

моментов генерального распределения, называются соответственно коэффициентами асимметрии и эксцесса. Из (3.22) следует, что если у распределения исходных случайных величин x_k асимметрия равна нулю, то распределение нормированной суммы этих случайных величин близко к нормальному с суперсостоятельной оценкой порядка $O(1/n)$, т.е. сходимость идет быстрее, чем $1/\sqrt{n}$. Если же и эксцесс равен нулю, то суперсостоятельная оценка имеет порядок $O(n^{-3/2})$. Отметим, что для нормального распределения оба указанных коэффициента равны нулю.

4.4. Отличие моментов стационарных рядов от нестационарных

Если временной ряд нестационарный, то доказательных утверждений о сходимости выборочных моментов к моментам генеральной совокупности нет, поскольку последняя отсутствует. Поэтому перед применением методики оценивания моментов бывает полезно провести экспресс-проверку ряда на стационарность. Для этого достаточно построить зависимость выборочной дисперсии, вычисленной в какой-нибудь (любой) момент времени, от объема выборки. Если рассматривается выборка пошаговых данных из «окна» $\Delta(t_0, t_0 + T)$ с началом в момент $t_0 + 1$, то при изменении объема выборки от 1 до T получается набор из T выбо-

рок увеличивающегося объема $\Delta(t_0, t_0 + n)$, где n меняется от 1 до T . Для каждой такой выборки можно вычислить среднее значение $m(t_0, n)$ и дисперсию $\sigma^2(t_0, n)$, определяемые формулами

$$m(t_0, n) = \frac{1}{n} \sum_{k=t_0+1}^{t_0+n} x_k, \quad \sigma^2(t_0, n) = \frac{1}{n} \sum_{k=t_0+1}^{t_0+n} (x_k - m(t_0, n))^2. \quad (4.1)$$

Поведение дисперсии (4.1) для стационарных и нестационарных рядов принципиально различно. Соответствующие примеры приведены на рис. 4.4.1. В качестве стационарного ряда был взят ряд, равномерно распределенный на $[0; 1]$ (см. рис. 4.2.1, значения с 1001 по 2000), а фрагмент нестационарного ряда образован приращениями тиковых цен фьючерсных контрактов на нефть WTI. Для удобства сравнения этот ряд также был нормирован на отрезок $[0; 1]$.

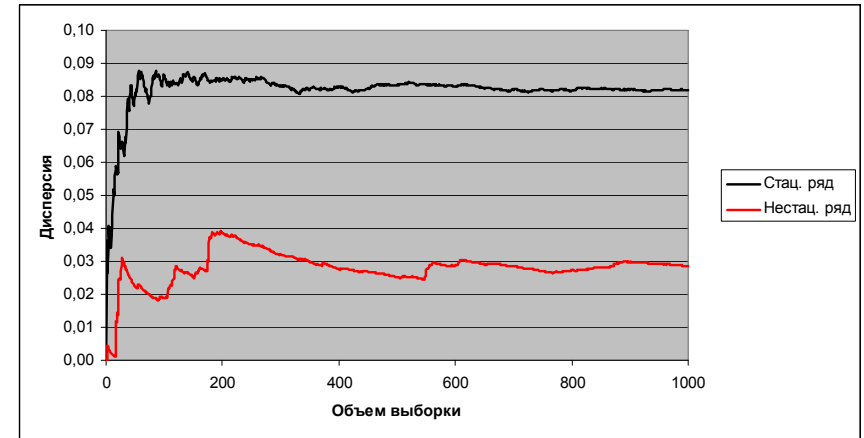


Рис. 4.4.1 — Зависимость выборочной дисперсии от объема выборки

При малых объемах выборки дисперсия для всех рядов сначала возрастает, пока не наберется достаточной репрезентативности данных, а затем ее поведение для разных типов рядов различно. Для стационарного ряда она стабилизируется, имея только «шумовые» колебания умень-

шающей амплитуды, много меньшей самой дисперсии, а в нестационарном случае дисперсия имеет статистически значимые экстремумы.

Наличие глобального минимума дисперсии на объемах выборки справа от первого значимого максимума во многих примерах нестационарных временных рядов позволяет поставить задачу оптимизации объема выборки, чему посвящен следующий параграф.

Отметим, что выборочная дисперсия для нестационарных, но равномерно ограниченных рядов, как и для стационарных, при большом объеме T выборки стабилизируется. Однако если взять такое стабильное, казалось бы, значение, полученное для выборки $\Delta(t_0, t_0 + T)$, и затем определить дисперсию по тому же объему выборки, но в другой момент времени t_1 , то окажется, что дисперсия заметно эволюционирует, тогда как для стационарного ряда колебания выборочной дисперсии описываются формулами (3.17) в соответствии с центральной предельной теоремой. Так, генеральное значение дисперсии равномерно распределенного ряда на $[0; 1]$ равно $1/12$, а дисперсия выборочной дисперсии по объему 1000 данных равна приблизительно $6 \cdot 10^{-6}$.

Пример поведения выборочной дисперсии, построенной по объему 1000 данных, для ненормированного ряда тиковых цен на нефть приведен на рис. 4.4.2.



Рис. 4.4.2 — Эволюция выборочной дисперсии нестационарного ряда

Из рис. 4.4.2 видно, что за промежуток времени в 1000 тиков выборочная дисперсия приростов цен, построенная по объему в 1000 тиков, изменилась более чем на 30% от своего минимального значения за этот период. Следовательно, дисперсию нестационарного ряда нельзя считать приближающейся к некоторой генеральной дисперсии с увеличением объема выборки, несмотря на то, что дисперсия в некоторый конкретный момент времени с увеличением объема выборки стабилизируется, что видно из рис. 4.4.2. Для нестационарных рядов отмеченная стабилизация связана не с предельной теоремой, а с тем фактом, что на практике эти ряды равномерно ограничены по времени.

4.5. Проблема оптимизации объема выборки

Если временной ряд нестационарный, то теоремы об эффективности, состоятельности и асимптотической нормальности выборочных оценок и их дисперсий в общем случае не выполняются. Рассмотрим в этой связи ограничения, присущие стандартным методам анализа и моделирования временных рядов.

В моделях регрессионного анализа вида средние величины (математическое ожидание, дисперсия, ковариация) постоянны во времени. Уточнение этих моделей в случае зависимости указанных величин от времени, т.е. от текущего значения t , может быть сделано посредством аналитического моделирования такой зависимости, либо переходом к первым, вторым и т.д. разностям в нестационарных временных рядах, выражающих зависимость средних величин от времени, либо тем же регрессионным анализом — но уже изучаемых величин на время. Окно усреднения становится при этом скользящим, однако остается невыясненным, какой ширины должно быть это окно.

Проблема промежутка усреднения возникает и при использовании различных авторегрессионных моделей. При этом возникают дополнительные трудности с анализом автокорреляционных функций. Например, необходимо отличать эффекты назначенной периодичности, связанной с суточным, недельным или иным циклом, и внутренне обусловленной зависимости между членами ряда. Увеличение промежутка усреднения в этом случае не приводит к успеху, поскольку зависимость, наблюдавшаяся в одной выборке, может исчезнуть в другой того же объема, но отнесенной к иному моменту времени. Усреднение коэффициентов модели по некоторому промежутку времени и переход к средней модели за период наблюдений увеличивает неточность прогноза на короткий промежуток времени, а на большом интервале такую задачу ставить вообще не очень

осмысленно. Как и в случае с регрессионными моделями, наилучший период усреднения не известен. На практике он определяется исходя из априорных представлений о ряде и на основе экспертных оценок.

Таким образом, регрессионные и автокорреляционные модели прогнозирования стационарных рядов нуждаются в адаптации при использовании их в нестационарном случае, поскольку тогда ошибка прогноза, получаемая этими методами, может не убывать с увеличением статистической базы.

В силу перечисленных трудностей нестационарные ряды изучаются в рамках принадлежности их определенным классам. Такие ряды обладают, как правило, гипотетически навязываемыми им априорными свойствами, и принадлежат тому или иному функциональному классу.

Если не накладывать ограничений на тип нестационарного процесса, то первоочередная задача связана с анализом скорости изменения его выборочной функции распределения и зависимости этой функции от объема выборки. Тем самым горизонт сдвига окна, в котором были заключены значения временного ряда, использованные для построения его ВПФР, зависит от того, какая величина изменения ВПФР допустима для того, чтобы приближенно считать ряд стационарным.

Объем выборки играет очень важную роль при анализе нестационарных временных рядов, и его увеличение чаще всего приводит не к повышению точности прогноза, а к ухудшению, что препятствует осознанию текущей динамики процесса. Уменьшение объема выборки тоже может привести к ухудшению качества прогноза, поскольку дисперсия по выборке малого объема обычно велика из-за недостаточной репрезентативности данных. Тогда можно ожидать, что для нестационарного ряда в каждый момент времени существует выборка оптимального объема, наиболее достоверно представляющая статистические свойства процесса на определенный временной горизонт.

Иначе ведут себя ВПФР стационарных процессов. В соответствии с законом больших чисел выборочные моменты с увеличением объема выборки стабилизируются к генеральным значениям, а эмпирические вероятности сходятся к теоретическим. Однако привести пример такого идеального ряда на практике не так легко. Практически все реальные временные ряды как последовательности значений, получаемых в результате конкретных измерений, являются нестационарными. Это связано с тем, что факторы, формирующие условия проведения случайного эксперимента, в своей основе физические, т.е. подвержены эволюции, причем необратимой. Кроме того, при физических измерениях, если они проводятся многократно, наступает усталость прибора, т.е. возникает определенный

дрейф его систематической ошибки. Все вышеперечисленное затрудняет выделить в чистом виде стационарный процесс, который представляет собой, таким образом, определенную идеализацию реальности. Степенью этой идеализации и является доверительная вероятность, с которой процесс можно отнести к тому или иному удобному классу.

Таким образом, при анализе нестационарного временного ряда необходимо задаться тремя величинами, находящимися между собой в определенной взаимосвязи наподобие того, как связаны между собой плотность, температура и давление в газах. Эти величины следующие: временной промежуток, на котором рассматривается меняющаяся плотность функции распределения (горизонт прогноза), величина изменения ВПФР на этом промежутке (точность прогноза) и объем выборки, обеспечивающий существование первых двух величин в требуемых пределах.

Чтобы конкретизировать идею о нахождении оптимального объема выборки, рассмотрим операцию усреднения данных по некоторому промежутку времени. Пусть $x(t)$ есть значение случайной величины ξ в момент времени t . Тогда текущие оценки статистических характеристик данного процесса (непрерывного или дискретного) используют усреднение определенных операторов $\Psi[x(t)]$ по некоторому промежутку $[t_0, t_0 + T]$:

$$\langle \Psi[x(t)] \rangle_{\Delta} = \frac{1}{T} \int_{t_0}^{t_0+T} \Psi[x(t)] dt. \quad (5.1)$$

В случае дискретных величин под интегралом подразумевается конечная сумма.

В качестве оператора Ψ могут выступать степенные функции (это требуется для нахождения выборочных моментов данного ряда на указанном промежутке), операторы разностного дифференцирования, ядра интегральных преобразований и др.

Применение преобразований вида (5.1) для вычисления текущих оценок статистических характеристик нестационарного процесса $x(t)$ приводит к появлению принципиально неустранимых погрешностей: погрешности, возникающей за счет конечности времени усреднения, т.е. вследствие недостаточной репрезентативности объема выборки, а также погрешности, возникающей за счет изменения статистики на интервале усреднения. Будем называть погрешность оценки статистики (5.1) за счет конечности выборки погрешностью первого типа и обозначать Σ_1 , а по-

грешность за счет нестационарности статистики — погрешностью второго типа и обозначать Σ_2 .

Важно понимать, что по результатам наблюдений нельзя статистически отделить в конечной выборке фактор нестационарности от фактора нерепрезентативности. Поэтому обозначения Σ_1 и Σ_2 соответствуют не строго определяемым величинам, а просто указывают на то, что существуют разные факторы, приводящие к ошибкам. В то же время для минимизации ошибки прогноза желательно уметь анализировать влияние этих факторов по отдельности. Идея метода такого анализа состоит в следующем.

Если случайный процесс стационарный в узком смысле, то две его ВПФР, построенные по любым двум различным выборкам, будут достаточно близки в силу критерия Колмогорова–Смирнова (3.7). Поэтому величину нестационарности можно оценить по близости между ВПФР. Разумеется, поскольку выборки конечны, то между соответствующими ВПФР будет некоторое различие независимо от того, является ряд стационарным или нет. Но с увеличением объема выборок это различие будет стремиться к нулю для стационарных распределений.

С другой стороны, величину статистического разброса естественно оценивать по дисперсии выборочного распределения. В ней также содержится влияние обоих факторов, но если окажется возможным зафиксировать допустимое расстояние между ВПФР и рассматривать только такие выборки, которые порождают достаточно близкие распределения, то дисперсия будет (приближенно) определяться в основном только квазистационарной статистикой ряда. Таким образом, расстояние между ВПФР и величину выборочной дисперсии при фиксированном значении этого расстояния можно рассматривать в качестве кандидатов для формального определения Σ_2 и Σ_1 .

Формализуем поставленную задачу в терминах выборочных функций распределения.

Итак, основной проблемой является минимизация функционала полной ошибки Σ в оценке статистики (5.1). Предположим, что процессы, приводящие к ошибкам первого и второго типов, независимы. Будем тогда определять функционал полной ошибки по формуле

$$\Sigma = \sqrt{\Sigma_1^2 + \Sigma_2^2}. \quad (5.2)$$

Задача оптимизации объема T выборки в (5.1) в терминах функционала ошибки (5.2) состоит в определении такого объема T_{opt} , при котором

$\Sigma(T_{opt}) = \min$, и построении соответствующего численного алгоритма нахождения T_{opt} для заданного нестационарного временного ряда. Оптимизационная задача возникает в силу того, что ошибки $\Sigma_1(T)$ и $\Sigma_2(T)$ как функции объема выборки имеют различное поведение. Именно, чтобы уменьшить ошибку $\Sigma_1(T)$ вследствие неполной статистической репрезентативности, следует увеличивать объем T , а для уменьшения ошибки $\Sigma_2(T)$ от влияния нестационарности на статистические характеристики временного ряда следует уменьшать объем выборки.

Требование соблюдения некоторой заданной величины изменения ВПФР приводит к постановке условия близости, которое объединяет две статистики данного временного ряда $x(t)$: статистику объема выборки $T_{opt}(t)$ и статистику горизонта прогноза $\tau(t)$. Именно, требуется определить $T_{opt}(t)$ так, чтобы ВПФР, построенные по выборкам этих оптимальных объемов на промежутках $[t, t + \tau(t)]$, различались бы меньше, чем на заданное число ε , характеризующее уровень нестационарности.

Ошибкой δ прогноза \hat{x}_i временного ряда $x(t_i) \equiv x_i$ будем называть величину среднеквадратичного отклонения прогнозных значений \hat{x}_i от фактических x_i на промежутке горизонта прогноза (прогнозные величины будем снабжать «крышкой», чтобы отличать их от фактических):

$$\delta = \sqrt{\frac{1}{\tau} \sum_{i=t+1}^{t+\tau} \delta_i^2}, \quad \delta_i = \dot{x}_i - x_i. \quad (5.3)$$

Пусть значения ряда принадлежат некоторому конечному промежутку $[a; b]$, который без ограничения общности можно принять за отрезок $[0; 1]$. Ошибкой прогноза ВПФР в момент времени t будем называть интегральное абсолютное отличие прогнозной ВПФР $\hat{f}(x, t)$ от фактической $f(x, t)$, построенных по выборкам равных объемов. Обозначим соответствующую величину $\varepsilon(t)$:

$$\varepsilon(t) = \int_0^1 \left| \hat{f}(x, t) - f(x, t) \right| dx. \quad (5.4)$$

Отметим, что ошибка прогноза ВПФР и ошибка прогноза временного ряда, по выборке из которого построена данная ВПФР, — это разные ошибки, определенные для разных математических объектов и в различных функциональных пространствах. Следовательно, чтобы использовать величины (5.3) и (5.4) совместно, надо придать корректный теоретико-вероятностный смысл получающемуся функционалу совокупной ошибки (5.2), поскольку сам по себе он лишь формализует некоторую идею, с помощью которой можно было бы оценить меру неточности прогноза нестационарных временных рядов. Такая формализация может быть проведена на основе нижеследующих оценок, получающихся из (5.3), если записать выборочную дисперсию в терминах ВПФР.

Предположим, что практическими соображениями диктуется некоторая максимально допустимая ошибка прогноза δ . В стационарном случае ошибка прогноза в смысле среднего квадратичного определяется дисперсией σ^2 :

$$\sigma^2(t) = \int_0^1 \left(x - \bar{x}(t) \right)^2 f(x) dx. \quad (5.5)$$

Различие же между прогнозными и фактическими средними величинами можно оценить из неравенства

$$\begin{aligned} \left| \bar{\hat{x}} - \bar{x} \right| &= \left| \int_0^1 x \hat{f}(x) dx - \int_0^1 x f(x) dx \right| \leq \\ &\leq \int_0^1 x \cdot \left| \hat{f} - f \right| dx = \int_0^1 |x| \left| \hat{f} - f \right| dx \leq \varepsilon. \end{aligned} \quad (5.6)$$

Последняя оценка в цепочке равенств (5.6) получается с учетом того, что на рассматриваемом промежутке $|x| \leq 1$, а расстояние между ВПФР удовлетворяет условию (5.4).

Рассмотрим ошибку (5.3) прогнозирования значения ряда в некоторый момент времени. Прогнозным значением ряда в момент времени t будем считать среднее значение $\hat{\bar{x}}(t)$, определенное по прогнозному распределению \hat{f} . При сравнении с фактом следует учесть, что фактическая

ВПФР изменилась по сравнению с прогнозной, так что оценка сверху квадрата фактической ошибки прогноза ряда составит величину

$$\begin{aligned} \delta^2 &= \int_0^1 \left(x - \hat{\bar{x}} \right)^2 f(x, t) dx = \int_0^1 \left(x - \hat{\bar{x}} + \bar{x} - \bar{x} \right)^2 f(x, t) dx = \\ &= \sigma^2 + (\bar{x} - \hat{\bar{x}})^2 \leq \sigma^2 + \varepsilon^2. \end{aligned} \quad (5.7)$$

Использованные при выводе оценки (5.7) величины ошибок имеют очевидный смысл: $\sigma = \Sigma_1$, $\varepsilon = \Sigma_2$. Таким образом, оценка (5.7) является обоснованием использования функционала (5.2) в качестве меры неточности прогноза нестационарного временного ряда.

Поскольку для стационарного ряда различие между двумя выборочными функциями распределения из одной и той же генеральной совокупности можно сделать сколь угодно малым, увеличивая объем выборки, то из формулы (5.7) следует, что ошибка прогноза будет тем меньше, чем точнее будет определена дисперсия ряда, т.е. чем большее количество членов этого ряда учитывается при построении ВПФР. Однако если ряд нестационарный, то увеличение объема выборки хотя и приводит к уменьшению ε , одновременно с этим в общем случае приводит к увеличению дисперсии σ^2 .

Таким образом, требуется определить такой объем выборки, при котором верхняя оценка величины в левой части (5.7) минимальна. В этом и состоит цель анализа нестационарного временного ряда. Далее в главе VI вводится индикативная статистика, на основе анализа которой можно провести требуемую оптимизацию.

Глава V. Согласованный уровень стационарности

5.1. Согласованный доверительный интервал

Теория стационарных случайных процессов позволяет по конечной выборке судить о том, с какой доверительной вероятностью ряд является стационарным в широком или узком смыслах. Соответствующая задача состоит в проверке статистических гипотез относительно того, насколько близко поведение выборочных средних значений к тому, что диктуется предельными теоремами. Более сложная задача оценки близости нестационарных распределений будет рассмотрена в следующих параграфах.

В стационарном случае выборочные средние с увеличением объема выборки должны стабилизироваться к соответствующим моментам генеральной совокупности, если последние существуют, а отклонения от этих постоянных величин оцениваются дисперсией выборочных моментов, вычисляемой по формулам (4.3.12) и (4.3.15). Распределение отклонений в стационарном случае является нормальным.

Проверка статистической гипотезы состоит в следующем. По имеющейся выборке вычисляется некоторая статистика, распределение вероятностей которой известно, если считать, что осуществлена именно проверяемая гипотеза. Эта статистика называется критерием. Затем вычисляется предполагаемая вероятность реализации критериальной статистики в данной выборке. Проверяемая гипотеза H_0 называется нулевой. Гипотеза H_1 , состоящая в том, что гипотеза H_0 неверна, называется альтернативной.

Подчеркнем, что статистическая гипотеза, в отличие от иных гипотез, относится к функции распределения случайной величины, а не к самой случайной величине. Например, гипотеза о том, что среднее значение случайной величины равно 0,5, является статистической, а гипотеза о том, что определенное будущее значение временного ряда равно 0,5, таковой не является. Статистика не предсказывает значения случайных величин, которые могут быть наблюдаемы в эксперименте, а оценивает параметры их распределений.

Уровень значимости — это априори заданное значение вероятности реализации критерия, несовместимое с признанием случайности. Событие называется значимым, а не случайным, если вероятность его появления

меньше, чем принятый уровень значимости. Соответствующее значение статистики критерия называется критическим. Дополнение уровня значимости до единицы называется уровнем достоверности. Если значение статистики критерия, вычисленное по имеющейся выборке, больше критического, то на выбранном уровне значимости проверяемая гипотеза отклоняется.

Поскольку принятие или отклонение гипотезы носит вероятностный характер, существует вероятность ошибочного решения. При этом различают ошибки двух типов. Ошибка первого рода состоит в отклонении верной гипотезы. Вероятность ошибки первого рода не выше уровня значимости. Ошибка второго рода состоит в принятии ложной гипотезы. Эффективность или мощность критерия, выбранного для проверки гипотезы H_0 , оценивается вероятностью отклонения ложной гипотезы. Чем выше эта вероятность, т.е. чем меньше вероятность ошибки второго рода, тем более мощным является критерий.

Обычно гипотеза указывает область, называемую доверительной, где могут находиться истинные параметры, характеризующие задачу. Зададим некоторое число $\alpha > 0$ как вероятность ошибки в определении этой области. Вероятность $1 - \alpha$ называется коэффициентом доверия. Для гипотез одностороннего оценивания границы ε , такой, что $P(x > \varepsilon) = \alpha$, доверительная область определяется $1 - \alpha$ -квантилем распределения проверяемой статистики относительно x , плотность распределения которой есть $f(x)$. Двусторонняя оценка интервала (a, b) , содержащего x , на уровне значимости α определяется из условия
$$\int_a^b f(x) dx = 1 - \alpha.$$
 Если распределение статистики близко к симметричному,

то полагают $a = -b$ и требуют выполнения условия
$$\int_{-b}^{+b} f(x) dx = \alpha / 2.$$

Для стационарных распределений отклонения средних значений сумм случайных независимых величин от их генеральных значений распределены асимптотически нормально, поэтому в статистике изучаются главным образом выборки из нормального распределения. Тогда отличие изучаемого распределения от стационарного можно обнаружить, анализируя

его выборочные моменты и определяя, насколько точно можно удовлетворить теореме Леви–Линдеберга.

Рассмотрим задачу оценки параметров нормального распределения (4.3.5). Пусть $\{x_1, \dots, x_n\}$ — выборка из нормального распределения с генеральными значениями μ и σ :

$$f_G(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Поставим задачу об определении среднего значения μ этого распределения по заданной выборке, считая, что генеральная дисперсия σ^2 известна. Обозначим выборочное среднее, определяемое по выборке $\{x_1, \dots, x_n\}$, через $m_1(n)$. В соответствии с теоремой Леви–Линдеберга величина

$$z_n = \frac{1}{\sigma\sqrt{n}} \left(\sum_{k=1}^n x_k - n\mu \right) = \frac{\sqrt{n}}{\sigma} (m_1(n) - \mu)$$

имеет распределение с плотностью $f_G(z; 0, 1)$. Тогда вероятность того, что $|z| \leq \varepsilon$, равна

$$P\{|z| \leq \varepsilon\} = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\varepsilon} e^{-z^2/2} dz = F(\varepsilon) - F(-\varepsilon),$$

причем в силу симметрии $1 - F(\varepsilon) = F(-\varepsilon)$, так что $P\{|z| \leq \varepsilon\} = 2F(\varepsilon) - 1$. Если мы хотим, чтобы эта вероятность была равна требуемому уровню доверия $1 - \alpha$, где α есть уровень значимости (вероятность ошибки первого рода), то следует взять такое значение ε , чтобы $\alpha = 2 - 2F(\varepsilon)$. Следовательно, $F(-\varepsilon) = \alpha/2$, $F(\varepsilon) = 1 - \alpha/2$. Таким образом, по заданному уровню доверия число ε находится из обращения равенства $F(\varepsilon) = 1 - \alpha/2$, т.е. представляет собой $1 - \alpha/2$ -квантиль нормального распределения, который будем для краткости обозначать $\varepsilon = u_\gamma$, $\gamma = 1 - \alpha/2$. В результате получаем, что

с вероятностью $1 - \alpha$ оцениваемое среднее значение распределения накрывается промежутком

$$|\mu - m_1(n)| \leq u_\gamma \frac{\sigma}{\sqrt{n}}. \quad (1.1)$$

Найдем теперь распределение выборочной дисперсии при известном среднем значении μ генеральной совокупности, т.е. распределение случайной величины

$$\chi^2 = \frac{n\hat{\sigma}^2(n)}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - \mu)^2 = \sum_{k=1}^n z_k^2, \quad (1.2)$$

$$\hat{\sigma}^2(n) = \frac{1}{n} \sum_{k=1}^n (x_k - \mu)^2,$$

где z_k — независимые нормально распределенные случайные величины с нулевым средним и единичной дисперсией. Это означает, что каждая из величин z_k имеет распределение с плотностью $f_G(z_k; 0, 1)$. Тогда совместная плотность вероятности для n таких величин равна $\prod_{k=1}^n f_G(z_k; 0, 1)$, а функция распределения событий, состоящих в том,

что $\chi^2 \leq r$, есть интеграл от этой совместной плотности по n -мерному шару радиуса \sqrt{r} :

$$P\{\chi^2 \leq r\} = \frac{1}{(2\pi)^{n/2}} \int_{|z|^2 \leq r} \exp\left(-\frac{1}{2} \sum_{k=1}^n z_k^2\right) dz_1 \dots dz_n.$$

Поскольку подынтегральная функция симметрична, то, переходя в сферические координаты, получаем

$$P\{\chi^2 \leq r\} = \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^r t^{n/2-1} e^{-t/2} dt.$$

Таким образом, плотность распределения величины χ^2 имеет вид

$$\varphi_{\chi^2}(t; n) = \frac{t^{n/2-1} e^{-t/2}}{2^{n/2} \Gamma(n/2)}. \quad (1.3)$$

Это распределение называется распределением χ^2 (или Пирсона) с n степенями свободы. Степенью свободы называется количество независимых величин в статистической задаче. Оно равно разности между общим числом величин (объемом выборки) и числом наложенных ограничений (связей).

Пусть теперь среднее значение μ и дисперсия σ^2 нормального распределения не известны, как это обычно и бывает на практике. Опять зададим доверительную вероятность $1 - \alpha$ того, что теоретическое значение μ лежит в некоторых пределах, которые требуется найти по выборке $\{x_1, \dots, x_n\}$. При анализе используется выборочное среднее $m_1(n)$, а также поправленная на смещенность дисперсия $s_2(n)$. Для нахождения распределения случайной величины

$$\eta^2 = \frac{(n-1)s_2(n)}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - m_1(n))^2 \equiv \sum_{k=1}^n c_k^2 \quad (1.4)$$

следует учесть, что из n величин в последней сумме не все независимы, поскольку для исходной выборки зафиксировано выборочное среднее, равное $m_1(n)$.

В общем случае плотность $\varphi_\eta(x)$ распределения случайной величины η , которая представляет собой некоторую функцию от n случайных величин $\eta = g(\xi_1, \dots, \xi_n)$ с плотностью совместного распределения $f_\xi(x_1, \dots, x_n)$, определяется формулой

$$\varphi_\eta(u) = \int f_\xi(x_1, \dots, x_n) \delta(g(x_1, \dots, x_n) - u) dx_1 \dots dx_n. \quad (1.5)$$

Тогда для случайной величины η^2 из (1.4) получаем, что она распределена по закону χ^2 (1.3) с $n-1$ степенью свободы. Заметим теперь, что величины $m_1(n)$ и $s_2(n)$ статистически независимы, так что величина $\xi = |\mu - m_1(n)| / \sqrt{s_2(n)}$ представляет собой отношение двух слу-

чайных величин с известными распределениями. Применяя (1.5), получаем, что эта величина имеет распределение Стьюдента с $n-1$ степенью свободы.

Плотность распределения Стьюдента для n степеней свободы дается формулой

$$\varphi(t; n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \quad (1.6)$$

Тогда доверительный промежуток для μ в задаче оценивания среднего при неизвестной дисперсии по выборке из нормального распределения дается неравенством

$$|\mu - m_1(n)| \leq q_\gamma^t \sqrt{\frac{s_2(n)}{n-1}}, \quad (1.7)$$

где q_γ^t есть $\gamma = 1 - \alpha/2$ -квантиль распределения Стьюдента с $n-1$ степенью свободы.

Если выборка не является нормально распределенной, то вышеприведенные формулы для получения доверительных интервалов неприменимы. Следует анализировать собственно выборочное распределение в условиях, когда генеральное распределение не известно. Как уже говорилось выше в п. 4.3, задача о принадлежности двух выборок одной генеральной совокупности решается непараметрической статистикой Колмогорова–Смирнова $S_N = \sup_x |F_{1,N}(x) - F_{2,N}(x)|$, для которой справедлива асимптотика

$$\lim_{N \rightarrow \infty} P\left\{0 < \sqrt{\frac{N}{2}} S_N < z\right\} = K(z). \quad (1.8)$$

Это означает, что если для двух ВФР было найдено значение S_N и вычислена величина $z = \sqrt{\frac{N}{2}} S_N$, то величина $1 - K(z)$ приближенно считается равной вероятности того, что $\sqrt{\frac{N}{2}} S_N \geq z$. Задав уровень зна-

чимости α , считаем, что если $1 - K(z) < \alpha$, то осуществилось маловероятное событие, несовместимое с понятием случайности, и эти выборки следует считать различными. Некоторая неопределенность вывода состоит в том, что надо априори задать желаемый уровень малости критерия $1 - K(z)$. Какую вероятность следует считать достаточной для того, чтобы признать выборки одинаковыми?

Чтобы практически ответить на данный вопрос, поставим статистический эксперимент по генерации выборок из нормального распределения и построим функцию распределения расстояний между независимыми выборками. На рис. 5.1.1 показан временной ряд расстояний между ВФР для двух выборок длины $N = 1000$ данных.

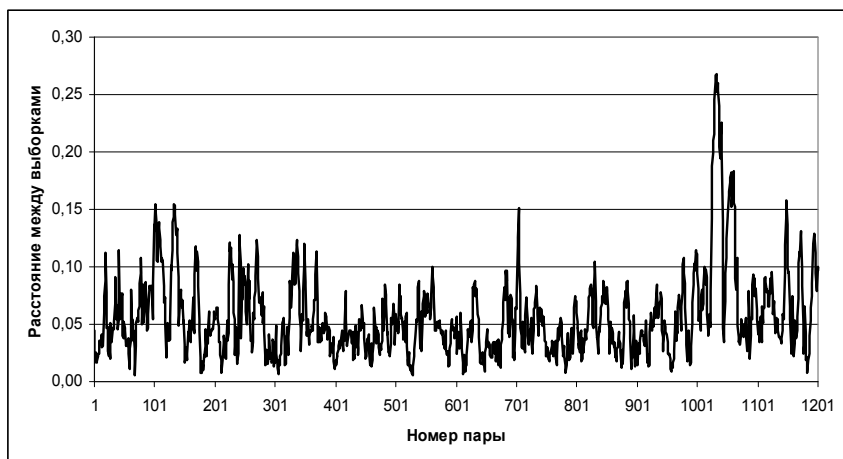


Рис. 5.1.1 — Фрагмент временного ряда расстояний между независимыми ВФР из нормальной совокупности длиной 1000 данных

Распределение расстояний между выборками, временной ряд которых показан на рис. 5.1.1, дано на рис. 5.1.2 по результатам 5000 экспериментов.

Поскольку все выборки в этом примере взяты из одного распределения, то ошибкой на уровне значимости ε будет непризнание выборок «одинаковыми», если расстояние между ними оказалось больше ε . Смысл же уровня значимости таков, что в среднем в доле случаев, равной уровню значимости, будет сделан ошибочный вывод из статистического

эксперимента. Но из рис. 5.1.1, в частности, следует, что при сравнении ВФР для выборок в 1000 данных бессмысленно задавать уровень значимости менее 0,01, так как тогда почти всегда придется признавать выборки различными, а на самом деле это не так. Следовательно, при анализе выборок определенной длины неправильно задавать априори желаемый уровень значимости, так как для заданной длины N выборки лишь при одном значении $\varepsilon = \varepsilon_0(N)$ вероятность превышения значения ε_0 равна значимости используемого для этой цели критерия, и это значение находится из уравнения

$$1 - \varepsilon_0 = K\left(\sqrt{\frac{N}{2}}\varepsilon_0\right). \quad (1.9)$$

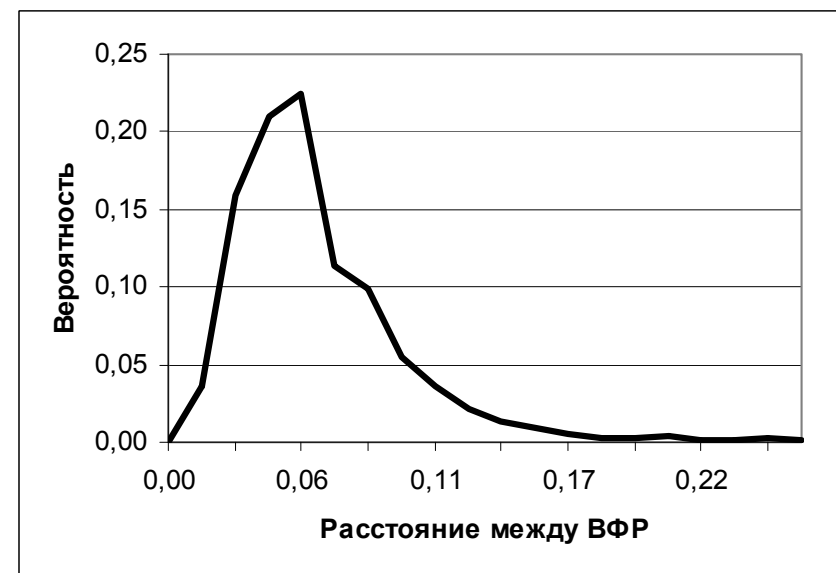


Рис. 5.1.2 — Распределение расстояний между независимыми выборочными распределениями из нормальной совокупности

Решение уравнения (1.9) единственно, поскольку правая часть как функция ε_0 монотонно возрастает от нуля до единицы, а левая монотонно убывает от единицы до нуля. График решения $\varepsilon_0(N)$ приведен на

рис. 5.1.3, а некоторые его значения — в табл. 5.1.1. Найденное решение будем называть согласованным уровнем значимости или согласованным отклонением между ВФР.

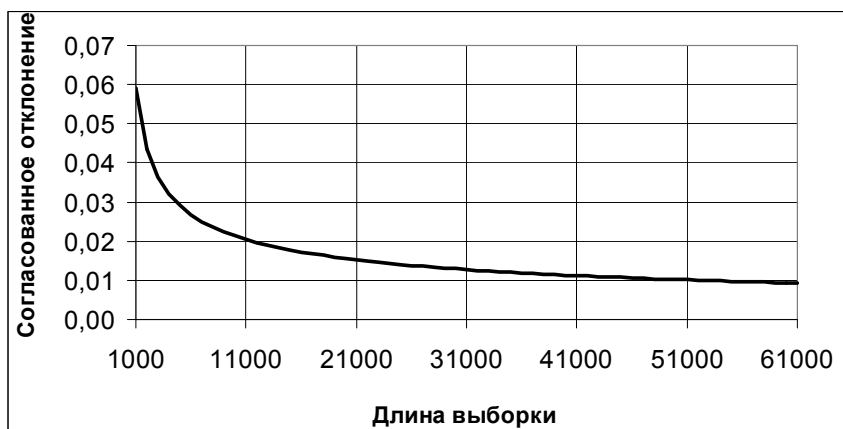


Рис. 5.1.3 — Согласованный уровень значимости для критерия Колмогорова–Смирнова

Из рис. 5.1.3 следует, что, например, вероятность того, что расстояние между двумя независимыми выборками длиной 1000 данных из одного распределения будет больше 0,06, равна 0,06; аналогично для выборок длиной 11 тыс. данных вероятность превышения расстояния 0,02 равна 0,02 и т.д.

Табл. 5.1.1. Значения согласованного уровня значимости критерия Колмогорова–Смирнова

Длина выборки N	Уровень значимости $\varepsilon_0(N)$
100	0,18
200	0,14
300	0,10
400	0,09
500	0,08
1000	0,05926
2000	0,04364
3000	0,03641
4000	0,03220

5000	0,02910
10000	0,02135
50000	0,01023

С дальнейшим увеличением N согласованный уровень значимости уменьшается очень слабо. Например, для $N = 100$ тыс. $\varepsilon_0(N) = 0,009$. Отметим, что на достаточно высоком уровне достоверности (выше, чем 0,99) численно найденное решение уравнения (1.9) при объемах выборок от 1 тыс. до 100 тыс. данных аппроксимируется эмпирической зависимостью

$$\varepsilon_0(N) = 0,06N^{-0,45}. \quad (1.10)$$

Аналогичные рассуждения можно в некоторых случаях провести и для задачи параметрического оценивания при построении доверительных интервалов. Какой уровень значимости α следует априори задать для получения практически разумных результатов? Подчеркнем, что точность оценки, например, среднего значения случайной величины дает не только уровень значимости, но и содержащий это значение интервал. Возможная неприемлемость оценки возникает тогда, когда доверительный интервал оказывается слишком широк, либо уровень значимости слишком велик. Поэтому важно согласовать обе этих величины.

Уменьшая уровень значимости, мы расширяем интервал доверия, но тем самым увеличивается ошибка практического применения полученных оценок точно так же, как она увеличивается при высоком уровне значимости и узком доверительном интервале. Так, практически ненужным является результат, состоящий в том, что с доверительной вероятностью 1,0 среднее значение некой величины заключено в промежутке $[0; 1]$, представляющем все множество допустимых значений этой величины. Точно так же не представляет интереса знание о том, что среднее значение заключено в узком промежутке с доверительной вероятностью, близкой к нулю.

С практической точки зрения выбранный уровень значимости α означает, что при проведении большого числа однотипных экспериментов по проверке статистической гипотезы в условиях, когда она заведомо верна, в доле α случаев формально будет допущена ошибка первого рода. Результаты экспериментов равновероятны, так как выборки берутся из одного и того же распределения. Поэтому практическая точность критерия принятия гипотезы равна наибольшему из двух чисел: априорному уровню значимости и доле меры доверительного интервала в мере множе-

ства допустимых исходов. Для равномерно ограниченной случайной величины удобно нормировать ее значения на промежуток $[0; 1]$. В этом случае можно определить согласованный доверительный интервал условием равенства указанной доли и уровня значимости. Пусть $f_T(x)$ есть выборочная плотность функции распределения изучаемого признака. Тогда согласованным доверительным интервалом $(a; b) \subset [0; 1]$ на уровне значимости ε называется интервал $(a; a + \varepsilon)$, где согласованный уровень значимости $\varepsilon(a)$ при заданном левом конце интервала находится из условия

$$\int_a^{a+\varepsilon} f_T(x) dx = 1 - \varepsilon. \quad (1.11)$$

Сам же левый конец интервала находится в общем случае неоднозначно. Имеет смысл выбрать такое значение a , при котором достигается наилучшая точность в оценке изучаемого показателя, т.е.

$$a_{opt} = \arg \min_a \varepsilon(a). \quad (1.12)$$

Описанный подход позволяет определить фактическую ошибку, совершаемую при практическом использовании того или иного статистического критерия.

5.2. Оценка плотности функции распределения

При проведении статистических экспериментов, совершаемых в дискретные моменты времени, удобно исходить из представления о некоторой непрерывной случайной величине, значения которой могут быть наблюдаемы в любой момент времени. Тогда представляется естественным оценить предполагаемую непрерывную функцию распределения этой величины через построение выборочной плотности распределения, агрегируя наблюдаемые данные в классовые интервалы. Ширина интервалов определяется двумя факторами: во-первых, точностью измерения наблюдаемых величин и, во-вторых, длиной выборки, по которой строится гистограмма. Чтобы можно было говорить о том, что распределение значимо изменилось за некоторый промежуток времени, следует определить статистическую неопределенность в оценке плотности распределения по гистограмме в предположении стационарности генеральной совокупности. Наибольшее из двух чисел — точности измерения случайной величины и

точности в оценке плотности — принимается за ширину интервала разбиения. Заметим, что точность в оценке плотности как раз и зависит от выбранной ширины интервала, поэтому задача определения этой ширины становится самосогласованной. Чтобы получить правило согласования, определим сначала статистическую ошибку в оценке плотности распределения при заданном разбиении гистограммы.

Рассмотрим последовательность из N значений случайной величины ξ , которые попали в определенные классовые интервалы Δ_j , число которых n . Элементы этой последовательности обозначим x_j , $j = 1, 2, \dots, N$. Пусть из них значение x_i встретилось k_i раз. Тогда выборочной плотностью функции распределения, оцениваемой по гистограмме по заданной выборке длины N , называется совокупность $f_N(i, t)$ величин

$$f_N(i, t) = \frac{k_i}{N}, \quad i = 1, 2, \dots, n. \quad (2.1)$$

Время t в аргументе $f_N(i, t)$ указывает на текущий момент времени, от которого назад отсчитывается выборка длины N . Далее для краткости аргумент t опускается, если анализ проводится в некоторый фиксированный момент времени или если конкретное значение момента времени на данном этапе не существенно.

Согласно теореме Леви–Линдберга отклонение выборочного среднего значения $\bar{\xi}(N)$ стационарной случайной величины, определяемого по выборке длины N , от генерального среднего распределено асимптотически нормально с нулевым средним и стремящейся к нулю дисперсией σ^2 / N , где σ^2 есть дисперсия этой величины по гипотетической генеральной совокупности. Рассмотрим в качестве такой случайной величины саму эмпирическую частоту $f(j) = f_N(j)$ попадания в j -ый классовый интервал. Формально эмпирическая частота есть среднее выборочное значение случайной величины, называемой индикатором I_{jk} принадлежности результата k -го наблюдения в j -ому классовому интервалу. Индикатор I_{jk} определяется формулой

$$I_{jk} = \begin{cases} 1, & x_k \in \Delta_j; \\ 0, & x_k \notin \Delta_j. \end{cases} \quad (2.2)$$

Как видно из формулы (2.2), эмпирическую частоту (2.1) можно представить в виде

$$f_N(j) = \langle I \rangle = \frac{1}{N} \sum_{k=1}^N I_{jk}.$$

Выборочная дисперсия эмпирической частоты $f_N(j)$ равна

$$\begin{aligned} s_2(N; j) &= \left\langle \left(I - \langle I \rangle \right)^2 \right\rangle = \langle I^2 \rangle - \langle I \rangle^2 = \langle I \rangle - \langle I \rangle^2 = \\ &= f_N(j) \cdot (1 - f_N(j)). \end{aligned} \quad (2.3)$$

Согласно (1.7), когда генеральная дисперсия не известна, а оценивается только по выборочной дисперсии $s_2(N; j)$, для получения доверительного интервала среднего значения следует рассматривать t -статистику Стьюдента, которая записывается в виде

$$t = \sqrt{N-1} \frac{|f_N(j) - f^*(j)|}{\sqrt{s_2(N; j)}}, \quad (2.4)$$

где $f^*(j)$ есть гипотетическая генеральная вероятность, если бы процесс был стационарным. Смысл статистики (2.4) в том, что на уровне значимости ε выражение $|f_N(j) - f^*(j)|$ не превосходит величины

$$t_{1-\varepsilon/2}(N-1) \sqrt{s_2(N; j)} / \sqrt{N-1},$$

где $t_\alpha(N-1)$ есть α -квантиль распределения Стьюдента с $N-1$ степенью свободы. При больших N можно считать $N-1 \approx N$ и число степеней свободы в квантиле распределения Стьюдента взять для простоты бесконечным (тогда это распределение совпадает с нормальным).

Заметим теперь, что, хотя формально уровень значимости может быть назначен любым, не очень осмысленно, чтобы он был выбран точнее (т.е. меньше) уровня неопределенности в позиционировании доверительного интервала $|f_N(j) - f^*(j)|$.

Таким образом, естественно потребовать выполнения условия

$$\sum_{j=1}^n |f_N(j) - f^*(j)| \leq \varepsilon. \quad (2.5)$$

Очевидно, условию (2.5) можно удовлетворить, потребовав выполнения более жесткого условия — близости каждой из частот в отдельности:

$$|f_N(j) - f^*(j)| \leq \varepsilon^*(j). \quad (2.6)$$

Если же теперь потребовать, чтобы для каждой эмпирической частоты было выполнено условие

$$t_{1-\varepsilon/2} \frac{\sqrt{s_2(N; j)}}{\sqrt{N}} \leq \varepsilon^*(j),$$

то одновременно с выполнением условия (2.6) будет достигнут и требуемый уровень значимости для статистики Стьюдента (2.4). Однако если некоторые вероятности в результате выбранного разбиения на классы оказались малы, много меньше ε , то нет необходимости требовать, чтобы и они были оценены с той же точностью. Поэтому уместно для каждой вероятности выбрать свою точность аппроксимации ε_j и считать, что требуемый в целом уровень значимости определяется средневзвешенной по разбиению точностью, так что

$$t_{1-\varepsilon/2} = \frac{1}{\Sigma_N(n)} \sum_{j=1}^n \sqrt{s_2(N; j)} t_{1-\varepsilon_j/2}, \quad (2.7)$$

где сумма, определяющая влияние мелкости разбиения гистограммы на точность оценки эмпирических вероятностей, равна

$$\Sigma_N(n) = \sum_{i=1}^n \sqrt{s_2(N; j)} = \sum_{i=1}^n \sqrt{f_N(i)(1 - f_N(i))}. \quad (2.8)$$

Сумма (2.8) выражает качество приближения плотности гистограммой, поскольку, чем меньше сумма, тем выше точность оценки ВПФР, т.е. тем меньше число ε . С увеличением числа интервалов сумма (2.8) возрастает, что означает снижение точности оценки ВПФР. В стационарном случае эта сумма представляет собой некий эффективный функционал учета особенностей графика функции плотности при аппроксимации плотности генерального распределения.

Теперь из (2.6), (2.7) получаем, что

$$\sum_{j=1}^n t_{1-\varepsilon_j/2} \frac{\sqrt{s_2(N; j)}}{\sqrt{N}} = t_{1-\varepsilon/2} \frac{\Sigma_N(n)}{\sqrt{N}} \leq \varepsilon \sum_{j=1}^n f^*(j) = \varepsilon,$$

откуда на уровне значимости ε следует оценка

$$\frac{t_{1-\varepsilon/2}}{\varepsilon} \leq \frac{\sqrt{N}}{\Sigma_N(n)}. \quad (2.9)$$

При заданной точности ε и способе разбиения гистограммы формула (2.9) для знака равенства дает оценку на минимальную длину выборки, при которой эта точность достигается в среднем. Поскольку функция $t_{1-\varepsilon}$ табулирована (см. [14]), то функция $t_{1-\varepsilon}/\varepsilon$ известна. Она монотонно убывает с ростом ε , поэтому к ней существует обратная, значение которой и дает верхнюю оценку точности определения эмпирических вероятностей по заданному разбиению гистограммы. Обозначим для краткости

$$\varphi(\varepsilon) = \frac{t_{1-\varepsilon}}{\varepsilon}, \quad \psi = \varphi^{-1}, \quad z \equiv z(N, n) = \frac{\sqrt{N}}{\Sigma_N(n)}. \quad (2.10)$$

Тогда точность оценки ВПФР определяется формулой

$$\varepsilon = 2\psi(2z). \quad (2.11)$$

Некоторые значения функции $\varphi(\varepsilon)$ приведены в табл. 5.2.1.

Табл. 5.2.1. Зависимость функции $t_{1-\varepsilon}/\varepsilon$ от ε

ε	$t_{1-\varepsilon}$	$t_{1-\varepsilon}/\varepsilon$
0,0005	3,291	6582,0
0,0025	2,807	1122,8
0,0050	2,576	515,2
0,0075	2,432	324,3
0,010	2,326	232,6
0,015	2,170	144,7
0,020	2,054	102,7
0,025	1,960	78,4
0,05	1,645	32,9
0,10	1,282	12,8
0,15	1,036	6,9
0,20	0,842	4,2

0,30	0,524	1,7
0,40	0,253	0,6

Отметим, что при $\varepsilon > 0,01$ имеет место аппроксимация [14] квантиля нормального распределения, с которым при больших N совпадает квантиль распределения Стьюдента:

$$t_{1-\varepsilon/2} = \sqrt{-\frac{\pi}{2} \ln(1 - (1-\varepsilon)^2)}. \quad (2.12)$$

Относительная ошибка аппроксимации (2.12) составляет 0,037. Графики функции $\varphi(\varepsilon)$ и ее аппроксимации по формуле (2.12) показаны на рис. 5.2.1. Видно, что практически во всем диапазоне изменения точности, кроме самой прецизионной ($\varepsilon < 0,01$), формула (2.12) может считаться достаточно точной.

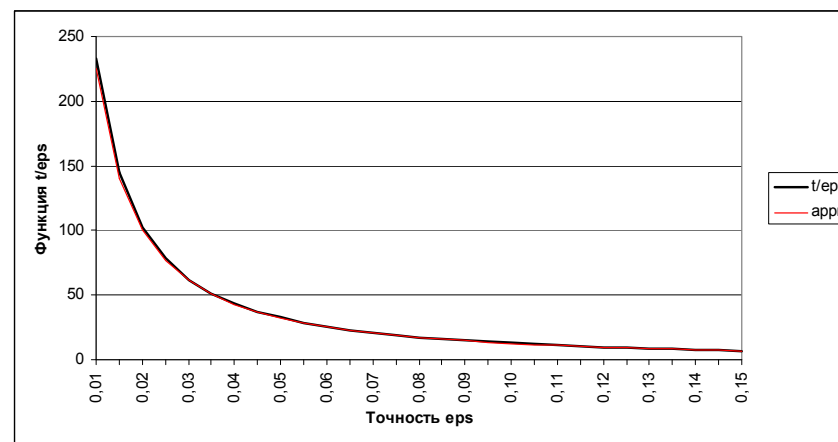


Рис. 5.2.1 — Вид функции $\varphi(\varepsilon)$ из (2.10)

Таким образом, при заданном разбиении гистограммы формула (2.10) позволяет определить точность, с которой оценивается предположительно стационарная ПФР. На практике, однако, разбиение гистограммы, т.е. выбор ширины классовых интервалов, есть задача, решаемая одновременно с оценкой точности ВПФР. В следующем параграфе будет рассмотрено построение оптимального равномерного разбиения гистограммы.

5.3. Оптимальное равномерное разбиение гистограммы

Целью построения оптимального разбиения гистограммы является решение проблемы наиболее точной аппроксимации выборочной плотности функции распределения временного ряда, который в общем случае не является стационарным. Очевидно, вид гистограммы зависит от того, как построены классовые интервалы принадлежности случайной величины. Даже в сильно упрощающем случае равномерного разбиения до сих пор нет удовлетворительного способа построения «правильного» разбиения области гистограммы, представляющей эмпирическую плотность вероятности. «Правильность» разбиения подразумевает, что в стационарном случае ошибка аппроксимации предположительно непрерывной плотности функции распределения кусочно-постоянной функцией минимальна. Трудность состоит в том, что оцениваемая плотность неизвестна, поэтому число интервалов сильно сказывается на виде распределения частот конечной выборки. С одной стороны, при фиксированной длине выборки укрупнение интервалов разбиения ведет к уточнению эмпирической вероятности попадания в них, но слишком сильно сглаживает изучаемое распределение. С другой стороны, измельчение интервалов делает вид распределения неоправданно изрезанным в силу малого количества данных, случайно попадающих в каждый из интервалов. Абсолютно точным является разбиение гистограммы на один интервал, но такое разбиение практически не интересно. Следовательно, задача построения гистограммы эквивалентна задаче оптимальной фильтрации шума в зависимости от объема данных.

Задача оптимального разбиения гистограммы довольно популярна среди специалистов по математической статистике, но нет единого устоявшегося мнения относительно ее решения, поскольку ответы носят полуэмпирический характер, либо требуют априорного знания закона распределения вероятностей, либо вообще имеют характер бытовых советов «группировать так, чтобы было не меньше 6 и не больше 20 интервалов». В классических же курсах по математической статистике эта проблема не обсуждается, видимо, как не решенная до конца или, может быть, как не существенная.

В справочных руководствах [23, 24] довольно часто упоминается эмпирическое правило Стерджеса, предложенное еще в 1926 г. Оно основано не на статистике, а на том формальном наблюдении, что если фактическое число попаданий k_i значений случайной величины в i -ый классовой интервал равно биномиальному коэффициенту

$$k_i = C_{n-1}^{i-1},$$

то, поскольку сумма по всем n интервалам разбиения должна быть равна количеству N элементов в выборке

$$\sum_{i=1}^n k_i = \sum_{i=1}^n C_{n-1}^{i-1} = 2^{n-1} = N,$$

соответствующее число классовых интервалов n зависит от длины выборки N по логарифмическому закону:

$$n = 1 + \log_2 N. \quad (3.1)$$

Это, конечно, слишком идеалистическое требование на группировку данных, не имеющее отношения к реальной ситуации.

В 1942 г. Манн и Вальд дали степенную оценку оптимального числа интервалов для применения критерия хи-квадрат:

$$n = 4 \left(\frac{3}{4} (N-1)^2 \right)^{1/5}. \quad (3.2)$$

Из (3.2) следует, что при больших N число интервалов растет как $N^{2/5}$. Заметим, что оценки (3.1) и (3.2) не зависят от вида распределения и уже потому не являются удовлетворительными, ибо для «плоских» распределений число интервалов должно быть явно меньше, чем для квази-сингулярных, с узкими пиками вероятностей.

В 1950 г. Смирнов показал, что уклонение гистограммы от графика неизвестной непрерывной плотности вероятности убывает как $N^{-1/3}$. Скотт в 1979 г. уточнил эту оценку для аппроксимации дифференцируемой плотности при больших длинах выборки и получил оптимальное число интервалов в виде

$$n = \left(\frac{N}{6} \int_{-\infty}^{+\infty} f'^2(x) dx \right)^{1/3}. \quad (3.3)$$

Применяется также оценка оптимального разбиения для аппроксимации дважды дифференцируемой плотности $f(x)$: число интервалов разбиения равно

$$n = \left(\frac{N |f''(x)|_{\max}}{4 f(x)_{\max}} \right)^{1/5}. \quad (3.4)$$

Формулы (3.3) и (3.4) обладают вырожденностью в отношении равномерного распределения, для которого производная плотности равна нулю. Это означает, что их нельзя применить к плоским трапецевидным распределениям.

В книге [24] приводится эмпирическая формула

$$n = \frac{1 + \kappa}{6} N^{2/5}, \quad (3.5)$$

где κ есть эксцесс распределения. Эта формула представляется наиболее подходящей из вышеперечисленных, поскольку она учитывает зависимость от формы распределения, но и она является полуэмпирической.

Таким образом, формулы (3.3–3.5) различаются по виду степенного закона, а вместе с (3.2) они отличаются от функционального закона (3.1).

Существует также подход, основанный на максимизации информационной энтропии. Он дает следующую оценку числа равных интервалов разбиения гистограммы:

$$n_{opt} = \arg \max_n \left(\sum_{i=1}^n k_i \ln k_i + N \ln n - n \right). \quad (3.6)$$

Однако не очевидно, какое вообще отношение имеет информационная энтропия к задаче об оптимальном разбиении, т.е. отношение-то она имеет в силу статистического характера размещения данных по промежуткам, но не ясно, что дает ее максимизация с точки зрения оптимальной фильтрации.

Мы видим, что предлагаемые формулы не только различны, но и не особенно хорошо обоснованы, т.е. обоснования приводятся косвенные по отношению к решаемой задаче. Тем не менее, после перечисления и, частично, анализа большого количества результатов по оптимальному разбиению гистограмм Новицкий и Зограф приходят в [24] к следующим выводам: «оптимальное число интервалов разбиения, безусловно, существует; оно зависит от эксцесса сильнее, чем от объема выборки; зависимость $n = AN^{2/5}$, где A — теоретически вычисляемый коэффициент, в дальнейших уточнениях, по-видимому, не нуждается». Несмотря на последнее оптимистичное заявление, ситуация с оптимальным разбиением

продолжает обсуждаться в издаваемой и в настоящее время обширной литературе по математической статистике.

Например, в справочнике [23] приведены «рекомендации» (именно в такой формулировке) выбирать число интервалов в соответствии с вышеупомянутыми работами, затем дана практическая рекомендация «выбирать как можно большее n , но не превышающее $N/5$ », а в заключение читатель отсылается к ГОСТу «Прикладная статистика» как источнику, содержащему «исчерпывающие рекомендации по методологии выбора числа классовых интервалов». В упомянутом ГОСТе, к сожалению, вновь приведены отнюдь не рекомендации в прямом смысле этого слова — в каких случаях какое разбиение выбирать, а набор несовпадающих между собой результатов все тех же работ и ряда других, менее известных, после чего приводится таблица рекомендованных ВНИИМ им. Д.И. Менделеева оценок, упоминаемых и в [24] (она короткая: если $N = 40 \div 100$, то $n = 7 \div 9$; если $N = 100 \div 500$, то $n = 8 \div 12$; если $N = 500 \div 1000$, то $n = 10 \div 16$; наконец, если $N = 1000 \div 10000$, то $n = 12 \div 22$). Перечень этих эмпирических правил завершается знакомым уже инженерным советом «выбирать как можно большее n , но не превышающее $N/(5 \div 10)$ », т.е. авторы ГОСТа как будто не доверяют сложным формулам, а предлагают обходиться практическим здравым смыслом.

На практике выбор оптимального равномерного разбиения гистограммы осложняется еще и тем, что для представляющих практический интерес нестационарных временных рядов не существует генеральной совокупности, которую якобы приближает гистограмма, и теоретически ей не от чего «уклоняться». Тем не менее, можно представлять себе и локально меняющуюся по времени плотность распределения.

Отличием предлагаемого подхода от всех вышеперечисленных является введение согласованного уровня значимости, на котором принимается решение об аппроксимации теоретического распределения выборочным: нельзя априори требовать определенной значимости независимо от вида распределения, поскольку им (видом) определяется, в частности, различие между выборочными плотностями распределений независимых выборок равных длин. Желательно, чтобы уровень значимости, теоретически предъявляемый к точности решения задачи, имел бы реальное практическое выражение в виде доли ошибок, допускаемых при многократном повторении соответствующего случайного эксперимента. Заметим здесь, что для функции распределения (не плотности, а интегральной функции) непараметрический критерий Колмогорова–Смирнова не зависит от спо-

соба разбиения гистограммы, но на оценку плотности этот результат, увы, не переносится.

На практике измеряемые величины всегда дискретны, поэтому можно было бы ввести естественное разбиение гистограммы в соответствии с точностью измерения. Например, если рассматривается временной ряд цен на некоторый финансовый инструмент и цены выражены в рублях с точностью до сотых, то естественным разбиением будет разбиение с шагом по цене, равным 0,01. Более мелкий шаг, естественно, делать бессмысленно. Однако для практического анализа такое разбиение будет неудобно, поскольку размах цен за период, например, в месяц может достигать до десятков единиц, что будет означать наличие тысяч классовых интервалов. В таком случае дискретная по существу случайная величина естественным образом начинает трактоваться как непрерывная, а приближенный график ее «истинного» распределения будет представлять собой гистограмму в объединенных классовых интервалах. Именно для такого случая задача оптимального разбиения гистограммы имеет практический смысл.

Заметим теперь, что если вероятности оценены с точностью ε , то, как легко показать, с той же точностью оценивается и среднее значение (см. формулу (4.5.6)). Поэтому знание фактических величин с большей точностью будет излишним (не вообще, а в смысле превышения точности). С другой стороны, если выбрано разбиение на n классовых интервалов, то среднее значение по выборке отличается от среднего значения по гистограмме на величину $1/n$, если считать, что внутри классового интервала распределение равномерно, что и показывает гистограмма.

Исходя из этих качественных соображений, можно сделать вывод, что на практике не имеет смысла выбирать мелкость разбиения гистограммы, превышающую точность оценки вероятностей по этой самой гистограмме. Так как число интервалов желательно выбрать наибольшим из возможных, то оптимальным равномерным разбиением гистограммы будем тогда считать ближайшее натуральное число n (классовых интервалов) к решению уравнения, согласующего уровень значимости (2.11) и мелкость разбиения:

$$\frac{1}{n} = 2\psi\left(\frac{2\sqrt{N}}{\Sigma_N(n)}\right). \quad (3.7)$$

Решение уравнения (3.7) может не существовать в натуральных числах, но оно всегда существует и единственно относительно точности ε , так как с уменьшением ε число разбиений $n = 1 + [1/\varepsilon]$ не убывает, как

и аргумент функции ψ в правой части (3.7), причем сама функция ψ при этом монотонно возрастает, а области значений обеих этих непрерывных функций ε и $\psi(\varepsilon)$ совпадают. Оптимальное число разбиений n_{opt} предполагается между n_0 (некоторым произвольно выбранным числом разбиений) и $1 + [1/\varepsilon(n_0)]$, или наоборот, в зависимости от того, какое из этих чисел меньше. Изменение числа интервалов в любую сторону приводит к противоположному изменению точности ε , так что с каждой следующей итерацией число n_{opt} заключено во все меньшем целочисленном промежутке. В результате за конечное число шагов находится разбиение, согласованное с точностью аппроксимации эмпирических вероятностей.

Применим эту методику к меняющейся во времени ВПФР. Тогда величина (2.8) является еще и функцией времени. Естественно, нет смысла на каждом временном шаге менять разбиение гистограммы, так как тогда будет несколько затруднительно сравнивать между собой распределения, меняющиеся во времени. Поэтому следует использовать не локально-оптимальное разбиение гистограммы, а ввести некоторое характерное разбиение в соответствии с характерным же уровнем точности эмпирических вероятностей. Последний строится следующим образом.

Пусть в результате произвольно выбранного разбиения гистограммы на n классовых интервалов получен временной ряд точностей оценки ВПФР в соответствии с (2.11). Построим плотность $\nu(\varepsilon, n)$ распределения этих точностей по всему доступному для анализа массиву данных и определим согласованный уровень точности $\varepsilon^*(n)$, определяемый как квантиль распределения $\nu(\varepsilon, n)$, равный собственному уровню значимости, т.е.

$$\int_0^{\varepsilon^*} \nu(\varepsilon, n) d\varepsilon = 1 - \varepsilon^*. \quad (3.8)$$

Заметим, что в силу интегрального характера условия (3.8) искомый уровень точности $\varepsilon^*(n)$ не зависит от способа разбиения гистограммы для построения распределения $\nu(\varepsilon, n)$, ибо фактически используется упорядоченный по возрастанию ряд значений $\varepsilon(n)$. Определим отве-

чающее этому согласованному уровню точности число промежутков разбиения $n^* = \lceil 1/\varepsilon^*(n) \rceil$. Если оказалось, что $n^* > n$, число интервалов увеличивается и процедура повторяется, строится новая плотность $\nu(\varepsilon, n)$ и т.д. Если же $n^* < n$, число интервалов уменьшается. Таким образом, процесс нахождения оптимального разбиения будет иметь итерационный характер. Наилучшее целочисленное приближение к решению уравнения (3.7) и даст оптимальную мелкость равномерного разбиения гистограммы. Эта мелкость в конечном итоге зависит от длины выборки и от суммы (2.8), интегрально характеризующей вид эмпирического распределения.

Применим описанный алгоритм сначала для аппроксимации стационарных распределений, когда результаты численного моделирования могут быть сравнены с теоретическими, поскольку сумма (2.8) вычисляется тогда в явном виде независимо от реализации временного ряда. В стационарном случае наибольшее значение суммы (2.8), как и любого функционала от плотности распределения при условии сохранения нормировки, достигается на равномерном распределении, и тогда $\Sigma_N(n) = \sqrt{n-1}$. Еще раз подчеркнем, что оптимальность понимается не в смысле минимального отклонения от генеральной плотности (для этой цели достаточно разбить весь промежуток на один-единственный интервал), а в смысле согласования точности попадания числа в тот или иной классовой интервал в зависимости от ширины шага гистограммы, с точностью оценки вероятности такого попадания.

Используя приближение (2.12), можно для равномерного распределения получить оптимальную точность ε его аппроксимации по гистограмме в зависимости от длины N выборки как обращение функции $N(\varepsilon)$ из равенства

$$N = -\frac{\pi \ln(1 - (1 - \varepsilon)^2)}{\varepsilon^3}. \quad (3.9)$$

Интересно сравнить получаемые результаты с другим предельным случаем — узким распределением типа «колокола», для чего возьмем нормальное распределение с центром в точке 0,5 и стандартным отклонением, равным $\sigma = 0,1$. Тогда на отрезке $[0;1]$ содержится почти все распределение, так как интеграл от плотности по этому отрезку равен $\text{erf}(1/(2\sigma\sqrt{2})) \approx \text{erf}(3,54) \approx 0,9995$.

Для полноты картины следует рассмотреть и промежуточный случай, когда плотность неравномерно «размазана» по всему отрезку $[0;1]$. Для этой цели возьмем стационарный ряд одномерной хаотической автономной динамической системы, задаваемой логистическим отображением

$$x(t+1) = 4x(t) \cdot (1 - x(t)). \quad (3.10)$$

Как было показано выше в главе III, динамически-инвариантная мера этого отображения дается формулой

$$d\mu = \frac{dx}{\pi \sqrt{x(1-x)}}. \quad (3.11)$$

Оптимальные равномерные разбиения отрезка $[0;1]$ в зависимости от длины выборки, отвечающие равномерному и нормальному распределениям, а также для плотности динамически-инвариантной меры (3.11) даны на рис. 5.3.1.

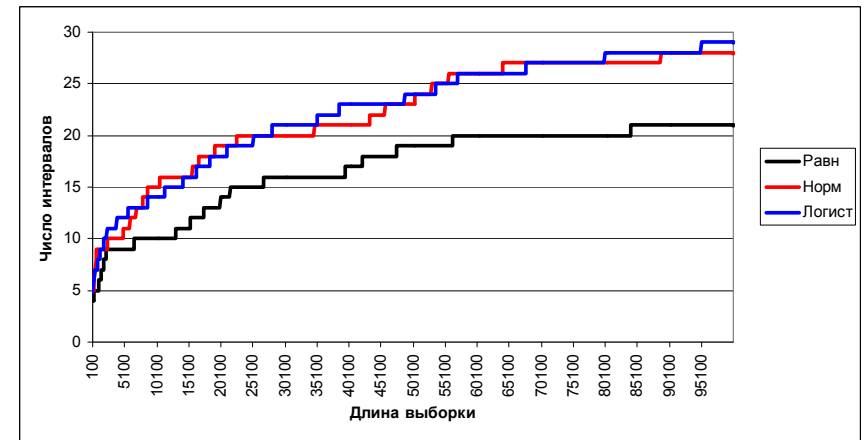


Рис. 5.3.1 — Оптимальное число интервалов в зависимости от длины выборки для некоторых стационарных распределений

Как и следовало ожидать, равномерное распределение требует меньше промежутков разбиения, чем нормальное, а логистическое отображение — чуть больше, чем нормальное. Качественно это связано с тем, что в тех областях, где генеральное распределение имеет исчезающе малую плотность, число интервалов и их ширина не имеют значения. Для распределений с большой дисперсией (а не эксцессом, в отличие от (3.5))

число интервалов может быть малым. Впрочем, на выборках относительно небольших длин (до 1000) точности приближения всех распределений близки.

На выборках средних длин порядка 5–10 тыс. различия в разбиениях уже заметны. Может показаться, что выборка в 10 тыс. данных достаточно представительна, в [24], например, выборка в 1000 данных называется «очень длинной». На самом деле это не так, поскольку наилучшая достижимая точность оценки выборочных вероятностей даже для выборки в 10 тыс. данных из нормального распределения составляет всего лишь 0,05. Для точности же 0,01 требуется более миллиона данных. Если раньше такое количество данных в практическом эксперименте по измерению, например, числа отказов электротехнического прибора получить было невозможно даже при использовании для этой цели всех выпущенных приборов, то в настоящее время существует область деятельности, представляющая данные десятками, а то и сотнями миллионов: это биржевые временные ряды, когда в сутки по отдельному финансовому инструменту проходит порядка 100 тыс. сделок. В результате статистический анализ можно было бы провести с очень высокой точностью, если бы не нестационарность изучаемых процессов.

Сравним точность приближения вышеприведенных непрерывных распределений по выборке в 10 тыс. точек гистограммой, разбитой по методике (16–19) на соответствующее количество промежутков, с вариантами, следующими из формул (1–5). Числа разбиений приведены в табл. 5.3.1.

Табл. 5.3.1. Число промежутков разбиения по разным методикам

Формула методики	Равн.	Норм.	Логист.
(3.1)	13	13	13
(3.2)	150	150	150
(3.3)	–	63	–
(3.4)	–	12	–
(3.5)	19	27	33
(3.7)	10	15	14

Для сравнения точности аппроксимаций были сгенерированы 2000 выборок из равномерного, нормального и логистического распределений на отрезке $[0;1]$, и для каждого из них вычислена интегральная ошибка отклонения

$$\varepsilon_k = \int_0^1 |f_{appr,k}(x) - f(x)| dx, \quad k = 1, \dots, 2000.$$

После этого для каждого метода была построена эмпирическая интегральная функция распределения ошибок по этим 2000 экспериментов.

Оказалось, что для аппроксимации нормального распределения с узким максимумом, как в данном случае, все методы дают близкие значения ошибок, среднюю ошибку (0,17) и стандартное отклонение ошибки (0,03). Различия проявляются в третьем знаке после запятой. Определенная позитивная особенность предложенного метода состоит в том, что вероятность малых ошибок в нем наибольшая: до значения $\varepsilon = 0,15$ интегральное распределение ошибки идет выше распределений, отвечающих другим методам. Впрочем, преимущество оказалось небольшим и проявилось во втором знаке после запятой.

Для равномерного распределения ситуация, очевидно, вырожденная: чем меньше число интервалов, тем точнее приближается плотность вероятности. Укажем, что, как и должно быть по определению согласованной точности аппроксимации, вероятность ошибки, меньшей 0,05, по оптимальному разбиению (3.7) равна 0,95, а для разбиения по формуле (3.5) она равна 0,91, т.е. точность разбиения (3.7) выше.

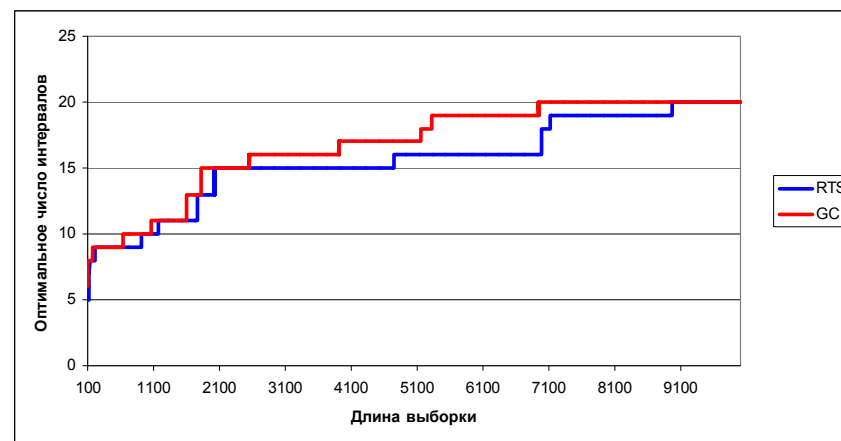


Рис. 5.3.2 — Оптимальное число интервалов в зависимости от длины выборки

Такое же наблюдение относится и к плотности меры (3.11). Все разбиения по иным методикам дают худшую аппроксимацию.

Рассмотрим теперь построение оптимального разбиения по описанной методике в нестационарном случае на примере биржевых рядов. Были взяты ряды индекса RTS по часовым ценам закрытия и тиковые данные фьючерсов: GC на золото и CL на нефть. Число интервалов в зависимости от длины выборки приведено на рис. 5.3.2. На рис. 5.3.3 оптимальное число интервалов вычислено для выборки достаточно большой длины в 100 тыс. данных (фьючерсы CL на нефть).

Различие в рассмотренных временных рядах проявляется, в частности, в том, что существует диапазон длин выборок в 3–7 тыс., где отличия в мелкости разбиения принципиальны. Разбиение на число интервалов, большее 20, для нестационарных рядов, как и для стационарных, оказывается актуальным при длинах выборки, больших 30 тыс. Последнее важно для тиковых рядов, характерное число данных в которых за достаточно короткий промежуток времени в одни сутки достигает 100 тыс. (рис. 5.3.3).

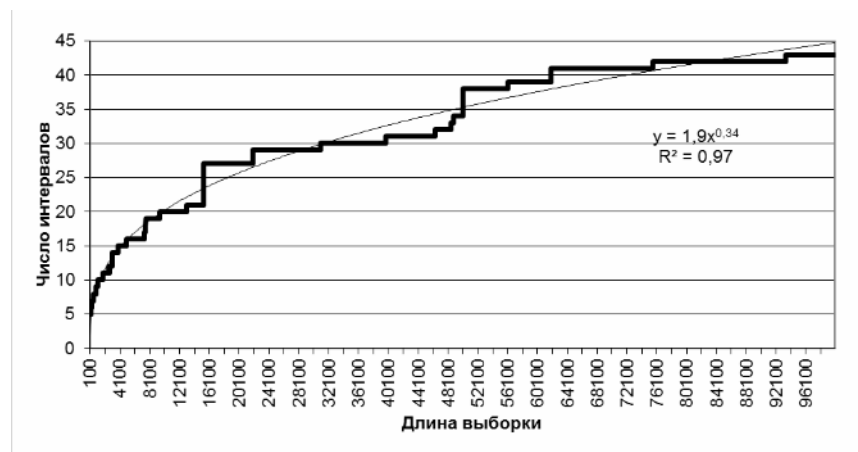


Рис. 5.3.3 — Оптимальное число интервалов в зависимости от длины выборки для тикового ряда фьючерсов CL и его степенная аппроксимация

Заметим, что в рассмотренных примерах полученные величины оптимального числа интервалов для небольших выборок 100–500 данных довольно близки ко всем вышеприведенным оценкам (3.1–3.5), включая рекомендации ВНИИМ [24]. Но, в отличие от последних, вместо просто ре-

комендованного диапазона мы имеем конкретные значения в зависимости от количества данных для каждого из временных рядов. Впрочем, значения суммы (2.8) для многих реальных временных рядов при одной и той же мелкости разбиения варьируются не особенно значительно, и потому оптимальное разбиение при небольшом числе данных (до нескольких тысяч) оказывается примерно одинаковым независимо от ряда. В этом смысле рекомендации [24] имеют практическое значение даже в большей степени, чем теоретические оценки (3.1–3.5). Однако все эти оценки вместе взятые заметно отклоняются от фактического оптимального разбиения на большом диапазоне изменения длины выборки, хотя наиболее близким к нему оказывается результат Смирнова, когда число интервалов растет как корень третьей степени из длины выборки. Соответствующий пример приведен на рис. 5.3.3.

Таким образом, численная итерационная процедура оптимального разбиения гистограммы с довольно высокой точностью привела к асимптотическому результату Смирнова.

5.4. Распределение расстояний между выборочными плотностями

Чтобы иметь возможность корректно анализировать нестационарный временной ряд, необходимо, во-первых, определить промежуток времени, на котором выборочная плотность распределения не успеет значительно измениться, и, во-вторых, на основе имеющихся данных оценить, как в среднем менялась ВПФР, построенная по выборке такого же объема, на разных промежутках времени, с тем, чтобы определить для нее эмпирический оператор эволюции.

Для того, чтобы определить, насколько изменяется ВПФР с течением времени, следует ввести норму в соответствующем пространстве функций. Поскольку мы анализируем плотности распределения вероятностей, являющиеся в каждый момент времени t суммируемыми по x функциями, то удобной для проведения оценок их близости является норма в пространстве L_1 :

$$\|f(x, t)\| = \int |f(x, t)| dx. \quad (4.1)$$

Если плотность оценивается по гистограмме, то под интегралом понимаем сумму по промежуткам разбиения.

Расстоянием между двумя ВПФР $f_1 = f_{T_1}(x, t_1)$ и $f_2 = f_{T_2}(x, t_2)$, построенными по выборкам длин T_1 и T_2 , называется величина

$$\rho(f_1, f_2)_{L_1} = \|f_{T_1}(x, t_1) - f_{T_2}(x, t_2)\| = \int |f_{T_1}(x, t_1) - f_{T_2}(x, t_2)| dx. \quad (4.2)$$

В пространстве непрерывных функций, в котором изучаются не плотности, а функции распределения, расстояние вводится как супремум модуля разности между функциями (применительно к ВФР будем называть его расстоянием по Смирнову):

$$\rho(F_1, F_2)_C = \sup_x |F_1 - F_2|. \quad (4.3)$$

Можно ввести и другие расстояния между функциями — например, в пространстве L_2 , но применительно к вероятностным распределениям оно не особенно интересно, поскольку не содержит вероятностной интерпретации. Среди «квазирасстояний» отметим расстояние Кульбака–Лэйблера между ВПФР, определяемое в соответствии с (1.3.22) формулой

$$\rho(f, g)_{KL} = \int f(x) \ln[f(x)/g(x)] dx. \quad (4.4)$$

В этом расстоянии функция g играет роль генеральной совокупности, а функция f — локального распределения. Если генеральная совокупность не известна, можно попытаться применить (4.4) для сравнения текущих ВПФР, однако в силу того, что расстояние (4.4) может быть любым положительным числом, сложно отфильтровать ошибку второго рода, когда случайный выброс в одной из крайних ячеек гистограммы приводит к очень большому отклонению между распределениями.

Распределение расстояний между плотностями или функциями распределения в различных нормах характеризует процесс с точки зрения типовых отклонений между выборочными распределениями. Поэтому принятие решения о том, что текущая выборка из временного ряда имеет значимое отличие в распределении от предыдущей выборки, должно основываться на изучении свойств распределения расстояний между выборками.

Рассмотрим сначала распределение расстояний между выборками из стационарных распределений. Для этого сгенерируем временной ряд с заданной функцией распределения и разобьем его на промежутки длины T , после чего определим расстояние между двумя выборками длины T ,

которые расположены встык. Такие выборки по построению состоят из независимых данных.

На рис. 5.4.1 показан ряд, образованный расстояниями между последовательными парами встык-выборок длиной 40 тыс. данных, взятых из равномерного распределения.

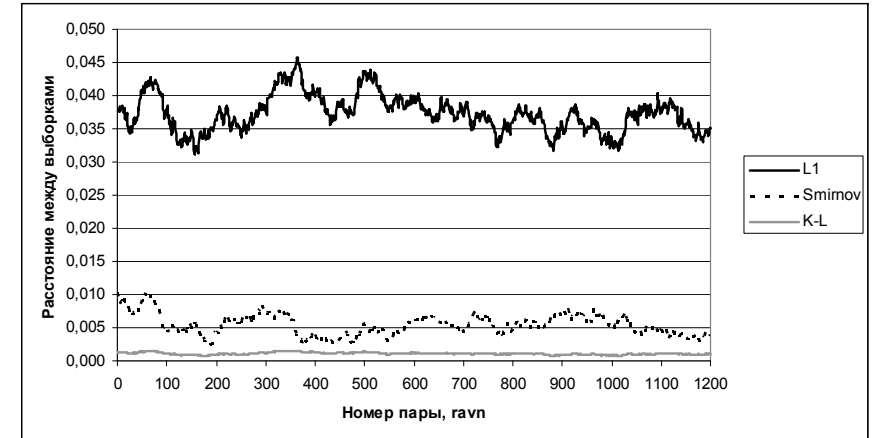


Рис. 5.4.1 — Ряды расстояний в разных нормах для выборок из равномерного распределения

На рис. 5.4.2 показан ряд, образованный расстояниями между последовательными парами встык-выборок длиной 40 тыс. данных, взятых из временного ряда логистической динамической системы $x_{n+1} = 4x_n(1 - x_n)$.

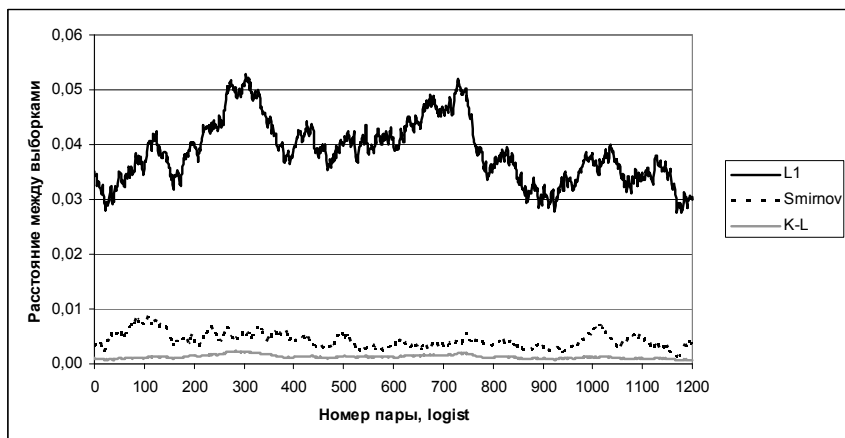


Рис. 5.4.2 — Ряды расстояний в различных нормах для выборок из логистического ряда

Для выборок, взятых из одного и того же стационарного распределения характерным является то, что расстояние в норме $L1$ больше расстояния в норме C , а то, в свою очередь, больше квазирасстояния относительной энтропии. Подчеркнем, что при вычислении расстояний между ВПФР гистограммы, представляющие оценку плотности, строились по разбиению, которое согласованным образом зависит от длины выборки, как это было определено выше в параграфе 5.3.

Как уже говорилось в параграфе 5.1, для выборки определенного объема T следует задать согласованный уровень значимости, на котором ее можно идентифицировать как выборку из стационарного распределения. Этот уровень определяется как такой квантиль распределения расстояний между выборками, который равен дополнению до уровня значимости. Если построено распределение $g(\rho)$ расстояний между выборками, причем ρ_{\max} есть априорное максимальное расстояние между ними, то согласованный уровень значимости ε^* по аналогии с (1.9) определим формулой

$$\int_0^{\varepsilon^*} g(\rho) d\rho = 1 - \varepsilon^* / \rho_{\max}. \quad (4.5)$$

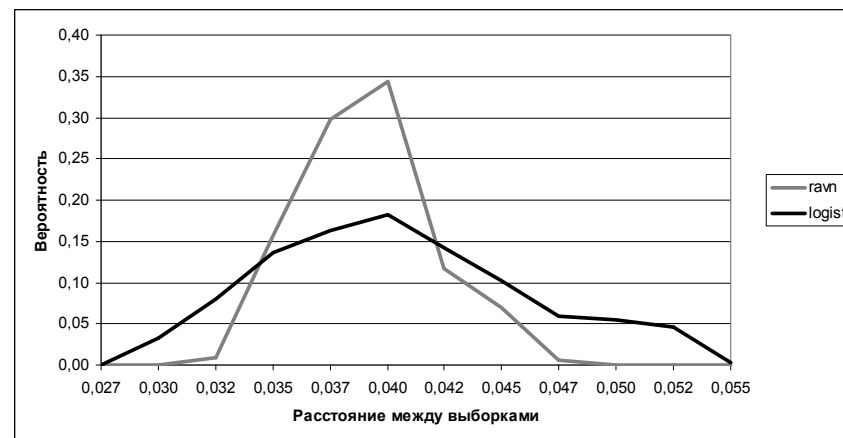


Рис. 5.4.3 — Распределение расстояний между ВПФР в норме $L1$.

Зависит ли распределение расстояний между выборочными распределениями от вида распределения генеральной совокупности? Построим, например, распределения для рядов расстояний, приведенных выше на рис. 5.4.1–5.4.2 по 1200 статистическим экспериментам. Эти распределения даны на рис. 5.4.3 — 5.4.5.

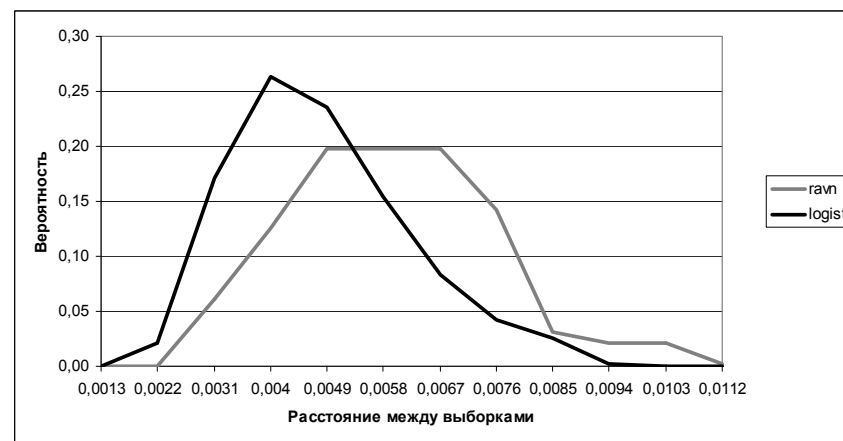


Рис. 5.4.4 — Распределение расстояний между ВПФР в норме C .

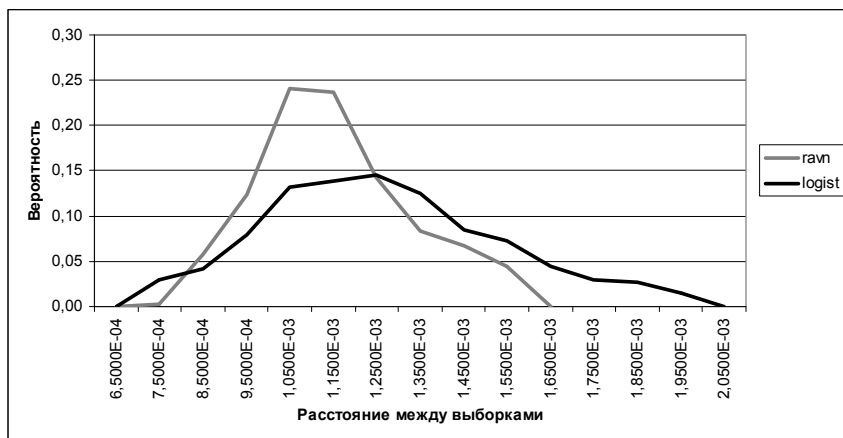


Рис. 5.4.5 — Распределение расстояний Кульбака–Лэйблера между ВПФР.

Из рис. 5.4.3–5.4.5 следует, что в разных нормах статистические свойства одного и того же случайного процесса могут заметно различаться. Так, расстояния в $L1$ и в квазинорме (4.4) между выборками из равномерного распределения распределены более компактно, чем для логистического ряда, а в норме C — наоборот. Последнее особенно интересно тем, что по критерию Колмогорова–Смирнова распределение расстояний для стационарных распределений не должно зависеть от вида распределения, если временной ряд состоит из статистически независимых величин. Поскольку для логистического ряда, как и любой рекуррентно задаваемой процедуры, последнее условие не выполняется, то распределение расстояний отличается от производной K -функции. Это отличие может служить индикатором нестационарности или статистической зависимости. Согласованный уровень значимости для этого критерия примерно равен 0,01, что близко к точному решению уравнения (1.9). Отметим, что если ВФР строить не непосредственно по выборке ряда, а на основе предварительно вычисленной плотности в виде гистограммы с оптимальным разбиением, то наблюдаемый на практике согласованный уровень будет примерно в два раза выше, чем это следует из решения уравнения (1.9). Это связано с тем, что критерий относится к непрерывным ВФР, а для гистограммы ограничением по точности является ширина промежутка разбиения. Поэтому, если анализируется именно ВПФР, то критерий Колмогорова–Смирнова к построенной на ее основе ВФР применять не

следует, так как он будет давать ошибку, превосходящую согласованный уровень значимости.

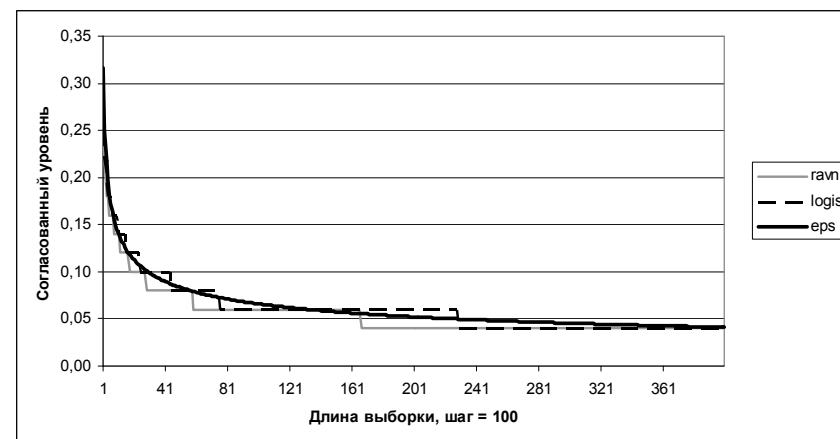


Рис. 5.4.6 — Согласованный уровень значимости для стационарных ВПФР

В отличие от распределения расстояний в норме C , интегральный индикатор стационарности в виде согласованного уровня значимости, введенный в п. 5.3, для всех стационарных распределений в норме $L1$ дает одинаковый результат. На рис. 5.4.6 приведены графики численного решения уравнения (4.5) в зависимости от длины выборки, где для сравнения показана также теоретическая кривая согласованной точности $2\varepsilon(T)$ для равномерного распределения, не выше которой по построению должен быть найденный численно согласованный уровень значимости. Это и наблюдается на графиках рис. 5.4.6.

Поскольку разбиения гистограмм для равномерного и логистического распределений различны, то теоретические согласованные уровни точности для этих процессов разные, хотя и достаточно близки. Поэтому сравнивать между собой две ВПФР для различных временных рядов на одной гистограмме можно только в том случае, если они имеют близкие разбиения. В этом случае можно взять единое разбиение, которое хотя и не будет оптимальным, тем не менее, сохранит приемлемую точность оценки выборочных вероятностей.

Норма, в которой сравнение между собой выборочных распределений дает наилучшую точность различения, априори не известна, но для каж-

дого конкретного случая она может быть определена из условия наиболее компактного распределения расстояний. Соответствующая характеристика — отношение стандартного отклонения к среднему значению. Чем оно меньше, тем более компактно распределение. Напомним, что значения случайной величины принадлежат отрезку $[0; 1]$, так что среднее значение отлично от нуля, за исключением тривиального случая ряда постоянных чисел, равных нулю. Оказалось, что для приведенных примеров равномерных и логистических выборок наименьшее отношение стандартного отклонения к среднему значению наблюдается для распределений расстояний между выборками в норме $L1$ (см. табл. 5.4.1). Тем самым эта норма выглядит более предпочтительной с практической точки зрения.

Табл. 5.4.1. Отношение $\sigma_\rho / \bar{\rho}$ для распределений расстояний между выборками

Норма	$L1$	C	$K-L$
Равномерное распределение	0,07	0,29	0,16
Логистический ряд	0,15	0,33	0,27

Как уже говорилось, квазирасстояние в виде относительной энтропии (4.4) для сравнения выборочных распределений использовать затруднительно, поскольку не известно, какой характерный уровень принять за верхнюю границу для ρ_{KL} . Заметим, что для достаточно длинных выборок из одного и того же распределения эта величина весьма мала. Тем не менее, хотя для выборок длины 40 тыс. отличие имеет порядок 0,001 (см. рис. 5.4.5), но выборки из разных совокупностей различаются вполне успешно: среднее расстояние между равномерной и логистической выборками тех же длин равно 0,06, т.е. в 60 раз больше. Для сравнения укажем, что расстояние в норме $L1$ между теми же выборками равно в среднем 0,4, что всего лишь в 20 раз больше, чем согласованный уровень значимости. Однако наличие самого этого уровня сильно увеличивает точность распознавания близких распределений за счет уменьшения ошибки второго рода. Далее по умолчанию расстояние между выборками понимается в смысле расстояния между соответствующими ВПФР в норме $L1$.

Построим теперь аналогичное распределение расстояний для выборок из нестационарного временного ряда. На рис. 5.4.7 показан пример ряда из значений функционала расстояния, вычисленного для двух встык-

выборки объемом 1000 данных с шагом в 1000 тиков для разностей дискретных тиковых цен на фьючерсы CL.

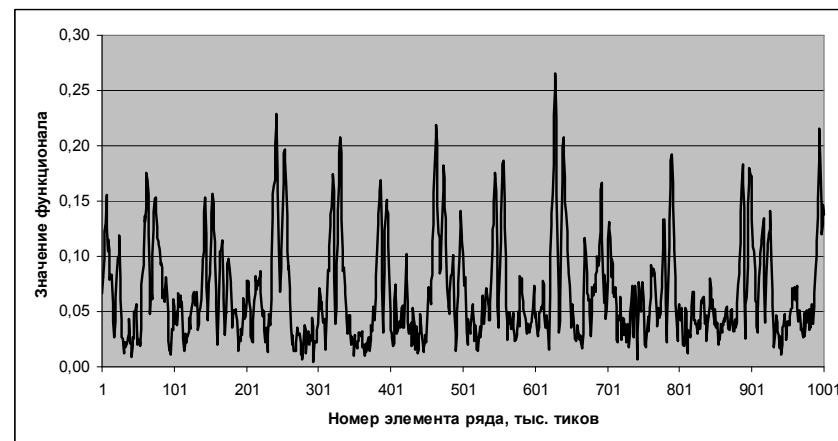


Рис. 5.4.7 — Ряд значений расстояний для двух встык-выборок в 1000 тиков

При анализе нестационарного ряда актуален вопрос о том, где провести уровень, ниже которого отклонения между распределениями с принятой точностью объясняются стационарным случайным процессом, а где провести уровень, выше которого распределения следует считать различными.

Для ответа на первый вопрос используется тот факт, что шаг оптимальной гистограммы совпадает с точностью оценки стационарных вероятностей. В частности, разбиение гистограммы для данного ряда приведено на рис. 5.3.3. В силу построения оптимального разбиения гистограммы стационарно объясняемые отклонения между выборками лежат в пределах от нуля до $\varepsilon_0 = 2/n(T)$, где $n(T)$ есть число интервалов равномерного разбиения, отвечающего выборке длины T .

В качестве верхнего уровня отсечения расстояний рекомендуется аналог решения уравнения (4.5) в нестационарном случае. Поскольку функционал расстояния между ВПФР представляет собой меру нестационарности ВПФР, то характерное значение $\rho_s(T)$, превышение которого трактуется как изменение распределения (или разладки системы), есть решение уравнения

$$\int_0^{\rho_s(T)} g_T(\rho) d\rho = 1 - \frac{\rho_s(T)}{2}. \quad (4.6)$$

Это значение зависит от распределения $g_T(\rho)$ расстояний между формально независимыми выборками длины T . Под формальной независимостью подразумевается то, что для ряда стационарных независимых величин выборки статистически независимы. Такому условию удовлетворяют, например, встык-выборки временного ряда. Численное решение уравнения (4.6) по имеющимся эмпирическим данным будем называть согласованным уровнем стационарности (далее СУС) временного ряда. В стационарном случае СУС совпадает с согласованным уровнем значимости. На рис. 5.4.8 показан пример распределения расстояний между тиковыми рядами индекса S&P длиной в 5000 тиков, построенного по встык-выборкам за трехлетний период с 2005 по 2007 гг.

Для этого инструмента СУС на 5000 данных равен примерно $\rho_s = 0,10$. Если бы данный ряд был стационарным, то этот показатель был бы равен 0,06. Это ряд сравнительно мало меняющийся.

Более нестационарный ряд тиков CL имеет распределение расстояний между встык-выборками в 10 тыс. тиков, показанное на рис. 5.4.9. Для него СУС равен примерно 0,14 вместо ожидаемого значения порядка 0,04 в стационарном случае.



Рис. 5.4.8 — Распределение расстояний между встык-выборками в 5000 тиков индекса S&P

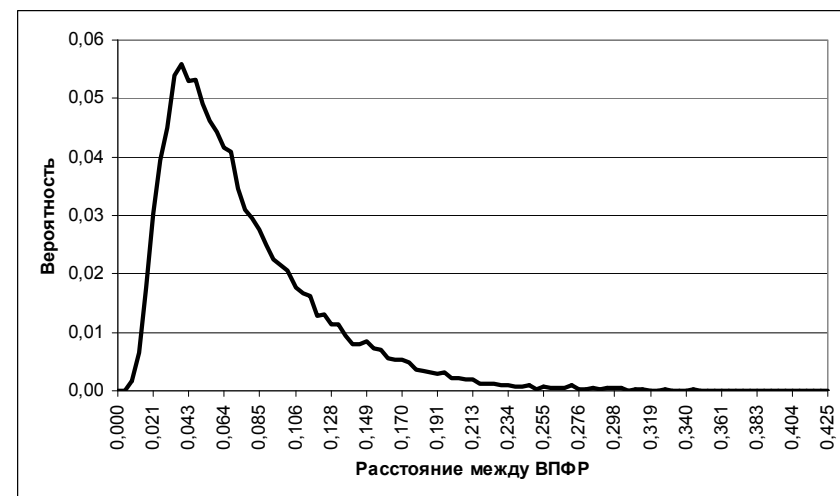


Рис. 5.4.9 — Распределение расстояний между встык-выборками в 10 тыс. тиков ряда CL

Приведенные примеры показывают, что для разных инструментов распределение $g_T(\rho)$ при одном и том же значении длины выборки T может заметно различаться. Чем более стационарен ряд, тем левее расположен максимум распределения функционала расстояния в норме $L1$.

Важную информацию относительно нестационарных свойств временного ряда несет зависимость СУС от длины выборки. Для многих нестационарных рядов характерно наличие статистически значимых локальных минимумов СУС в зависимости от длины встык-выборок, тогда как для стационарного ряда наблюдается монотонно убывающая зависимость, как на рис. 5.4.6. Аргументы этих локальных минимумов можно трактовать как типовые для данного ряда промежутки времени, на которых происходит смена режима работы наблюдаемой системы. Аналогично, аргументы локальных максимумов отвечают наибольшей разладке между распределениями. Пример зависимости СУС $\rho_s(T)$ от длины выборки для тикового ряда CL и для часовых цен закрытия индекса RTS приведен на рис. 5.4.10. Так, для ряда CL характерными размерами выборок являются 500, 1100, 2600 данных, а для ряда «часовиков» RTS — 200 и 1100. Для того, чтобы удостовериться, что эти характерные длины выборки дают статистически значимый эффект, следует сравнить максимумы СУС с теоретической кривой $2\varepsilon(T)$ точности оценки вероятностей в стационарном

случае. Если колебания уровня СУС идут ниже теоретической кривой, наблюдаемый эффект не является значимым.

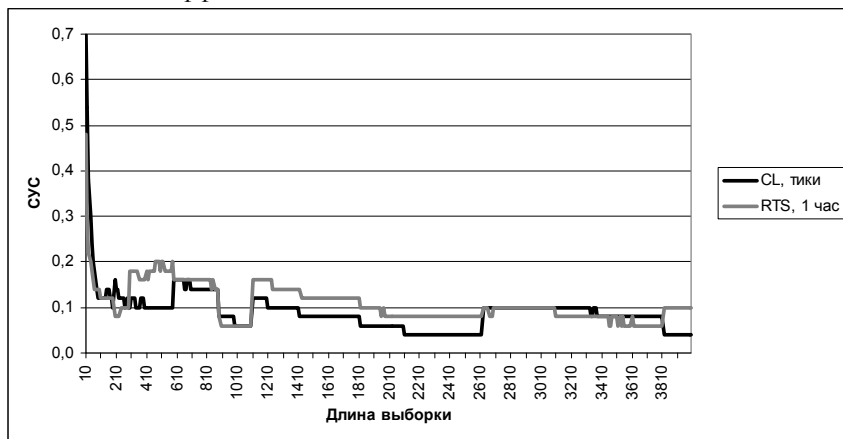


Рис. 5.4.10. Зависимость СУС от длины выборки

На рис. 5.4.11 приведены графики СУС в широком диапазоне зависимости от длины выборки в сравнении со стационарным уровнем оценки точности. Инструменты — тиковые ряды фьючерсов на нефть CL и на золото GC. Они показаны на одном графике с единым уровнем точности, так как разбиение гистограммы для них примерно совпадает. Но, строго говоря, оптимальное разбиение гистограммы для каждого инструмента — свое.

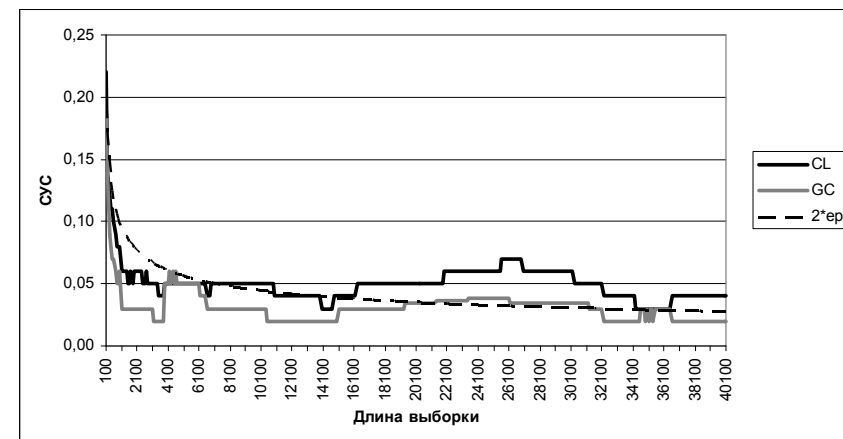


Рис. 5.4.11. Зависимость СУС в сравнении со стационарным уровнем

Из рис. 5.4.11 видно, что на малых длинах выборки минимумы СУС не являются значимыми как для CL, так и для GC, поскольку соответствующие графики проходят ниже кривой стационарной точности. Иными словами, минимум в области 4 тыс. тиков для СУС ряда GC интересен, но последующий рост расстояний между ВПФР не может, строго говоря, считаться значимым. Лишь на длинах, превышающих 15 тыс. тиков для ряда CL и 20 тыс. тиков для ряда GC имеет смысл искать доказательств разладки, причем ряд GC сам по себе весьма близок к стационарному. На длинах свыше 50 тыс. тиков СУС обоих рядов стабилизируется на относительно низком уровне, слабо превышающем уровень стационарной точности. Таким образом, как малые, так и большие длины выборок не дают возможности распознать значимые изменения в текущем распределении вероятностей нестационарного ряда. В первом случае недостаточно точности в оценке ВПФР, а во втором случае ряды имеют квазистационарное поведение без резких изменений СУС. Интерес представляют выборки средних длин, каковыми в данном случае следует считать длины порядка 15–25 тыс. данных.

Наряду с тиковыми рядами практическую важность имеют ценовые ряды, характеризующие процесс в определенном временном промежутке: это цена открытия промежутка, максимальная и минимальная цены, цена закрытия. Эти ряды имеют иные статистические свойства по сравнению с тиковыми данными. При анализе, например, приростов цен закрытия по

соседним временным промежуткам на тот или иной финансовый инструмент следует изучить поведение СУС не только как функцию длины выборки, но и в зависимости от тайм-фрейма (минутки, 5-минутки и т.д.). Опорным рядом этих цен является ряд с 1-минутным тайм-фреймом, из которого получаются остальные ряды с большим промежутком агрегирования. Статистические свойства рядов укрупненных данных отличаются от «минуток». Обычно отличия проявляются в виде скэйлинга характерных параметров. Приведем некоторые примеры скэйлингового анализа статистики СУС для биржевых рядов.

Во-первых, для каждого тайм-фрейма имеет место свое оптимальное разбиение гистограммы, в связи с чем одинаковая теоретическая точность в оценке вероятностей распределений приростов цен достигается на разных длинах выборок. С увеличением тайм-фрейма τ , в котором считаются приросты цен закрытия, растет длина выборки, на которой достигается одинаковый уровень точности в оценке вероятностей. Иными словами, нестационарность ряда с увеличением тайм-фрейма растет. Обозначим T_1 длину выборки на единичном тайм-фрейме, отвечающем ряду минуток ($\tau_1 = 1$ мин). Тогда наблюдается следующий скэйлинг длины выборки для инвариантной точности, т.е. той, которая равна точности оценки вероятностей для ряда минуток длины T_1 :

$$\sqrt{\frac{T(\tau)}{T_1}} = 1 + a \cdot \ln\left(\frac{\tau}{\tau_1}\right), \quad (4.7)$$

где a есть параметр, зависящий от типа инструмента. Типичные его значения лежат в промежутке от 0,2 до 0,5. Достоверность скэйлинговой аппроксимации (4.7) достаточно высокая — она равна $0,97 \div 0,99$.

Во-вторых, практический интерес представляет первый минимум СУС, который трактуется как достижение рядом точки своей локальной стационарности. С увеличением тайм-фрейма τ реальное время $t_{\min}(\tau)$ достижения первого минимума (в минутах, а не в количестве элементов выборки) растет экспоненциально по эмпирическому закону

$$t_{\min}(\tau) = t_{\min}(\tau_1) \exp\left(\frac{\tau}{2\tau_1}\right), \quad (4.8)$$

где $\tau_1 = 1$ мин, а $t_{\min}(\tau_1) \approx 240 \div 300$ мин. Сама же величина этого первого минимума СУС растет по степенному закону вида

$$s_{\min}(\tau) = s_{\min}(\tau_1) \left(\frac{\tau}{\tau_1}\right)^{4/5}. \quad (4.9)$$

Достоверность скэйлингов (4.8–4.9) составляет 0,95.

В-третьих, важной является информация о скорости установления распределения с заданным уровнем стационарности. Числовой характеристикой этой величины является длина выборки, на которой СУС впервые достигает заданной величины. Скэйлинг длины выборки в реальном времени τ в минутах для первого достижения определенного уровня СУС s в зависимости от тайм-фрейма имеет зависимость с достоверностью 0,9 вида

$$T(\tau, s) = A(s) \cdot \tau^{\alpha(s)},$$

где модельный параметр амплитуды в диапазоне практически интересных значений СУС $0,2 \leq s \leq 0,3$ имеет степенной вид

$$A(s) = \frac{b}{s^{5/2}}, \quad b \approx 1 \div 2, \quad \text{а показатель степени } \alpha(s) \text{ уменьшается с}$$

увеличением уровня СУС по зависимости $\alpha(s) = \frac{c}{s^2}, \quad c \approx 0,1 \div 0,3$.

В результате получаем, что скэйлинговая зависимость длины выборки для достижения уровня s СУС приближенно имеет вид (T, τ выражены в минутах):

$$T(\tau, s) = \frac{b \cdot \tau^{c/s^2}}{s^{5/2}}. \quad (4.10)$$

5.5. Индекс нестационарности выборочных распределений

Введенный в параграфе 5.4 согласованный уровень стационарности позволяет определить характерную величину неопределенности в выборочных распределениях временного ряда. Подчеркнем, что ВПФР ряда строится для оптимального разбиения гистограммы. Поскольку же для разных временных рядов оптимальное разбиение, вообще говоря, различно, то сравнивать на одном графике два уровня нестационарности не корректно. Чтобы иметь возможность такого сравнения, следует ввести инвариантную характеристику нестационарности временного ряда.

Рассмотрим более общую статистику расстояний между ВПФР — не только между встык-выборками, как это было сделано выше для СУС, но и между выборками, сдвинутыми одна относительно другой на произвольное число τ шагов, но не большее, чем само окно выборки. Длину выборки, как и раньше, обозначаем через T , а число промежутков гистограммы через n . Введем

$$\rho(T, \tau; t) = \|f_T(x, t) - f_T(x, t + \tau)\| = \sum_{i=1}^n |f_T(i, t) - f_T(i, t + \tau)|. \quad (5.1)$$

Варьируя величину смещения τ от 1 до T , получаем, что возможные значения расстояний между выборками образуют следующий дискретный набор $\tau + 1$ чисел: $\rho \in \{0; 2/T; 4/T; \dots; 2\tau/T\}$. Обозначим через $G_{T, \tau}(k)$ вероятность того, что расстояние ρ между выборками при сдвиге на τ равно $2k/T$. Тогда величина

$$F_{T, \tau}(\rho) = \sum_{l=0}^{k(\rho)} G_{T, \tau}(l), \quad k(\rho) = [T\rho/2] \quad (5.2)$$

представляет эмпирическую вероятность того, что расстояние между распределениями не больше ρ . Наибольшее возможное расстояние между выборками равно двум, когда мера пересечения носителей соответствующих распределений равна нулю. Определим СУС $\rho^*(T, \tau)$ для таких сдвинутых выборок по тому же правилу, что и в п. 5.3: практическая вероятность его превышения равна уровню значимости, т.е.

$$F_{T, \tau}(\rho^*) = 1 - \rho^* \frac{T}{2\tau}. \quad (5.3)$$

Равенство (5.3) определяет функцию $\rho^*(T, \tau)$, которая обладает тем свойством, что при проведении достаточно большого числа экспериментов по вычислению расстояний между двумя выборочными распределениями длины T , сдвинутыми на промежуток τ , в доле $\rho^* T/(2\tau)$ случаев будет наблюдаться превышение расстояния, равного ρ^* . Тем самым ρ^* можно трактовать как характерное расстояние между распределениями на уровне значимости, не превосходящем этого расстояния.

Введем индекс нестационарности $J(T, \tau)$ ряда, положив его равным отношению доли расстояний, не превосходящих фактический СУС, построенный по имеющимся эмпирическим данным, к уровню шума, за который естественно принять долю расстояний между ВПФР, не превосходящих $2\varepsilon^* \tau/T$ при заданной длине выборки T и сдвиге τ :

$$J(T, \tau) = \frac{F_{T, \tau}(\rho^*)}{F_{T, \tau}(2\varepsilon^* \tau/T)}. \quad (5.4)$$

Напомним, что $\varepsilon^* = \varepsilon^*(T)$ есть точность оценки эмпирических вероятностей по стационарному критерию (см. 5.2).

Если $J(T, \tau) \leq 1$, ряд считается стационарным, а если $J(T, \tau) > 1$, то ряд нестационарный.

Рассмотрим пример сравнения двух рядов приростов тиковых цен на фьючерсы CL и GC за период 2008–2010 гг. с помощью индекса нестационарности для встык-выборок, т.е. при $\tau = T$. При произвольном τ картина качественно та же. Число интервалов разбиения гистограммы для этих двух рядов в зависимости от длины выборки приведено на рис. 5.5.1.

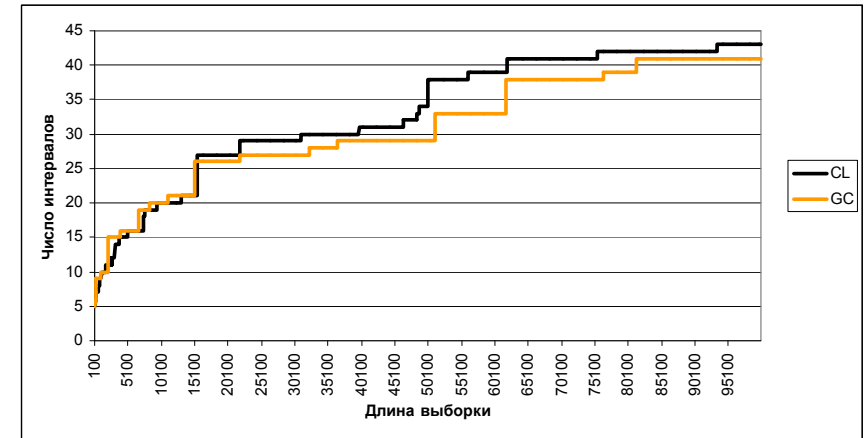


Рис. 5.5.1 — Оптимальное число интервалов разбиения в зависимости от длины выборки для тиковых рядов CL и GC

Индексы нестационарности рассматриваемых рядов приведены на рис. 5.5.2.

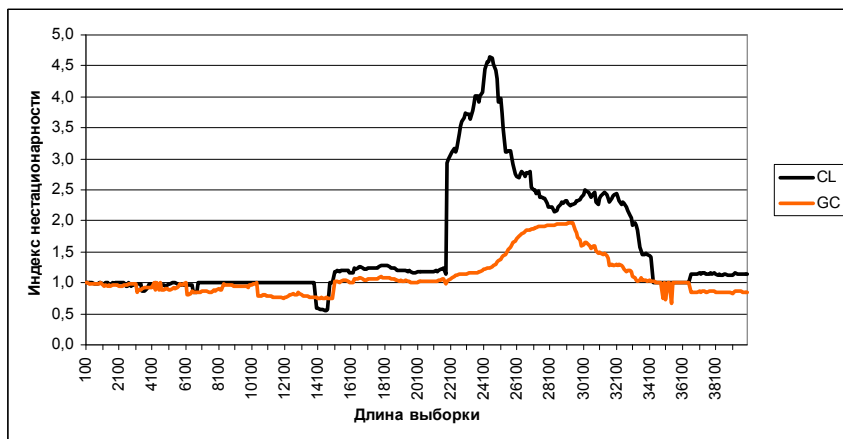


Рис. 5.5.2 — Зависимость индекса нестационарности от длины встык-выборки

Как видно, разбиения довольно похожи. Отметим, что с достоверностью 0,97 число интервалов в нашем примере растет как $T^{0,34}$, что, как уже отмечалось в п. 5.3, близко к результатам Смирнова [3], показавшего, что уклонение гистограммы от графика непрерывной плотности функции распределения обратно пропорционально корню третьей степени из длины выборки. Здесь численным методом получен приблизительно тот же результат, но он не требует наличия генеральной совокупности, что позволяет его применить и к нестационарным временным рядам.

Существенно, что стационарные и нестационарные ряды при малых выборках (до нескольких тысяч) неотличимы в силу низкой точности оценки эмпирических вероятностей, и лишь с увеличением длины выборки, когда $\rho^* > 2\varepsilon^*$, такое разделение достоверно. Так, по индикатору (5.4) нестационарный тиковый ряд CL начинает отделяться от стационарного уровня с длин выборок порядка 8 тыс., но значимо нестационарность для обоих рядов начинает проявляться после 22 тыс. данных.

Глава VI. Горизонтный ряд

6.1. Определение горизонтного ряда

Перейдем к построению основной индикативной статистики для анализа нестационарных временных рядов — горизонтного ряда. Чтобы определить оптимальный объем выборки в текущий момент времени, необходимо понять, выборки какого объема обеспечивают квазистационарность ВПФР. Пусть две ВПФР, сдвинутые одна относительно другой на τ шагов, интегрально отличаются не более чем на заданную величину ε . Соответствующий интегральный функционал обозначим через $V(T, \tau; x, t)$:

$$V(T, \tau; x, t) \equiv \rho(f_T(x, t + \tau), f_T(x, t)) = \int_0^1 |f_T(x, t + \tau) - f_T(x, t)| dx \leq \varepsilon. \quad (1.1)$$

Далее для краткости аргумент x в функционале $V(T, \tau; x, t)$ будем опускать, если это не будет искажать смысл утверждений, и писать $V(T, \tau; t)$. ВПФР $f_T(x, t)$ временного ряда $x(t)$ будем называть ε -стационарной (или квазистационарной) на временном промежутке θ , если

$$\forall \tau: 1 \leq \tau \leq \theta, \quad \forall t \quad V(T, \tau; t) \leq \varepsilon. \quad (1.2)$$

Подчеркнем, что квазистационарность не означает возможности использования асимптотических (по времени или объему выборки) оценок. Она нужна только для того, чтобы можно было зафиксировать допустимое различие между плотностями распределений. Основное утверждение, показывающее возможность равномерного ограничения по времени для оценки неточности в определении ВПФР, состоит в том, что для функционала $V(T, \tau; t)$ близости двух ВПФР имеет место оценка

$$0 \leq V(T, \tau; t) \leq \min(2\tau/T; 2). \quad (1.3)$$

Действительно, из неотрицательности произвольной ВПФР (в том числе и нестационарной) и ее нормированности на единицу в любой момент времени следует тривиальная оценка неравенства треугольника, примененного к норме (1.1):

$$0 \leq \rho(f, \varphi) = \|f(x, t_1) - \varphi(x, t_2)\| \leq \\ \leq \|f\| + \|\varphi\| = \int (|f| + |\varphi|) dx = \int (f + \varphi) dx = 2.$$

В случае, если $2\tau < T$, оценка (1.3) может быть уточнена посредством изучения распределения расстояний между ВПФР. Рассмотрим сдвиг выборки на 1 шаг по времени, т.е. рассмотрим две ВПФР, построенные в окнах $\Delta_T(t)$ и $\Delta_T(t+1)$. Для них все значения $x(k)$, на основе которых были оценены соответствующие эмпирические вероятности p_i , совпали, кроме, быть может, первого в окне $\Delta_T(t)$ и последнего в окне $\Delta_T(t+1)$. Если значения $x(t-T+1)$ и $x(t+1)$ попали в разные промежутки Δ_i разбиения R_N отрезка $[0; 1]$, то изменение ВПФР составит $2/T$. Если же эти значения попали в один и тот же промежуток, то ВПФР не изменилась. Аналогично, при сдвиге на τ шагов возможное изменение ВПФР изменяется дискретно с шагом $2/T$ от нуля до $2\tau/T$ в зависимости от количества совпадений между набором $x(t-T+1), \dots, x(t-T+\tau)$ и набором $x(t+1), \dots, x(t+\tau)$. Таким образом, в этом случае $V(T, \tau; t)$ не превосходит $2\tau/T$.

Таким образом, мы можем сделать важный вывод о том, что при фиксированном τ функционал (1.1) равномерно ограничен по t . Поэтому, если нельзя сравнить выборочную функцию распределения со стационарным распределением, то всегда можно добиться близости двух нестационарных выборочных распределений.

При увеличении точности в определении ε -стационарности, т.е. при уменьшении ε , объем h выборки, при котором достигается условие $V(T, \tau; t) \leq \varepsilon$, растет. В силу равномерной ограниченности для каждого момента времени t и для каждого значения τ существует такое минимальное значение $h(t, \tau; \varepsilon)$, что при всех $T \geq h(t, \tau; \varepsilon)$ значения функционала $V(T, \tau; t)$ не превосходят ε . Например, для любого $\varepsilon > 0$ всегда можно подобрать такой объем выборки $T > 2\tau/\varepsilon$, что ВПФР будет ε -стационарной.

▼ **Определение 1.1.** Горизонтным рядом $h(t, \tau; \varepsilon)$ для ряда $x(t)$ при сдвиге на промежуток τ называется такой минимальный объем выборки

$h(t, \tau; \varepsilon)$, что при всех $T \geq h(t, \tau; \varepsilon)$ выполнено условие $V(T, \tau; t) \leq \varepsilon$. ▲

Рассмотрим величину

$$H(\tau; \varepsilon) = \max_t h(t, \tau; \varepsilon). \quad (1.4)$$

Из (1.3) следует, что верхней оценкой минимально допустимого объема выборки $H(\tau; \varepsilon)$ является $2\tau/\varepsilon$. Отсюда также следует, что если взять $H(\tau; \varepsilon) = 2\tau/\varepsilon$, то для всех $T \geq H(\tau; \varepsilon)$ и для всех $\tau' \leq \tau$ выполняется $V(T, \tau; t) \leq \varepsilon$, т.е. такая выборка гарантированно будет ε -стационарной.

Эта оценка для $H(\tau; \varepsilon)$ может быть затем уточнена путем исследования статистических свойств конкретного ряда $x(t)$. В частности, может оказаться, что распределение величин $h(t, \tau; \varepsilon)$, представляющих при заданных τ и ε самостоятельный временной ряд, имеет выборочное среднее по времени t значение $\langle h(t, \tau; \varepsilon) \rangle$, значимо меньшее равномерной по t оценки (1.4).

Введем плотность распределения $\psi_{\tau, \varepsilon}(T)$ значений горизонтного ряда $h(t, \tau; \varepsilon)$, т.е. определим вероятность того, что расстояние между двумя ВПФР, построенными по выборке объема T и сдвинутыми по времени на фиксированный промежуток τ , не превосходит ε для всех $T' \geq T$. Для краткости зависимость от момента времени t в аргументах $\psi_{\tau, \varepsilon}(T)$ опущена. В силу (1.4) достаточно рассмотреть конечное множество значений T' , не превосходящих величины $2\tau/\varepsilon$.

Эта плотность $\psi_{\tau, \varepsilon}(T)$ строится по имеющимся данным к текущему моменту времени следующим образом. Для каждого момента времени t' : $1 \leq t' < t - \tau$, строятся ВПФР по выборкам объемов $T = 1, 2, \dots, t'$, после чего для каждой из этих ВПФР вычисляется функционал $V(T, \tau; t')$. Промежуток значений $[0; 2]$, принимаемых функционалом $V(T, \tau; t')$, разбивается при этом на некоторое количество K отрезков, например, равномерно, так что правый конец k -го отрезка есть $v_k = 2k/K$, $k = 1, 2, \dots, K$. Номер отрезка, фиксирующий заданную

точность ε , есть $k_\varepsilon = [K\varepsilon/2]$. Промежуток $[1; t - \tau]$ возможных значений объемов выборок содержит $t - \tau$ целочисленных точек, так что плоскость $\{T \times V\}$ покрыта $1 \times \frac{2}{T}$ -сетью. Отдельную ячейку сети нумеруем индексами $\{i, k\}$. Если значение $V(T, \tau; t')$ соответствует номерам i, k некоторой ячейки, помещаем в нее индикатор «1». Ячейки, номерам которых не отвечают значения функционала $V(T, \tau; t')$, заполняем нулями. Совокупность ячеек для данного i или k образует полосу. В каждый момент времени t' k -ая полоса состоит, таким образом, из нулей и единиц. Значением $h_k(t')$ горизонтного ряда в k -ой полосе называется индекс i ячейки, следующей за ячейкой с наибольшим индексом, содержащей ноль. Если нулевая ячейка — последняя, то значение горизонтного ряда в этой полосе отсутствует. Плотность $\psi_{\tau, \varepsilon}(T)$ представляет собой выборочную плотность функции распределения временного ряда $h_{k_\varepsilon}(t')$, построенную по выборке объема $t - \tau$, доставляющей исходному ряду условие ε -стационарности.

Определим интегральную функцию распределения

$$\Psi_{\tau, \varepsilon}(T) = \sum_{n=1}^T \psi_{\tau, \varepsilon}(k), \quad \Psi_{\tau, \varepsilon}(2\tau/\varepsilon) = 1. \quad (1.5)$$

Смысл функции распределения (1.5) в том, что если вместо величины $H(\tau; \varepsilon)$, определенной в (1.4), во многих случаях оказывающейся равной своему теоретическому максимуму $2\tau/\varepsilon$, взять некоторое меньшее значение T^* , то с вероятностью $\Psi_{\tau, \varepsilon}(T^*)$ будут выполнены условия ε -стационарности. Выборочные квантили распределения (1.5) могут служить оценками точности в установлении соответствующих объемов выборок.

▼ **Теорема 1.1.** Если ВПФР $f_T(x, t)$ ε -стационарна, то и ВПФР $\psi_{\tau, \varepsilon}(T)$ соответствующего горизонтного ряда также ε -стационарна.

Доказательство. Коль скоро для некоторых τ и ε определена указанная эмпирическая вероятность $\psi_{\tau, \varepsilon}(T)$ распределения минимальных объе-

мов, то из ее построения следует, что объем N имеющихся данных во всяком случае больше, чем $[2\tau/\varepsilon]$. Пусть в этом горизонтном распределении значение элемента $h(t, \tau; \varepsilon)$ горизонтного ряда, равное $T_k = k$, $k = 1, 2, \dots, [2\tau/\varepsilon]$, встретилось n_k раз, так что $\psi_{\tau, \varepsilon}(T_k) = n_k/T$. По смыслу величин $h(t, \tau; \varepsilon)$ каждый элемент исходного ряда $x(t)$ может быть первым элементом только в одной горизонтной выборке. Действительно, если оказалось, что выборка объемом k с начальным элементом $x(t_1)$ оказалась горизонтной в некоторый момент t , то никакая выборка большего объема с тем же начальным элементом не будет горизонтной для момента t в силу того, что $h(t, \tau; \varepsilon)$ есть минимальное значение объема соответствующей выборки. Точно так же не будет горизонтной и выборка меньшего объема с тем же начальным элементом, т.к. это будет противоречить тому, что исходная выборка объемом k является горизонтной. Следовательно, при сдвиге всего ряда данных, участвующих в построении ВПФР $\psi_{\tau, \varepsilon}(T)$, на τ шагов вперед изменение совокупности данных составит не более чем 2τ членов, что приведет к тому, что сумма абсолютных величин изменений чисел n_k также будет не более, чем 2τ , так что изменение ВПФР $\sum_{n=1}^{[2\tau/\varepsilon]} |\psi_{\tau, \varepsilon}(k; t) - \psi_{\tau, \varepsilon}(k; t + \tau')|$ для всех $\tau' \leq \tau$ не превзойдет величины $2\tau/N < \varepsilon$. Таким образом, ВПФР $\psi_{\tau, \varepsilon}(T)$ является ε -стационарной. Теорема 1.1 доказана. ▲

Укажем еще одно свойство распределения $\psi_{\tau, \varepsilon}(T)$, позволяющее оптимизировать объем выборки. Оно состоит в том, что если ВПФР $f_T(x, t)$ ε -стационарна, то можно уменьшить объем выборки без ухудшения точности в оценке уровня стационарности.

▼ **Теорема 1.2.** Функционал $V(T^*, \tau; t)$, где T^* есть $1 - \varepsilon$ -квантиль распределения $\psi_{\tau, \varepsilon}(T)$ в среднем отличается от функционала $V(T, \tau; t)$ на величину $o(\varepsilon)$.

Доказательство. Выберем некоторое $T^* < 2\tau/\varepsilon$. Для него функционал $V(T^*, \tau; t)$ не превосходит ε с вероятностью, определяемой распределением $\Psi_{\tau, \varepsilon}(T^*)$, где $\Psi_{\tau, \varepsilon}(k) = \sum_{n=1}^k \psi_{\tau, \varepsilon}(n)$, причем $\Psi_{\tau, \varepsilon}(2\tau/\varepsilon) = 1$. Пусть T^* есть $1 - \varepsilon$ -квантиль распределения $\psi_{\tau, \varepsilon}(T)$, т.е. $\Psi_{\tau, \varepsilon}(T^*) = 1 - \varepsilon$. В качестве объема выборки T^* возьмем ближайшее сверху целое число к решению относительно T^* этого уравнения. Тогда математическое ожидание функционала $V(T^*, \tau; t)$ будет не больше, чем

$$\begin{aligned} E[V(T, \tau; t)] &\leq \varepsilon' = \varepsilon \Psi_{\tau, \varepsilon}(T^*) + \frac{2\tau}{T^*} (1 - \Psi_{\tau, \varepsilon}(T^*)) = \\ &= \varepsilon \left(1 + \frac{2\tau}{T^*} - \varepsilon \right) = \varepsilon + o(\varepsilon). \end{aligned}$$

Последнее равенство в вышеприведенной цепочке следует из того, что при $\varepsilon \rightarrow 0$ $T^* \rightarrow 2\tau/\varepsilon$, поскольку 1-квантиль распределения есть такое значение T^* , при котором $\Psi_{\tau, \varepsilon}(T^*) = 1$. Следовательно, $\frac{2\tau}{T^*} - \varepsilon \rightarrow 0$ при $\varepsilon \rightarrow 0$, а это и означает, что $\varepsilon \left(\frac{2\tau}{T^*} - \varepsilon \right) = o(\varepsilon)$.

Теорема 1.2 доказана. ▲

Используя доказанное утверждение, можно провести минимизацию ошибки в оценке статистических свойств исследуемого ряда, поскольку для этой цели желательно иметь как можно более низкую оценку для $H(\tau; \varepsilon)$ с тем, чтобы иметь больший диапазон вариации объема выборки для достижения минимума суммы выборочной дисперсии и квадрата нормы разности ВПФР.

Аналогично проведенному рассмотрению для статистики минимально достаточных объемов выборки, для каждой выборки объема T не меньше, чем $[2/\varepsilon]$, существует максимальное значение горизонта сдвигов

$s(t, T; \varepsilon)$, такое, что при всех $\tau \leq s(t, T; \varepsilon)$ значения функционала $V(T, \tau; t)$ гарантированно не превосходят ε . Введем величину

$$S(T; \varepsilon) = \min_t s(t, T; \varepsilon). \quad (1.6)$$

Функция $S(T; \varepsilon)$ представляет эмпирическую оценку максимально-го горизонта прогноза, внутри которого распределение остается τ - ε -стационарным. Из (1.3) следует, что если взять $S(T; \varepsilon) = [\varepsilon T/2]$, то любые $\tau \leq [\varepsilon T/2]$ являются допустимыми горизонтами прогноза ВПФР с указанной точностью.

Пусть $\varphi_{T, \varepsilon}(\tau)$ представляет плотность функции распределения ряда $s(t, T; \varepsilon)$, т.е. вероятность того, что при фиксированном объеме выборки T расстояние между двумя ВПФР, сдвинутыми на интервал времени τ' , не превосходит ε для всех $\tau' < \tau$. Эта величина может быть построена по аналогичному алгоритму, как и величина $\psi_{\tau, \varepsilon}(T)$, только вместо ячеек на плоскости $\{T \times V\}$ рассматриваются ячейки на плоскости $\{\tau \times V\}$. Значением максимального прогнозного горизонта $s_k(t')$ в k -ой полосе называется индекс i ячейки, предшествующей ячейке с наименьшим индексом, содержащей ноль. Если нулевая ячейка — первая, то значение ряда горизонта прогноза в этой полосе отсутствует. Описанная процедура позволяет определить вероятность ε -близости ВПФР исходного ряда для фиксированного объема выборки.

Определим теперь интегральную функцию распределения горизонтов прогноза

$$\Phi_{T, \varepsilon}(\tau) = \sum_{k=1}^{\tau} \varphi_{T, \varepsilon}(k), \quad \Phi_{T, \varepsilon}(\varepsilon T/2) = 1. \quad (1.7)$$

Величина $\Phi_{T, \varepsilon}(\tau)$ представляет собой вероятность того, что при всех $T' \geq T$ и при всех $\tau' \leq \tau$ расстояние между различными ВПФР ряда $x(t)$ не превосходит ε . Поскольку эта вероятность имеет тот же самый смысл, что и вероятность (1.5), то справедливо утверждение о двойственности между введенными интегральными функциями распределения. Именно, значения интегральных функций распределения рядов минимальных допустимых объемов выборки и максимальных допусти-

мых горизонтов прогноза при соблюдении ε -стационарности ВФР совпадают, если аргумент одной из функций является параметром другой:

$$\Psi_{\tau, \varepsilon}(T) = \Phi_{T, \varepsilon}(\tau). \quad (1.8)$$

Из (1.8) следует, что для нахождения плотности $\varphi_{T, \varepsilon}(\tau)$ нет необходимости строить отдельный алгоритм. Поскольку по определению $\varphi_{T, \varepsilon}(\tau) = \Phi_{T, \varepsilon}(\tau) - \Phi_{T, \varepsilon}(\tau - 1)$, то из (1.8) следует, что

$$\varphi_{T, \varepsilon}(\tau) = \Psi_{\tau, \varepsilon}(T) - \Psi_{\tau-1, \varepsilon}(T). \quad (1.9)$$

Таким образом, ВФР горизонтной статистики $\Psi_{\tau, \varepsilon}(T)$ является достаточным инструментом для исследования обоих типов задач — «прямой» (определения объема выборки для прогнозирования на заданный горизонт) и «обратной», по определению горизонта прогнозирования при заданном объеме выборки.

Третья статистика — распределение расстояний между ВПФР при заданных величинах объема выборки и окна сдвига — была рассмотрена ранее в главе V. Она используется для построения индикаторов нестационарности и для классификации временных рядов по типу нестационарности в соответствии с величиной их согласованного уровня стационарности. Эта статистика в определенном смысле независима от горизонтных статистик, поскольку не требует выполнения каких-либо условий оптимальности.

6.2. Распределение горизонтного ряда стационарного процесса

Для анализа нестационарных рядов с помощью горизонтной статистики необходимо представлять основные отличия от стационарного случая. Именно мера отклонения от стационарных распределений и показывает, насколько данный процесс нестационарный. Изучение стационарной горизонтной статистики позволит построить специальный индикатор нестационарности как меры нарушения стационарного горизонтного распределения. Исследуем поэтому сначала свойства временного ряда $h(t, \tau; \varepsilon)$, являющегося горизонтным рядом для некоторого стационарного временного ряда $x(t)$.

Построим ПФР статистики $h(t, \tau; \varepsilon)$, если известно распределение вероятностей p_i того, что элемент исходного ряда попал в промежуток

Δ_i некоторого заданного разбиения отрезка $[0; 1]$ на n равных частей. Рассмотрим сначала простейший случай, когда делается сдвиг на один шаг, т.е. когда $\tau = 1$.

Сравним ВПФР $\psi_{1, \varepsilon}(T; t)$ горизонтного ряда, построенную по выборке $(t - T, t]$, с ВПФР $\psi_{1, \varepsilon}(T; t + 1)$, построенной по выборке того же объема, но сдвинутой вперед на 1 шаг по времени, т.е. по выборке $(t - T + 1, t + 1]$. При таком сдвиге из исходной выборки исключается первый ее элемент $x(t - T + 1)$ и добавляется один новый элемент $x(t + 1)$.

Возможны две ситуации: крайние элементы $x(t - T + 1)$ и $x(t + 1)$ принадлежат одному и тому же промежутку Δ_i , и тогда $V(T, 1; t) = 0$; либо $x(t - T + 1)$ и $x(t + 1)$ принадлежат разным промежуткам, и тогда $V(T, 1; t) = 2/T$. Поэтому, если $\varepsilon \geq 2/T$, то в пределах одного шага по времени любое выборочное распределение с объемом T является ε -стационарным. Если же $\varepsilon < 2/T$, то условие ε -стационарности применительно к данному моменту времени означает, что должно быть $V(T, 1; t) = 0$. Вероятность реализации последнего условия равна вероятности попадания двух фиксированных элементов временного ряда — $x(t - T + 1)$ и $x(t + 1)$ — в один и тот же, но произвольный, промежуток Δ_i . Для конкретного промежутка Δ_i эта вероятность равна p_i^2 . Тогда в силу предполагаемой независимости событий вероятность попадания двух значений временного ряда в один произвольный промежуток равна $\sum_{i=1}^n p_i^2$.

Определим далее теоретическую вероятность $P\{h(t, 1; \varepsilon) = m\}$ того, что $h(t, 1; \varepsilon) = m$, где m есть некоторое натуральное число. Обозначим эту вероятность через $q_{1, \varepsilon}(m)$ или, если это не будет существенно для рассуждений, просто $q(m)$.

Прежде всего заметим, что если $m > [2/\varepsilon]$, то $q(m) = 0$. Если же $m < [2/\varepsilon]$, то минимально допустимый объем выборки в данный мо-

мент времени t должен обладать тем свойством, что при сдвиге на 1 шаг выборок объемов $m, m+1, m+2, \dots, [2/\varepsilon]-1$ выполнено условие $V(m, 1; t) = 0$, т.е. первый элемент рассматриваемого ряда (этот элемент имеет номер $t - T + 1$) и еще $[2/\varepsilon] - m$ элементов принадлежат одному и тому же произвольному промежутку Δ_i , а элемент с номером $t - T + m + 1$ ему не принадлежит. Следовательно,

$$q(m) = \sum_{i=1}^n p_i^{M-m+1} (1 - p_i), \quad 1 < m < M,$$

где для краткости обозначено $M = [2/\varepsilon]$.

Вероятность же того, что $h(t, 1; \varepsilon) = 1$, означает, что реализовалось событие, когда все M элементов выборки попали в один промежуток разбиения, т.е.

$$q(1) = \sum_{i=1}^n p_i^M.$$

Наконец, вероятность того, что $h(t, 1; \varepsilon) = M$, определяется из условия нормировки $\sum_{m=1}^M q(m) = 1$, что дает

$$q(M) = 1 - \sum_{m=2}^{M-1} \sum_{i=1}^n p_i^{M-m+1} (1 - p_i) - \sum_{i=1}^n p_i^M = 1 - \sum_{i=1}^n p_i^2.$$

Поскольку вероятность того, что два фиксированных значения временного ряда попали в один промежуток, равна $\sum_{i=1}^n p_i^2$, то полученное

выше выражение есть, очевидно, вероятность того, что два фиксированных значения временного ряда не попали в один промежуток. Одновременно с этим доказано, что сумма вероятностей всех допустимых исходов для ряда $h(t, 1; \varepsilon)$ равна единице. Легко видеть, что вероятность события $\{h(t, 1; \varepsilon) = M\}$ может быть представлена в виде

$$P\{h(t, 1; \varepsilon) = M\} = 1 - \sum_{i=1}^n p_i^2 = \sum_{i=1}^n p_i (1 - p_i), \text{ т.е. в виде, аналогичном}$$

вероятности события $\{h(t, 1; \varepsilon) = m\}$ при $1 < m < M$.

Итак, при $\tau = 1$ получено теоретическое распределение $q_{\tau, \varepsilon}(m)$ случайной величины $h(t, \tau; \varepsilon)$, принимающей целочисленные значения от 1 до $M = [2/\varepsilon]$:

$$q_{1, \varepsilon}(m) \equiv P\{h(t, 1; \varepsilon) = m\} = \begin{cases} \sum_{i=1}^n p_i^M, & m = 1 \\ \sum_{i=1}^n p_i^{M-m+1} (1 - p_i), & 1 < m \leq M = [2/\varepsilon] \\ 0, & m > M \end{cases} \quad (2.1)$$

Распределение (2.1) можно назвать по аналогии с известным геометрическим распределением, задаваемым формулой $p_k = a^k (1 - a)$, $0 < a < 1$, $k = 1, 2, \dots$, конечно-аддитивным геометрическим распределением.

Укажем некоторые свойства стационарного распределения горизонтного ряда для сдвига на 1 шаг.

Свойство 1. Распределение горизонтного ряда (2.1) имеет максимум в точке $m = M$, поскольку оно монотонно возрастает независимо от вида распределения вероятностей p_i исходного ряда.

Действительно, из (2.1) следует, что

$$q_{1, \varepsilon}(m) - q_{1, \varepsilon}(m-1) = \sum_{i=1}^n p_i^{M-m+1} (1 - p_i)^2 > 0.$$

Свойство 2. Распределение (2.1) обладает следующим экстремальным свойством: функционал, сопоставляющий исходному распределению $\{p_i\}$ супремум распределения горизонтного ряда, достигающийся, со-

гласно свойству 1, в точке $m = M$, имеет абсолютный максимум при равномерном распределении вероятностей $p_i = 1/n$.

Свойство 3. Относительное среднеквадратичное отклонение для распределения (2.1) при любом распределении вероятностей p_i , отделенном от единицы, имеет при $\varepsilon \rightarrow 0$ порядок ε .

В общем случае сдвига на τ шагов область определения горизонтной статистики представляет собой целочисленный набор точек от 1 до $[2\tau/\varepsilon]$. Эта область разбивается на полосы шириной $[2/\varepsilon]$, внутри которых распределение горизонтного ряда является монотонно возрастающей функцией. Кусочно-монотонный рост распределения в последней (крайней справа) полосе, отвечающей наибольшим значениям горизонтного ряда, имеет определенную специфику (см. [25]), которая может быть использована для построения специального критерия стационарности ряда.

Пусть ряд $x(t)$ стационарный, и горизонтный ряд рассматривается в области значений $\frac{2(\tau-1)}{\varepsilon} < h(t, \tau; \varepsilon) \leq \frac{2\tau}{\varepsilon}$. Тогда условная вероятность $A_k \equiv A_k(i_1, \dots, i_\tau)$ того, что значения горизонтного ряда не превосходят заданного числа $m = [2\tau/\varepsilon] - \tau + k$ из указанного диапазона при условии попадания значений исходного ряда $x(t)$ в промежутки $\Delta_{i_1}, \Delta_{i_2}, \dots, \Delta_{i_\tau}$, выражается через интеграл Коши по контуру, охватывающему нули характеристического многочлена $Q(\lambda)$ рекуррентной последовательности для определения A_k :

$$A_k = 1 - \frac{Q(1)}{2\pi i} \int \frac{z^k dz}{Q(z) \cdot (z-1)}. \quad (2.2)$$

Многочлен $Q(\lambda)$ имеет вид

$$Q(\lambda) = \frac{\lambda^{\tau+1} - \lambda^\tau + a(1-a)^\tau}{\lambda - (1-a)}, \quad (2.3)$$

где $a \equiv a(i_1, \dots, i_\tau)$ есть вероятность того, что τ значений временного ряда попали соответственно в промежутки $\Delta_{i_1}, \Delta_{i_2}, \dots, \Delta_{i_\tau}$:

$$a(i_1, \dots, i_\tau) = \sum_{k=1}^{\tau} p_{i_k} - \sum_{\substack{k,l=1 \\ k \neq l}}^{\tau} p_{i_k} p_{i_l} + \dots + (-1)^{\tau-1} \prod_{k=1}^{\tau} p_{i_k}. \quad (2.4)$$

Представление (2.2) позволяет получить несколько важных следствий относительно свойств горизонтной статистики для указанного диапазона значений. Эти утверждения оказываются полезными во многих практических случаях при оценке того, насколько исследуемый ряд нестационарный.

Во-первых, можно определить вероятность максимального значения горизонтного ряда. Она равна

$$q_{\tau, \varepsilon}([2\tau/\varepsilon]) = \left(\sum_i (1-p_i)^\tau p_i \right)^\tau. \quad (2.5)$$

Действительно, из вида величин $a(i_1, \dots, i_\tau)$, определенных в (2.4), получаем цепочку равенств, приводящую к (2.5):

$$\begin{aligned} q_{\tau, \varepsilon}([2\tau/\varepsilon]) &= \sum_{i_1, \dots, i_\tau=1}^n p_{i_1} \dots p_{i_\tau} (1 - A_1(i_1, \dots, i_\tau)) = \\ &= \sum_{i_1, \dots, i_\tau=1}^n \left(\prod_i (1-p_i) \right)^\tau p_{i_1} \dots p_{i_\tau} = \\ &= \left(\sum_{i_1=1}^n (1-p_{i_1})^\tau p_{i_1} \right) \cdot \dots \cdot \left(\sum_{i_n=1}^n (1-p_{i_n})^\tau p_{i_n} \right) = \left(\sum_i (1-p_i)^\tau p_i \right)^\tau. \end{aligned}$$

Во-вторых, можно построить новый — горизонтный — критерий ε -стационарности ВПФР исследуемого ряда $x(t)$. Рассмотрим распределение горизонтного ряда в крайней справа полосе $\frac{2(\tau-1)}{\varepsilon} < h(t, \tau; \varepsilon) \leq \frac{2\tau}{\varepsilon}$ шириной $[2/\varepsilon]$. Оказывается, что в этой полосе распределение $q_{\tau, \varepsilon}([2\tau/\varepsilon] - m)$ при $1 \leq m \leq \tau$ является равномерным.

Действительно, поскольку

$$A_k = a \cdot (A_{k-1} + (1-a) \cdot A_{k-2} + \dots + (1-a)^{\tau-1} \cdot A_{k-\tau}), \text{ то}$$

$$\begin{aligned}
q_{\tau, \varepsilon}([2\tau / \varepsilon] - m) &= \\
&= \sum_{i_1, \dots, i_n=1}^n p_{i_1} \dots p_{i_\tau} (A_m(i_1, \dots, i_\tau) - A_{m+1}(i_1, \dots, i_\tau)) = \\
&= \sum_{i_1, \dots, i_n=1}^N \left(a(i_1 \dots i_\tau) (1 - a(i_1 \dots i_\tau))^\tau \right) p_{i_1} \dots p_{i_\tau}.
\end{aligned} \quad (2.6)$$

Последнее выражение не зависит от m , что и требовалось доказать.

Таким образом, для проверки того, что ВПФР ε -стационарна, надо построить горизонтную статистику для достаточно большого параметра сдвига, если это позволяет сделать имеющийся набор данных, и проверить выборочное распределение $\psi_{\tau, \varepsilon}([2\tau / \varepsilon] - m)$ в последней полосе шириной $[2 / \varepsilon]$ на равномерность: именно, в конце этой полосы, но не включая ее крайнюю правую точку, должен существовать равномерно распределенный участок длины $\tau - 1$. Отсутствие этого участка означает наличие нестационарности.

На рис. 6.2.1 приведен горизонтный ряд стационарного временного ряда, представляющего собой выборку из нормального распределения, для значений $\tau = 10$, $\varepsilon = 0,05$.

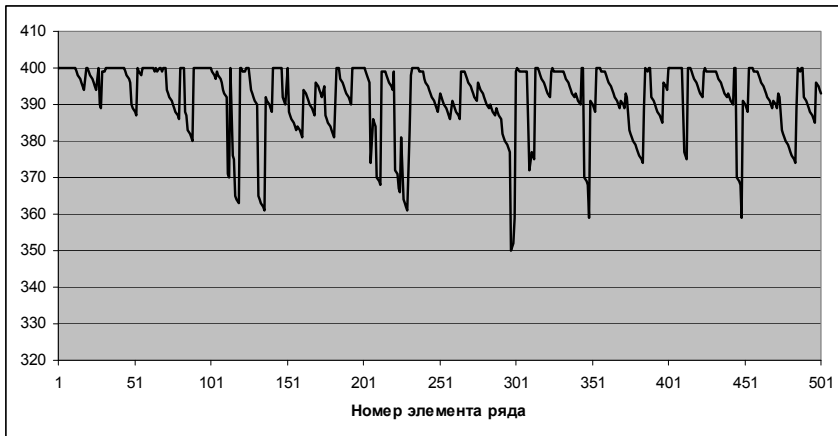


Рис. 6.2.1 — Горизонтный ряд стационарного процесса для $\tau = 10$, $\varepsilon = 0,05$

На рис. 6.2.2 показано распределение этого горизонтного ряда, построенное по выборке в 2000 значений.

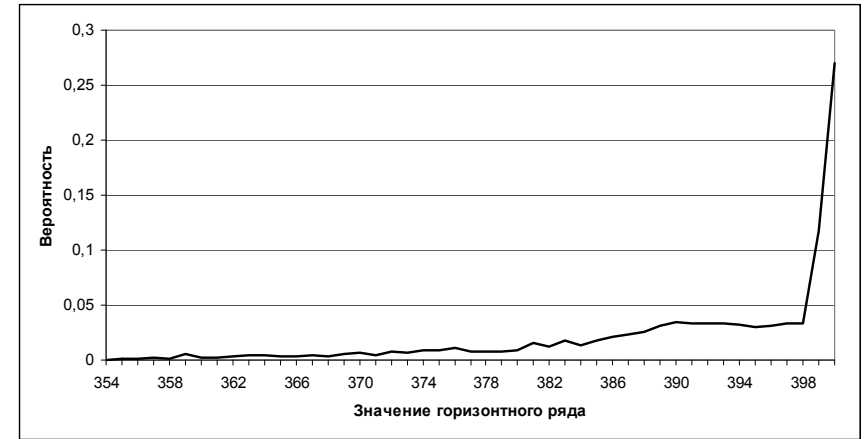


Рис. 6.2.2 — Распределение стационарного горизонтного ряда для $\tau = 10$, $\varepsilon = 0,05$

Характерность горизонтной статистики, показанной на рис. 6.2.2, состоит в том, что, кроме равномерно распределенного «плато», имеется преобладающее количество значений горизонтного ряда, которые равны максимальному значению (в данном случае 400). Для статистически зависимых случайных величин последнее не имеет места, так что горизонтная статистика, как мы увидим ниже, позволяет обнаружить зависимость (не обязательно линейную) между элементами ряда.

6.3. Горизонтный ряд хаотической динамической системы

Динамические системы часто используются для моделирования временных рядов, поскольку известно, что каждой консервативной (т.е. сохраняющей свой фазовый объем) ДС можно поставить в соответствие стационарный случайный процесс. Это связано не только с тем, что существуют ДС, демонстрирующие хаотическое поведение, но главным образом с тем, что многие модели стационарных временных рядов — регрессионные, авторегрессионные, скользящих средних и др. — являются примерами динамических систем с дискретным временем. Так, авторегрессионная модель вида $\hat{x}_{n+1} = ax_n + b$, которая в линейном приближении

устанавливает статистическую связь между двумя последовательными элементами временного ряда, является примером ДС, траектория которой, состоящая из точек \hat{x}_n , определяется рекуррентно. Шляпка над \hat{x}_n указывает на то, что это модельное значение, строящееся по известному предыдущему члену ряда.

Интерес к хаотическим ДС вызван тем, что, как оказалось, многие явления, воспринимаемые как случайные, могут быть на самом деле детерминированными, и тогда модели типа (3.1) являются средством для изучения и познания хаоса. С одной стороны, точность реальных измерений всегда конечна, поэтому окончательно решить вопрос о причинах, приводящих к тем или иным следствиям, невозможно. С другой стороны, наблюдая за выборкой из некоторого распределения, можно попытаться с некоторой точностью построить ДС с похожим поведением.

Возникает вопрос: похожи ли горизонтные ряды для статистического и динамического хаосов? Между этими двумя типами рядов существует то принципиальное различие, что для статистического белого шума не только линейная, но и любая нелинейная автокорреляция равна нулю, тогда как для динамического хаоса существует функциональная корреляционная связь между элементами ряда, а именно: та, которая определяется логистическим отображением.

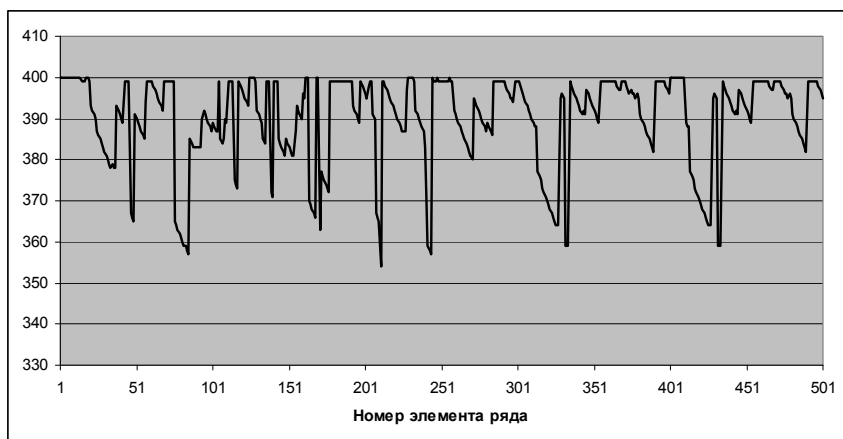


Рис. 6.3.1 — Горизонтный ряд логистической системы для $\tau = 10, \varepsilon = 0,05$

На рис. 6.3.1 приведен горизонтный ряд динамического хаоса, порожденного логистическим отображением, а на рис. 6.3.2 показано распределение этого горизонтного ряда. Параметры построения горизонтного ряда были выбраны такими же, как и для стационарного ряда на рис. 6.2.1: $\tau = 10, \varepsilon = 0,05$. Принципиальных различий между рядами на рис. 6.2.1 и 6.3.1, на первый взгляд, нет. Но распределения этих рядов (рис. 6.2.2 и 6.3.2) качественно различны.

Главное отличие распределения горизонтного ряда логистической системы — в более сильной немонотонности поведения в окрестности предельных значений и, в частности, в том, что максимум распределения приходится не на максимальное, а на предпоследнее значение.

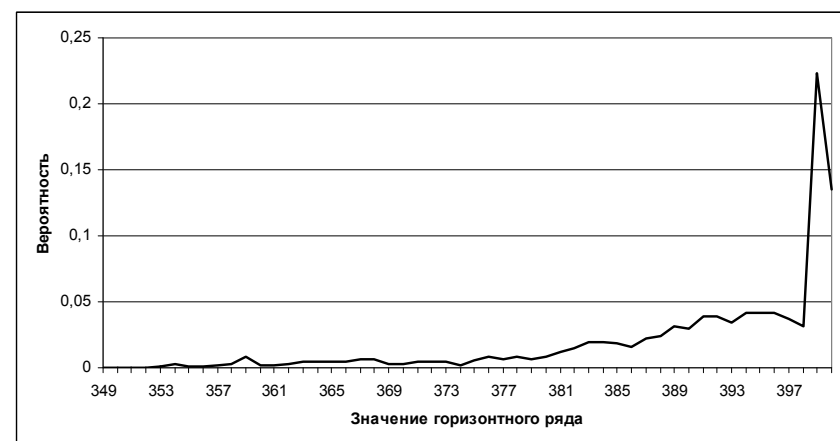


Рис. 6.3.2 — Распределение горизонтного ряда логистической системы

То, что вероятность максимального значения горизонтного ряда существенно ниже, чем предыдущего значения, свидетельствует о корреляциях между элементами ряда, хотя линейные корреляции, как мы видели ранее, близки к нулю. Таким образом, горизонтная статистика позволяет выявить наличие скрытых зависимостей между значениями временного ряда ДС, которые не находятся средствами линейного корреляционного анализа.

Приведем еще один пример горизонтного ряда для системы, родственной рассмотренной выше логистической ДС — это трехмерная система с квадратичной правой частью, известная как система Лоренца:

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = \lambda x - y - xz, \\ \dot{z} = -\mu z + xy. \end{cases} \quad (3.1)$$

Она часто возникает в различных физических приложениях. Например, к ней приводится система газодинамических уравнений в одной из моделей погоды. Хаотическое поведение численного решения системы (3.1) наблюдается, в частности, при значениях параметров $\sigma = 10$, $\lambda = 28$, $\mu = 8/3$ и шаге дискретизации по времени в пределах 0,05–0,06. Горизонтный ряд и его распределение для второй компоненты этой системы при начальных условиях $x(0) = 0,71$, $y(0) = 0,28$, $z(0) = 0,32$ и шаге дискретизации 0,06 показаны на рис. 6.3.3 и 6.3.4.

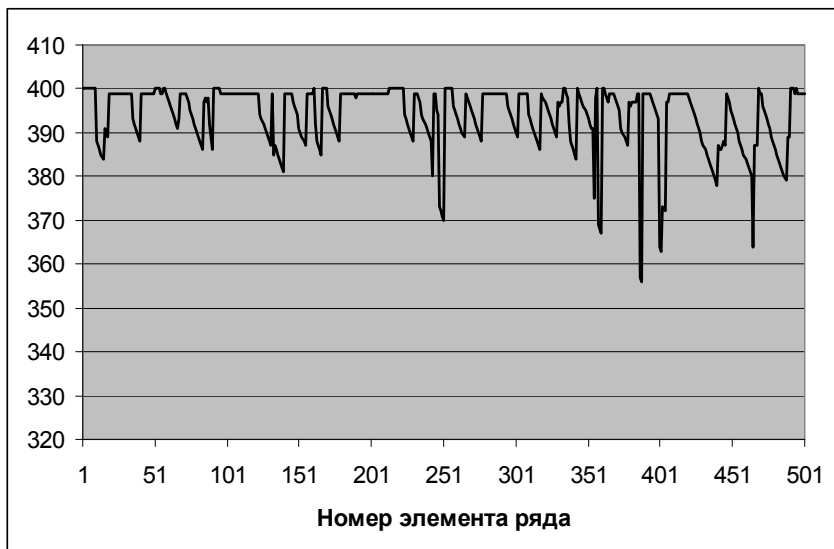


Рис. 6.3.3 — Горизонтный ряд системы Лоренца для $\tau = 10$, $\varepsilon = 0,05$



Рис. 6.3.4 — Распределение горизонтного ряда системы Лоренца

Распределения горизонтных рядов на рис. 6.3.2 и 6.3.4 весьма похожи. Этот факт трактуется в том смысле, что тип и величина хаоса в исходных динамических системах примерно одинаковы. Подчеркнем, что, несмотря на стационарность ряда, горизонтальный участок в последних $\tau = 10$ значениях не наблюдается, что свидетельствует о коррелированности значений ДС.

6.4. Распределения горизонтных рядов в нестационарном случае

Для нестационарных временных рядов распределение горизонтного ряда само не является квазистационарным, однако важность для исследователя представляет не то, что это распределение эволюционирует, а его отличие как таковое от стационарных распределений. Статистика горизонтного ряда представляет собой инструмент для анализа исходного ряда, но инструмент не универсальный, а требующий определенной «настройки». В чем-то его применение схоже с увеличительным стеклом, которое четко показывает не все вообще предметы, а лишь те, которые находятся в фокусе. Так и в распределении горизонтного ряда наличие, например, нелинейных корреляций может наиболее заметно проявиться

при определенных значениях параметров сдвига τ и квазистационарности ε . Это означает, что анализ статистики необходимо становится многомерным даже в том случае, когда исходный временной ряд одномерный. Сами значения горизонтного ряда зависят от трехмерного массива переменных — текущего времени t и параметров τ и ε . Распределение же этого ряда, построенное по некоторой временной выборке, двумерно. Те значения τ , при которых обнаруживаются отклонения горизонтного распределения от картины, типичной для стационарного белого шума, покажут дальность корреляционной связи.

В последующих главах будут рассмотрены на товарно-сырьевых рынках, рынках ценных бумаг, в задачах прогнозирования погоды и в некоторых других областях человеческой деятельности. Здесь мы приведем примеры горизонтных распределений для некоторых из нестационарных рядов, возникающих в практической деятельности, чтобы продемонстрировать необходимость проведения такого достаточно трудоемкого по вычислительным средствам анализа.

На рис. 6.4.1 и 6.4.2 показаны фрагмент горизонтного ряда и его распределение по объему в 2500 данных для приростов минутных цен закрытия на акции компании General Electric, а на рис. 6.4.3 и 6.4.4 — аналогичная пара для приростов суточных данных обменного курса евро/доллар. Как и в предыдущем параграфе, будем рассматривать эти ряды для значений $\tau = 10$, $\varepsilon = 0,05$.

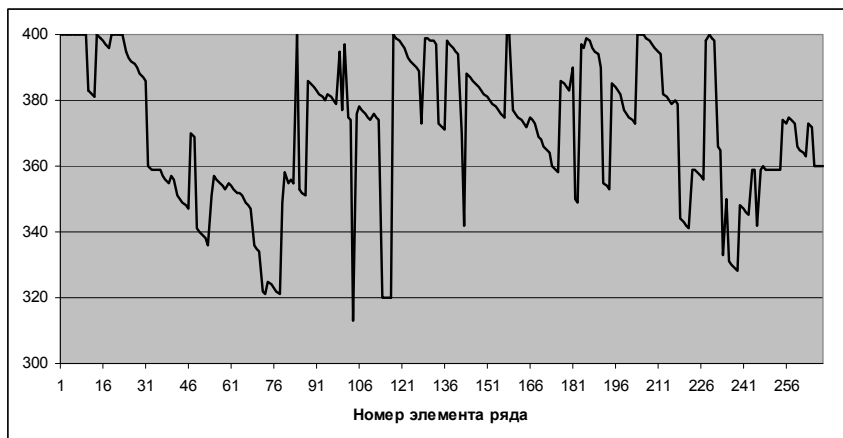


Рис. 6.4.1 — Горизонтный ряд для приростов минутных цен закрытия на акции General Electric, $\tau = 10$, $\varepsilon = 0,05$



Рис. 6.4.2 — Распределение горизонтного ряда приростов цен на акции General Electric

Распределение на рис. 6.4.2 существенно отличается от графика на рис. 6.2.2. Оно заметно немонотонно и имеет четкий максимум в конце предыдущей полосы ширины в 40 значений.

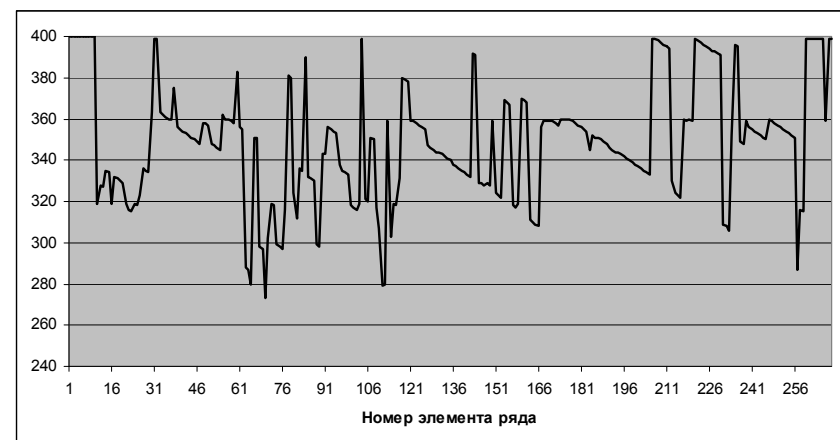


Рис. 6.4.3 — Горизонтный ряд для приростов суточных данных курса евро/доллар, $\tau = 10$, $\varepsilon = 0,05$

Этот ряд также заметно отличается от горизонных рядов, рассмотренных в предыдущих двух параграфах. Главные отличия состоят в значительно более широкой полосе изменения и в том, что максимальные значения встречаются весьма редко. Это свидетельствует о тесной связи между элементами ряда, хотя линейный корреляционный анализ такую связь выявить не позволяет.



Рис. 6.4.4 — Распределение горизонтного ряда приростов курса евро/доллар

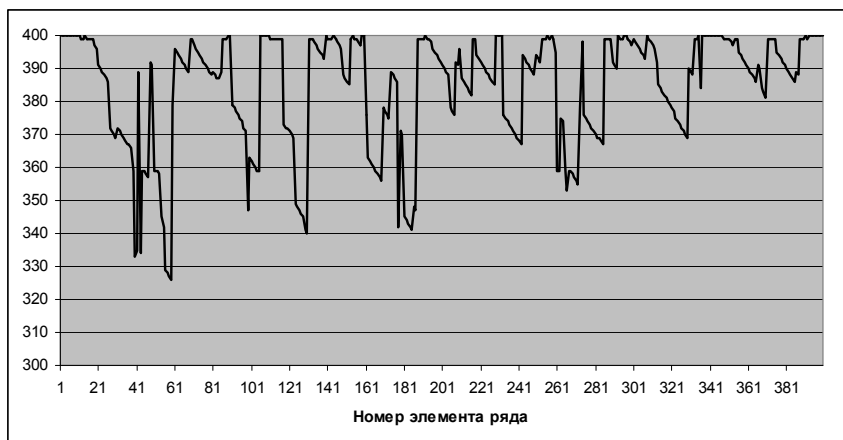


Рис. 6.4.5 — Горизонтный ряд для отклонений среднесуточной температуры в Москве

Приведем также на рис. 6.4.5 и 6.4.6 горизонтный ряд для ряда отклонений среднесуточной температуры в Москве за период 2000–2010 гг. от среднегодового профиля, построенного за десятилетний период с 1990 по 1999 гг., и распределение этого ряда.

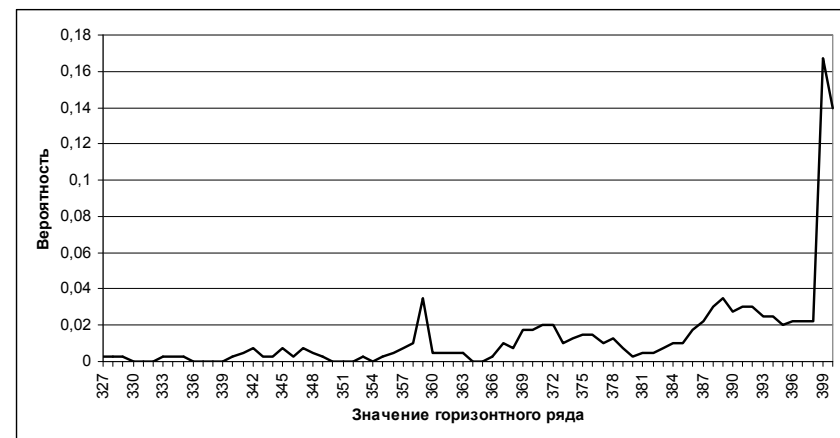


Рис. 6.4.6 — Распределение горизонтного ряда отклонений среднесуточной температуры

Как оказалось, что ряд температурных отклонений гораздо более стационарный, чем ряды биржевых котировок, причем он проявляет черты динамической системы, которая похожа на систему Лоренца.

Рассмотренные примеры позволяют сделать следующие выводы об отличиях горизонтных рядов стационарных и нестационарных случайных процессов. Отметим, что анализ уровня нестационарности некоторого временного ряда с помощью распределения горизонтного ряда полезно проводить совместно с анализом автокорреляционной функции исходного ряда. Возможны следующие типичные варианты качественного поведения АКФ и горизонтной ВПФР.

Первый случай. АКФ с увеличением лага быстро падает до условно незначительной величины, близкой к нулю, а горизонтная ВПФР имеет монотонный рост с равномерным плато в последней полосе значений. Тогда ряд представляют собой выборку из стационарного распределения и может быть исследован стандартными методами.

Второй случай. АКФ быстро падает до незначительной величины, но при некоторых τ и ε горизонтная ВПФР имеет заметно немонотон-

ное поведение внутри промежутка ширины $2/\varepsilon$. Тогда это свидетельствует о том, что ряд содержит нелинейные корреляции. Этот вариант реализуется в рядах, генерируемых динамическими системами.

Третий случай. АКФ спадает медленно, а горизонтная ВПФР монотонно возрастает. Тогда ряд существенно нестационарный, и требуется его предобработка с целью снижения влияния временного тренда.

Четвертый случай. АКФ имеет локальные максимумы, а горизонтная ВПФР немонотонна. Тогда следует провести пред-обработку ряда по авторегрессионной модели с лагом порядка τ .

6.5. Алгоритмы нахождения оптимального объема выборки

Из предыдущего рассмотрения становится понятно, что оптимальность объема выборки существует не сама по себе, а в зависимости от горизонта прогнозирования и желаемой близости между ВПФР. Кроме того, поскольку дисперсия служит оценкой точности прогноза даже в нестационарном случае, то следует также найти точки локального минимума выборочной дисперсии ряда $x(t)$ как функции объема выборки.

В зависимости от метода прогнозирования можно предложить два основных алгоритма оптимизации объема выборки. Первый алгоритм используется для построения наиболее точного прогноза методом кинетических уравнений. Второй алгоритм применяется для уменьшения ошибки прогнозирования в рамках некоторой стандартной модели временного ряда.

В первом случае последовательность действий для нахождения оптимального объема выборки для описания ряда на горизонте τ состоит в следующем.

Сначала задается желаемая близость ε между двумя ВПФР, сдвинутыми одна относительно другой на τ шагов, и вычисляется функционал разности $V(T, \tau; t)$ между ними. Затем в каждый момент времени t строится статистика дисперсий $\sigma^2(t, T)$ как функций объема выборки.

Напомним, что если ряд, для которого прогнозируется ВПФР, стационарный, то минимальный объем выборки, который должен использоваться для такого прогноза, равен $T_0 = [2\tau/\varepsilon]$ данных. Вообще, чем больше данных используется для прогноза стационарного ряда, тем этот прогноз точнее.

Горизонтный ряд и минимумы выборочной дисперсии определяются по имеющемуся периоду наблюдений. Этот период должен быть достаточно большим, чтобы число минимумов дисперсии и количество данных горизонтной статистики было бы не менее, чем $T_0/4$.

Если же нестационарность существенна, то объем выборки объема T_0 может оказаться слишком большим, учитывающим и те данные, которые в силу удаленности по времени де-факто не влияют на текущий процесс. Найдя $1 - \varepsilon$ -квантиль распределения горизонтного ряда, мы определим объем выборки T_{\min} , достаточный для прогнозирования ВПФР с указанной точностью. Этот объем может быть значительно меньше, чем T_0 . Затем строится эмпирическое распределение минимумов выборочных дисперсий ряда как функций объема выборки, и находится $1 - \varepsilon$ -квантиль их распределения, дающий некоторый объем выборки T_σ .

Если оказалось, что $T_\sigma \geq T_{\min}$, то оптимальный объем для прогнозирования ВПФР на выбранные τ шагов с заданной точностью есть $T_{opt} = \max\{T_\sigma, T_0\}$.

Если же $T_\sigma < T_{\min}$, то $\forall T \in [T_\sigma; T_{\min}]$ рассматривается величина

$$\delta^2(T, \tau; t) = \sigma^2(T, t) + V^2(T, \tau; t), \quad (5.1)$$

где $V(T, \tau; t)$ есть функционал расстояния между двумя ВПФР. Составляется статистика объемов выборки $T_{opt}(\tau, t)$, дающих минимум величины $\delta^2(T, \tau; t)$ в каждый момент времени t и при фиксированном горизонте прогноза τ .

Среднее значение

$$T_\delta(\tau) = \langle T_{opt}(\tau, t) \rangle \quad (5.2)$$

принимается за оптимальный объем выборки для построения прогноза.

Смысл оценки (5.1) в том, что она минимизирует верхнюю грань совокупного отклонения $\sigma^2 + \varepsilon^2$ аппроксимации значений временного ряда от фактических данных по причинам нерепрезентативности и нестационарности. Идея метода опирается на отклонение оценок дисперсий выборочных моментов в нестационарном случае от их поведения в соответствии с центральной предельной теоремой для достаточно

ветствии с центральной предельной теоремой для достаточно больших выборок. Именно, во многих примерах временных рядов, возникающих на практике, выборочная дисперсия как функция объема выборки стабилизируется с ростом последней слабее, чем $1/T$. Факт стабилизации объясняется принципиальной ограниченностью наблюдаемых данных (цен, спроса на товар, курса валюты и т.п.).

Отметим, что неверно было бы определять T_{opt} просто как $\arg \min \sigma(T)$, поскольку для нестационарного ряда выборочная дисперсия может быть заниженной оценкой ошибки прогноза.

Подчеркнем, что введенное понятие оптимальности относится исключительно к минимизации величины $\sigma^2 + \varepsilon^2$ в оценке верхней грани ошибки прогноза, а не к самой ошибке прогноза δ . Если для анализа статистических свойств ряда и последующего прогноза берется выборка объемом \bar{T}_{opt} , то можно гарантировать, что ошибка прогноза на горизонт τ не превзойдет (в среднем по серии прогнозов) наименьшего значения суммы квадратов $\sigma^2 + \varepsilon^2$ как функции объема выборки T , однако это не означает, что при любом другом значении T' среднеквадратичная ошибка $\delta(T')$ будет не меньше, чем $\delta(\bar{T}_{opt})$.

Использование функционала ошибки в виде $\sigma^2 + \varepsilon^2$, а не в виде, скажем, статистики ошибок прогнозов $\delta(T)$ предпочтительнее тем, что функционал $\delta(T)$ представляет собой временной ряд, зависящий от многих факторов, в том числе и от применения той или иной прогнозной модели, тогда как верхняя грань $\sigma^2 + \varepsilon^2$ этого функционала относится непосредственно к анализируемому временному ряду, причем для нее можно получить априорные оценки сверху. В то же время на практике зачастую применяются конкретные модели — например, авторегрессионные или регрессионные, которые либо дают в среднем приемлемую точность, либо по иным причинам, устраивающим пользователя. В таком случае точность этих моделей может быть повышена прямой оптимизацией объема выборки, причем в результате можно будет оценить и верхнюю грань ошибки прогнозной модели. Опишем этот второй алгоритм и получим выражение для ошибки прогноза.

Рассматривая ошибку прогноза по некоторой модели как функцию объема выборки, в каждый прогнозный момент времени находится значе-

ние объема $T_{opt}(\tau, t)$, на котором относительная ошибка прогноза исходного ряда $x(t)$ на τ шагов вперед минимальна:

$$\delta^2(T, \tau, t) = \frac{1}{\tau} \sum_{n=1}^{\tau} (\hat{x}(T; t+n) / x(t+n) - 1)^2 \rightarrow \min. \quad (5.3)$$

Затем строится ВПФР этих объемов $T_{opt}(\tau, t)$ и находится их среднее по времени значение $\bar{T}(\tau)$ и дисперсия $\sigma_{\bar{T}}^2(\tau)$. Если отношение σ_T / \bar{T} (или, более точно, согласованный доверительный интервал для среднего значения) есть величина порядка δ_{\min} , определенной из (5.3) для $\bar{T}(\tau)$, то величина \bar{T} считается таким оптимальным объемом. Если отношение σ_T / \bar{T} велико, то следует выяснить, какова чувствительность ошибки прогноза к вариации объема выборки. Для этого на промежутке $\Delta_{\delta}(t) = (\delta(T_{opt}(\tau, t)) - \sigma_T, \delta(T_{opt}(\tau, t)) + \sigma_T)$ вычисляется величина

$$\lambda(\tau, t) = \frac{T_{opt}(\tau, t)}{2\sigma_T \delta(T_{opt}(\tau, t))} \left(\max_{\Delta_{\delta}(t)} \delta(T, \tau, t) - \delta(T_{opt}(\tau, t)) \right), \quad (5.4)$$

которая представляет собой чувствительность ошибки прогноза к объему выборки. Если средняя по времени чувствительность $\bar{\lambda}(\tau)$ мала, т.е. порядка δ_{\min} , то величину \bar{T} по-прежнему можно считать оптимальным объемом. Если же чувствительность велика, то ошибка прогноза при использовании объема \bar{T} как оптимального приведет к ошибке большей, чем давалось статистикой (5.3). Оценим эту ошибку.

Пусть $\bar{\delta}(\tau)$ есть среднее значение минимальной ошибки прогноза на горизонт τ , т.е. среднее значение величины $\delta(T_{opt}(\tau, t))$, а $D^2(\tau)$ — дисперсия минимальной ошибки по имеющемуся набору наблюдений. Тогда оценка дисперсии $\sigma_{\bar{\delta}}^2(\tau)$ ошибки прогнозирования при использовании во все моменты времени среднего оптимального объема \bar{T} вместо локально оптимального $T_{opt}(\tau, t)$ может быть найдена из следующих

соображений. Предположим, что оба случайных фактора — случайность самой минимальной ошибки прогноза и вариация ее при отклонении используемого объема \bar{T} от оптимального — проявляются независимо. При этом минимальная ошибка прогноза как функция оптимального объема выборки изменяется в силу своей собственной случайной природы, а также в силу возможной функциональной зависимости от этого объема. Пусть $\rho(\tau)$ есть коэффициент корреляции между ошибкой $\delta(T_{opt}(\tau, t))$

и объемом $T_{opt}(\tau, t)$. Тогда получаем

$$\begin{aligned}\sigma_{\delta}^2 &= \left\langle (\delta(t) - \bar{\delta})^2 \right|_{T=T_{opt}(t)} + (1 - \rho^2) \left(\frac{\partial \delta(t)}{\partial T} \right)^2 (T_{opt}(t) - \bar{T})^2 \rangle = \\ &= D^2 \left(1 + (1 - \rho^2) \frac{\bar{\lambda}^2 \sigma^2}{\bar{T}^2} \right).\end{aligned}\quad (5.5)$$

Формула (5.5) позволяет оценить ошибку прогноза ряда, если использовать выборку объема \bar{T} . Правый конец интервала, содержащего ожидаемую ошибку прогноза, оценивается величиной

$$\hat{\delta} = \bar{\delta} + \sigma_{\delta} = \bar{\delta} + D \cdot \sqrt{\left(1 + (1 - \rho^2) \frac{\bar{\lambda}^2 \sigma^2}{\bar{T}^2} \right)}.\quad (5.6)$$

Следует сказать, что при сравнении двух методов прогнозирования — кинетического по оптимальному объему выборки, определяемому минимумом совокупной ошибки (5.1), и линейного статистического метода в скользящем по времени окне меняющихся коэффициентов модели с оптимальным объемом (5.3), берущимся, естественно, с предыдущего шага по времени, — ошибка первого метода неизменно оказывалась меньше, чем второго.

Глава VII. Индикаторы разладки

7.1. Скачки значений горизонтного ряда

В этой главе будут рассмотрены некоторые индикаторы изменения выборочной функции распределения одномерного временного ряда. Следует подчеркнуть, что выборочное распределение, строго говоря, не имеет отношения к гипотетической генеральной совокупности, из которой «как бы» берутся значения нестационарного ряда в тот или иной момент времени. С другой стороны, именно такая интерпретация наблюдаемых статистик и делается на практике, хотя осмысленной эта процедура является только для стационарных процессов. Тем не менее, если, например, считать, что в каждый момент времени на бирже генеральная совокупность порождена ожиданиями игроков, то сигнал о том, что распределение этих ожиданий (неизвестное, заметим, в принципе) изменилось, сам по себе является важным.

Ситуация, когда распределение значимо изменилось, называется разладкой. Если процесс сильно нестационарный, то разладка наблюдается все время. В таком случае ошибка применения текущей статистики к прогнозированию будущего распределения велика, и для ее снижения требуется пред-обработка ряда, состоящая в удалении слишком явного временного тренда или сезонных колебаний. Если же дело не в тренде, а колебания носят нерегулярный характер, то предсказательная способность статистического анализа оказывается весьма слабой.

Другой вариант задачи о разладке реализуется тогда, когда система переходит из одного относительно устойчивого (т.е. квазистационарного) состояния в другое, тоже устойчивое. Тогда задача исследователя состоит в том, чтобы в течение переходного процесса опознать изменение распределения. Опознать же его после установления нового стационарного распределения труда не составляет, но время для принятия решений оказывается при этом упущенным.

В качестве индикатора разладки выступает некий функционал от выборочного распределения временного ряда. Обычно исходят из того, что если значение этого функционала превосходит некоторое пороговое, то ситуация считается изменившейся. При этом следует быть готовым к тому, что каждый индикатор неизбежно имеет некоторую вероятность ложного срабатывания, когда статистический шум принимается за разладку, а также вероятность ложного пропуска реальной разладки. Чтобы построить эффективный индикатор разладки в нестационарном ряде, следует

иметь в своем распоряжении два функционала, один из которых будет иметь нестационарное распределение (собственно индикатор разладки), а второй должен иметь стационарное распределение или близкое к таковому, на основе которого должны определяться параметры выборки для формирования первого индикатора.

Например, нестационарным индикатором может быть значение выборочной дисперсии, которое попало в определенную критическую полосу. Стационарным же функционалом должен быть объем выборки для вычисления этой дисперсии.

Одним из эффективных индикаторов разладки является горизонтный ряд, т.е. его значения, которые характеризуют наблюдаемое распределение выборочной вероятности значений исходного ряда. Конкретное значение горизонтного ряда показывает, выборкой какого объема определяется вероятностное распределение случайной величины с заданной точностью и изменение этого распределения. Именно, значение горизонтной статистики в момент времени t показывает, сколько значений исходного ряда надо взять к моменту $t - \tau + 1$, чтобы выборочное распределение изменилось за время τ не более чем на фиксированное значение ε . Вследствие этого большие значения горизонтного ряда свидетельствуют о высокой хаотичности процесса, а относительно малые — о включении упорядоченности.

Как следует из определения горизонтного ряда, этот индикатор имеет фиксированное запаздывание τ . Индикативный смысл этой статистики в том, что она позволяет оценить уровень консолидации участников рынка, или, иными словами, определить возникновение дальних корреляций между значениями временного ряда.

Примеры горизонтных рядов, приведенные в параграфе 6.4 предыдущей главы, имеют характерную особенность, которая проявляется как пилообразный резкий спад со значений, близких к максимальному $h_{\max} = [2\tau / \varepsilon]$, и последующий монотонный рост. Также наблюдаются и противоположные процессы — резкий рост после квазимонотонного снижения. Прямолинейный рост или снижение горизонтного ряда означает, что каждое последующее значение не содержалось (или наоборот, содержится) в прошлой выборке. Резкий спад означает, что выборки на заданном горизонте сдвига внезапно стали иметь много общих элементов, что трактуется как проявление системности.

Периодичность указанных пилообразных промежутков определяет временное расстояние между консолидацией значений временного ряда, а рост свидетельствует о нарастании хаоса в системе, т.е., например, об от-

сутствии тренда на выбранном временном промежутке. В этом смысле информация о времени, нужном для возникновения согласованного поведения или, напротив, для релаксации, содержится в распределении промежутков между локальными минимумами и максимумами горизонтного ряда и представляет несомненный практический интерес.

В этой связи возникают две задачи. Первая — определить характерные промежутки между хаосом и консолидацией, т.е. между максимальными и минимальными значениями горизонтного ряда. Вторая задача — определить локальную периодичность в тренде поведения случайного процесса, т.е. распределение промежутков нарастания и убывания хаоса.

Чтобы построить указанные распределения, необходимо, во-первых, найти такие τ и ε , для которых поведение горизонтного ряда обладает нужными свойствами, т.е. информативно с точки зрения исследователя, и, во-вторых, провести уровень отсечки значений, претендующих на «звание» локального экстремума. Последнее требование типично при анализе поведения любой эмпирической статистики случайного процесса: она всегда изрезана, и почти каждое ее значение — точка локального экстремума. Поэтому необходимо исключить те локальные минимумы и максимумы, которые отвечают просто статистическому шуму.

Поскольку из свойств распределения стационарного горизонтного ряда следует выделенность последней полосы значений ширины τ , то на роль максимальных значений может претендовать не только последнее, равное $[2\tau / \varepsilon]$, но и все значения в промежутке $[2\tau / \varepsilon - \tau + 1; 2\tau / \varepsilon]$. К минимальным же значениям отнесем все те, которые меньше ε -квантили распределения горизонтного ряда. Итак, положим

$$\begin{aligned} H^* &= \min h_{\max} = [2\tau / \varepsilon] - \tau, \\ H_1 &= \max h_{\min} : \int_0^{H_1} \psi(h) dh = \varepsilon. \end{aligned} \quad (1.1)$$

где $\psi(h)$ есть выборочная плотность распределения горизонтного ряда.

Если оказалось, что $H_1 > H^*$ (такое возможно, если, например, все значения горизонтного ряда равны $[2\tau / \varepsilon]$ или, наоборот, нет максимальных значений), то анализ ряда не даст интересных результатов. В первом случае ряд сильно нестационарный и требует исключения временной зависимости, а во втором — напротив, сильно коррелированный, что также требует предобработки. Далее будем полагать, что $H_1 < H^*$.

Проведем прямые $h = H_1$ и $h = H^*$ через область значений горизонтного ряда и определим точки пересечения этих прямых с линией, соединяющей последовательные фактические значения горизонтного ряда, т.е. найдем множество точек пересечения прямых и горизонтной траектории. Эти точки разделят область определения $[1; 2\tau/\varepsilon]$ горизонтной траектории на отрезки, по каждому из которых можно найти глобальный минимум, если он относится к нижней части траектории, и глобальный максимум, если к верхней части. Части траектории, находящиеся выше и ниже определенной прямой, очевидно, перемежаются, так как возможная точка касания прямой и траектории автоматически относится к предыдущему промежутку.

При этом будем применять следующий алгоритм. Укажем первый промежуток по моменту появления экстремума горизонтной траектории, т.е. первый промежуток, содержащий значение, которое либо больше H^* , либо меньше H_1 . Пусть для определенности это будет максимум. Он принадлежит некоторой части траектории, находящейся выше прямой $h = H^*$. Если следующая часть траектории, находящаяся ниже прямой $h = H^*$, не пересекает прямую минимумов $h = H_1$, то эта следующая часть объединяется с предыдущей, и к этой объединенной части присоединяется и следующая часть, находящаяся по построению выше прямой $h = H^*$. В этом объединении трех участков траектории ищется максимум. Если одинаковых максимальных значений оказывается несколько, то за единственный максимум принимаем тот, у которого наибольший аргумент. Аналогично поступаем с нахождением локального минимума. В результате получаем совокупность чередующихся максимумов и минимумов траектории.

Для изучения текущей динамики нарастания или уменьшения доли хаоса в системе достаточно одной прямой, проведенной на уровне среднего значения горизонтного ряда. Алгоритм при этом несколько меняется, поскольку требуется исключить эффект статистического шума. Для этого исключения потребуем, чтобы величина размаха между соседними максимумом и минимумом была бы больше 2τ . Если размах оказался не превосходящим этой пороговой величины, то соответствующий промежуток объединяется с предыдущим.

Рассмотрим применение описанной методики на примере горизонтного ряда, отвечающего приращению валютного курса евро/доллар (рис.

6.4.3). Для этого ряда размах составил величину 127, линия максимумов проходит на уровне 390, а линия минимумов — на уровне $H_1 = 308$, чему отвечает 0,05-квантиль горизонтного распределения. Среднее значение горизонтного ряда оказалось равным $h_0 = 349$. Выяснилось, что для рассматриваемого ряда переход от хаоса к консолидации, характеризующийся расстоянием по времени между максимумом и соседним минимумом, меньшим, чем 308, требует от 7 до 100 дней, максимум распределения расстояний приходится на промежуток 20–25 дней. Обратный же процесс рассогласования идет от 4 до 40 дней с максимумом в промежутке 30–35 дней. Эти наблюдения относятся к долгосрочной динамике. Краткосрочное поведение описывается локальными экстремумами траектории относительно средней линии $h = h_0$. Это поведение также несимметрично. Чаще всего (с вероятностью 0,38) консолидация занимает промежуток от 3 до 5 дней, а хаотизация с вероятностью 0,44 происходит за 1–2 дня, хотя возрастание горизонтного ряда в течение 3–5 дней также имеет достаточно высокую вероятность, равную 0,33. Вероятность перехода от хаоса к порядку не более чем за 5 дней составила 0,50, а в обратную сторону за тот же период оказалась равной 0,77, что существенно выше. При этом средний скачок горизонтного ряда в обоих направлениях оказался равным одной и той же величине 69.

Разобранный пример показывает, как использовать горизонтный ряд в качестве индикатора разладки. Но чтобы использовать его наиболее эффективно, следует рассмотреть все возможные варианты для горизонта сдвига и задаваемой близости распределений. Для каждого варианта надо определить по фактическим данным ошибки первого и второго родов срабатывания индикатора и построить функцию распределения ошибки в зависимости от этих параметров. Это распределение квазистационарно на промежутке времени, которое выбирается в качестве величины сдвига в горизонтной статистике, поскольку за целочисленный промежуток време-

ни τ можно поставить $2^\tau - 1$ статистических экспериментов по получению данных для построения распределения ошибок, и потому возможное изменение используемого распределения можно сделать в достаточной степени малым, если собрать статистику за период $T \gg \tau$. Следовательно, выбрав из практических соображений функционал от этого стационарного распределения (мода, медиана, среднее значение и т.п.), определяем оптимальные параметры сдвига и точности для горизонтной статистики, с которыми и следует проводить вышеописанный анализ.

7.2. Статистическая добротность временного ряда

В параграфах 5.4 и 5.5 были введены статистики согласованного уровня стационарности и индекса нестационарности временного ряда. Эффективность введенных понятий как индикаторов нарастания хаоса зависит от объема выборки T . При малых объемах индикатор сам является весьма нестационарным, как это видно из рис. 5.4.11 и 5.5.2. При больших же объемах ряд функционала становится излишне стационарным. Следовательно, требуется оптимизировать объем выборки, по которому надо вычислять индекс нестационарности. Сформулируем подходящий критерий оптимизации.

При анализе нестационарного поведения ряда превышение определенного порогового значения $\rho_s(T)$ для расстояний между ВПФР является полезным сигналом. Этот сигнал проявляется на фоне стационарного шума, что хорошо видно на рис. 5.4.7. Очевидно, чем выше проведена отсечка ρ_s , тем меньше вероятность принять шум за полезный сигнал. С другой стороны, если провести уровень ρ_s слишком высоко, то будут ошибочно отвергнуты многие полезные сигналы. Таким образом, наряду с $\rho_s(T)$ надо рассмотреть еще две характеристики временного ряда функционалов: максимальное значение $\rho_{\max}(T)$ по выборке объема T и уровень $\rho_0(T)$, который может быть объяснен в рамках стационарной модели. Для ряда расстояний в норме $L1$ максимальное расстояние $\rho_{\max}(T)$ равно двум. Аналогичные статистики вводятся и для выборок, сдвинутых одна относительно другой на горизонт τ .

Важную информацию об анализируемой системе можно извлечь из анализа индикатора, который называется статистической добротностью временного ряда. Этот индикатор представляет собой статистический аналог отношения «сигнал-шум», т.е. отношение доли выборок, расстояние между которыми превосходит СУС, к доле выборок, расстояние между которыми не превосходит отклонений, объясняемых в рамках стационарной теории. Этот индикатор, обозначаемый далее как $q(T, \tau)$, является функцией длины выборки и величины сдвига и определяется формулой

$$q(T, \tau) = \frac{\rho^* T / (2\tau)}{F_{T, \tau}(2\varepsilon^* \tau / T)} = \frac{1 - F_{T, \tau}(\rho^*)}{F_{T, \tau}(2\varepsilon^* \tau / T)}. \quad (2.1)$$

Здесь величина $F_{T, \tau}(k)$ определена в (5.5.2) и представляет собой вероятность того, что расстояние ρ между выборками при сдвиге на τ не превосходит $2k/T$, ρ^* есть СУС выборок длины T при сдвиге на τ , а ε^* есть уровень неопределенности при стационарной оценке вероятностей.

Максимум по T функции добротности при фиксированном параметре сдвига τ мог бы дать оптимальную длину выборки для прогнозирования свойств системы на горизонте τ , если бы не одно «но»: если ряд сильно нестационарный, то его СУС очень большой, находящийся вблизи максимальных значений расстояний между выборками, а тогда доля событий, когда $\rho \geq \rho^*$, окажется весьма малой, и индикатор будет давать большую ошибку второго рода, т.е. редко срабатывать. Чтобы уменьшить долю несрабатываний, введем корректирующий множитель, равный доле меры носителя распределения $F_{T, \tau}(k)$, для которой расстояния превосходят СУС, так что мы будем далее использовать скорректированную статистическую добротность, определяемую формулой

$$Q(T, \tau) = q(T, \tau) \left(1 - \rho^* T / (2\tau)\right) = \frac{(1 - F_{T, \tau}(\rho^*)) \cdot F_{T, \tau}(\rho^*)}{F_{T, \tau}(2\varepsilon^* \tau / T)}. \quad (2.2)$$

Смысл индикатора (2.2) в том, что на промежутке $\rho \geq \rho^*$ должно быть достаточно много событий. Если индекс нестационарности $J(T, \tau)$, определяемый формулой (5.5.4), большой, но доля идентифицируемых событий мала, то индикатор (2.2) не обязательно будет иметь максимум на соответствующих значениях T и τ . Поэтому может оказаться, что оптимальным вариантом будет не тот, которому отвечает наибольший индекс нестационарности (5.5.4) или максимальное значение $q(T, \tau)$ в (2.1), а тот, где и ряд достаточно нестационарен, и доля превышения уровня нестационарности заметно отлична от нуля. Оптимальной длиной сканирования ряда и длиной сдвига являются такие пары T, τ , при которых $Q \rightarrow \max$. Само же по себе численное значение величины Q не играет роли. Оно может быть как больше, так и меньше единицы.

Исходный временной ряд может обладать квазипериодической разладкой, проявляющейся не только локально по времени, но и в среднем на

уровне СУС в немонотонном его поведении. Такие примеры для встык-выборок были приведены на рис. 5.4.10 и 5.4.11. Но зависимость СУС от для параметра сдвига τ также обладает локальными экстремумами, хотя в целом расстояние между двумя ВПФР с увеличением длины выборки при фиксированном окне сдвига уменьшается, а с увеличением окна сдвига при фиксированной длине выборки растет.

Большой практический интерес представляет возможность идентифицировать разладку на меньших, чем «встык-выборки», длинах. Как уже говорилось, желательно построить индикатор, имеющий минимальное время запаздывания по отношению к моменту, который постфактум будет опознан как начало переходного процесса разладки. В то же время ошибочная сигнализация о разладке в отсутствие таковой не должна быть слишком частой. Например, первый максимум расстояния между встык-выборками мог бы служить индикатором разладки, причем он приходится на относительно небольшие длины выборки порядка 1–5 тыс., но этот максимум оказывается статистически незначимым, поскольку находится ниже кривой уровня стационарного шума.

На рис. 7.2.1 приведены линии уровня индикатора добротности (2.2) в зависимости от длин выборки и сдвига для тикового ряда CL, который был рассмотрен в параграфах 5.4–5.5. Величина сдвига в расчетах не превосходила длину выборки. Оказалось, что для него существует область параметров «длина выборки — окно сдвига», где индикатор добротности имеет локальный максимум.

Графики на рис. 7.2.1 начинаются с добротности, отвечающей встык-выборке, т.е. ситуации, когда параметр τ из легенды равен длине выборки, отложенной по оси абсцисс. Далее параметр сдвига для каждой кривой остается тем же самым, а длина выборки возрастает. На графиках хорошо видна характерная особенность индикатора добротности (2.2): для ряда CL оказалось, что второй локальный максимум всегда выше первого. Это означает, что встык-выборки не являются наиболее оптимальным инструментом, а лучше всего разладку искать на выборках, имеющих общую часть длиной 1–2 тыс. тиков, что составляет 5–10 % от длины выборки. Наилучшим представляется вариант, когда выборки длин 19,5–20,5 тыс. сдвинуты на 19 тыс. тиков (сплошная черная линия). В этом случае локальный максимум наиболее широк, что свидетельствует о его устойчивости в пределах примерно 1,2 тыс. тиков. На выборках других длин и(или) с другой величиной сдвига максимум добротности либо заметно меньше, либо весьма острый, что не годится в силу неустойчивости такого индикатора. Аналогично выглядят графики добротности и для ряда GC.

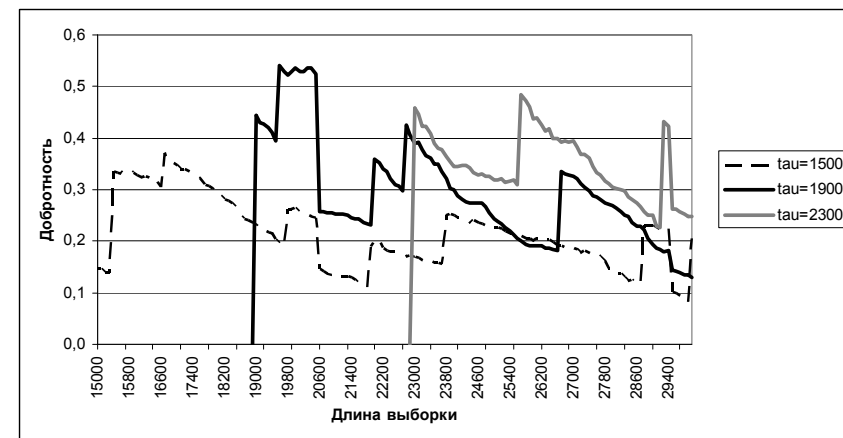


Рис. 7.2.1 — К определению оптимальных длин сканирования, ряд CL

Заметим, что индикатор добротности может и не привести к нахождению оптимальной длины сканирования выборки, если нестационарность ряда проявляется на слишком больших длинах, поскольку в таких случаях может не быть локального максимума добротности, а будет наблюдаться его монотонный рост с увеличением длины выборки и длины сдвига. В таких случаях для выбора оптимального окна рекомендуется использовать локальные максимумы СУС «встык-выборок».

Подчеркнем, что построенный индикатор не универсален в плане эффективности своего применения. В большинстве случаев временные ряды, встречающиеся на практике, имеют малые выборки, которые не позволяют построить распределение вероятностей с точностью, различающей нестационарное поведение на фоне шума. Малость выборки в нашем понимании — это длины менее 10 тыс. данных. Лишь тиковые биржевые ряды представляют интерес в плане анализа разладки в терминах ВПФР. Минутные отсчеты или, «хуже» того, часовые не особенно интересны, поскольку для построения выборки большого объема надо погрузиться в неактуальное прошлое. Для таких рядов более эффективен горизонтный анализ, рассмотренный в главе VI.

Но даже для выборок надлежащих длин можно довольно часто обнаружить лишь тривиальный эффект: на малых длинах распределение неотличимо от стационарного в силу грубости оценки эмпирических вероятностей, а на больших длинах индекс нестационарности и статистическая

добротность монотонно возрастают. Этот эффект связан с тем, что СУС многих нестационарных рядов перестает убывать после определенного объема данных, тогда как СУС стационарных рядов (т.е. согласованная значимость) монотонно убывает, поэтому отношение «сигнал-шум» в нестационарном случае может неограниченно возрастать с увеличением выборки. Интерес представляет именно локальный максимум индикатора добротности, который к тому же имеет еще и устойчивое поведение, т.е. достаточно пологую форму.

7.3. Относительная энтропия и другие расстояния между ВПФР

Рассмотрим индикативные свойства других расстояний между выборочными распределениями. Кроме расстояний в традиционно используемых функциональных пространствах $(C, L1, L2)$, для сравнения близости плотностей распределений применяются и другие неотрицательные функционалы. В отличие от первых, эффективность применения вторых зависит от конкретно решаемой задачи.

В качестве индикатора изменчивости ВПФР в задачах установления стационарного распределения или в задачах распознавания текущего распределения $f(x, t)$ как элемента некоторого кластера с известным эталонным распределением $g(x)$, часто используется информационное расстояние между распределениями, называемое относительной энтропией или расстоянием Кульбака–Лэйблера:

$$\rho_{KL}(f, g) = \int f(x) \ln(f(x)/g(x)) dx. \quad (3.1)$$

В дискретной форме это расстояние определяется как суммирование по разбиению гистограммы, общему для обоих распределений.

Из примеров п. 5.4 по вычислению ряда расстояний в нормах $C, L1$ и $K-L$ следует, что расстояние $\rho_{KL}(f, g)$ наименьшее из них. Это связано с тем, что, будучи неотрицательным и не содержащим в своем определении функции модуля, для близких ε -стационарных распределений оно имеет порядок не ε , а ε^2 .

Пусть, в частности, f и g — плотности нормальных распределений с одинаковыми средними и с различающимися дисперсиями α^2 и β^2 . Тогда из (3.1) получаем, что расстояние Кульбака–Лэйблера определяется отношением дисперсий:

$$\rho_{KL}(f, g) = \frac{1}{2}(\gamma - 1 - \ln \gamma), \quad \gamma = \frac{\alpha^2}{\beta^2}. \quad (3.2)$$

Из (3.2) следует, что если дисперсии близки, т.е. $\gamma = 1 \pm \varepsilon$, то $\rho_{KL} = \varepsilon^2/4 + o(\varepsilon^2)$.

Такое же по виду выражение для расстояния ρ_{KL} получается не только для экспоненциально спадающих распределений, но и для распределений степенного вида. Для распределений Парето с плотностями

$$f(x) = \frac{p}{x^{p+1}}, \quad g(x) = \frac{q}{x^{q+1}}, \quad p, q > 1, \quad x \geq 1$$

также получаем, что

$$\rho_{KL}(f, g) = \gamma - 1 - \ln \gamma, \quad \gamma = \frac{q}{p}. \quad (3.3)$$

Тогда при близких значениях параметров p и q расстояние между этими распределениями имеет второй порядок малости.

Аналогичный вывод можно сделать и для плотностей, представленных в виде гистограммы. Пусть, например, имеется всего два состояния системы, так что вероятности этих состояний равны p и $q = 1 - p$. При оценке этих вероятностей по другой выборке получилось ε -стационарное распределение, так что новые вероятности оказались равными $p' = p(1 - \varepsilon)$ и $q' = 1 - p'$. Тогда

$$\rho_{KL}(f, g) = -p \ln(1 - \varepsilon) - q \ln\left(1 + \frac{p}{q} \varepsilon\right). \quad (3.4)$$

Если вероятность q не очень мала, так что $p\varepsilon/q < 1$, то расстояние (3.4) равно $\rho_{KL}(f, g) = \frac{p}{2q} \varepsilon^2 + o(\varepsilon^2)$, а если $q \rightarrow 0$, то из (3.4) по-

лучаем, что $\rho_{KL}(f, g) \rightarrow p\varepsilon^2/2$.

Причиной, по которой расстояние Кульбака–Лэйблера не особенно эффективно при анализе эволюции нестационарных ВПФР, является неизвестность характерного уровня, который отвечает стационарному распределению: этот уровень не априорный, а свой для каждого распределения. Поэтому распределение расстояний ρ_{KL} для набора выборок из

нестационарного временного ряда не позволяет провести устойчивую линию стационарно объясняемых отклонений.

К квазирасстоянию (3.1) близко по смыслу так называемое расстояние χ^2 , определяемое в дискретной форме при ненулевых значениях ПФР формулой

$$\rho(f, g) = \sum_j \frac{(f(j) - g(j))^2}{f(j)}. \quad (3.5)$$

Как и расстояние Кульбака–Лэйблера, выражение (3.5) не симметрично относительно перестановки входящих в него функций. Тем не менее, оно используется при оценке неизвестного значения параметра распределения, в предположении существования генерального распределения, зависящего от исследуемого параметра, когда оценка параметра проводится по критерию χ^2 .

Близость гистограмм можно оценить также по расстоянию Хеллингера, которое симметрично и определяется формулой

$$\rho_H(f, g) = \sum_j (\sqrt{f(j)} - \sqrt{g(j)})^2 = 2 - 2 \sum_j \sqrt{f(j)g(j)}. \quad (3.6)$$

Для \mathcal{E} -стационарных распределений оно имеет порядок \mathcal{E}^2 .

Сравним эффективность рассмотренных расстояний на примере решения задачи об определении принадлежности текущего распределения одному из заданных кластеров, сформированных близкими в соответствующих нормах распределениями.

Формализуем задачу идентификации текущей ВПФР. Она состоит в следующем. Имеется библиотека эталонов, представленных в виде ПФР для A типов поведения нестационарного временного ряда. Пусть K_a — имеющееся количество фрагментов a -го типа, и $N_{i,a}$ — объем данных в i -ом фрагменте этого типа, $i = 1, 2, \dots, K_a$. Обозначим $f_{i,a}(j)$ — ВПФР соответствующего фрагмента, где аргумент j отвечает нумерации промежутка гистограммы. Каждому типу квазистационарности поставим в соответствие средневзвешенную ПФР:

$$\bar{f}_a(j) = \frac{1}{N_a} \sum_{i=1}^{K_a} f_{i,a}(j) N_{i,a}, \quad N_a = \sum_{i=1}^{K_a} N_{i,a}. \quad (3.7)$$

Пусть имеется текущий фрагмент «0» неизвестного типа, который надо идентифицировать внутри данной библиотеки. Типом фрагмента «0» считается тот из типов « a », для которого норма $\rho_a^0 = \|f_0 - \bar{f}_a\|$ разности между ВПФР $f_0(j)$ фрагмента «0» и эталонной ПФР $\bar{f}_a(j)$ минимальна:

$$\rho_a^0 = \|f_0 - \bar{f}_a\|, \quad a^0 = \arg \min_a \rho_a^0. \quad (3.8)$$

Процедура отбора элементов в кластер здесь не обсуждается, считается, что кластеры уже есть. Применительно к данной задаче эта процедура — экспертная, т.е. «на глаз» выделяются интересные с точки зрения исследователя фрагменты траектории временного ряда приблизительно равных длин по 20–25 тыс. тиков и строится их средневзвешенная ПФР. Можно было бы строить кластеры по принципу попарной близости ВПФР фрагментов в выбранной норме, но тогда разным нормам могут отвечать содержательно разные кластеры. Эффективность норм с точки зрения кластеризации ВПФР анализируется далее в главе X части III, п. 10.4.

При экспертном отборе фрагментов временного ряда естественно возникают три кластера: с общим трендом вверх (up-кластер), с трендом вниз (down-кластер) и боковик (flat-кластер). Рассмотрим три таких эталона, сформированные выборками из ряда приростов дистинктивных тиков для фьючерсов CL. Они показаны на рис. 7.3.1.

Из рис. 7.3.1 видны характерные отличия между эталонами: up-кластер имеет эталон с превышением правого плеча (доля положительных приростов в один «пипис») над левым (доля отрицательных приростов в один «пипис»), down-кластер имеет превышение левого плеча над правым, а flat имеет эталон с примерно равными плечами.

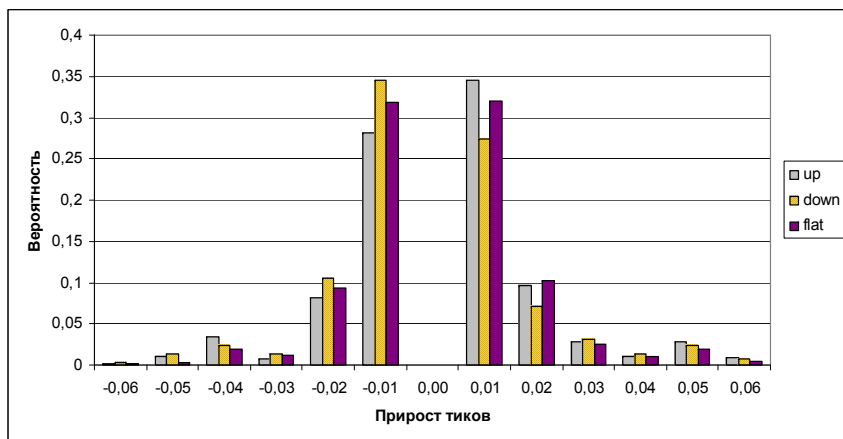


Рис. 7.3.1 — Эталонные ПФР кластеров для приростов тиков CL

Вошедшие в кластеры фрагменты имеют следующие статистические свойства своих ВПФР, представленные в табл. 7.3.1.

Табл. 7.3.1. Статистические свойства up-down-flat-кластеров

Расстояние	$L1$	$K-L$	Хеллинггер	χ^2
Между эталонами up-down	0,20	0,021	0,015	0,046
Между эталонами up-flat	0,15	0,027	0,021	0,041
Между эталонами down-flat	0,16	0,033	0,020	0,049
Среднее $\bar{\rho}$ между ВПФР своего кластера	0,08	0,012	0,013	0,033
Среднее $\bar{\rho}$ между ВПФР чужих кластеров	0,25	0,035	0,030	0,063
$\sigma_{\rho} / \bar{\rho}$ для своих кластеров	0,09	0,10	0,14	0,23
$\sigma_{\rho} / \bar{\rho}$ для чужих кластеров	0,15	0,14	0,15	0,26
Индекс разделения «свой-чужой»	1,2	1,0	0,9	0,8

Из табл. 3.7.1 видно, что наиболее узкое распределение по отношению $\sigma_{\rho} / \bar{\rho}$ имеют расстояния в $L1$. Затем идет расстояние $K-L$, а расстояние Хеллингера и χ^2 имеют заметно худшие качества.

Индексом разделения кластеров будем называть величину, равную отношению разности средних значений расстояний соответственно между чужими и своими элементами кластеров к сумме соответствующих среднеквадратичных отклонений:

$$r = \frac{\bar{\rho}_{out} - \bar{\rho}_{in}}{\sigma_{out} + \sigma_{in}}. \quad (3.9)$$

Величина (3.9) показывает, насколько велика вероятность ошибки идентификации: если $r < 1$, то оба распределения имеют заметное перекрытие графиков, и вероятность ошибки первого и второго родов велика; если же $r > 1$, то перекрытие графиков происходит только в «хвостах» распределений, и вероятность ошибки мала. Сам по себе индекс разделения не является точной характеристикой вероятности ошибки. Он показывает лишь, какой вариант расчетов предпочтительнее, но не позволяет численно оценить вероятность ошибки идентификации. Это — ранжирующий индекс.

В данном случае в норме $L1$ любые два элемента из разных кластеров с вероятностью 0,95 находятся на расстоянии большем, чем расстояние между каждым из этих элементов и любым элементом «его» кластера. Аналогичные вероятности для других норм равны: KL 0,85, Хеллинггер 0,75, χ^2 0,70. Таким образом, вероятности ошибки в разных нормах расположились в соответствии с индексом разделения. Наилучшей нормой оказалась норма в $L1$.

Таким образом, характерные расстояния между ВПФР в разных нормах можно использовать для построения индикаторов тренда. Следует только учитывать, что длина выборки сканирования должна быть примерно в два раза меньше, чем выборка среднего фрагмента, участвовавшего в построении эталона. Это требуется из тех практических соображений, что идентификация изменения распределения должна быть сделана в процессе разладки, а не тогда, когда она по сути закончилась.

7.4. Оптимальный дефлятор коинтеграции

Индикаторы нестационарного поведения временного ряда могут быть построены на основе функционала разности не только между ВПФР само-

го этого ряда или его первых (вторых и т.п.) разностей, как в разделах 7.1–7.3, но и между ВПФР решений некоторой задачи оптимизации, т.е. может анализироваться ряд, вторичный по отношению к исходному. Рассмотрим следующий практический пример.

На рынке энергоресурсов часто возникает задача определения влияния средневзвешенной за некоторый период времени (неделя, месяц и т.п.) цены одного вида топлива на цену другого, например, нефти на природный газ. Оба ряда цен нестационарны, но может оказаться, что между ними устанавливается линейная зависимость с коэффициентами, постоянными на определенном временном промежутке. Иными словами, регрессия одного нестационарного ряда на другой может оказаться квазистационарной. Подробно регрессионные модели рассматриваются в третьей части книги в главе VIII. Здесь будут обсуждаться некоторые следствия из модели регрессии безотносительно к корректности самой такой модели.

Два временных ряда называют коинтегрированными на определенном уровне достоверности, если с этой достоверностью между ними установлена регрессионная связь.

Отметим, что выборки для построения регрессии могут быть взяты с лагом одна относительно другой. Лаг может быть сравнительным индикатором нестационарности между разными коинтегрированными парами рядов. Пусть для одной пары рядов на выборке заданной длины T в момент времени t наилучшая коинтеграция (т.е. имеющая наименьшую ошибку аппроксимации) достигается на лаге $l_T(t)$. По имеющимся историческим данным строится плотность распределения оптимального лага $p_T(l)$, которая затем исследуется на стационарность. Если это распределение стационарно с заданной точностью, то можно выбрать в качестве оптимального лага среднее значение распределения. Построив затем аналогичное распределение для пары других рядов, из сравнения двух таких распределений можно сделать вывод о большей или меньшей нестационарности коинтеграционной связи.

Во многих практических случаях вид коинтеграционной связи имеют контрактные формулы на природный газ, цена на который часто привязывается к текущей или запаздывающей цене нефти или нефтепродуктов. С течением времени коэффициенты в этих формулах приходится менять, т.е. пересматривать условия контрактов. Причины этого изменения могут быть весьма различны. Нестационарность коинтеграционной связи может отражать реальное изменение спроса у конечного потребителя, может быть обусловленной разным уровнем отраслевой инфляции, на нее могут влиять экономические интересы поставщика и оптового покупателя, по-

литические и иные аспекты. С точки зрения внешнего наблюдателя нестационарность приводит к снижению уровня достоверности для коинтеграции двух рядов. Тогда возникает вопрос о корректном сравнении цен и построении индикатора, отражающего изменения в инфляционных процессах.

При вычислении скользящего коэффициента корреляции между двумя рядами цен за достаточно большой промежуток времени (физического, не «шагового») необходимо отметить, что сравнение цен сегодня и два или три года назад как цен одинакового качества из-за инфляции не вполне корректно. Можно предположить, что предобработка сравниваемых рядов путем введения дефлятора повысит точность анализа. Опишем метод введения соответствующего скользящего фактора.

Будем говорить, что временной ряд $X_q(i; j)$ является q -преобразованием ряда $x(i)$ в точке j , если задана некоторая последовательность $q(k)$, $k = 0, 1, 2, \dots$, такая, что $q(0) = 1$ и значения исходного ряда $x(i)$ в предыдущие моменты времени по отношению к моменту j умножаются на $q(j-i)$:

$$X_q(i; j) = x(i)q(j-i), \quad i \leq j. \quad (4.1)$$

В частности, если положить $q(k) = q^k$, где $q \in R$, то такое преобразование будем называть q -преобразованием ряда величин $x(i)$:

$$X_q(i; j) = x(i)q^{j-i}, \quad i \leq j. \quad (4.2)$$

Пусть теперь имеются два ряда: $x(i)$ и $y(i)$. Зададим некоторый момент времени j и рассмотрим объем выборки n , состоящий из элементов каждого из этих рядов в моменты $j, j-1, \dots, j-n+1$. Производящей функцией $Z_n^x(q; j)$ ряда $x(i)$ на выборке длины n в точке j называется частичная сумма q -деформированного ряда:

$$Z_n^x(q; j) = \sum_{k=0}^{n-1} x(j-k)q^k. \quad (4.3)$$

Скользящее среднее деформированного ряда по выборке длины n равно

$$\langle X_q(i; j) \rangle_n = \frac{Z_n^x(q; j)}{n}. \quad (4.4)$$

Построим одновременную скользящую регрессию между модифицированными рядами $X_q(i; j)$ и $Y_q(i; j)$ (модель регрессии с лагом строится аналогично). Для этого введем производящую функцию q -деформированного произведения

$$Z_n^{xy}(q; j) = \sum_{k=0}^{n-1} x(j-k)y(j-k)q^{2k}. \quad (4.5)$$

Тогда аппроксимация величин $Y_q(i; j)$ на рассматриваемой выборке дается формулой

$$Y_q^{app}(i; j) = \frac{1}{n} Z_n^y(q; j) + a \cdot \left(X_q(i; j) - \frac{1}{n} Z_n^x(q; j) \right), \quad (4.6)$$

$$a = \frac{n Z_n^{xy}(q; j) - Z_n^x(q; j) Z_n^y(q; j)}{n Z_n^{xx}(q; j) - (Z_n^x(q; j))^2}.$$

Прогноз на 1 шаг вперед строится теперь путем замены деформированного значения $X_q(i; j)$ в правой части (4.6) на значение ряда $x(j+1)$:

$$\hat{y}(j+1) = \frac{1}{n} Z_n^y(q; j) + a \cdot \left(x(j+1) - \frac{1}{n} Z_n^x(q; j) \right). \quad (4.7)$$

Интерпретируем величину q в (4.7) как шаговый дефлятор. Например, если данные идут с шагом в месяц, то в предположении, что инфляция составила p % в год, получаем $q = (1 + p/100)^{1/12}$. Можно считать инфляцию переменной, и тогда надо будет использовать общее определение (4.1) модифицированного ряда.

Величину q , при которой прогноз по формуле (4.7) имеет наименьшую ошибку на заданном горизонте, естественно назвать оптимальным дефлятором модели. Как и для лагов, в каждый момент времени t определяется оптимальный дефлятор $q(t)$ — свой для каждого лага и длины выборки. По имеющимся данным строится распределение $D(q)$ и исследуется на стационарность. Заметное изменение распределения, т.е. раз-

ладка системы в терминах оптимального дефлятора, служит указанием на необходимость пересмотра коинтеграционной связи.

7.5. Плотность распределения показателя Херста

В анализе нелинейных хаотических динамических систем в качестве одного из индикаторов меры персистентности часто используют показатель Херста [26]. Он является асимптотическим индикатором, поскольку выделяет главную степенную часть нормированного размаха изучаемой динамической системы в зависимости от времени. В связи с тем, что методы анализа хаотических динамических систем активно применяются на практике для того, чтобы аппроксимировать наблюдаемый случайный процесс, например, биржевой ряд цен на какой-либо финансовый инструмент, подходящей динамической системой, показатель Херста применяется и к таким временным рядам. Даже если прогноз собственно ряда по динамической модели оказывается недостаточно точным, системные свойства ряда, такие как горизонт памяти, оцениваемый по показателям Ляпунова, величина квазипериода, уровень нестационарности могут быть оценены вполне адекватно, что представляет важность для практикующего аналитика.

Чаще всего используются две традиционных взаимно-дополняющих трактовки смысла показателя Херста H применительно к временным рядам. Первая состоит в том, что если $H > 1/2$, то ряд демонстрирует свойство преимущественного сохранения тренда (персистентность), а при $H < 1/2$ напротив, его поведение антиперсистентно. Вторая трактовка заключается в том, что, чем больше показатель Херста, тем меньшая длина выборки требуется для того, чтобы вычислять тот или иной скользящий индикатор (среднее значение ряда, дисперсию и т.п.).

Существует также самостоятельная область математики, развитая «вокруг» показателя Херста. Это модели самоподобных случайных процессов и фрактального блуждания [20]. Несмотря на отсутствие таких процессов в реальности, практическая ценность этой науки в том, что если реальный временной ряд удастся с достаточной точностью приблизить самоподобным процессом, то строго доказанные утверждения о свойствах последнего переносятся и на первый.

Насколько же точно можно сопоставить временному ряду (чаще всего нестационарному) постоянное число — показатель Херста? Анализу такого сопоставления посвящено большое число публикаций, суть которых весьма противоречива. В одних работах демонстрируется ряд показателей Херста, вычисленных по определенному объему выборки в разные мо-

менты времени, и размах этого ряда оказывается весьма широк — от 0,2 до 0,9. В других показано, как показатель Херста меняется в данный момент времени в зависимости от объема выборки. В третьих работах он вообще считается постоянным.

Следует сказать, что локальный по времени, т.е. вычисленный в скользящем окне фиксированной длины показатель Херста не является настолько надежным индикатором поведения ряда, чтобы строить на нем, например, торговую стратегию на финансовом рынке. Заметим, что ни в одной из работ, ратующих за показатель Херста как индикатор персистентности, не приведены оценки условных вероятностей его позитивного срабатывания как предиктора тренда или перелома, а они оказываются весьма незначительными, причем зависящими от длины выборки.

В чем же причины практической неэффективности этого индикатора? Для ответа на этот вопрос следует провести анализ временных рядов из показателей Херста $H_N(n)$ как статистик, зависящих от длины выборки N , отнесенных к текущему номеру события n . Качественные соображения об ограниченной применимости этого показателя основаны на том, что персистентность ряда оценивается постфактум с окном запаздывания N в соответствии с длиной выборки, и потому она не должна, вообще говоря, выполняться в последующие моменты времени. Более того, скорее всего, наблюдаемая в прошлом тенденция не будет выполняться в ближайшем будущем, ибо в силу асимптотического характера показателя Херста его значения корректны лишь при весьма больших выборках, а тогда тренд с высокой вероятностью остался в прошлом. Для более строгих выводов следует изучить распределения рядов из показателей Херста с целью выяснения того, насколько эти ряды являются стационарными. Именно, надо провести исследование эволюции выборочных распределений показателя Херста для различных биржевых рядов, взятых по ценам закрытия в минутном диапазоне: валютных, сырьевых фьючерсов, индексов, акций, а также и для тиковых рядов.

В результате обнаружилось, что, начиная с выборок весьма скромных объемов в 300 точек, выборочные распределения показателей Херста для всех рассмотренных временных рядов оказались стационарными с той точностью, с которой можно говорить об оценке эмпирических вероятностей по выборке конечной длины. Эти стационарные распределения с высокой точностью определились как нормальные. Следует подчеркнуть, что сами ценовые ряды были при этом существенно нестационарными. Для разных инструментов параметры нормального распределения показателей Херста оказались близки в одинаковых тайм-фреймах (минутки, 5-

минутки и т.п.), причем они характеризуются практически не зависящей от длины выборки дисперсией. Дисперсия уменьшается с увеличением тайм-фрейма, а среднее значение распределения смещается с увеличением длины выборки в область значений 0,50–0,55. Этот факт может служить объяснением того, почему на самом деле показатель Херста не является надежным индикатором текущего состояния временного ряда. Дело в том, что его распределение стационарно на больших выборках, и потому не имеет отношения к текущей реальности, а на малых выборках достоверность оценки самого показателя Херста мала.

Итак, рассматривается одномерный временной ряд $x(n)$, где номер n некоторого элементарного события совпадает с временем наступления события, если ряд эквидистантный, либо отвечает случайному моменту времени, если маркированный. В качестве ряда $x(n)$ обычно берется ряд первых разностей изучаемого процесса

$$x(n) = c(n+1) - c(n) \quad (5.1)$$

или, как вариант, ряд логарифмических приростов $x(n) = \ln(c(n+1)/c(n))$.

Показатель Херста для ряда $c(n)$ определяется следующим образом [26]. Вводится скользящее среднее приростов $x(n)$ на шаге n по выборке длины k

$$\bar{x}(n, k) = \frac{1}{k} \sum_{i=n-k+1}^n x(i) \quad (5.2)$$

и вычисляется накопленное отклонение от среднего для выборки длины k :

$$R(n, k) = \max_{j \leq n} \left(\sum_{i=n-k+1}^j (x(i) - \bar{x}(n, k)) \right) - \min_{j \leq n} \left(\sum_{i=n-k+1}^j (x(i) - \bar{x}(n, k)) \right). \quad (5.3)$$

Вычисляется также скользящая дисперсия рассматриваемого временного ряда по выборке длины k :

$$\sigma_x^2(n, k) = \frac{1}{k} \sum_{i=n-k+1}^n (x(i) - \bar{x}(n, k))^2, \quad (5.4)$$

после чего вычисляется логарифм отношения размаха к шуму:

$$\xi(n, k) = \ln \left(\frac{R(n, k)}{\sigma_x(n, k)} \right). \quad (5.5)$$

Фиксируется некоторая минимальная длина выборки k_0 и максимальная длина N , которая отвечает временному горизонту анализа. Описанная процедура (5.1–5.5) делается для выборок длин k таких, что $1 < k_0 \leq k \leq N$. Показатель Херста $H_N(n)$ по выборке длины N на шаге n определяется как коэффициент регрессии величины $\xi(n, k)$ на логарифм длины выборки:

$$\xi(n, k) - \bar{\xi}_N(n) = H_N(n) \cdot \left(\ln k - \frac{\ln(N!/(k_0 - 1)!)}{N - k_0 + 1} \right) + \varepsilon(k), \quad (5.6)$$

$$\bar{\xi}_N(n) = \frac{1}{N - k_0 + 1} \sum_{k=k_0}^N \xi(n, k),$$

где $\varepsilon(k)$ представляет собой остаток регрессии. Угол наклона прямой регрессии и равен показателю Херста ряда. На практике следует выбрать минимальную длину выборки достаточно большой $1 \ll k_0 \ll N$, чтобы детерминация показателя Херста как параметра регрессии была высокой. Если в формуле (5.6) использовать формулу Стирлинга для асимптотической аппроксимации факториала, то для показателя Херста получаем компактное выражение

$$H_N(n) = \frac{1}{N} \sum_{k=1}^N (\xi(n, k) - \bar{\xi}_N(n)) (1 + \ln(k/N)). \quad (5.7)$$

Представление (5.6–5.7) будем использовать для численного нахождения рядов показателей Херста в зависимости от длины выборки N и текущего момента времени n . Пример ряда показателя Херста $H_N(n)$ для тикового ряда CL в терминах номеров событий, вычисленный по формулам (5.1–5.7) в окне $N = 20$ тыс. тиков, показан на рис. 7.5.1. На этом графике отсчет начинается с первого значения после выборки в 20 тыс. тиков.

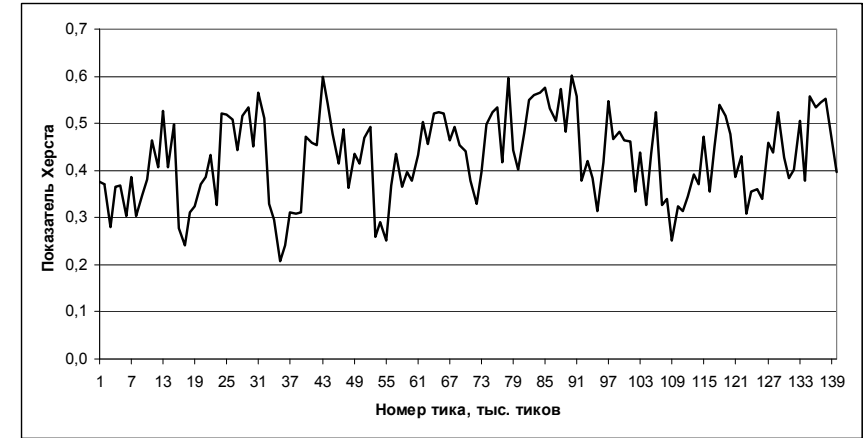


Рис. 7.5.1 — Динамика показателя Херста $H_N(n)$ для $N = 20$ тыс. тиков

Из рис. 7.5.1 видно, что в разные моменты времени, трактуемые как номера событий, показатели Херста существенно отличаются. В общем-то, для нестационарного тикового ряда в этом нет ничего удивительного. Интересно, однако, то, что временной ряд показателей Херста оказывается стационарным в узком смысле, начиная с выборок длины 300 точек, т.е. две встык-выборки показателей Херста длиной 300 отличаются одна от другой менее, чем на удвоенную точность оценки эмпирических частот соответствующего распределения. С увеличением окна N для вычисления $H_N(n)$ этот ряд становится все более стационарным. Более того, если рассмотреть ряд показателей Херста длиной более 1 тыс. точек, то с высокой точностью (на уровне достоверности более 0,99) его распределение оказывается нормальным. Центр этого распределения $\bar{H}(N)$ приблизительно линейно смещается вправо с увеличением окна N от 0,41 для $N = 10$ тыс. тиков до 0,53 для $N = 100$ тыс. тиков и далее не меняет своего значения. Стандартное отклонение практически не зависит от N и составляет 0,09. Таким образом, это распределение аппроксимируется нормальным:

$$f_N(H) = \frac{1}{\sqrt{2\pi}\sigma_H} \exp \left(-\frac{(H - \bar{H}(N))^2}{2\sigma_H^2} \right), \quad \sigma_H = 0,09. \quad (5.8)$$

Такой же вывод относительно показателя Херста оказывается возможным сделать и для других биржевых рядов. Его распределения стационарны и нормальны. Примеры распределений показателей Херста для разных инструментов приведены на рис. 7.5.2. При этом окно для тикового ряда взято равным $N = 20$ тыс. тиков, а для минутных цен закрытия — равным $N = 500$ минут, чтобы сравнить ряды в одном временном диапазоне. Подчеркнем, что на этих длинах распределение показателя Херста наименее стационарно, с увеличением длины окна стационарность распределения возрастает.

Из рис. 7.5.2 видно, что распределения для разных инструментов очень похожи, а численные оценки показывают, что распределения близки и в смысле близости их ПФР. Для выборок данных длин СУС этих инструментов разный: 0,12 для тиков WTI, 0,40 для минутных цен Brent, 0,16 для индекса DAX, и 0,44 для курса валюты. Точность же оценки их эмпирических частот во всех случаях одинаковая, равная 0,05. Стандартные отклонения распределений параметра Херста для минутных тайм-фреймов разных инструментов примерно одинаковы и равны 0,08, причем с увеличением промежутков агрегирования исходного ряда они уменьшаются. Для тикового ряда стандартное отклонение наибольшее, равное 0,09.

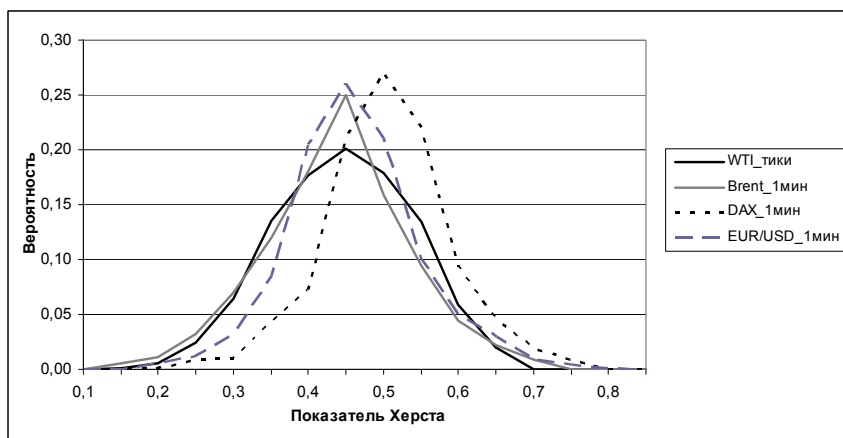


Рис. 7.5.2 — Распределения показателя Херста для разных инструментов

Таким образом, обнаружено, что показатель Херста в определенном смысле не является индикатором персистентности: его распределение отвечает стационарному гауссовому белому шуму. Поэтому практический смысл имеет среднее значение по этому распределению, а оно оказывается близким к 0,5. И хотя в отдельные моменты времени могут наблюдаться весьма разные значения показателя Херста, они не имеют того смысла, который ему традиционно приписывается. Более корректной представляется та точка зрения, что исходный ряд характеризуется не показателем Херста, а распределением этого показателя.

Этот же вывод относится ко всем вообще индикаторам временных рядов. Практическая применимость индикаторов определяется в первую очередь плотностью их распределения, и лишь затем фактическими их значениями в зависимости от того, стационарно это распределение, или нет. Если распределение индикатора стационарно, этот индикатор относится не к текущей ситуации, а отражает общие свойства ряда, которые могут быть использованы для отбора квазистационарных параметров, таких, например, как длина выборки, которые участвуют в построении индикаторов первого уровня, т.е. индикаторов разладки. Распределение же самих индикаторов разладки для нестационарного ряда также должно быть нестационарным. В противном случае относительно текущей ситуации они ничего существенного не показывают.

ЧАСТЬ III. КИНЕТИЧЕСКИЕ УРАВНЕНИЯ И МОДЕЛИ ВРЕМЕННЫХ РЯДОВ

Глава VIII. Регрессионные модели

8.1. Стационарная линейная регрессия

Наиболее часто используемой моделью временного ряда является регрессионная, как правило, линейная модель, поэтому ее свойства следует разобрать подробно.

Пусть в результате наблюдений за случайными величинами ξ и η были получены ряды значений x_1, \dots, x_n и y_1, \dots, y_n соответственно. Пусть также $\rho(y, y')$ — функция расстояния в пространстве Y , которому принадлежат значения случайной величины η . Пространство значений случайной величины ξ обозначим через X .

Функцией теоретической регрессии случайной величины ξ на случайную величину η называется такая функция $\varphi: X \rightarrow Y$, для которой математическое ожидание расстояния $M\rho(y, \varphi(x))$ минимально. Переменная x называется регрессором или объясняющей переменной, а переменная y — откликом или объясняемой переменной.

Если X и Y — линейные пространства с евклидовой нормой, то метод нахождения функции регрессии называется методом наименьших квадратов (МНК). Если при этом рассматривается линейная зависимость между величинами ξ и η , то такая модель называется линейной регрессионной моделью (ЛРМ).

Чтобы отличать значения случайной величины η , полученные в результате аппроксимации $\varphi(x)$, от фактических, будем снабжать их тильдой (\tilde{y}), а прогнозные значения — шляпкой (\hat{y}).

Пусть значения x и y случайных величин ξ и η предположительно связаны линейным соотношением, и пусть N — длина анализируемой

выборки. Найдем такую прямую $\tilde{y}_n = a + bx_n$, которая бы минимально в среднем квадратичном отклонялась бы от фактических значений y_n , т.е. доставляла бы минимум функционалу

$$\rho(\tilde{y}, y) = \sum_{n=1}^N (\tilde{y}_n - y_n)^2 = \sum_{n=1}^N (a + bx_n - y_n)^2 \rightarrow \min. \quad (1.1)$$

Условия экстремальности задачи (1.1) имеют вид:

$$\begin{aligned} \frac{\partial \rho}{\partial a} &= 2 \sum_{n=1}^N (a + bx_n - y_n) = 0, \\ \frac{\partial \rho}{\partial b} &= 2 \sum_{n=1}^N x_n (a + bx_n - y_n) = 0 \end{aligned} \quad (1.2)$$

Обозначая

$$\begin{aligned} \bar{x} &= \frac{1}{N} \sum_{n=1}^N x_k, \quad \bar{y} = \frac{1}{N} \sum_{n=1}^N y_k, \quad \sigma_x^2 = \frac{1}{N} \sum_{n=1}^N (x_k - \bar{x})^2, \\ \text{cov}_{xy} &= \frac{1}{N} \sum_{n=1}^N (x_k - \bar{x})(y_k - \bar{y}). \end{aligned} \quad (1.3)$$

преобразовываем линейную систему (1.2) из двух уравнений относительно двух неизвестных a и b к виду

$$\begin{cases} a + b\bar{x} = \bar{y} \\ a\bar{x} + b(\sigma_x^2 + \bar{x}^2) = \text{cov}_{xy} + \bar{x}\bar{y} \end{cases} \quad (1.4)$$

Отсюда находим

$$b = \frac{\text{cov}_{xy}}{\sigma_x^2}, \quad a = \bar{y} - b\bar{x}. \quad (1.5)$$

Покажем, что найденные решения (1.5) дают именно минимум функционалу (1.1). Предположим, что рассматривается какая-нибудь другая прямая с параметрами $a' = a + \alpha$, $b' = b + \beta$. Тогда значение функционала ρ будет равно

$$\begin{aligned} \rho &= \sum_n ((a + bx_n - y_n) + (\alpha + \beta x_n))^2 = \\ &= \sigma^2 + \sum_n (\alpha + \beta x_n)^2 + 2 \sum_n (a + bx_n - y_n)(\alpha + \beta x_n). \end{aligned}$$

Поскольку последнее слагаемое в написанном выше выражении равно нулю в силу выполнения условий экстремальности (1.2), то $\rho \geq \sigma^2$, что и требовалось доказать.

Таким образом, найденные параметры ЛРМ по имеющимся N парам значений (x_n, y_n) дают выборочную прямую регрессии, которая записывается в виде:

$$\tilde{y}_n - \bar{y} = \frac{\text{cov}_{xy}}{\sigma_x^2} \cdot (x_n - \bar{x}). \quad (1.6)$$

Если ввести нормированные величины $x' = \frac{x - \bar{x}}{\sigma_x}$ и $y' = \frac{y - \bar{y}}{\sigma_y}$, то

для них уравнение линейной регрессии (1.6) принимает вид

$$\tilde{y}' = R_{xy}(N) x', \quad (1.7)$$

где

$$R_{xy}(N) = \frac{\text{cov}_{xy}}{\sigma_x \sigma_y} \quad (1.8)$$

есть коэффициент корреляции между двумя рядами данных по рассматриваемой выборке длины N .

В результате применения этой модели случайная величина η приближенно связывается со случайной (или, возможно, неслучайной) величиной ξ . Остаток, т.е. разность между фактическим значением y_i и аппроксимацией $\tilde{y}(x_i)$, представляет собой некоторый случайный процесс $\varepsilon = y - \tilde{y}$, значения которого определяются из равенства

$$y_i = \bar{y} + R_{xy} \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}) + \varepsilon_i. \quad (1.9)$$

Квадрат дисперсии σ^2 остатка ε определяет ошибку аппроксимации данных:

$$\sigma^2 = \langle \varepsilon_i^2 \rangle = \sigma_y^2(n) (1 - R_{xy}^2(n)). \quad (1.10)$$

Коэффициентом детерминации R называется отношение средне-квадратичного отклонения Q_R аппроксимационных значений от среднего значения к дисперсии σ_y фактических величин:

$$R = \frac{Q_R}{\sigma_y} = \sqrt{\frac{\sum_i (\tilde{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}}. \quad (1.11)$$

Для определения коэффициента детерминации надо вычислить Q_R . Имеем

$$\begin{aligned} N\sigma_y^2 &= \sum_i (y_i - \bar{y})^2 = \sum_i ((\tilde{y}_i - \bar{y}) + (y_i - \tilde{y}_i))^2 = \\ &= NQ_R^2 + N\sigma^2 + 2\sum_i (\tilde{y}_i - \bar{y})(y_i - \tilde{y}_i). \end{aligned}$$

но в силу (1.5)

$$\sum_i (\tilde{y}_i - \bar{y})(y_i - \tilde{y}_i) = b \text{cov}_{xy} - b^2 \sigma_x^2 = 0,$$

так что получаем

$$\sigma_y^2 = Q_R^2 + \sigma^2. \quad (1.12)$$

Сопоставляя (1.12) с (1.10) и определением R из (1.11), находим, что для линейной парной регрессионной модели

$$R^2 = R_{xy}^2. \quad (1.13)$$

Пусть теперь имеется $s+1$ случайных величин ξ_1, \dots, ξ_{s+1} , наблюдение за которыми дает $s+1$ рядов x_{ri} , $r=1, 2, \dots, s+1$; $i=1, 2, \dots$. Рассмотрим линейную аппроксимацию $s+1$ -го ряда через остальные s регрессоров, т.е. будем искать наилучшее в среднем квадратичном линейное приближение вида

$$\tilde{x}_{s+1,i} - \bar{x}_{s+1} = \sum_{r=1}^s a_r (x_{r,i} - \bar{x}_r) \quad (1.14)$$

по выборке длины N . Искомыми являются коэффициенты a_r в (1.14). Для этого следует поступить так же, как и при выводе формул (1.4) для парной регрессии, но технически проще непосредственно решить систему (1.14) в смысле среднего квадратичного. Для этого умножим правую и левую части (1.14) на $x_{r,i} - \bar{x}_r$, где индекс r пробегает значения от 1 до s , и просуммируем полученные равенства по i от 1 до N . В результате получаем систему

$$\sum_{k=1}^s C_{rk} a_k = C_{r,s+1}, \quad (1.15)$$

где $C_{rk} = \text{cov}(\xi_r, \xi_k) = C_{kr}$ есть ковариация между соответствующими рядами — в данном случае выборочная, по выборке длины N . Если ситуация невырожденная, то

$$a_r = \sum_{k=1}^s (C^{-1})_{rk} C_{k,s+1}. \quad (1.16)$$

В частности, для тройной регрессии величины z на x и y формулы аппроксимации имеют вид

$$\begin{aligned} \tilde{z} &= \bar{z} + a(x - \bar{x}) + b(y - \bar{y}), \\ a &= \frac{C_{xz}\sigma_y^2 - C_{xy}C_{yz}}{\Delta}, \quad b = \frac{C_{zy}\sigma_x^2 - C_{zx}C_{xy}}{\Delta}, \quad (1.17) \\ \Delta &= \sigma_x^2\sigma_y^2 - C_{xy}^2. \end{aligned}$$

Найдем для этого примера коэффициент детерминации. Регрессионно объясняемая часть дисперсии величины z равна

$$\begin{aligned} Q_R^2 &= \langle (\tilde{z} - \bar{z})^2 \rangle = \frac{C_{yz}^2\sigma_x^2 + C_{xz}^2\sigma_y^2 - 2C_{xy}C_{yz}C_{zx}}{\Delta} = \\ &= \sigma_z^2 \frac{R_{xz}^2 + R_{zy}^2 - 2R_{xy}R_{yz}R_{zx}}{1 - R_{xy}^2}. \end{aligned} \quad (1.18)$$

Тогда коэффициент детерминации равен

$$R = \frac{Q_R}{\sigma_z} = \sqrt{\frac{R_{xz}^2 + R_{zy}^2 - 2R_{xy}R_{yz}R_{zx}}{1 - R_{xy}^2}}, \quad (1.19)$$

причем $Q_R^2 + \sigma^2 = \sigma_z^2$, где σ^2 есть квадрат дисперсии остатков регрессии, равный

$$\sigma^2 = \frac{\sigma_x^2\sigma_y^2\sigma_z^2(1 - R_{xy}^2 - R_{yz}^2 - R_{zx}^2 + 2R_{xy}R_{yz}R_{zx})}{1 - R_{xy}^2}. \quad (1.20)$$

Увеличение коэффициента детерминации трактуется обычно как уточнение корреляционной связи между величинами. Но если объясняющих переменных много и нет эффекта мультиколлинеарности (матрица ковариаций невырождена), то детерминация при увеличении числа регрессоров, как правило, растет. Тем не менее, применяя затем полученную аппроксимационную модель как прогнозную, на практике часто наблюдается ухудшение точности прогноза при увеличении числа регрессоров. Это связано с тем, что сами регрессоры не обязательно являются статистически независимыми, причем появляется значимая корреляция между ними и остатком регрессионной аппроксимации. Наиболее точная регрессионная модель подбирается эмпирически.

В общем случае соотношение (1.14) для центрированных величин может быть записано в матричном виде (для удобства обозначим $y_i = x_{s+1,i} - \bar{x}_{s+1}$):

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1s} \\ x_{21} & x_{22} & \cdots & x_{2s} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Ns} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_s \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix} \Leftrightarrow \vec{y} = X\vec{a} + \vec{\varepsilon}, \quad (1.21)$$

т.е. матрица X составлена из наблюдаемых значений регрессоров с условием их центрированности по выборке. Тогда

$$\|\vec{\varepsilon}\|^2 = \|\vec{y}\|^2 - (\vec{y}, X\vec{a}) - (\vec{a}, X^T \vec{y}) + (\vec{a}, X^T X \vec{a}) \rightarrow \min.$$

Матрица $X^T X$ есть умноженная на N матрица ковариаций

C . Из условия $\frac{\partial \|\vec{\varepsilon}\|^2}{\partial a_i} = 0$ следует решение (1.16):

$$\vec{a} = C^{-1} X^T \vec{y}. \quad (1.22)$$

С практической точки зрения смысл, например, парной регрессии состоит в том, что если коэффициент детерминации достаточно близок к единице, то шумовыми остатками ε_i в (1.9) можно пренебречь, и тогда это уравнение как бы выявляет функциональную связь между двумя рядами данных. Считается, что в реальном эксперименте эта связь искажается шумом. Однако на самом деле регрессионная связь не является истинно функциональной, т.е. если делать различные преобразования, например, выражать в формулах регрессии одни переменные через другие и подставлять полученные выражения в интересующие нас связи, то точность полученных выражений будет хуже, чем при непосредственном построении регрессии для аппроксимации этих связей.

Действительно, построим, например, обратную модель к (1.6), в которой x будет объясняемой переменной, а y — регрессором. Тогда имеем

$$\tilde{x}_i - \bar{x} = \frac{\text{cov}_{xy}}{\sigma_y^2} \cdot (y_i - \bar{y}). \quad (1.23)$$

В пренебрежении остатками ε_i можно опустить тильду у аппроксимационных значений \tilde{x}_i в (1.23), предполагая, что это-то и есть искомая функциональная (в данном случае линейная) связь. Пусть ситуация невырожденная, т.е. ковариация между рассматриваемыми величинами отлична от нуля. Тогда из (1.23) можно y_i выразить через x_i , что дает

$$y_i = \bar{y} + \frac{\sigma_y^2}{\text{cov}_{xy}} (x_i - \bar{x}). \quad (1.24)$$

Если теперь уравнение (1.24) трактовать как регрессию y на x , т.е. вернуть тильду, повесив ее на y_i , то квадрат дисперсии остатка для такой модели будет равен $\sigma_y^2(n)(1 - R_{xy}^2(n))/R_{xy}^2(n)$, что не меньше, чем величина (1.10). Это означает, что нельзя трактовать статистическую

связь как истинно функциональную. Обе прямых, определяемых уравнениями (1.24) и (1.9) без ε_i , совпадут только в том случае, если $\text{cov}_{xy}^2 = \sigma_x^2 \sigma_y^2$, т.е. если $R_{xy}^2 = 1$.

Другой важный пример недопустимости функциональной трактовки регрессионных зависимостей относится к задаче регрессионного анализа для многих переменных. Пусть, например, имеются две случайных величины, принимающих значения y_i и z_i , и одна объясняющая переменная x_i . Построив парные регрессии вида (1.6), получим уравнения прямых

$$\tilde{y} - \bar{y} = \frac{C_{xy}}{\sigma_x^2} \cdot (x - \bar{x}), \quad \tilde{z} - \bar{z} = \frac{C_{xz}}{\sigma_x^2} \cdot (x - \bar{x}). \quad (1.25)$$

Если теперь опустить тильду над z и исключить x из системы (1.25), то получится прямая

$$\tilde{y} - \bar{y} = \frac{C_{xy}}{C_{xz}} \cdot (z - \bar{z}), \quad (1.26)$$

которая отнюдь не является прямой регрессии y на z , потому что, если ее трактовать в таком смысле, то зависимость (1.26) будет иметь более высокую ошибку аппроксимации, чем собственно регрессия

$$\tilde{y} - \bar{y} = \frac{C_{zy}}{\sigma_z^2} \cdot (z - \bar{z}). \quad (1.27)$$

Действительно, ошибка аппроксимации (1.26) равна

$$\sigma^2 = \sigma_y^2 + \frac{C_{xy}^2 \sigma_z^2 - 2C_{xy} C_{yz} C_{zx}}{C_{xz}^2}, \quad (1.28)$$

тогда как ошибка регрессионной аппроксимации (1.27) равна

$\sigma_y^2 - \frac{C_{yz}^2}{\sigma_z^2}$. Вычитая ее из σ^2 (1.28), получаем неотрицатель-

ную величину $\frac{(C_{xy} \sigma_z^2 - C_{yz} C_{zx})^2}{C_{xz}^2 \sigma_z^2}$.

Рассмотренные примеры важны для понимания возможных причин плохих результатов прогнозирования по регрессионным моделям. Прежде всего, это касается допустимой малости шума ε_i в аппроксимационных формулах. Если шум, т.е. ошибки регрессии, не коррелируют с регрессорами, то оценка параметров регрессии (величин a_r в (1.14)) по выборке является несмещенной и состоятельной: средние значения оценок ковариационной матрицы по стационарной генеральной совокупности, предположительно существующей для изучаемых случайных величин, равны генеральным ковариациям, а при увеличении объема выборки $N \rightarrow \infty$ выборочная ковариационная матрица сходится к генеральной. Если же есть корреляция между ошибками и регрессорами, то оценки могут быть несостоятельны. Такая ситуация возникает, если на изучаемую величину и на регрессоры влияет один и тот же фактор, который, однако, скрыт от наблюдателя.

Например, пусть изменение случайной величины η обусловлено двумя случайными факторами — величинами ζ и ξ , так что следовало бы рассмотреть тройственную регрессию $\tilde{y} - \bar{y} = \alpha(z - \bar{z}) + \beta(x - \bar{x})$. Кроме того, если величины ζ и ξ не являются независимыми, то в регрессионной модели $\tilde{z} - \bar{z} = \gamma(x - \bar{x})$ коэффициент γ будет значимо отличаться от нуля. Предположим теперь, что фактор x недоступен для наблюдения, и используется модель $\tilde{y} - \bar{y} = \alpha(z - \bar{z})$. В этом случае ошибки упрощенной модели оказываются коррелированными с регрессорами:

$$\langle z - \bar{z}, y - \bar{y} - \alpha(z - \bar{z}) \rangle = \langle \gamma \Delta x, \beta \Delta x \rangle = \gamma \beta \sigma_x^2.$$

Как правило, к выводу о коррелированности ошибок и регрессоров приводят наблюдения за регрессорами, выполненные с определенной неточностью, т.е. к этому эффекту приводят ошибки измерения наблюдаемых величин, вызванные третьим фактором. Само по себе влияние третьего фактора неустранимо, но можно попытаться снизить корреляцию объясняющей переменной и ошибки модели введением новых переменных — это так называемый метод инструментальных переменных или двухшаговый МНК.

Итак, пусть рассматривается модель $y(n) = \bar{y} + \alpha(z(n) - \bar{z}) + \varepsilon(n)$, связывающая две стационарных случайных величины. Если выяснилось, что между z и ε существует значимая корреляция, надо подобрать некоторые другие регрессоры, от которых требуется, чтобы они хорошо коррелировали бы с z , но давали бы близкую к нулю корреляцию с ε . Выберем некоторую переменную w (если, конечно, она найдется среди наблюдаемых величин), обладающую свойством нулевой корреляции с ε , и построим регрессию старых регрессоров на эту переменную: $\tilde{z}(n) = \bar{z} + \beta(w(n) - \bar{w})$. Затем эту аппроксимационную величину \tilde{z} принимаем за инструментальную переменную и строим регрессию y на \tilde{z} . По построению, остатки регрессии не коррелируют с регрессорами.

Точность регрессионного приближения и обоснование его применимости даются теоремой Гаусса-Маркова.

▼ **Теорема 1.1.** Если парная регрессионная модель $y_n = a + bx_n + \varepsilon_n$ такова, что ε_n есть нормально распределенная случайная величина с нулевым средним и постоянной дисперсией, то оценки параметров a и b по формулам МНК имеют наименьшую дисперсию в классе всех линейных несмещенных оценок. ▲

Пусть регрессия строилась по выборке N данных. Из теоремы Гаусса-Маркова и стационарности регрессии, влекущей выполнение центральной предельной теоремы для выборочных средних, следует, что статистика $z = \frac{\tilde{y} - \bar{y}(N)}{\sigma_y(N)}$ имеет распределение Стьюдента с $N - 2$ степенями

свободы. В результате доверительный интервал для прогнозных значений \hat{y}_k на уровне значимости α определяется статистикой Стьюдента $t_{1-\alpha}(N - 2)$:

$$\tilde{y}_k - t_{1-\alpha}(N - 2)\sigma_y(N) \leq \hat{y}_k \leq \tilde{y}_k + t_{1-\alpha}(N - 2)\sigma_y(N). \quad (1.29)$$

Если остатки регрессии образуют гауссов белый шум (нормально распределенный процесс с независимыми приращениями), то можно теоретически определить значимость коэффициента корреляции R . Исходным положением является то, что при отсутствии корреляции между объясняемой и объясняющей переменными статистика

$$\frac{|R(N)|\sqrt{N-2}}{\sqrt{1-R^2(N)}} > t_{1-\alpha}(N-2). \quad (1.30)$$
$$\hat{y}_n = \sum_{k=1}^p U_k y_{n-k} + \varepsilon, \quad \hat{y}_n = \hat{x}_n - m_{1,x}(N). \quad (1.31)$$

268

[illegible]

AP первого порядка имеет вид

$$\hat{y}_n = A(1)y_{n-1}. \quad (1.33)$$
$$\hat{y}_n = \frac{A(1)(1 - A(2))y_{n-1} + (A(2) - A^2(1))y_{n-2}}{1 - A^2(1)}, \quad (1.34)$$

Отметим, что на практике важную роль часто играет не абсолютная ошибка или аналогичная ей ошибка нормированных величин, а относительная среднеквадратичная ошибка аппроксимации δ , при условии, что изучаемая величина принимает ненулевые значения. Квадрат этой ошибки определяется формулой

$$\delta^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i}{y_i} \right)^2 \quad (1.35)$$

$$m_{-k,y}(n) = \frac{1}{n} \sum_{i=1}^n \frac{1}{y_i^k}, \quad y_i \neq 0. \quad (1.36)$$

269

$$m_{-2,y}(n) = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{y_i} - m_{-1,y}(n) \right)^2 + (m_{-1,y}(n))^2 \equiv \\ \equiv \mu_{-2,y}(n) + (m_{-1,y}(n))^2$$

и аналогично

$$\left\langle \frac{x}{y} \right\rangle = \text{cov} \left(x, \frac{1}{y} \right) + m_{1,x}(n) m_{-1,y}(n) = \\ = R_{x,1/y}(n) \sqrt{\mu_{2,x}(n) \mu_{-2,y}(n)} + m_{1,x}(n) m_{-1,y}(n).$$

то для ошибки (1.18) получаем выражение

$$\delta^2 = a^2 \left(\text{cov} \left(x^2, \frac{1}{y^2} \right) + m_{2,x} m_{-2,y} \right) + \\ + 2ab \left(\text{cov} \left(x, \frac{1}{y^2} \right) + m_{1,x} m_{-2,y} \right) + b^2 m_{-2,y}, \quad (1.37)$$

где

$$a = R_{xy} \sqrt{\frac{\mu_{2,y}}{\mu_{2,x}}}, \quad b = m_{1,y}(1-a).$$

Как и для абсолютной ошибки аппроксимации, если $R_{xy} \rightarrow 1$, то $\delta \rightarrow 0$.

8.2. Линейная регрессия в скользящем окне

При вычислении прогнозного значения по модели аппроксимации в стационарном случае полагают, что коэффициенты регрессии постоянны. Тогда прогнозной моделью является продолжение аппроксимационной зависимости по индексу объясняющей переменной за пределы промежутка аппроксимации. Например, для парной регрессии y на x по выборке объема n в момент времени t аппроксимация \tilde{y} дается формулой (1.6), которую, опуская для краткости зависимость от объема выборки, перепишем в виде

$$\tilde{y}(i) = \bar{y}(t) + \frac{C_{xy}(t)}{\sigma_x^2(t)} \cdot (x(i) - \bar{x}(t)), i = t - n + 1, \dots, t. \quad (2.1)$$

В этой формуле средние, дисперсия и ковариация вычисляются по выборке объема n , отсчитанной назад от момента времени t . Тогда консервативный прогноз \hat{y}_c , т.е. прогноз в предположении постоянства коэффициентов регрессии, в момент времени $t+1$ строится по формуле

$$\hat{y}_c(t+1) = \bar{y}(t) + \frac{C_{xy}(t)}{\sigma_x^2(t)} \cdot (x(t+1) - \bar{x}(t)). \quad (2.2)$$

Концепция консервативного прогноза корректна для однократного прогнозирования. Если же мы прогнозируем в скользящем окне, то, очевидно, прогноз, в предположении стационарности ряда, должен представлять собой аппроксимацию данных в этом скользящем окне. Следовательно, более правильной является модель прогноза, в которой прогнозное значение является оптимальным с точки зрения будущей регрессии в сдвинутом на один шаг окне. Как и в обычной стационарной модели, предполагаем, что скользящая связь вида (2.1) между y и x остается справедливой и в следующий момент времени. Тогда прогнозируемое значение само входит в правую часть (2.2), так как средние должны вычисляться, согласно скользящим средним, в тот же момент времени $t+1$. В этом случае вместо (2.2) следует писать

$$\hat{y}(t+1) = \bar{y}(t+1) + \frac{C_{xy}(t+1)}{\sigma_x^2(t+1)} \cdot (x(t+1) - \bar{x}(t+1)). \quad (2.3)$$

В формуле (2.3) средние величины в момент времени $t+1$ можно выразить через средние в момент t и через новые значения $x(t+1)$ и $\hat{y}(t+1)$, причем $\hat{y}(t+1)$ трактуется как будущее фактическое значение. Тогда используем рекуррентные формулы:

$$\bar{y}(t+1) = \bar{y}(t) + \frac{\hat{y}(t+1) - y(t-n+1)}{n}; \quad (2.4)$$

$$\begin{aligned}
C_{xy}(t+1) = C_{xy}(t) + \\
+ \frac{1}{n} \{ \hat{y}(t+1) \cdot (x(t+1) - \bar{x}(t)) - \\
- \bar{y}(t) \cdot (x(t+1) - x(t-n+1)) + \bar{x}(t) \cdot y(t-n+1) \} - \\
- \frac{(\hat{y}(t+1) - y(t-n+1)) \cdot (x(t+1) - x(t-n+1))}{n^2}.
\end{aligned}
\quad (2.5)$$

Подставляя выражения (2.4-2.5) в (2.3), получаем линейное уравнение относительно прогнозного значения $\hat{y}(t+1)$, разрешая которое, находим

$$\begin{aligned}
& \bar{y}(t) - \frac{1}{n} y(t-n+1) + \frac{x(t+1) - \bar{x}(t+1)}{\sigma_x^2(t+1)} \left(C_{xy}(t) + \frac{a}{n} + \frac{b}{n^2} \right) \\
& = \frac{1 - \frac{1}{n} - \frac{x(t+1) - \bar{x}(t+1)}{\sigma_x^2(t+1)} \left(\frac{x(t+1) - \bar{x}(t)}{n} + \frac{x(t+1) - x(t-n+1)}{n^2} \right)}{1 - \frac{1}{n} - \frac{x(t+1) - \bar{x}(t+1)}{\sigma_x^2(t+1)} \left(\frac{x(t+1) - \bar{x}(t)}{n} + \frac{x(t+1) - x(t-n+1)}{n^2} \right)}, \\
& a = \bar{x}(t)y(t-n+1) - \bar{y}(t)(x(t+1) - x(t-n+1)), \\
& b = y(t-n+1)(x(t+1) - x(t-n+1)).
\end{aligned}
\quad (2.6)$$

При $n \rightarrow \infty$ формулы (2.6) переходят в формулы обычной ЛРМ.

Для множественной регрессии этот подход также применим, если только коэффициенты регрессии записаны в виде явных выражений для средних значений, т.е. если известна ковариационная матрица. В противном случае надо численно решать задачу о регрессионной аппроксимации в виде итерационного процесса, когда прогнозное значение, полученное на предыдущем шаге, подставляется в выборку, по которой строилась аппроксимация, в качестве последнего значения. Этот процесс строится следующим образом.

Пусть имеются $p+1$ выборок данных объема n каждая, и строится множественная регрессия выборки y на выборки x_1, \dots, x_p . Соответствующий ряд для j -ой выборки есть набор данных $\{x_j(1), \dots, x_j(n)\}$. Обозначим средние по этим выборкам через \bar{x}_j , ковариации между ними

C_{ij} , ковариацию между выборками y и x_j через K_j . Перейдем к рядам отклонений от средних по выборке, что обозначим штрихом:

$$x'_j(k) = x_j(k) - \bar{x}_j, \quad y'(k) = y(k) - \bar{y}. \quad (2.7)$$

Тогда аппроксимация отклонений имеет вид

$$\tilde{y}'(k) = \sum_{j=1}^p A_j x'_j(k), \quad (2.8)$$

где коэффициенты регрессии A_j являются решением системы уравнений

$$\sum_{j=1}^p C_{ij} A_j = K_i. \quad (2.9)$$

Если нет эффекта мультиколлинеарности, ковариационная матрица невырождена, что позволяет найти решение системы (2.9). Таким образом, коэффициенты A_j являются некоторыми функционалами от элементов выборок:

$$A_j = \sum_{i=1}^p (C^{-1})_{ji} K_i. \quad (2.10)$$

Поставим теперь следующую задачу. Пусть заданы n элементов выборок y и x_j , $j = 1, \dots, p$. Требуется найти такое значение $\tilde{y}(n)$, чтобы гиперплоскость регрессии (2.8) прошла бы через эту последнюю точку. Дадим для иллюстрации точное решение этой задачи при $p = 1$.

Среднее значение $\tilde{\bar{y}}$ по выборке, содержащей искомое значение $\tilde{y}(n)$, отличается от среднего \bar{y} по исходной выборке на величину $(\tilde{y}(n) - y(n))/n$, т.е.

$$\tilde{\bar{y}} = \bar{y} + \frac{\tilde{y}(n) - y(n)}{n}. \quad (2.11)$$

Аналогично ковариация искомой выборки с выборкой x равна

$$\tilde{C}_{xy} = C_{xy} + \frac{(x(n) - \bar{x}) \cdot (\tilde{y}(n) - y(n))}{n}. \quad (2.12)$$

Следовательно, значение $\tilde{y}(n)$ находится из уравнения прямой регрессии

$$\tilde{y}(n) = \tilde{\bar{y}} + \tilde{C}_{xy} \frac{x(n) - \bar{x}}{\sigma_x^2},$$

т.е.

$$\tilde{y}(n) = \frac{\tilde{y}_c(n) - \frac{y(n)}{n} \left(1 + \frac{(x(n) - \bar{x})^2}{\sigma_x^2} \right)}{1 - \frac{1}{n} \left(1 + \frac{(x(n) - \bar{x})^2}{\sigma_x^2} \right)}, \quad (2.13)$$

где $\tilde{y}_c(n)$ есть аппроксимация $y(n)$ по исходной выборке,

т.е.

$$\tilde{y}_c(n) = \bar{y} + C_{xy} \frac{x(n) - \bar{x}}{\sigma_x^2}. \quad (2.14)$$

Рассмотрим теперь следующий итерационный процесс по нахождению $\tilde{y}(n)$. Зададим нулевое приближение, полагая $\tilde{y}_0(n) = \tilde{y}_c(n)$, где $\tilde{y}_c(n)$ определена в соответствии с (2.14). На следующем шаге заменим в исходной выборке $y(n)$ на $\tilde{y}_0(n)$ и построим аппроксимацию для этой новой выборки. Соответствующее значение y обозначим $\tilde{y}_1(n)$. Повторяем процесс аппроксимации с этим новым значением и т.д. Найдем условия, при которых этот итерационный процесс сходится и его пределом является $\tilde{y}(n)$ из (2.13). На k -ом шаге итерации имеем линейное рекуррентное соотношение, коэффициенты которого определяются неизменяемыми элементами выборок, т.е. постоянны:

$$\begin{aligned} \tilde{y}_k(n) &= \alpha + \beta \tilde{y}_{k-1}(n), \\ \alpha &= \tilde{y}_c(n) - \frac{y(n)}{n} \left(1 + \frac{(x(n) - \bar{x})^2}{\sigma_x^2} \right), \quad (2.15) \\ \beta &= \frac{1}{n} \left(1 + \frac{(x(n) - \bar{x})^2}{\sigma_x^2} \right) > 0. \end{aligned}$$

Выражение для β определено, если дисперсия объясняющей переменной отлична от нуля. Тогда оператор рекурсии сжимающий, если $\beta < 1$. В этом случае предел последовательности $\tilde{y}_k(n)$ существует и равен $\tilde{y}(n) = \alpha / (1 - \beta)$, т.е. совпадает с (2.13).

8.3. Коинтегрированные регрессионные модели

Коинтегрированными называются такие временные ряды, которые, может быть, сами по себе являются нестационарными, но существует их линейная комбинация, или комбинация разностей определенного порядка элементов этих рядов, являющаяся стационарным временным рядом. Для формализации используемых в таких случаях моделей дадим некоторые определения.

▼ **Определение 3.1.** Спектральным представлением действительного скалярного случайного процесса с АКФ $A(t)$ называется представление

$$A(t) = \int_{-\pi}^{\pi} e^{i\lambda t} dQ(\lambda) \quad (3.1)$$

для дискретного времени и

$$A(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dQ(\lambda) \quad (3.2)$$

для непрерывного времени. Функция $Q(\lambda)$ называется спектральной функцией, а порожденная ей мера — спектральной мерой. ▲

▼ **Определение 3.2.** Пусть $\xi(t)$ стационарный в широком смысле процесс, $A(t)$ и $Q(\lambda)$ его автокорреляционная и спектральная функции. Преобразование Φ процесса называется линейным фильтром, если процесс $\eta(t) = \Phi \xi(t)$ представляется в виде

$$\eta(t) = \int_{-\infty}^{\infty} h(t-s) \xi(s) ds. \quad (3.3)$$

Функция $h(t)$ называется импульсной переходной функцией фильтра, а ее Фурье-образ называется частотной характеристикой фильтра. ▲

▼ **Определение 3.3.** Процесс $\xi(t)$ называется процессом с независимыми приращениями, если для любого n и для всех моментов времени $t_1 < t_2 < \dots < t_n$ случайные величины $\xi(t_1)$, $\xi(t_2) - \xi(t_1)$, ..., $\xi(t_n) - \xi(t_{n-1})$ независимы. Процесс с независимыми приращениями называется также белым шумом. ▲

▼ **Теорема 3.1.** (Разложение Вольда). Всякий стационарный процесс $\xi(t)$ может быть представлен единственным образом в виде суммы некоррелированных между собой процессов $\xi(t) = \xi^S(t) + \xi^R(t)$, где $\xi^S(t)$ — сингулярный (или детерминированный) процесс, а $\xi^R(t)$ — регулярный (полностью случайный). ▲

Сингулярная составляющая стационарного случайного процесса извлекается из него по бесконечно удаленному прошлому, а регулярная составляющая ищется на основе следующего утверждения.

▼ **Теорема 3.2.** Для того чтобы стационарный процесс $\xi(t)$ был регулярным, необходимо и достаточно, чтобы он представлял собой преобразование фильтрации (3.3) некоторого процесса с независимыми приращениями. ▲

Теорема Вольда в терминах фильтрации дискретного вида может быть сформулирована следующим образом: всякий стационарный временной ряд с конечными математическим ожиданием и дисперсией может быть представлен в виде

$$x(t) = \mu + \sum_{k=0}^{\infty} h_k \varepsilon(t-k), \quad \sum_{k=0}^{\infty} |h_k| < \infty, \quad (3.4)$$

где μ есть среднее значение процесса, а $\varepsilon(t)$ есть стационарный процесс белого шума с нулевым средним и конечной дисперсией σ_ε^2 .

▼ **Определение 3.4.** Процесс называется процессом скользящего среднего порядка q и обозначается $CC(q)$, если он представляется в виде

$$x(t) = \mu + \sum_{k=0}^q h_k \varepsilon(t-k). \quad \blacktriangle (3.5)$$

Процесс $CC(q)$ имеет следующие свойства:

$$Mx = \mu, \quad Dx = \sigma_\varepsilon^2 \sum_{k=0}^q h_k^2, \quad A_x(\tau) \Big|_{\tau \leq q} = \sum_{k=0}^{q-\tau} h_k h_{k+\tau}, \quad (3.6)$$

$$A_x(\tau) \Big|_{\tau > q} = 0.$$

Как видно из свойств (3.6), модель (3.5) имеет смысл использовать в тех случаях, когда выборочная АКФ процесса близка к нулю, начиная с определенного значения $\tau = q$.

Рассмотрим представление (3.5) как результат применения операторного полинома от оператора \hat{L} сдвига на 1 шаг:

$$y(t) = x(t) - \mu = \sum_{k=0}^q h_k \hat{L}^k \varepsilon(t), \quad \hat{L} \varepsilon(t) = \varepsilon(t-1). \quad (3.7)$$

Характеристическим многочленом процесса $CC(q)$ называется полином

$$Y_q(\lambda) = \sum_{k=0}^q h_k \lambda^{q-k}. \quad (3.8)$$

▼ **Определение 3.5.** Процесс $CC(q)$ (3.7) называется обратимым, если существует такой процесс фильтрации с коэффициентами ψ_k , что

$$\varepsilon(t) = \sum_{k=0}^{\infty} \psi_k \hat{L}^k y(t). \quad \blacktriangle (3.9)$$

Пусть $\{\lambda_k\}$ суть различные корни характеристического полинома $Y_q(\lambda)$. Тогда этот полином представим в виде

$$Y_q(\lambda) = h_0 \prod_{k=1}^q (\lambda - \lambda_k). \quad \text{Поскольку норма оператора сдвига равна}$$

единице, необходимым условием обратимости, т.е. сходимости по λ разложения в ряд дроби $1/Y_q(\lambda)$, является условие $|\lambda_k| < 1$. Если подставить разложение (3.9) в исходную модель (3.7), то получится представление, в котором значение случайной величины y в момент времени t выражено через значения этой же величины во все предыдущие моменты времени. Если теперь ограничиться в нем несколькими первыми p слагаемыми, получится АР порядка p , рассмотренная выше в (1.31).

▼ **Определение 3.6.** Процесс называется процессом авторегрессии порядка p АР(p), если он представляется в виде

$$y(t) = \varepsilon(t) + \sum_{k=1}^p U_k y(t-k). \quad \blacktriangle \quad (3.10)$$

По аналогии с процессом СС(q) (3.7), для процесса АР(p) также можно написать характеристическое уравнение и рассмотреть вопрос об обратимости этого процесса. Видно, что эти два типа процессов в некотором смысле взаимно обратные: именно, обратным к СС(q) является некоторый АР(∞), а обратным к АР(p) является некоторый СС(∞). В последнем случае будет осуществлена декомпозиция Вольда, поэтому обратимый процесс АР(p) является стационарным. Общее решение уравнения (3.10) представляется в виде суммы частного решения неоднородного уравнения, которое дается формулами Юла-Уокера (1.32), и общего решения однородного уравнения. Последнее ограничено, если корни характеристического уравнения для процесса (3.10) по модулю меньше единицы, т.е. это условие совпадает с условием обратимости процесса СС(q). Если же хотя бы один корень имеет действительную часть, по модулю больше либо равную единице, декомпозиция Вольда невозможна, и такой процесс не будет стационарным.

Характерным примером нестационарного авторегрессионного процесса является так называемое случайное блуждание, представляющее собой процесс АР(1), имеющий единичный корень. Модель случайного блуждания дается формулой

$$y(t) = y(t-1) + \varepsilon(t) = y(0) + \sum_{k=1}^t \varepsilon(t-k). \quad (3.11)$$

Поскольку (3.11) можно переписать в виде $\varepsilon(t) = (\hat{L}^0 - \hat{L}^1)y(t)$, характеристическое уравнение правой части этого представления есть

$1 - \lambda$, которое имеет единственный единичный корень. Для модели (3.11) математическое ожидание постоянно $My(t) = 0$, но дисперсия оказывается зависящей от времени:

$$Dy(t) = D \sum_{k=1}^t \varepsilon(t-k) = \sum_{k=1}^t D\varepsilon(t-k) = \sum_{k=1}^t \sigma_\varepsilon^2 = t\sigma_\varepsilon^2.$$

Таким образом, на практике при использовании моделей типа АР(p) необходимо исследовать проблему единичного корня. Если соответствующая статистическая гипотеза решается положительно, т.е. единичный корень весьма вероятен, то ряд нестационарный. Тест модели вида (3.34) $y(t) = \lambda y(t-1) + \varepsilon(t)$ на нестационарность, т.е. на вероятность того, что $\lambda = 1$, называется тестом Дики-Фуллера. Поскольку проверяемая статистика может быть нестационарной, критерии типа Стьюдента неприменимы, т.к. они оказываются несостоятельными. Для оценки λ составляется статистика Стьюдента, но критические значения не имеют распределения Стьюдента, а определяются по таблицам, полученным на основе численного моделирования случайных блужданий методом Монте-Карло.

Обобщая описанную схему, вместо случайного блуждания (3.11) можно рассмотреть стационарный процесс $z(t) = y(t) - y(t-1) = \varepsilon(t)$, перейдя к его первым разностям. Если и этот новый процесс оказался нестационарным, можно перейти ко вторым разностям и т.д. В результате получаются модели относительно разностей определенного порядка, а возвращение к исходным случайным величинам дается тогда суммированием вычитавшихся переменных, так что такие модели называются интегрируемыми.

▼ **Определение 3.7.** Если в результате d -кратного применения разностного дифференцирования $x^{(d)}(t) = \Delta^d x(t)$, где оператор $\Delta = 1 - \hat{L}$ действует по правилу $\Delta x(t) = x(t) - x(t-1)$, получился стационарный процесс, то модель относительно $x^{(d)}(t)$ называется интегрируемой моделью порядка d . \blacktriangle

Переход к разностным случайным величинам является стандартным методом исключения полиномиального по времени тренда. Он основан на том, что разностная производная полинома порядка d является полиномом порядка $d-1$. Разумеется, в случае трендов других видов такой

подход не является продуктивным. Определение тренда представляет собой задачу подгонки данных под кривую того или иного типа, а выбор кривой оставляется на усмотрение исследователя. От удачности подгонки зависит точность анализа остатка, обусловленная главным образом тем, можно ли считать остаток стационарным.

▼ **Определение 3.8.** Комбинированным стационарным процессом авторегрессии — скользящего среднего $AR(p)$ и $CC(q)$ называется процесс $ARCC(p, q)$ вида

$$y(t) = x(t) - \mu = \sum_{k=1}^p \psi_k \hat{L}^k y(t) + \sum_{k=0}^q h_k \hat{L}^k \varepsilon(t), \quad (3.12)$$

допускающий также представления вида $AR(\infty)$ и $CC(\infty)$. Если процесс после d -кратного применения разностного дифференцирования приводится к процессу типа $ARCC(p, q)$, он называется процессом авторегрессии — интегрированного скользящего среднего, обозначаемого $ARISCC(p, d, q)$. ▲

На практике встречаются ряды, сами по себе нестационарные, но в долгосрочном аспекте связанные один с другим. Таковы, например, ряды цен на сырую нефть и нефтепродукты, а также цены на природный и сжиженный газ и т.п. Может оказаться, что даже линейные модели типа ЛРМ продуктивны для описания взаимозависимости между такими случайными величинами, т.е. в этом случае некоторая их линейная комбинация представляет стационарный ряд. Обобщим определение, данное в начале этого параграфа.

▼ **Определение 3.9.** Ряды $x_1(t), x_2(t), \dots, x_n(t)$ называются коинтегрированными, если существует такая линейная комбинация с постоянными коэффициентами их разностей порядка $d_0, d_1, d_2, \dots, d_n$, которая представляет стационарный процесс белого шума:

$$\sum_{k=0}^n c_k \Delta^{d_k} x_k(t) = \varepsilon(t). \quad \blacktriangle$$

Основной проблемой при использовании моделей типа (3.12) является то, что порядки (p, d, q) могут меняться со временем. Кроме того, чтобы анализ был более точным, желательно знать, какой белый шум следует рассматривать в качестве остатка в каждом конкретном случае — нор-

мальный, равномерный, пуассоновский и т.д. В сущности, гипотеза о распределении процесса $\varepsilon(t)$, как и о типе временного тренда, является внешней по отношению к собственно статистическому анализу данного временного ряда.

Насколько адекватны могут быть описанные регрессионные модели при практическом применении, зависит от достоверности коэффициентов лаговой корреляции или автокорреляции. Для нестационарных временных рядов само определение этих коэффициентов нуждается в уточнении, поскольку сходимости к генеральным значениям с увеличением длины выборки может не быть. Такое уточнение дается в следующем параграфе.

8.4. Оценка корреляции нестационарных временных рядов

Во многих задачах долгосрочного прогнозирования величин, модель изменения которых во времени предполагает наличие случайных процессов, требуется выделить главные векторы корреляционной матрицы, чтобы определить группы величин, находящиеся между собой в статистической зависимости. Таковы модели и прогнозы макроэкономических показателей, модели факторов, влияющих на экологическое состояние, статистические модели в социологии и др.

Традиционно практические исследования с использованием статистических методов опираются на модель реакции (линейной или нелинейной) сложной системы при изменении параметров, представляемых как внешние или управляющие воздействия. При этом цель анализа данных за определенный исторический период состоит в определении коэффициентов в такой модели, предполагая, что они образуют стационарный временной ряд. Чаще всего анализируются линейные функции реакции, когда предполагается, что два наблюдаемых нестационарных ряда данных находятся между собой в стационарной статистической взаимосвязи. В реальности стационарность связи выполняется лишь приближенно, поскольку в сложных системах коэффициенты в параметрических моделях являются неизвестными функционалами от анализируемых рядов данных. Поэтому все модели коинтеграции нестационарных рядов содержат коррелированные остатки регрессий. В результате оказывается, что очевидный качественный результат о том, что некоторые два ряда находятся в корреляционной зависимости — например, уровень загрязненности атмосферы определенным типом вещества и заболеваемость населения тем или иным видом заболевания, не имеет четкого численного подтверждения. Именно, если зафиксировать лаг корреляции и момент времени наблюдения, и на-

чать затем варьировать длину выборки, то часто обнаруживается, что выборочный коэффициент корреляции изменяется при этом в весьма широких пределах и даже может менять знак. И наоборот, если зафиксировать длину выборки, то корреляция в разные моменты времени может быть сильно нестационарной. Следовательно, актуальной является задача корректного определения интересующей исследователя корреляционной связи.

Построим методику определения величины корреляции между двумя нестационарными временными рядами. Математически проблема состоит в том, что выборочный коэффициент корреляции между двумя рядами данных стабилизируется к своему генеральному значению с увеличением объема выборки только для стационарных в широком смысле процессов, среднее и дисперсия которых не зависят от времени. Для нестационарных процессов наблюдается непостоянство корреляции как функции времени для выборки определенного объема, а также и непостоянство ее в один и тот же момент времени, но для выборок разных объемов. Как определить величину корреляции и ее достоверность для нестационарных рядов? Необходимым условием для корректного определения нестационарной корреляции является нахождение оптимальной длины выборки, на которой следует вычислять выборочные моменты распределений. Именно, требуется найти длину выборки, на которой определенное значение корреляции между двумя рядами проявляется с наибольшей достоверностью.

Формально нестационарный выборочный коэффициент корреляции представляет собой обычный выборочный коэффициент корреляции, зависящий от длины выборки и текущего момента времени, от которого назад в прошлое отсчитывается эта выборка. Ограничимся далее рассмотрением двух случайных величин, равномерно ограниченных по времени на отрезке $[0; 1]$. Пусть $x(t)$ и $y(t)$ — два исследуемых временных ряда на промежутке времени $[1; T]$, и $R(k, t)$ есть их выборочный коэффициент корреляции по выборке объема k в момент времени t , т.е. коэффициент корреляции, вычисленный по данным, взятым в моменты времени $t - k + 1, t - k + 2, \dots, t$:

$$R(k, t) = \frac{k \sum_{i=t-k+1}^t x_i y_i - \left(\sum_{i=t-k+1}^t x_i \right) \left(\sum_{i=t-k+1}^t y_i \right)}{\sqrt{k \sum_{i=t-k+1}^t x_i^2 - \left(\sum_{i=t-k+1}^t x_i \right)^2} \sqrt{k \sum_{i=t-k+1}^t y_i^2 - \left(\sum_{i=t-k+1}^t y_i \right)^2}}. \quad (4.1)$$

Для каждого интервала, ширина которого определяет точность определения коэффициента корреляции, можно найти длину выборки, на которой доля корреляций, попадающих в этот промежуток, наибольшая. Эта доля представляет собой уровень достоверности корреляционной связи. Тот интервал, для которого эта доля абсолютно наибольшая, принимается за промежуток, содержащий коэффициент корреляции.

Определим два взаимодополняющих временных ряда: ряд $r(t)$ максимальных алгебраических значений корреляции $R(k, t)$, и ряд $n(t)$ соответствующих объемов выборки, на которых корреляция в данный момент времени t максимальна:

$$r(t) = \max_k R(k, t), \quad n(t) = \arg \max_k R(k, t). \quad (4.2)$$

При вычислении $r(t)$ объем выборки должен превосходить некоторый минимальный объем, на котором вообще разумно вычислять корреляцию, поэтому будем считать, что $k \geq 3$.

Определим также среднее значение \bar{r} и дисперсию σ_r^2 максимальных значений корреляции по всем моментам времени, и аналогично \bar{n} и σ_n^2 . Пусть также γ есть коэффициент корреляции рядов $r(t)$ и $n(t)$ на промежутке времени $[1; T]$. Этот коэффициент подсчитывается по количеству данных $T - N + 1$, где $N \leq T$ есть максимальный объем среди всех $n(t)$: $N = \max_t n(t)$.

Величина σ_r / \bar{r} характеризует относительную ширину разброса средних по времени значений максимумов корреляций, и для стационарных рядов является основным показателем точности оценки $R(n, t) = \bar{r}$. Если же объем выборки $n(t)$ также меняется со временем, то фактиче-

ская неточность в оценке корреляции $r(t)$ должна определяться с учетом вариации σ_n / \bar{n} . Однако при вариации объема выборки может оказаться, что корреляция на отрезке $\Delta_\sigma(n(t)) = [n(t) - \sigma_n; n(t) + \sigma_n]$ меняется незначительно. Поэтому, чтобы учесть вариацию объема выборки, необходимо оценить чувствительность корреляции по объему выборки.

Введем величину

$$\lambda(t) = \frac{n(t)}{2\sigma_n r(t)} \left(\max_{k \in \Delta_\sigma(n(t))} R(k, t) - \min_{k \in \Delta_\sigma(n(t))} R(k, t) \right), \quad (4.3)$$

которая представляет собой разностный аналог логарифмического размаха корреляции по логарифму объема выборки в окрестности точки максимума, который вычисляется на отрезке $\Delta_\sigma(n(t)) = [n(t) - \sigma_n; n(t) + \sigma_n]$ ширины $2\sigma_n$. Среднее по времени значение $\bar{\lambda}$ характеризует среднюю крутизну графика корреляции как функции объема выборки в окрестности своего максимального значения. Если $\bar{\lambda}$ близко к нулю, т.е. график $R(n(t), t)$ представляет слабо меняющуюся функцию объема выборки, приблизительно параллельную оси абсцисс, то знание именно того объема $n(t)$, на котором достигается максимум корреляции, не обязательно, поскольку близкие к нему значения корреляции достигаются и на других объемах. Тем самым $\bar{\lambda}$ можно трактовать как чувствительность максимума корреляции к объему выборки. Тогда в качестве ориентировочной оценки относительной неточности O^2 в установлении корреляционной связи на уровне \bar{r} между двумя рядами можно использовать показатель

$$O^2 = \left(\frac{\sigma_r}{\bar{r}} \right)^2 + (1 - \gamma^2) \bar{\lambda}^2 \left(\frac{\sigma_n}{\bar{n}} \right)^2. \quad (4.4)$$

Величина γ^2 определяет ту долю дисперсии максимальной корреляции, которая может быть «объяснена» дисперсией объема выборки в регрессионном приближении, т.е. γ есть коэффициент корреляции между

рядами $r(t)$ и $n(t)$. Остаток представляет собственный вклад вариации объема выборки в неточность оценки корреляции с учетом коэффициента чувствительности. Тогда ширина доверительного интервала, содержащего оценку максимальной корреляции на уровне значимости O , равна $2O\bar{r}$.

Критерий « O^2 » (4.4) дает уровень значимости оценки корреляционной связи, характеризуемой числом \bar{r} . Но объем выборки, на котором следует вычислять эту корреляцию, не фиксируется критерием (4.4), этот объем может быть любым в границах $[\bar{n} - \sigma_n; \bar{n} + \sigma_n]$, что учитывается введением коэффициента чувствительности $\bar{\lambda}$. Поскольку же не обязательно наибольшая корреляция с наибольшей же достоверностью достигается на среднем объеме \bar{n} , дающем максимумы корреляций, то следует найти соответствующий оптимальный объем n_{opt} . Для этого поступим следующим образом. Введем достоверность α в определении объема выборки, при котором корреляция между рядами не ниже определенного уровня. Определим множество $M_n(t)$ тех объемов выборки n , корреляция по которым в момент времени t лежит в определенной области Δ значений, т.е.

$$M_n(t) : \bigcup_n \{n : R(n, t) \in \Delta\}. \quad (4.5)$$

▼ **Определение 4.1.** Корреляционная связь между двумя рядами на промежутке времени $[1; T]$ называется достоверной по объему выборки на

уровне α с абсолютной ошибкой δ , если пересечение $\bigcap_{t=1}^T M_n^\alpha(t)$ не

менее чем $[\alpha(T - n(t) + 1)]$ носителей $n(t)$ значений корреляции этих рядов как функции объема выборки, содержащихся в полосе минимальной ширины δ ($[1; 1 - \delta]$ для максимальной корреляции, $[-1; -1 + \delta]$ для минимальной, $[-\delta/2; \delta/2]$ для нулевой корреляции), не пусто. ▲

Интерес представляют три случая: достоверная корреляция, достоверная антикорреляция, а также достоверное отсутствие корреляции (нулевая корреляция). Разберем подробно случай достоверной корреляции, т.е. устойчивого по времени достаточно высокого положительного коэф-

фициента выборочной корреляции, вычисленного по одинаковому объему данных. Остальные случаи рассматриваются аналогично.

В каждый момент времени $t \in [1; T]$ строится выборочный коэффициент корреляции $R(n, t)$ как функция объема выборки n , $n \leq t$. Задается произвольное значение $0 < \delta < 1$ и рассматриваются те и только те значения $R(n, t)$, которые попали в полосу $[1; 1 - \delta]$. Пусть $\{n_i(t)\}$ — соответствующие значения объемов выборок. Множество $M_n(t)$ определяется тогда как $M_n(t) = \bigcup_i n_i(t)$. Значение $n_i(t) = m$ как элемент

множества M_n может появиться в различные моменты времени не более чем $T - m + 1$ раз. Если оно появилось именно столько раз, то в 100% случаев (с достоверностью $\alpha = 1$) объем выборки m удовлетворяет условию принадлежности корреляции множеству $[1; 1 - \delta]$.

Зададим некоторое $0 \leq \alpha \leq 1$. Пусть значение $n_i(t) = m$ появилось k раз. Если $k / (T - m + 1) \geq \alpha$, то пересечение соответствующей части $\bigcap_{t=1}^T M_n^\alpha(t)$ будем считать непустым с достоверностью α . Взяв затем объединение всех таких непустых пересечений, получим множество (не обязательно связное) объемов выборки, на каждом из которых удовлетворяется условие наличия корреляционной связи на выбранном уровне значимости. Далее для определенности берется тот объем выборки n_{opt} , для которого достоверность, т.е. отношение $k / (T - m + 1)$ выше, а при равных значениях таких отношений — наибольший из объемов.

Минимальное значение $\delta(\alpha)$, при котором вышеописанное объединение пересечений не пусто, и будет представлять абсолютную ошибку в определении корреляции, а величина $1 - \delta(\alpha)$ — наименьшую из возможных оценок максимума корреляции. Относительная ошибка в определении максимума корреляции будет тогда не больше, чем $\delta / (1 - \delta)$.

Устойчивость корреляционной связи по отношению к объему выборки характеризуется достоверностью α , которую желательно сделать как можно больше. Однако с увеличением α начинает возрастать и мини-

мальное $\delta(\alpha)$, так что оптимальным следует считать такое значение α , при котором

$$U^2 = \delta^2 + (1 - \alpha)^2 \rightarrow \min. \quad (4.6)$$

Подчеркнем, что критерий (4.6) не обобщает критерий (4.4), а уточняет объем, по которому следует вычислять выборочную корреляцию. Величина критерия U не является уровнем значимости корреляционной связи, т.е. его числовое значение не имеет самостоятельного смысла, а важность имеют лишь аргументы, при которых этот критерий минимален. Доверительная вероятность того, что корреляция при выбранном объеме n_{opt} принадлежит промежутку $[1; 1 - \delta(\alpha)]$, определяется выборочным распределением $F(R, n_{opt})$, и по построению равна α .

Описанный подход может быть обобщен для определения наиболее вероятной корреляции между двумя рядами. Зададим некоторое произвольное число $-1 < r < 1$ и рассмотрим отрезок, его содержащий, шириной δ :

$$\Delta_\delta(r) = [a; b] \subset [-1; 1], \quad b - a = \delta. \quad (4.7)$$

Для каждого r рассматриваются те и только те значения $R(n, t)$, которые попали в отрезок $\Delta_\delta(r)$. Пусть $\{n_i(t)\}$ — соответствующие значения объемов выборок. Корреляцию R между двумя рядами на промежутке времени $[1; T]$ называем достоверной по объему выборки на уровне β в промежутке $\Delta_\delta(r)$, если пересечение $\bigcap_{t=1}^T M_n^\beta(t)$ не менее чем

$[\beta(T - n(t) + 1)]$ носителей $n(t)$ значений корреляции этих рядов как функции объема выборки, не пусто. Из всех возможных пар a, b , для которых это пересечение не пусто, выбираются те, которые определяют отрезок (4.7) наименьшей длины. Варьируя затем величину β , находим минимум критерия (4.6): $U^2(r) = \delta^2 + (1 - \beta)^2 \rightarrow \min$. В результате находим $\beta(r)$, $\delta(\beta)$ и $n_{opt}(r)$. Наиболее достоверной будем считать ту корреляцию, для которой уровень доверия наибольший:

$$r_{opt} = \arg \max \beta(r). \quad (4.8)$$

Итак, построена методика определения нестационарной корреляционной связи между двумя рядами, которая формально без изменений обобщается на случай большего числа случайных величин. Для краткости изложения оптимизация проведена для фиксированного лага между рядами, поскольку методика не зависит от этого параметра. Если требуется определить также и наиболее достоверный лаг, то, параметризуя результат (4.8), находим максимальное значение из наибольших уровней доверия, которому и отвечает этот лаг.

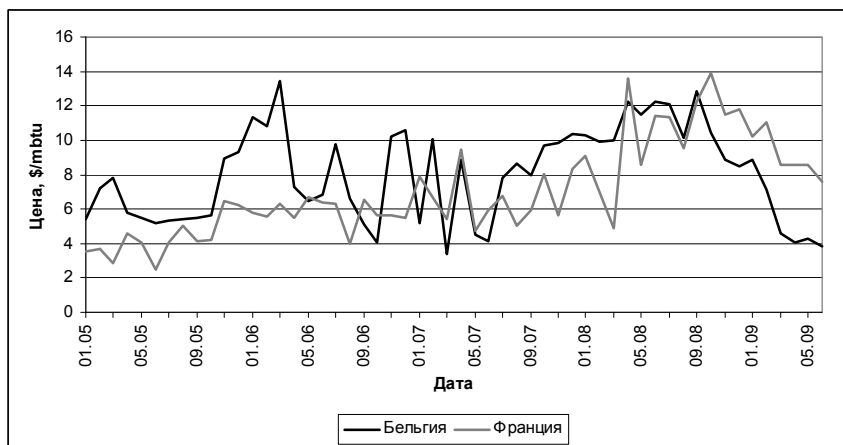


Рис. 8.4.1 — Динамика среднемесячных контрактных цен на СПГ в Бельгии и Франции

Покажем применение этой методики на примере ценовой динамики европейского рынка сжиженного природного газа (СПГ). Именно, рассмотрим в качестве иллюстративного примера задачу о коррелированности цен на СПГ для рынков Франции и Бельгии. Графики средневзвешенных месячных цен импортируемого СПГ на терминалах этих стран приведены на рис. 8.4.1. Другие примеры анализа цен на мировых газовых рынках содержатся в [27].

Корреляция между ценами на этих двух рынках в зависимости от объема выборки показана на рис. 8.4.2 для момента времени на июнь 2009 г..

Максимальное значение корреляции было достигнуто на объеме в 9 месяцев и оказалось равным 0,97. С увеличением объема выборки корреляция

в целом убывает, но, начиная с объема в 24 месяца, стабилизируется в полосе $0,3 \div 0,5$.

Непосредственно по графику рис. 8.4.2 нельзя сказать, какова корреляция между этими рынками, поскольку в один и тот же момент времени, но на разных объемах, могут быть сделаны различные выводы. Интерес, естественно, представляет ситуация, когда корреляция статистически значима. В частности, рис. 8.4.2 позволяет предложить такую «рабочую» гипотезу: рынки высоко коррелированы по данным на объеме 9 месяцев, а на промежутке в 24 месяца корреляция между ними снижается до мало-значимой величины. Если окажется, что и в другие моменты времени наблюдается похожая картина, то такая гипотеза станет уже рабочим инструментом анализа.

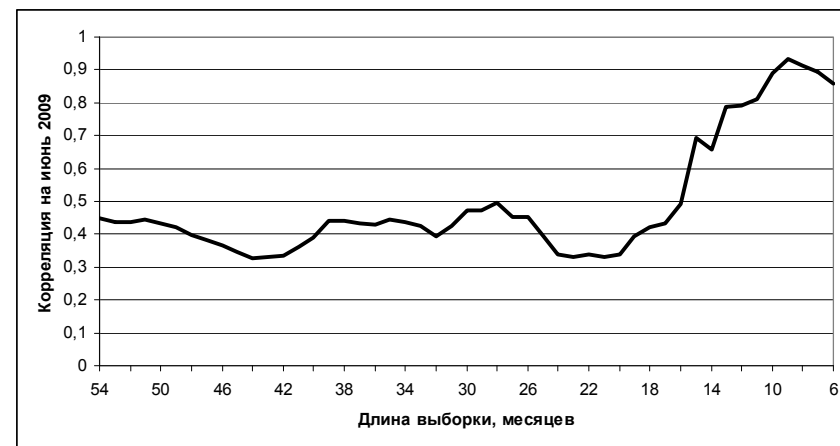


Рис. 8.4.2 — Корреляция цен на СПГ для пары «Франция-Бельгия» в зависимости от объема выборки для июня 2009 г.

Возьмем тогда объем выборки, на котором корреляция была максимальна в данный момент времени (т.е. $n = 9$), и пройдем с ним по другим моментам времени. Соответствующий график значений $R(9, t)$ показан на рис. 8.4.3. Видно, что выдвинутая наивная гипотеза оказалась грубо неверной: поведение сильно нестационарно, корреляция на фиксированном объеме заведомо не максимальна во все моменты времени, причем в ее изменениях не просматривается даже и сезонной периодичности.

Следовательно, максимум корреляции и соответствующий объем выборки надо определять в каждый момент времени. Эмпирические распре-

деления максимумов корреляций $f(r)$ и объемов выборки $F(n)$, отвечающих этим максимумам, показаны на рис. 8.4.4-8.4.5.



Рис. 8.4.3 — Корреляция цен СПГ на рынках Франции и Бельгии на объеме в 9 данных в различные моменты времени

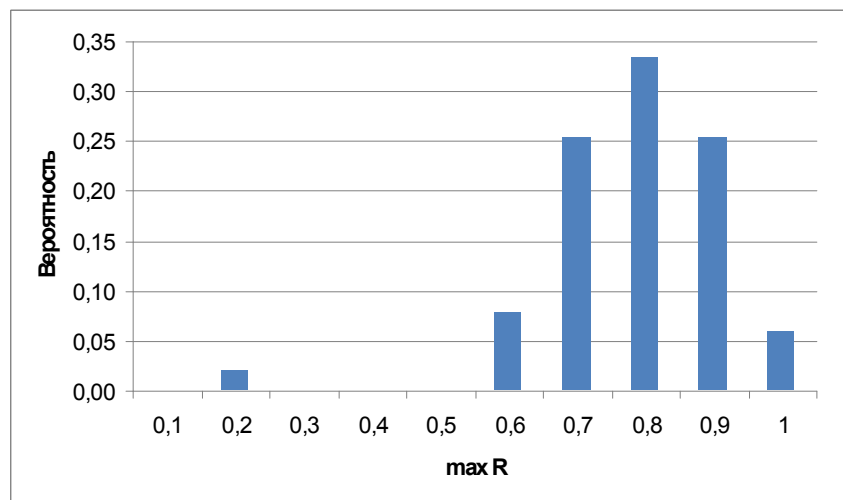


Рис. 8.4.4 — Распределение максимумов выборочной корреляции

Распределение корреляций показывает, что максимальная корреляция между данными рядами не всегда больше, чем 0,9. Чаще всего, с вероятностью 0,34, она находится в интервале 0,7-0,8. Кроме того, в отдельные моменты времени она находится в пределах 0,5-0,6 и даже 0,1-0,2. Среднее значение максимума составило $\bar{r} = 0,73$, причем $\sigma_r = 0,12$.

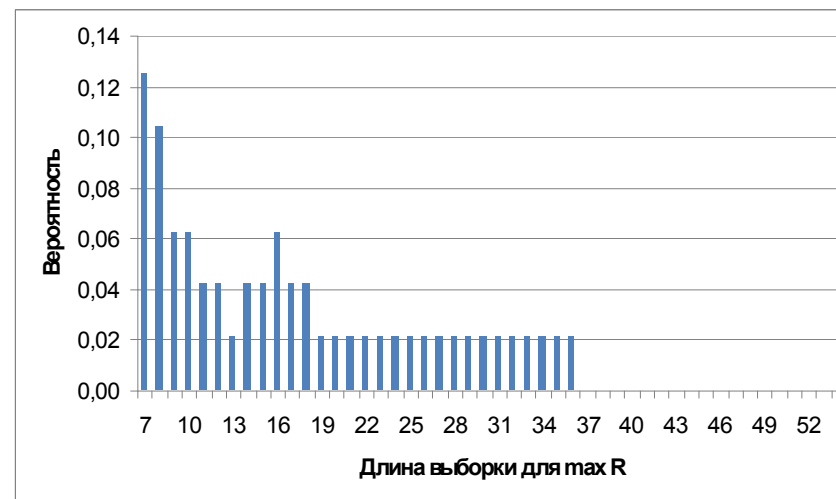


Рис. 8.4.5 — Распределение длин выборок для максимумов выборочной корреляции

Распределение объемов выборки имеет очень большой разброс. Объемы в 6 и 7 данных самые вероятные, но эта вероятность, тем не менее, незначительна. На промежутке от полутора до трех лет распределение объемов равномерно. Среднее значение объема, отвечающего максимальной корреляции, равно $n = 17$, а среднее квадратичное отклонение равно $\sigma_n = 9$.

Проведем оценку достоверности корреляционной связи методом «О-критерия». Расчет показал, что коэффициент корреляции между рядами $r(t)$ и $n(t)$ равен $\gamma = -0,52$, а средний коэффициент чувствительности равен $\bar{\lambda} = 0,69$. Тогда по формуле (4.4) получаем, что значимость оценки корреляции числом $\bar{r} = 0,73$ равна $O = 0,35$. Подчеркнем, что 0,35 — это совокупная значимость, учитывающая как разброс собственно кор-

реляции, так и нестационарность объема выборки для расчета максимальной корреляции. Поскольку дисперсия корреляции равна $\sigma_r = 0,12$, то полученный результат означает, что вероятность принадлежности корреляции промежутку $[\bar{r} - \sigma_r; \bar{r} + \sigma_r] = [0,61; 0,85]$ равна величине $1 - O = 0,65$. Это показывает, что, как и ожидалось, рынки коррелированы весьма слабо.

Применим теперь метод «U-критерия» (4.6) для определения одно-стороннего доверительного интервала и уточнения объема выборки, на котором следует вычислять корреляцию. Так, оказалось, что если потребовать, чтобы корреляция была не ниже 0,8 (т.е. $\delta = 0,2$), то оптимальным будет объем выборки $n = 6$, но при крайне низком уровне доверия $\alpha = 0,14$. Значению $\delta = 0,3$ отвечает объем $n = 10$ и уровень доверия $\alpha = 0,31$. Наилучшим является вариант, когда $\delta = 0,4$. Отметим, что величина $1 - \delta$ практически совпадает с границей найденного выше доверительного интервала для \bar{r} . При этом на уровне доверия $\alpha = 0,62$ объем выборки определяется величиной $n_{opt} = 15$, в отличие от среднего значения максимума $\bar{n} = 17$.

Итак, в результате применения данной методики удалось корректно определить уровень значимости и величину нестационарной корреляционной связи. Графики цен на рис. 8.4.1 показывают, что в среднем на рассматриваемом промежутке времени корреляция должна быть невысока, что и подтверждается малым уровнем доверия $1 - O$. Графики же на рис. 8.4.2-8.4.3 демонстрируют нетривиальность самой постановки задачи, когда ни объем выборки, ни сама корреляция не образуют стационарных рядов.

Выше мы обсуждали одновременную корреляцию, но метод может быть применен и к анализу лаговой корреляции между рядами. Ввиду практической значимости лаговой корреляции проанализируем, какой рынок из двух рассматриваемых следует за другим, с каким лагом, и какова точность корреляционной связи.

Будем для определенности считать, что в паре «Франция-Бельгия» первая страна идет вслед за второй с лагом в l месяцев. Отрицательный лаг будет означать, что, наоборот, вторая страна идет вслед за первой.

Применим для анализа корреляционной связи «O-критерий» (4.4). Проводя расчеты по той же схеме, что и выше, но для случаев, когда один объем выборки сдвинут относительно другого на l месяцев, получаем за-

висимость средней по времени максимальной по объему выборки корреляции от лага, а также значимость O оценки величины корреляции, которые показаны на графиках рис. 8.4.6.

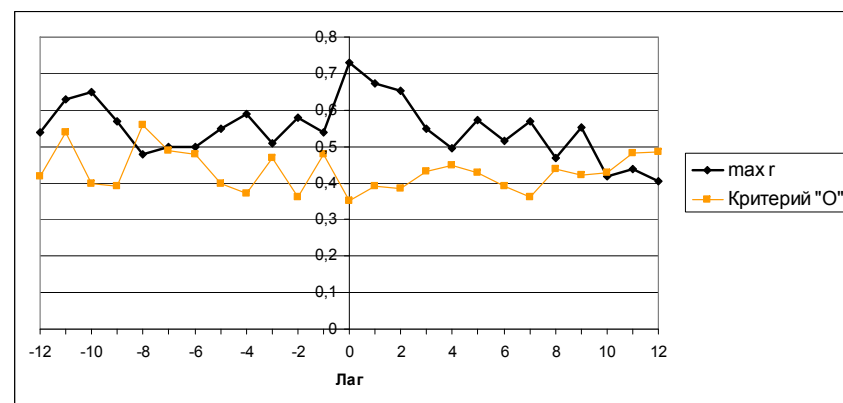


Рис. 8.4.6 — Средние по времени максимумы выборочной корреляции и значимость их оценки в зависимости от лага для рынков СПГ «Франция-Бельгия»

Проведенный анализ позволяет сделать вывод о том, что оба рынка «управляются» набором одновременных факторов (например, цена нефти, курс доллара к евро), причем рынки слабо зависят один от другого. Это следует из того, что максимум корреляции достигается при нулевом лаге, и достоверность этой оценки наибольшая, т.к. критерий значимости ошибки минимальный. С увеличением лага в обе стороны корреляционная связь ослабевает и возрастает уровень значимости, т.е. снижается точность оценки корреляции. Но даже максимальная корреляция 0,73 не является настолько большой, чтобы можно было говорить о статистически значимой связи между рынками, тем более, что значимость этой связи равна 0,35, а не 0,05, как принято в достоверных оценках.

Снижение корреляции с увеличением лага происходит немонотонно. Это присуще вообще всем сырьевым рынкам: имеется определенная «рефлексия» на ситуации, которые развивались на других рынках. Эта рефлексия несимметрична. Рынок Франции реагирует на ситуацию в Бельгии через пять месяцев, а рынок Бельгии на ситуацию во Франции — через два и четыре месяца, а также через десять месяцев.

Если провести анализ лаговой корреляции для всех пар стран-импортеров СПГ, то можно выделить рынки, которые теснее других свя-

заны между собой, и определить точность такой оценки. Этот анализ можно провести для любого набора временных рядов. Тем самым построенная методика может быть положена в основу практического применения корреляционного анализа в нестационарном случае.

8.5. Метод численного распознавания нелинейной динамической системы

Предположим, что между элементами x и y двух временных рядов имеется прямая функциональная связь, т.е. $y = \varphi(x)$. Тогда, построив совместное распределение вероятностей $F_N(x, y, t)$ по выборке длины N в момент времени t , мы обнаружим, что с точностью $1/\sqrt{N}$ в норме суммируемых функций оно не меняется, и его носитель находится в квадратах $y_i = \varphi(x_j)$, в соответствии с разбиением гистограммы. Точность, с которой мы можем говорить о такой функциональной связи, равна точности позиционирования случайных величин в классовых интервалах, т.е. $1/n$. Действительно, точность оценки функциональной связи — это мера носителя совместного распределения в единичном квадрате. При фиксированной мелкости разбиения эта доля, т.е. точность оценки, не может быть лучше, чем $1/n$ (n квадратов с площадью $1/n^2$). Уровень достоверности полученной оценки равен интегралу от плотности распределения по выбранной доле носителя. Поскольку в данном случае вне отмеченных квадратов нет точек носителя совместного распределения, уровень достоверности равен единице.

Если функциональной связи нет, то при фиксированном номере j интервала по первому аргументу мы обнаружим отличные от нуля значения функции $F_N(j, i, t)$ для нескольких номеров i интервалов по второму аргументу. При этом с увеличением длины выборки носитель совместного распределения занимает все большую долю области разбиения гистограммы. Это означает, что путем потери точности можно получить достоверную оценку функциональной связи даже в отсутствие таковой, но будет ли это удовлетворять исследователя? Насколько точно нужно позиционировать искомое значение, чтобы и вероятность его принадлежности определенному интервалу была не исчезающе малой, и сам интервал существенно отличался бы от всего множества значений случайной величины? Как и в главе V, вместо априори задаваемого уровня значимости сле-

дует ввести согласованный критерий совместной оценки точности и уровня значимости.

Пусть $\delta = \int_{\Omega} dx dy$ есть мера множества $\Omega(x, y)$, принадлежащего

носителю совместного распределения, на котором можно однозначно говорить о связи между x и y . Величина δ будет точностью, с которой установлена эта связь, а величина $\alpha = \int_{\Omega} F(x, y) dx dy$ будет давать уро-

вень достоверности найденной связи. И δ , и α зависят от множества $\Omega(x, y)$. Тогда введем функционал, минимизирующий совокупную ошибку оценки корреляционной связи (не обязательно линейной), которая находится на том множестве $\Omega(x, y)$, где

$$U^2 = (1 - \alpha(\Omega))^2 + \delta(\Omega)^2 \rightarrow \min. \quad (5.1)$$

Заметим, что условие (5.1) не позволяет в общем случае однозначно определить множество $\Omega(x, y)$. Для унимодальных распределений $F(x, y, t)$ это множество при фиксированном x содержит $y = \arg \max F(x, y)$. Будем поэтому для определенности считать, что $\Omega(x, y)$ содержит полосу разбиения гистограммы, содержащую локальные (при фиксированных номерах x -ячеек) максимумы распределения $F_N(j, i, t)$. Собственно значением $i(j)$ будет называться номер ячейки, содержащей условное среднее значение номера i по множеству ячеек с j -ым номером вертикальной полосы, которое с точностью δ охватывает $\arg \max F(j, i)$.

Функционал U зависит от длины выборки N . В стационарном случае с увеличением N происходит лишь уточнение множества $\Omega(x, y)$ и, соответственно, уменьшение значения U . В нестационарном случае увеличение длины выборки сверх оптимального значения может привести к увеличению функционала U . Пусть $U(N, t)$ есть результат оптимизации (1) в данный момент времени t по выборкам произвольных объемов, при котором находятся локально-оптимальные значения $\delta(N, t)$ и $\alpha(N, t)$. Тогда

$$N_{opt}(t) = \arg \min U(N, t). \quad (5.2)$$

Подчеркнем, что функциональная связь, определяемая множеством $\Omega(x, y; t)$, полученном в результате оптимизации (5.1), в разные моменты времени может быть совершенно различной. Также и оптимальные длины $N_{opt}(t)$ образуют в совокупности некоторое распределение с плотностью $\nu(N)$. В результате глобально оптимальным за рассматриваемый промежуток времени будет некоторое среднее из оптимальных длин и некоторая средняя функциональная связь. Точность этой связи будет определяться ее условной дисперсией (при условии, что имеется заданная неточность в определении локально-оптимальной длины выборки) и чувствительностью функциональной связи к длине выборки.

По этой методике можно определить, насколько точно конкретный временной ряд моделируется с помощью ДС определенной размерности. Для этой цели надо построить выборочную плотность многомерной функции распределения этого ряда, разбив множество принимаемых им значений на количество ячеек в соответствии с оптимальным разбиением гистограммы (п. 5.3).

Например, чтобы определить возможность представления данного временного ряда x_n посредством одномерной ДС вида $x_n = g(x_{n-1})$, надо сформировать двумерный массив совместных плотностей вероятности $f_T(x_{n-1}, x_n)$ по выборке объема T . Если оказалось, что с принятой точностью для каждого интервала разбиения, содержащего x_n , существует только один интервал, содержащий x_{n-1} , то временной ряд может быть смоделирован автономной ДС первого порядка. Соответствующий график функции $g(x)$ является носителем характеристического множества ненулевых значений построенной плотности $f_T(x_{n-1}, x_n)$.

Если для интервала Δx_{n-1} таких интервалов Δx_n хотя бы для одного n оказалось несколько, то при выбранной мелкости разбиения такая ДС либо не автономна, либо имеет более высокую размерность (если, конечно, это вообще ДС). В таком случае дальнейшие действия зависят от требований к точности модели. Всегда можно так ухудшить точность, что любой ряд будет описываться одномерной ДС (например, регрессионной моделью), но это, разумеется, не самое интересное решение. Более продуктивным является выяснение того, что система одномерна, но не автономна. Этот вариант реализуется в том случае, если множество значений временного ряда можно разбить на непересекающиеся выборки, покрывающие все это множество, объемы которых достаточны для того, чтобы

строить выборочные парные плотности вероятностей, причем внутри этих выборок можно построить свои одномерные ДС. Если это оказалось невозможным (чаще всего по причине того, что требуемое разбиение оказывается разбиением на выборки объема $T = 1$), то следует перейти к рассмотрению плотностей вероятностей следующей размерности, именно, к $f_T(x_{n-2}, x_{n-1}, x_n)$. Как и в предыдущем случае, если оказалось, что для каждой пары интервалов разбиения, содержащих x_{n-2} и x_{n-1} , существует только один интервал, содержащий x_n , то временной ряд может быть смоделирован автономной ДС второго порядка, и т.д.

После того, как определена размерность ДС, можно переходить к задаче определения конкретной функциональной зависимости, т.е. наиболее подходящей интерполяции. Численное решение этой задачи для нелинейных систем весьма трудно, поскольку нелинейность при фиксированной точности вычислений может существенно исказить совместное распределение образов и прообразов: если величины x и $x + \varepsilon$ принадлежат одному промежутку разбиения, то значения $g(x)$ и $g(x + \varepsilon)$ могут принадлежать разным промежуткам даже для линейных систем, как, например, для геометрической прогрессии. Ширина полосы, в которой будет заключен возможный график функции $g(x)$, очень часто в несколько раз превосходит точность, с которой случайные величины размещаются по промежуткам области определения двумерной (или многомерной) ВПФР. В случае не слишком сильной нелинейности можно пренебречь некоторым ухудшением точности, если изначально величина допустимой ошибки была задана достаточно малой.

Рассмотрим пример логистической системы $x_{n+1} = 4x_n(1 - x_n)$ с точностью в определении x_n , равной 0,01. Ее двумерная выборочная плотность $f_T(x_n, x_{n+1})$ по выборке в $T = 10000$ точек приведена на рис. 8.5.1.

Из рис. 8.5.1 видно, что множество аргументов, при которых двумерная ВПФР отлична от нуля, образует кривую параболического типа, но вычисления показывают, что ширина заключающей ее полосы равна не 0,01, а 0,04. Это означает, что с ухудшением точности до 0,04 временной ряд, порожденный этой системой, может быть численно опознан по виду двумерной ВПФР как траектория одномерной ДС. Уравнение характеристического множества легко строится по координате вершины и имеет

вид параболы. Таким же образом могут быть опознаны и другие ДС похожего вида с невысоким порядком нелинейности и малой размерностью.

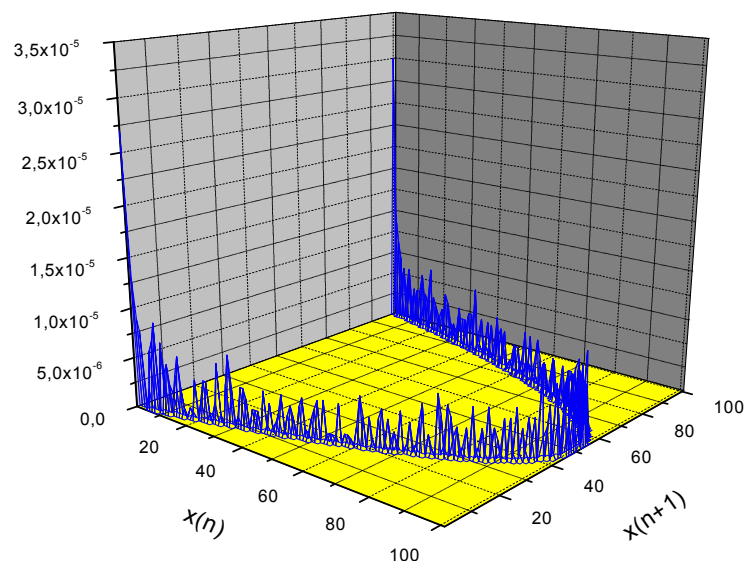


Рис. 8.5.1 — Двумерная ВПФР логистической системы

Нахождение носителя совместной k -мерной ВПФР $f_T(x_{n-k+1}, \dots, x_{n-1}, x_n)$ означает фактически возможность исключения нелинейной корреляционной связи вида $x_n = g(x_{n-k+1}, \dots, x_{n-1})$, что чрезвычайно важно на этапе пред-обработки временного ряда. На практике определение функциональной зависимости $x_n = g(x_{n-k+1}, \dots, x_{n-1})$ может быть приближенно найдено численно следующим простым способом. Поскольку реально при численных расчетах используются не непрерывные функции, а их дискретные аналоги в k -мерном фазовом пространстве, то ВПФР определена как гистограмма на k -мерных кубах. Тогда для каждого промежутка разбиения Δx_n оси x_n находится тот $k-1$ -мерный куб, на котором значение совместной k -мерной ВПФР максимально. Гиперплоскости, соединяющие центры ото-

бранных k -мерных кубов, и представляют собой линейную интерполяцию зависимости $x_n = g(x_{n-k+1}, \dots, x_{n-1})$. Доверительная вероятность того, что связь между последовательными значениями ряда имеет вид именно $x_n = g(x_{n-k+1}, \dots, x_{n-1})$, дается интегралом от $f_T(x_{n-k+1}, \dots, x_{n-1}, x_n)$ по отмеченным k -мерным кубам, а ошибка интерполяции равна доле суммарного объема отмеченных k -мерных кубов в объеме носителя совместного распределения. Если доверительная вероятность слишком мала, то надо расширить k -мерную «полосу» объемов, взяв все соседние по граням кубы к отмеченным ранее. Это повысит доверительную вероятность и, естественно, ухудшит точность интерполяции. Такая ситуация типична для нахождения согласованных доверительных интервалов, но уже в многомерном случае.

Заметим в заключение этой главы, что к регрессионным моделям — линейным или нелинейным — приводит трактовка появления случайной величины как следствие зашумленности некоторой неслучайной величины. Это — физическая точка зрения на возникновение шума, сводящаяся в конце концов к погрешности инструмента измерения. Но существует также и другой шум, состоящий в неупорядоченности проявлений многомерной динамической системы. При этом сами значения бывают известны абсолютно точно, как, например, выпавшие грани кубика или цена на тот или иной финансовый инструмент на бирже. Динамическая неопределенность, появляющаяся вследствие нерегулярности последовательно воспринимаемых значений, возможно, что и не случайных порознь величин, имеет несколько иной характер, чем шум, добавленный к стационарному среднему значению.

Например, пусть имеется два стационарных временных ряда $\{x_n\}$ и $\{y_n\}$, средние значения которых равны \bar{x} и \bar{y} , дисперсии равны σ_x^2 и σ_y^2 , а корреляции с любым лагом, как и автокорреляции, равны нулю. Тогда ряд, составленный из величин $z_n = x_n + y_n$, будет иметь среднее $\bar{z} = \bar{x} + \bar{y}$ и дисперсию $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$, и также нулевую автокорреляцию с любым лагом. Если же мы рассмотрим ряд z_n , составленный из последовательно перемежающихся элементов x_n и y_n , т.е.

$z_{2n-1} = x_n$, $z_{2n} = y_n$, то в результате получим в общем случае автокоррелированный ряд. Легко проверить, что в этом случае

$$\sigma_z^2 = \frac{\sigma_x^2 + \sigma_y^2}{2} + \frac{(\bar{x} - \bar{y})^2}{4},$$

$$\text{cov}(z_n, z_{n-1}) = -\text{cov}(z_n, z_{n-2}) = -\frac{(\bar{x} - \bar{y})^2}{4}.$$

В более общем случае, если

$$\text{cov}(x_n, y_n) = r_0, \quad \text{cov}(x_n, y_{n-1}) = r_{-1}, \quad \text{cov}(x_n, y_{n+1}) = r_1,$$

то для таким же образом построенного ряда z_n получается

$$\text{cov}(z_n, z_{n-1}) = -\frac{(\bar{x} - \bar{y})^2}{4} + \frac{r_0 + r_{-1}}{2},$$

$$\text{cov}(z_n, z_{n-2}) = \frac{(\bar{x} - \bar{y})^2}{4} + \frac{r_0 + r_1}{2}.$$

Если теперь вместо двух белых шумов рассмотреть двухкомпонентный динамический белый шум, каждая компонента которого ведет себя, например, по типу логистической системы, то при определенной упорядоченности действия каждой компоненты можно, по крайней мере, в принципе, разделить эти компоненты и понизить дисперсию процесса, разложив его не в сумму, а в подходящую очередность двух или более процессов. Этот метод может быть использован в случае, когда АКФ ряда после удаления сезонных и иных трендовых компонент осциллирует около нуля с небольшой амплитудой. Тогда остаток может быть смоделирован как упорядоченная последовательность ортогональных белых шумов.

Глава IX. Кинетический метод прогнозирования

9.1. Уравнение эволюции выборочной плотности распределения

Исследуем теперь вопрос о том, какими уравнениями имеет смысл описывать эволюцию эмпирического распределения. Из анализа горизонтной статистики нестационарных временных рядов следует, что для них имеются две различных временных области, в каждой из которых действует некоторая своя эволюция. Первая область — это область \mathcal{E} -стационарности ВПФР, а вторая — все моменты времени, превосходящие τ . Будем рассматривать эволюцию в первой временной области.

С одной стороны, поскольку ВПФР строится по конечной выборке данных, то уравнение ее эволюции по своему существу должно быть дискретным. С другой стороны, качественные черты модели более заметны в непрерывном описании, дискретная форма которого будет представлять собой численную схему расчетов в каждом конкретном случае. Поэтому для удобства восприятия чаще будет использоваться запись уравнения эволюции ВПФР в дифференциальной форме.

Предположительно, модель эволюции ВПФР должна основываться на некотором аналоге уравнения Лиувилля, выражающем закон сохранения вероятности со временем. Тем самым выборочной функции распределения будет локально по времени сопоставлена некоторая динамическая система. Однако, имея в виду применение теории к нестационарным процессам, следует указать на существенное отличие того уравнения неразрывности, которое будет строиться в фазовом пространстве координат (значений ряда $x(t)$), скоростей (приращений этого ряда $v(t) \equiv \dot{x} = x(t+1) - x(t)$), и, возможно, ускорений высших порядков, от уравнения Лиувилля классической статистической механики.

В случае динамических систем, заданных аналитически, во многих случаях удастся в явном виде определить динамически-инвариантную меру. В случае же численной реализации некоторой системы (возможно, что и не динамической) в виде временного ряда априорно выбираемая мера не обязана быть инвариантной. Напомним, что вероятностная трактовка уравнения Лиувилля как уравнения, сохраняющего нормировку, есть просто переформулированная теорема Лиувилля классической механики о сохранении фазового объема для определенного класса динамических систем (консервативных).

Для статистической выборки фазовый объем, определяемый разностями производными, доступными по этой выборке, может и не сохраняться точно. Чтобы устранить это неудобное обстоятельство, мы и вводим требование равномерной ограниченности временного ряда. Тем не менее, выборочная скорость изменения значений ряда даже и в этом случае может не обладать требуемым свойством, поскольку численное определение дивергенции фазовой скорости проводится по конечному, а не бесконечному набору данных. В то же время ВПФР любой размерности нормирована на единицу по построению, поэтому оператор, переводящий ВПФР из одного момента времени в другой в том же самом фазовом пространстве, должен иметь вид закона сохранения нормировки. Но изучаемый временной ряд не обязательно порожден динамической системой, для которой и выводится уравнение Лиувилля. Следовательно, необходимо согласовать «выборочную фазовую скорость», появляющуюся в математической статистике, со скоростью в статистической механике, для чего следует провести подробный анализ того, как возникает уравнение эволюции для выборочной плотности функции распределения.

Итак, предположим, что имеется достаточно большое количество данных $\{x(t)\}$, анализ которых позволяет определить оптимальный объем T выборки для построения ВПФР $f_T(x, t)$ и описания ее эволюции, скажем, на один шаг вперед. Рассматривая набор таких ВПФР $f_T(x, t)$, $f_T(x, t-1)$, ... в предшествующие моменты времени, можно определить соответствующие эмпирические изменения этих плотностей за один шаг:

$$\frac{\partial f_T(x, t-k)}{\partial t} = f_T(x, t-k+1) - f_T(x, t-k), \quad k=1, 2, \dots \quad (1.1)$$

Частная производная по времени в (1.1) написана для удобства восприятия. Фактически это есть обозначение для выражения, стоящего в правой части (1.1).

Допустим, что знаний о таких изменениях первого порядка по времени оказалось достаточно для того, чтобы сложилось представление об эволюции ВПФР за один шаг по времени с точностью ε , для которой и был найден оптимальный объем выборки T . Определим тогда подходящую «скорость» изменения ВПФР $u_T(x, t)$ так, чтобы изменение (1.1) можно было записать в виде уравнения

$$\frac{\partial f_T(x, t)}{\partial t} + \frac{\partial}{\partial x} u_T(x, t) f_T(x, t) = 0. \quad (1.2)$$

Скорость $u_T(x, t)$, удовлетворяющую уравнению (1.2) совместно с (1.1), будем называть эмпирической лиувиллевой скоростью (ЭЛС).

Далее считаем объем выборки T фиксированным, равным оптимальному объему, определенному на предыдущих этапах анализа, и соответствующий индекс для краткости опускаем.

ЭЛС можно определить для каждой дискретизации области значений $x(t)$. При этом удобно считать, что на границах этой области ВПФР равна нулю. Для равномерно ограниченного ряда $x(t)$ этому условию всегда можно удовлетворить, введя нужное количество дополнительных граничных ячеек разбиения, в которых плотность распределения равна нулю во все моменты времени.

В результате ЭЛС в каждой ячейке разбиения, где существуют значения временного ряда, определяется из соотношений (1.1) и (1.2) по формулам

$$u(i+1, t) = \frac{u(i, t)f(i, t) - f(i, t) + f(i, t+1)}{f(i+1, t)}. \quad (1.3)$$

Если же в i -ой ячейке плотность распределения $f(i, t)$ оказалась равной нулю, то формально скорость в такой ячейке не определена. В фиктивных граничных ячейках скорость естественно положить равной нулю. Во внутренних ячейках с нулевой плотностью скорость также можно положить равной нулю. Это не приведет к противоречиям, поскольку уравнение (1.2) содержит скорость в комбинации с произведением на соответствующую плотность.

Из (1.3) следует, что скорость $u(i, t)$ изменения ВПФР в i -ой ячейке в момент t может быть представлена в виде

$$u(i+1, t)f(i+1, t) = \sum_{k=1}^i (f(k, t+1) - f(k, t)) = F(i, t+1) - F(i, t). \quad (1.4)$$

Таким образом, правая часть (1.4) представляет собой изменение по времени выборочной функции распределения в ячейке, предшествующей той, в которой вычисляется значение эмпирической скорости. Видно, что скорость изменения ВПФР в данный момент времени определяется значениями распределения в этот и последующий моменты. Это означает, что в

текущий момент времени t ЭЛС известна для предыдущего момента $t-1$, что следует учитывать при составлении эволюционной модели. Именно, чтобы определить ВПФР в момент времени $t+1$, надо построить ЭЛС в момент времени t .

Выясним, какой статистический смысл можно придать введенной эмпирической скорости изменения ВПФР. Для этого рассмотрим уравнение эволюции первого выборочного момента, определяемого как

$$m(t) = \langle x \rangle_t = \int x f(x, t) dx \quad (1.5)$$

Из (1.2) следует, что эволюция первого момента записывается в виде

$$\frac{dm(t)}{dt} = \int x \frac{\partial f(x, t)}{\partial t} dx = - \int x \frac{\partial}{\partial x} u(x, t) f(x, t) dx.$$

После интегрирования по частям с учетом того, что в граничных ячейках плотность распределения равна нулю, получаем

$$\frac{dm(t)}{dt} = \int u(x, t) f(x, t) dx = \langle u \rangle_t = U(t), \quad (1.6)$$

где $U(t)$ есть среднее значение эмпирической скорости

$u(x, t)$ по распределению, которое задано в момент времени t .

С другой стороны, то же самое изменение первого выборочного момента, вычисляемое непосредственно по элементам выборки временного ряда, есть

$$\begin{aligned} \frac{dm(t)}{dt} &= m(t+1) - m(t) = \\ &= \frac{1}{T} \sum_{k=1}^T x(t-T+k+1) - \frac{1}{T} \sum_{k=1}^T x(t-T+k) = \frac{1}{T} \sum_{k=1}^T \dot{x}(t-T+k), \end{aligned} \quad (1.7)$$

т.е. $U(t)$, как показывает последнее равенство, есть среднее значение выборочной скорости, определяемой по приращениям значений исходного временного ряда.

Выражения (1.6) и (1.7) представляют среднее значение одной и той же скорости. Но, в отличие от (1.6), среднее в (1.7), будучи скоростью \dot{x} , должно вычисляться не через распределение $f(x, t)$ значений ряда, а через совместное распределение ряда и его приращений. Введем тогда

двумерную ВПФР $F(x, \dot{x}, t)$ совместного распределения случайных величин x и \dot{x} исходного ряда и ряда его производной. Производная $\dot{x} = x(t+1) - x(t)$ трактуется как разность между значениями ряда в соседние моменты времени. Эта разность и есть аналог микроскопической скорости системы из T «частиц» (значений ряда). Для построения двумерной ВПФР $F(x, \dot{x}, t)$ в момент времени t необходимо иметь значение случайной величины в момент $t+1$, чтобы можно было построить микроскопическую скорость в момент времени t . Заметим здесь, что если ряд неэквилибриантный, то разностная скорость определяется как

$$\dot{x} = \frac{x(n+1) - x(n)}{t(n+1) - t(n)}. \quad (1.8)$$

Одномерная ВПФР $f(x, t)$ определяется затем по $F(x, \dot{x}, t)$ формулой

$$f(x, t) = \int F(x, \dot{x}, t) d\dot{x}. \quad (1.9)$$

Заметим, что среднее значение эмпирической скорости из правой части (1.7) равно

$$U(t) = \int \dot{x} F(x, \dot{x}, t) dx d\dot{x},$$

и эта же величина равна выражению $\int u(x, t) f(x, t) dx$ в (1.6).

Следовательно, средняя локальная скорость $u(x, t)$ определяется как

$$u(x, t) f(x, t) = \int \dot{x} F(x, \dot{x}, t) d\dot{x}. \quad (1.10)$$

Аналогия построенной эволюции ВПФР с динамическими системами в том, что если рассмотреть неавтономную ДС, задаваемую уравнением

$$\frac{dx}{dt} = u(x, t), \quad x(0) = \int x f(x, 0) dx, \quad (1.11)$$

то среднее значение этой ДС будет эволюционировать как среднее значение исходного временного ряда.

9.2. Модели замыкания эмпирического уравнения Лиувилля

Уравнение Лиувилля (1.2) пока еще не является замкнутым: кроме неизвестной плотности вероятности, относительно которой оно записано, в нем фигурирует и неизвестная скорость $u(x, t)$. Если исходить только

из требования сохранения нормировки ВПФР, то скорость $u(x, t)$ может быть, вообще говоря, любой функцией, поскольку на нормировку распределения ее конкретный вид не влияет.

Например, при прогнозировании на τ шагов вперед можно исходить из последовательно усложняющихся прогнозных моделей для ЭЛС. В простейшем варианте скорость во все последующие моменты времени может быть принята равной скорости в последний момент времени, т.е. $\hat{u}(x, t+k) = u(x, t-1)$, $k = 0, 1, \dots, \tau-1$. Более сложная модель использует среднее значение скорости по прошлому промежутку длительности τ :

$$\hat{u}(x, t) = \bar{u}(x, t-1) = \frac{1}{\tau} \sum_{t'=1}^{\tau} u(x, t-t'). \quad (2.1)$$

Как вариант, можно предложить модель с автокоррекцией, когда прогнозная скорость включает самое себя в усреднение вида (1.1):

$$\hat{u}(x, t) = \bar{u}(x, t) = \frac{1}{\tau} \sum_{t'=0}^{\tau-1} u(x, t-t') = \frac{\hat{u}(x, t)}{\tau} + \frac{1}{\tau} \sum_{t'=1}^{\tau-1} u(x, t-t'),$$

откуда следует, что

$$\hat{u}(x, t) = \frac{1}{\tau-1} \sum_{t'=1}^{\tau-1} u(x, t-t'), \quad \tau > 1. \quad (2.2)$$

Прогноз скорости $\hat{u}(x, t)$ на промежутке времени τ может быть построен с применением линейной регрессии скорости на время, если в качестве промежутка усреднения по времени взять промежуток $\Delta_{\tau}(t)$:

$$\hat{u}(x, t+k) = \bar{u}(x, t) + 6c(x, t) \frac{2k - \tau - 1}{\tau^2 - 1}, \quad 1 < \tau; \quad k \in [1, \tau];$$

$$c(x, t) = \frac{1}{2\tau} \sum_{t'=1}^{\tau} (u(x, t-t') - \bar{u}(x, t))(2t' - \tau - 1). \quad (2.3)$$

Вышеописанные модели основаны на эвристических соображениях, которые на практике в одних случаях могут быть полезными, а в других — нет. Однако важно, чтобы вычисленная плотность $f(x, t)$ была бы

неотрицательна, что накладывает на функцию $u(x, t)$ определенные ограничения. Чтобы более корректно подойти к вопросу замыкания уравнения Лиувилля (1.2), будем его решать совместно с уравнением (1.10). Полученная система все еще не является замкнутой в смысле кинетического подхода, поскольку возникшая в (1.10) двумерная ВПФР $F(x, \dot{x}, t)$ пока еще не определена как функция времени. Для нее также следует записать уравнение Лиувилля, в котором x и \dot{x} формально считаются независимыми переменными.

По аналогии с одномерным случаем, производную скорости по времени следует рассматривать в некотором усредненном смысле, для чего надо расширить пространство фазовых переменных. Тогда уравнение Лиувилля для $F(x, \dot{x}, t)$ должно иметь следующий вид, вытекающий из тех же представлений о сохранении нормировки:

$$\frac{\partial F}{\partial t} + \frac{\partial}{\partial x}(F\dot{x}) + \frac{\partial}{\partial \dot{x}}\left(F\left\langle \frac{\ddot{x}}{\delta t} \right\rangle\right) = 0. \quad (2.4)$$

Смысл этого уравнения определяется тем, как трактуется величина $\langle \ddot{x} / \delta t \rangle$. Если рассматривается динамическая система, для которой известна зависимость ускорения $\ddot{x} = w(x, \dot{x}, t)$ от координат, скоростей и времени, то следует считать, что $\langle \ddot{x} / \delta t \rangle$ и есть это ускорение. В таком случае из (2.4) вытекает

$$\frac{\partial F}{\partial t} + \frac{\partial}{\partial x}(F\dot{x}) + \frac{\partial}{\partial \dot{x}}(Fw) = 0. \quad (2.5)$$

Если же считать, что $\ddot{x} / \delta t$ является случайной величиной, то в уравнении Лиувилля аналог динамического ускорения если и можно ввести, то в некотором усредненном смысле, т.е. в виде $\langle \ddot{x} / \delta t \rangle$. Разумеется, это не «вывод» уравнения эволюции, а всего лишь эвристические соображения относительно того, каким могло бы быть кинетическое уравнение выборочной плотности функции распределения для нестационарного временного ряда.

Тем самым для определения величины $\langle \ddot{x} / \delta t \rangle$ требуется дальнейшее расширение фазового пространства. Введем среднее ускорение как функцию координаты, скорости и времени:

$$w(x, \dot{x}, t)F(x, \dot{x}, t) = \int \ddot{x} F_3(x, \dot{x}, \ddot{x}, t) d\ddot{x}. \quad (2.6)$$

Видно, что в этом случае возникает зацепление многочастичных распределений. Замыкание цепочки можно провести, задав априори выражение для интеграла от неизвестной величины $\int \ddot{x} F_3(x, \dot{x}, \ddot{x}, t) d\ddot{x}$ в правой части уравнения (2.6). Характерно, что такое замыкание совершенно не повлияет на вид уравнения для одномерной ВПФР в силу дивергентного характера слагаемых.

Если же нет оснований считать интеграл $\int \ddot{x} F_3(x, \dot{x}, \ddot{x}, t) d\ddot{x}$ известной величиной, то цепочка уравнений Лиувилля выписывается для ВПФР следующего порядка, и т.д. Однако, чем выше порядок ВПФР, тем труднее найти разумное основание для того или иного способа обрыва цепочки кинетических уравнений. Следовательно, желательно построить систему уравнений, замкнутую относительно одномерной ВПФР. Чтобы замкнуть уравнение Лиувилля (1.2), которое используется для прогнозирования ВПФР $f(x, t)$, следует независимо определить величину $u(x, t)$ как значение средней локальной скорости.

Для получения замкнутой теории будем исходить из определения локальной скорости (1.10) и уравнения эволюции (2.5). Дифференцируя (1.10) частным образом по времени, получаем эволюционное уравнение для средней локальной скорости:

$$\left(\frac{\partial u}{\partial t} - u \frac{\partial u}{\partial x} \right) f - u^2 \frac{\partial f}{\partial x} = - \frac{\partial(ef)}{\partial x} + Wf, \quad (2.7)$$

где введены новые величины $e(x, t)$ и $W(x, t)$ согласно формулам

$$e(x, t) = \langle \dot{x}^2 \rangle, \quad W(x, t) = \langle w(x, \dot{x}, t) \rangle. \quad (2.8)$$

Здесь угловыми скобками обозначено усреднение по всем фазовым переменным, кроме x , по распределению соответствующей размерности:

$$e(x, t) f(x, t) = \int \dot{x}^2 F(x, \dot{x}, t) d\dot{x}, \quad (2.9)$$

$$W(x, t) f(x, t) = \int w(x, \dot{x}, t) F(x, \dot{x}, t) d\dot{x}.$$

Таким образом, выясняется, что для определения эволюции первого момента двумерной ВПФР по скорости надо знать эволюцию ее второго момента по той же переменной и эволюцию среднего ускорения. Для этих величин аналогичными вычислениями получаем:

$$\left(\frac{\partial e}{\partial t} - e \frac{\partial u}{\partial x} \right) f - ue \frac{\partial f}{\partial x} = - \frac{\partial}{\partial x} \left(\langle \dot{x}^3 \rangle f \right) + \langle \dot{x} w \rangle f, \quad (2.10)$$

$$\left(\frac{\partial W}{\partial t} - W \frac{\partial u}{\partial x} \right) f - uW \frac{\partial f}{\partial x} = - \frac{\partial}{\partial x} \left(\langle \dot{x} \ddot{x} \rangle f \right) + Af, \quad (2.11)$$

где $A(x, t)$ есть среднее ускорение следующего порядка, определяемое через F_3 и F_4 :

$$A(x, t) f(x, t) = \int a(x, \dot{x}, \ddot{x}, t) F_3(x, \dot{x}, \ddot{x}, t) d\ddot{x} d\dot{x}, \quad (2.12)$$

$$a(x, \dot{x}, \ddot{x}, t) F_3(x, \dot{x}, \ddot{x}, t) = \int \ddot{x} F_4(x, \dot{x}, \ddot{x}, \ddot{\ddot{x}}, t) d\ddot{\ddot{x}}.$$

Видно, что точная система уравнений в принципе не может быть получена в замкнутой форме, поскольку в ней возникают зацепления более высокого порядка вида (2.12). Кроме того, появились и моменты более высокого порядка — третьего по скорости в уравнении (2.10).

Возникшая ситуация полностью аналогична классическому выводу уравнений гидродинамики из кинетических уравнений [6], чему была посвящена глава I. Следует каким-либо образом замкнуть моментную систему, т.е. определить независимо правые части эволюционных уравнений (2.7), (2.10) и (2.11). Это можно сделать различными способами, которые, однако, не могут быть основаны на точных математических следствиях, а должны исходить из более или менее правдоподобных практических предпосылок. Каждый из способов замыкания будет приводить к новой модели эволюции, т.е. к новому методу прогнозирования. Рассмотрим некоторые из таких способов.

Например, если рассматривать промежутки времени, на которых ВПФР квазистационарна, то естественным условием замыкания является нулевое ускорение: $w(x, \dot{x}, t) = 0$. Напомним, что на таких промежутках времени можно использовать более сильное приближение нулевой скорости $u(x, t) = 0$, поскольку даже в этом случае гарантирована квазистационарность одномерного распределения. Уравнение Лиувилля с ненулевой скоростью является следующим приближением, уточняющим эволюцию ВПФР. Таким образом, приближение нулевого ускорения вполне согласовано с квазистационарностью распределения. Тогда уравнение (2.11) можно отбросить, а для $e(x, t)$ из (2.10) получить

$$f \frac{\partial e}{\partial t} - e \frac{\partial(uf)}{\partial x} = - \frac{\partial(e_3 f)}{\partial x}, \quad e_3 \equiv \langle \dot{x}^3 \rangle.$$

Переносим слагаемое $e \frac{\partial(uf)}{\partial x}$ в правую часть, преобразуем это уравнение к виду

$$\left(\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} \right) f = \frac{\partial}{\partial x} (ue - e_3) f.$$

Объединяя полученное уравнение с уравнениями эволюции (1.2) для ВПФР и (2.7) для ЭЛС, получаем следующую систему уравнений:

$$\begin{aligned} \frac{\partial f(x, t)}{\partial t} + \frac{\partial}{\partial x} f(x, t) u(x, t) &= 0; \\ \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} \right) f &= \frac{\partial}{\partial x} (u^2 - e) f; \\ \left(\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} \right) f &= \frac{\partial}{\partial x} (ue - e_3) f. \end{aligned} \quad (2.13)$$

В этой системе момент e_3 третьего порядка по скорости пока еще не определен. Для этого требуется выразить его через моменты низших порядков, т.е. провести собственно замыкание. Ниже приводятся некоторые варианты такого замыкания как примеры прогнозных моделей «гидродинамического» типа, т.е. моделей, основанных на анализе эволюции моментов выборочного распределения большего порядка, чем то, которое непосредственно прогнозируется.

Первый вариант. Положим $e_3 = 0$, т.е. будем считать эту величину пренебрежимо малой. Тогда система замкнется, но, решив ее, следует убедиться в том, что найденное решение совместимо с этим условием малости. Замыкающее последнее уравнение в системе (2.13) при этом имеет вид

$$f \frac{\partial e}{\partial t} = e \frac{\partial(uf)}{\partial x}. \quad (2.14)$$

Этот подход справедлив в том случае, если скорости $\dot{x} = x(t+1) - x(t)$ малы, что характерно для случая слабо меняющегося ряда $x(t)$.

Второй вариант. Будем считать пренебрежимо малой величину ковариации, т.е. положим $ue - e_3 = 0$. Эта ситуация возникает, например, тогда, когда скорость $u(x, t)$ как функция от x приблизительно постоянна, за исключением, возможно, точек вблизи границ промежутка изменения величины x . В таком случае можно считать, что $\partial u / \partial x$ с принятой точностью является пренебрежимо малой величиной, так что замыкающее уравнение имеет вид первого уравнения системы — уравнения Лиувилля:

$$\frac{\partial e}{\partial t} + \frac{\partial(ue)}{\partial x} = 0. \quad (2.15)$$

Третий вариант. Заметим, что, как и скорость $u(x, t)$, ее второй момент $e(x, t)$ известен по исходной выборке не в текущий, а в предыдущий момент времени. Подход к замыканию в духе наивного прогноза состоит в том, чтобы оценить величину $e(x, t)$ по ее значению в момент $t-1$, которое известно по имеющейся статистике к этому времени. Положим тогда $e(x, t) = e(x, t-1) \equiv g(x, t)$, где $g(x, t)$ — в данном случае известная функция, определяемая непосредственно по значениям временного ряда. Тогда замыкающее уравнение примет вид

$$f \frac{\partial e}{\partial t} = e \frac{\partial(uf)}{\partial x} - \frac{\partial(gf)}{\partial x}. \quad (2.16)$$

Итак, каждый из этих трех способов, будучи применен к системе (2.13), либо к первым двум ее уравнениям или только к первому, приведет к замкнутой модели эволюции одночастичной ВПФР.

Например, если применить второй метод (нулевой ковариации) ко второму уравнению в (2.13), то он в данном случае будет означать, что средний квадрат скорости и квадрат средней скорости совпадают. Следовательно, средняя скорость $u(x, t)$ как функция от x равна постоянной. Тогда достаточно использовать только первое уравнение в (2.13), общее решение которого в этом случае имеет вид $f(x, t) = f(x - u, t - 1)$. Поскольку в граничных промежутках, по условию, плотность всегда нулевая, то решение в виде сдвига вдоль характеристик возможно, только если скорость нулевая. Тогда $f(x, t)$ не меняется во времени, т.е. в этом приближении мы получили наивный прогноз относительно самой выборочной плотности распределения.

Если же применить ко второму уравнению в (2.13) метод наивного прогноза, касающийся величины $e(x, t)$, то система замыкается следующим образом: правая часть этого уравнения считается известной по данным в предыдущий момент времени, после чего определяется $u(x, t)$ в текущий момент, что позволяет определить $f(x, t)$ из первого уравнения.

Отметим, что с точки зрения формальной обоснованности эти замыкания ничем не лучше по сравнению с вариантами замыкания (2.1) - (2.3) на первом уравнении, но на практике бывает лучше замыкать систему на моменте более высокого порядка, чем первый или второй. Это связано с тем, что для рядов, которые равномерно ограничены по времени, неточность от аппроксимации высших моментов дает уменьшающийся вклад в уравнение эволюции квазистационарных ВПФР с увеличением порядка момента.

Пусть, например, выполнено условие $\|f(x, t) - f(x, t+1)\| \leq \varepsilon$. Тогда для момента n -го порядка имеем оценку

$$\begin{aligned} |m_n(t+1) - m_n(t)| &= \left| \int x^n f(x, t+1) dx - \int x^n f(x, t) dx \right| \leq \\ &\leq \int x^n |f(x, t+1) - f(x, t)| dx \leq \varepsilon \int_0^1 x^n dx = \frac{\varepsilon}{n}. \end{aligned}$$

Поэтому замыкание уравнения эволюции для момента, например, третьего порядка в духе наивного прогноза внесет в это уравнение погрешность порядка $\varepsilon/3$. С другой стороны, если распределение не квазистационарно, то увеличение порядка момента, на котором производится замыкание, не обязательно приведет к уточнению модели. Кроме того, и для квазистационарных распределений увеличение числа уравнений сопряжено с ростом ошибок в результате численного счета, так что многочисленные модели также имеют ограниченное применение.

Вышеприведенные примеры моделей, в которых подразумевается, что эволюция ВПФР данной размерности определяется эволюцией нескольких первых моментов ВПФР расширенной размерности, можно назвать гидродинамическими по аналогии с аналогичными моментными уравнениями классической статистической механики.

Как уже говорилось, подход с использованием квазистационарных распределений имеет много общего с концепцией локально-равновесного приближения в статистической механике [6]. Рассмотрим отмеченную

аналогию более подробно. С этой целью запишем первые два уравнения цепочки кинетических уравнений для ВПФР:

$$\begin{aligned} \frac{\partial F(x, \dot{x}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F(x, \dot{x}; t) + \frac{\partial}{\partial \dot{x}} w(x, \dot{x}; t) F(x, \dot{x}; t) &= 0, \\ F &\equiv F_2, \\ \frac{\partial f(x, t)}{\partial t} + \frac{\partial}{\partial x} u(x, t) f(x, t) &= 0. \end{aligned} \quad (2.17)$$

В цепочке уравнений (2.17) функция $u(x, t)$ определяется через $F(x, \dot{x}; t)$ соотношением $u(x, t) f(x, t) = \int \dot{x} F(x, \dot{x}; t) d\dot{x}$, а ускорение $w(x, \dot{x}; t)$ определяется через выборочную плотность более высокого порядка $w(x, \dot{x}; t) F(x, \dot{x}; t) = \int \ddot{x} F_3(x, \dot{x}, \ddot{x}; t) d\ddot{x}$.

Если замыкание цепочки проводится на функции F_3 , то, согласно этому предположению, замыкающая функция должна удовлетворять уравнению Лиувилля без зацепления с F_4 , т.е.

$$\frac{\partial F_3(x, \dot{x}, \ddot{x}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F_3 + \frac{\partial}{\partial \dot{x}} \ddot{x} F_3 = 0. \quad (2.18)$$

В таком приближении замыкание уравнения (2.17) состоит в выборе модели для эволюции ускорения $w(x, \dot{x}; t)$. Дифференцируя по времени соотношение, определяющее ускорение $w(x, \dot{x}; t)$, и используя (2.17) и (2.18), получаем для $w(x, \dot{x}; t)$ следующее уравнение:

$$F \frac{\partial w}{\partial t} = \left(\frac{\partial(\dot{x} F)}{\partial x} + \frac{\partial(w F)}{\partial \dot{x}} \right) w - \int \ddot{x} \left(\frac{\partial(\dot{x} F_3)}{\partial x} + \frac{\partial(\ddot{x} F_3)}{\partial \dot{x}} \right) d\ddot{x}.$$

Полагая теперь результат усреднения величин $\dot{x} \frac{\partial \ddot{x}}{\partial x}$ и $\ddot{x} \frac{\partial \ddot{x}}{\partial \dot{x}}$ пренебрежимо малым, т.е. действуя по аналогии с идеей исключения величин такого рода из уравнений (2.10) - (2.11), получаем отсюда уравнение

$$\left(\frac{\partial w}{\partial t} + \dot{x} \frac{\partial w}{\partial x} + w \frac{\partial w}{\partial \dot{x}} \right) F = \frac{\partial}{\partial x} (\dot{x} w - \pi_{11}) F + \frac{\partial}{\partial \dot{x}} (w^2 - \pi_{02}) F, \quad (2.19)$$

где введены обозначения $\pi_{kl}(x, \dot{x}, t)$ для средних величин, определяемых формулой

$$\pi_{kl}(x, \dot{x}, t)F(x, \dot{x}, t) = \int \dot{x}^k \dot{x}^l F_3(x, \dot{x}, \ddot{x}; t) d\ddot{x}. \quad (2.20)$$

Введем далее величины Π_{kl} так, что

$$\begin{aligned} \Pi_{kl}(x, t)f(x, t) &= \iint \dot{x}^k \dot{x}^l F_3(x, \dot{x}, \ddot{x}) d\ddot{x} d\dot{x} = \\ &= \int \pi_{kl}(x, \dot{x}, t)F(x, \dot{x}, t) d\dot{x}. \end{aligned} \quad (2.21)$$

Тогда для эволюции одномерной ВПФР $f(x, t)$ получаем цепочку уравнений

$$\begin{aligned} \frac{\partial f}{\partial t} &= -\frac{\partial(uf)}{\partial x}; \\ \left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right)f &= \frac{\partial}{\partial x}(u^2 - e)f + \Pi_{01}f; \\ \left(\frac{\partial e}{\partial t} + u\frac{\partial e}{\partial x}\right)f &= \frac{\partial}{\partial x}(ue - e_3)f + 2\Pi_{11}f; \dots \end{aligned} \quad (2.22)$$

Эта система уточняет систему (2.13), полученную в предположении обрыва цепочки на двумерной ВПФР. Как и сама цепочка уравнений для многомерных ВПФР, моментная система (2.22) бесконечна. Замыкание моментной системы и будет являться прогнозной моделью для эволюции одномерного выборочного распределения.

Прямым следствием замкнутой системы эволюционных уравнений относительно одномерной ВПФР и моментов двумерной ВПФР является то, что в такой модели зависимость выборочного распределения от времени входит только через несколько первых моментов. Тогда эти моменты с принятой точностью определяют и само распределение, что составляет содержание одного из положений функциональной гипотезы Боголюбова [6] относительно локально-равновесных распределений в классической статистической механике.

Например, если оборвать (2.22) на третьем уравнении, то такой обрыв будет означать, что принята следующая функциональная гипотеза о временной зависимости высших ВПФР: зависимость от времени в $F(x, \dot{x}; t)$, $F_3(x, \dot{x}, \ddot{x}; t)$ и т.д. входит только через $f(x, t)$, $u(x, t)$ и $e(x, t)$. Тогда задача моделирования будет состоять в том, чтобы выдвигать

дополнительные предположения, с помощью которых можно будет выразить e_3 , Π_{01} , Π_{11} через f , u , e , что позволит замкнуть систему (2.22). Продемонстрируем, как может выглядеть один из вариантов такого обрыва.

Пусть в выражении для e_3 можно пренебречь величиной $\langle (\dot{x} - u(x, t))^3 \rangle$. Тогда считаем, что $\int (\dot{x} - u)^3 F d\dot{x} \approx 0$, после чего получаем

$$e_3 f = \int ((\dot{x} - u) + u)^3 F d\dot{x} = 3uef - 2u^3 f,$$

откуда следует $e_3 = 3ue - 2u^3$.

Предположим далее, что $\langle w \rangle \equiv \pi_{01} \approx \partial u / \partial t$, и, кроме того, приближенно выполняется условие факторизации $\Pi_{11} \approx u\Pi_{01}$. Тогда сделанные три допущения замыкают систему (2.22), которая в этом случае переписывается в виде:

$$\begin{aligned} \frac{\partial f}{\partial t} &= -\frac{\partial(uf)}{\partial x}; \quad uf \frac{\partial u}{\partial x} = \frac{\partial}{\partial x}(u^2 - e)f; \\ f \frac{\partial(e - u^2)}{\partial t} &= e \frac{\partial(uf)}{\partial x} - \frac{\partial}{\partial x}(3ue - 2u^3)f. \end{aligned} \quad (2.23)$$

Аналогично можно строить модели в пространствах большей размерности. Фактически задачей прогноза по уравнению Лиувилля является определение локальной скорости изменения ВПФР методом, наиболее подходящим для конкретного временного ряда. Разумеется, описанные методы не исчерпывают все способы замыкания системы прогнозных уравнений, получаемых в рамках кинетического подхода. Эти методы носят эвристический характер и не являются универсальными для любых временных рядов. Задача исследователя — подобрать наиболее адекватную модель для описания конкретного процесса. Заметим, что такая же ситуация сложилась и в самой статистической механике. Например, нелинейное кинетическое уравнение Больцмана применяется в основном для описания разреженных газов с короткодействующим потенциалом взаимодействия между молекулами, самосогласованное уравнение Власова — для систем со слабыми дальнедействующими силами (электромагнитные, гравитационные), линейное уравнение переноса — для излучения, проходящего сквозь среду с приближенно постоянными свойствами. Кинетиче-

ские уравнения служат инструментом для корректного вывода моделей, позволяя строго сформулировать вероятностные свойства, присущие той или иной физической системе.

9.3. Уравнение Фоккера-Планка для выборочных распределений

Обратимся вновь к уравнению Лиувилля (1.2) и выясним, как эволюционируют в силу этого уравнения высшие моменты одночастичной ВПФР. Из (1.6) следует, что эволюция первого момента совпадает с эмпирической средней скоростью изменения значений ряда, как это и должно быть. Поскольку выборочные моменты ряда определяются по ВПФР как

$$m_r(t) = \int x^r f(x, t) dx, \quad (3.1)$$

то, дифференцируя (3.1) по времени и выражая производную $\partial f / \partial t$ через уравнение Лиувилля (1.2), получаем уравнение для эволюции моментов ВПФР:

$$\begin{aligned} \frac{dm_r}{dt} &= \int x^r \frac{\partial f}{\partial t} dx = - \int x^r \frac{\partial u f}{\partial x} dx = r \int x^{r-1} u f dx = \\ &= r \operatorname{cov}(x^{r-1}, u) + r m_{r-1} U. \end{aligned} \quad (3.2)$$

При $r = 1$ уравнение (3.2) переходит в (1.6), которое, как было показано, совпадает с уравнением эволюции «истинного» среднего, вычисленного непосредственно по элементам временного ряда. Однако при $r \geq 2$ уравнение (3.2) не является непрерывным аналогом соответствующей разностной производной выборочных моментов. Выборочные моменты, вычисленные по элементам ряда, будем снабжать верхним индексом e , чтобы отличать их от моментов ВПФР. Тогда имеем

$$m_r^e(t) = \frac{1}{T} \sum_{k=1}^T x_{t-T+k}^r, \quad (3.3)$$

$$\dot{m}_r^e = m_r^e(t+1) - m_r^e(t) \neq r \operatorname{cov}(x, \dot{x}) + r m_{r-1}^e(t) U(t).$$

Например, при $r = 2$ эволюция второго выборочного момента непосредственно по элементам ряда дается формулой

$$\begin{aligned} \dot{m}_2^e(t) &= \frac{1}{T} \sum_{k=1}^T (x_{t-T+k+1}^2 - x_{t-T+k}^2) = \\ &= \frac{x_{t+1}^2 - x_{t-T+1}^2}{T} = U(t) (x_{t+1} + x_{t-T+1}). \end{aligned}$$

В то же время эволюция второго момента ВПФР согласно (1.2) имеет вид

$$\begin{aligned} \dot{m}_2 &= 2 \operatorname{cov}(x, \dot{x}) + 2mU = 2 \langle x \dot{x} \rangle = \\ &= \frac{2}{T} \sum_{k=1}^T x_{t-T+k} \dot{x}_{t-T+k} = \frac{2}{T} \sum_{k=1}^T x_{t-T+k} (x_{t-T+k+1} - x_{t-T+k}), \end{aligned}$$

что, очевидно, не равно выражению для \dot{m}_2^e , написанному строкой выше.

Точно так же и эволюция центральных выборочных моментов ряда, определяемых по ВПФР как

$$\mu_r(t) = \int (x - m(t))^r f(x, t) dx, \quad r \geq 2, \quad (3.4)$$

отличается от эволюции моментов, вычисленных по элементам ряда:

$$\begin{aligned} \mu_r^e(t) &= \frac{1}{T} \sum_{k=t-T+1}^t (x(k) - m(t))^r, \\ m(t) &= \frac{1}{T} \sum_{k=t-T+1}^t x(k), \quad \dot{\mu}_r(t) \neq \dot{\mu}_r^e(t). \end{aligned} \quad (3.5)$$

Найдем изменение величин $\mu_r(t)$ из (3.4) в силу уравнения Лиувилля (1.2). Имеем следующую цепочку равенств:

$$\begin{aligned}
\frac{d\mu_r}{dt} &= -r\mu_{r-1} \frac{dm}{dt} + \int (x-m)^r \frac{\partial f}{\partial t} dx = \\
&= -r\mu_{r-1} \frac{dm}{dt} - \int (x-m)^r \frac{\partial uf}{\partial x} dx = \\
&= -r\mu_{r-1} \frac{dm}{dt} + r\mu_{r-1}U + r \int (x-m)^{r-1} (u-U) f dx = \\
&= r \left\langle (\Delta x)^{r-1} \Delta u \right\rangle.
\end{aligned}
\tag{3.6}$$

В частности, для уравнения эволюции выборочной дисперсии получаем из (3.6)

$$\dot{\mu}_2 = 2 \text{cov}(x, u). \tag{3.7}$$

С другой стороны, изменение тех же величин (3.6) может быть вычислено в момент времени t непосредственно по имеющимся выборкам. Предположительное знание эволюции эмпирических моментов позволяет провести коррекцию прогнозной модели, основанной на уравнении Лиувилля (1.2) так, чтобы и высшие моменты ВПФР эволюционировали бы в соответствии с выборочными данными. Способ такой коррекции состоит в следующем.

Обозначим через $\lambda(t)$ невязку между эмпирически вычисленной производной \dot{m}_2^e непосредственно по элементам ряда, и производной \dot{m}_2 , отвечающей формуле (3.2) согласно уравнению Лиувилля (будем снабжать ее верхним индексом L). Именно, положим

$$\begin{aligned}
\lambda(t) &= \dot{m}_2^e - \dot{m}_2^L = \dot{m}_2^e - 2 \text{cov}(x, u) - 2mU = \\
&= \frac{x_{t+1}^2 - x_{t-T+1}^2}{T} - \frac{2}{T} \sum_{k=1}^T x_{t-T+k} (x_{t-T+k+1} - x_{t-T+k}).
\end{aligned}
\tag{3.8}$$

Рассмотрим для ВПФР вместо уравнения Лиувилля (1.2) следующее уравнение:

$$\frac{\partial f}{\partial t} + \frac{\partial(uf)}{\partial x} - \frac{\lambda(t)}{2} \frac{\partial^2 f}{\partial x^2} = 0. \tag{3.9}$$

Уравнение (3.9) отличается от уравнения Лиувилля (1.2) добавкой, которая имеет дважды дивергентную форму. В результате уравнение эволюции первого момента остается тем же самым, что и для уравнения Лиувилля, а уравнение эволюции второго момента, как легко проверить путем интегрирования по частям, совпадает с соответствующей эмпирической производной \dot{m}_2^e в соответствии с (3.8):

$$\begin{aligned}
\frac{dm_2}{dt} &= \int x^2 \frac{\partial f}{\partial t} dx = \\
&= \int x^2 \left(-\frac{\partial(uf)}{\partial x} + \frac{\lambda}{2} \frac{\partial^2 f}{\partial x^2} \right) dx = 2 \int x \left(uf - \frac{\lambda}{2} \frac{\partial f}{\partial x} \right) dx = \\
&= 2 \text{cov}(x, u) + 2mU + \lambda = \dot{m}_2^e.
\end{aligned}
\tag{3.10}$$

Уравнение эволюции (3.9) с диффузионной добавкой будем называть моделью с дивергентной невязкой [25]. По структуре дифференциального оператора это уравнение является уравнением Фоккера-Планка, которое при независимых от времени величинах u и λ описывает перенос и диффузию вероятности для гауссовского белого шума. Отметим, что уравнение Фоккера-Планка относительно плотности теоретического распределения является не приближенным, а точным, если процесс имеет независимые приращения, которые распределены нормально. Для нестационарного процесса оно играет роль модельного и является модификацией эмпирического уравнения Лиувилля. Смысл этой модификации в том, что к динамической системе, приближенно моделирующей эволюцию рассматриваемой выборки, добавляются случайные блуждания. Тем самым такое уравнение становится более адекватным для описания случайных процессов, природа которых может быть отличной от чисто динамического хаоса.

В результате оказалось, что можно уточнить модель (1.2), не увеличивая существенно объем вычислений, т.е. оставаясь в пространстве той же размерности, чтобы некоторые, по крайней мере, первые центральные моменты ВПФР действительно эволюционировали бы в соответствии с уточненным уравнением. Однако порядок этих уточнений ограничен. Поскольку выборочный момент порядка k для временного ряда имеет $k-1$ степень свободы, то для выборки фиксированного объема T осмысленно можно ставить вопрос об исследовании только первых $T-1$ моментов соответствующей ВПФР как независимых величин. Более того, чтобы получающиеся оценки были ε -несмещенными, требуется выполнение

условия $k/T \leq \varepsilon$. Это означает, что корректными оценками моментов нестационарного распределения конечной выборки являются только первые несколько выборочных моментов. Именно, если при прогнозировании на τ шагов при заданном уровне ε -стационарности ВПФР определен оптимальный объем выборки $T \leq 2\tau/\varepsilon$, то при построении уточненной модели эволюции ВПФР можно использовать оценки только для таких моментов, порядок которых не превосходит 2τ .

Рассмотрим теперь более общий вопрос о том, каким уравнением следует описывать эволюцию ВПФР, чтобы несколько первых ее моментов изменялись бы в соответствии с выборочными данными.

С целью обобщения уравнения (3.9) при построении модели эволюции ВПФР, которое должно приводить к верным эмпирическим значениям производной по времени для выборочных моментов вплоть до заданного порядка r , вводим эмпирически определяемые величины

$$\lambda_k(t) = \dot{m}_k^e - \dot{m}_k^L \quad (3.11)$$

и составляем уравнение

$$\frac{\partial f}{\partial t} + \frac{\partial(uf)}{\partial x} + \sum_{k=2}^r (-1)^{k-1} \frac{\lambda_k(t)}{k!} \frac{\partial^k f}{\partial x^k} = 0. \quad (3.12)$$

Легко проверяется, что в силу этого уравнения $\dot{m} = U$, $\dot{m}_k = \dot{m}_k^e$.

Количество учитываемых эмпирических выборочных моментов в (4.12) зависит от горизонта и желаемой точности прогноза. Если оказывается, что при этом близость между ВПФР должна быть не больше заданного числа ε , то, как указывалось выше, количество независимых моментов не может превосходить εT , которое, в свою очередь, не превосходит 2τ . Например, при $\varepsilon \sim (1/T - 2/T)$ корректной моделью эволюции ВПФР является уравнение Лиувилля (1.2). Если $\varepsilon \sim (2/T - 4/T)$, то можно использовать уравнение Фоккера-Планка (3.9), при $\varepsilon \sim (4/T - 6/T)$ можно включить в рассмотрение третью пространственную производную и т.д.

Подчеркнем, что, как и в случае с лиувиллевой скоростью U , эволюция моментов ВПФР определяется в предыдущий по отношению к прогнозируемому момент времени. Следовательно, модель прогнозирования должна включать в себя некоторую идею о том, как эволюционируют моменты ВПФР. В этом смысле уравнение (3.12) является уравнением адаптивной модели эволюции, когда на следующем расчетном шаге прогноз-

ная дисперсия и другие моменты подстраиваются к своим фактическим величинам, известным с предыдущего шага. Как и в случае с уравнением Лиувилля, его можно дополнить уравнениями эволюции моментов в пространстве скоростей и ускорений.

Несмотря на кажущееся исправление ситуации с эволюцией выборочных моментов, такая модель не является внутренне согласованной, поскольку уравнения эволюции моментов выводились из уравнения Лиувилля, а использоваться стало уравнение Фоккера-Планка. Поэтому представляется более естественным стартовать с самого уравнения Фоккера-Планка, которое было выведено в части I как следствие уравнения Чепмена-Колмогорова. Пусть, как и ранее, $f(x, t)$ — одномерная ВПФР, построенная по выборке некоторого объема, который далее в этом параграфе считается постоянным и потому не указывается в аргументах ВПФР. Обозначим также $F(x, v, t)$ двумерную ВПФР, а $\Phi(x, v, w, t)$ трехмерную ВПФР значений ряда, его первых и, соответственно, вторых разностей. Тогда в приближении Фоккера-Планка функция $F(x, v, t)$ удовлетворяет двумерному уравнению диффузионного типа. Если матрица диффузионных коэффициентов не зависит от x, v , а зависит только от времени t , то уравнение Фоккера-Планка для $F(x, v, t)$ имеет вид

$$\begin{aligned} \frac{\partial F}{\partial t} + \frac{\partial}{\partial x}(vF) + \frac{\partial}{\partial v}(WF) - \\ - \frac{\lambda(t)}{2} \frac{\partial^2 F}{\partial x^2} - \chi(t) \frac{\partial^2 F}{\partial x \partial v} - \frac{\mu(t)}{2} \frac{\partial^2 F}{\partial v^2} = 0, \end{aligned} \quad (3.13)$$

$$W(x, v, t) = \frac{1}{F(x, v, t)} \int w \Phi(x, v, w, t) dw.$$

Эта модель дополняет уравнение Лиувилля эффектом случайного нестационарного блуждания. Как и ранее, используя связь

$$f(x, t) = \int F(x, v, t) dv,$$

и учитывая обращение в ноль ВПФР всех порядков на границе области интегрирования, получаем из (3.13) одномерное уравнение Фоккера-Планка:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x}(uf) - \frac{\lambda(t)}{2} \frac{\partial^2 f}{\partial x^2} = 0, \quad (3.14)$$

$$u(x, t) = \frac{1}{f(x, t)} \int v F(x, v, t) dv.$$

В отличие от (3.10), в уравнении (3.13) коэффициент λ не является невязкой, а должен определяться по элементам ряда в соответствии с уравнениями эволюции моментов в силу этого самого уравнения. Важно учитывать, что, хотя одномерное уравнение (3.14) не содержит в явном виде некоторых коэффициентов, которые входят в двумерное уравнение (3.13), эти коэффициенты влияют на эволюцию моментов распределения $F(x, v, t)$ по скоростям.

Если мы обрываем цепочку совместных распределений на бинарной ВПФР $F(x, v, t)$, то среднее ускорение $W(x, v, t)$ следует задать независимо от трехмерной ВПФР. Это можно сделать в духе наивного прогноза, считая ускорение известным с предыдущего шага по времени (производная по времени трактуется в разностном смысле): $W(x, v, t) = W(x, v, t-1)$. Это же замечание относится и к коэффициентам λ, χ, μ . Если окажется, что такое приближение слишком грубо, то и обрывать на двумерной ВПФР нельзя, а требуется использовать аналогичное уравнение для $\Phi(x, v, w, t)$, и тогда моменты четырехмерного распределения считать известными, и т.д.

Подчеркнем, что нашей целью является сведение задачи к одномерному уравнению (3.14), а двумерное уравнение (3.13) будет использоваться для того, чтобы замыкать моментную систему. Рассмотрим уравнение эволюции первого выборочного момента, определяемого как

$$m(t) = \langle x \rangle_t = \int x f(x, t) dx,$$

но эволюция будет следовать не уравнению Лиувилля, как в первоначальной простой постановке, а уравнению Фоккера-Планка (3.14). Из (3.14) получаем

$$\begin{aligned} \frac{dm(t)}{dt} &= \int x \frac{\partial f(x, t)}{\partial t} dx = \\ &= - \int x \frac{\partial}{\partial x} u(x, t) f(x, t) dx + \frac{\lambda(t)}{2} \int x \frac{\partial^2 f(x, t)}{\partial x^2} dx. \end{aligned}$$

После интегрирования по частям с учетом того, что в граничных ячейках плотность распределения равна нулю, получаем результат, аналогичный (3.2):

$$\frac{dm(t)}{dt} = \int u(x, t) f(x, t) dx = \langle u \rangle_t = U(t), \quad (3.15)$$

где $U(t)$, как и выше, есть среднее значение эмпирической скорости $u(x, t)$ по одномерному распределению в момент времени t . Рассмотрим теперь эволюцию самой эмпирической скорости. С одной стороны, из (3.14) следует, что

$$\begin{aligned} \frac{\partial}{\partial t}(f(x, t)u(x, t)) &= u \frac{\partial f}{\partial t} + f \frac{\partial u}{\partial t} = \\ &= u \left(-\frac{\partial(uf)}{\partial x} + \frac{\lambda}{2} \frac{\partial^2 f}{\partial x^2} \right) + f \frac{\partial u}{\partial t}. \end{aligned} \quad (3.16)$$

С другой стороны, используя (3.13), получаем, что та же самая правая часть (3.16) равна

$$\begin{aligned} \frac{\partial(uf)}{\partial t} &= \int v \frac{\partial F(x, v, t)}{\partial t} dv = \\ &= \int v \left(-v \frac{\partial F}{\partial x} - \frac{\partial(WF)}{\partial v} + \frac{\lambda}{2} \frac{\partial^2 F}{\partial x^2} + \chi \frac{\partial^2 F}{\partial x \partial v} + \frac{\mu}{2} \frac{\partial^2 F}{\partial v^2} \right) dv. \end{aligned}$$

После интегрирования по частям последнее выражение преобразуется к виду

$$\begin{aligned} \frac{\partial(uf)}{\partial t} &= -\frac{\partial}{\partial x} \int v^2 F(x, v, t) dv + \int W(x, v, t) F(x, v, t) dv + \\ &\quad + \frac{\lambda}{2} \frac{\partial^2(uf)}{\partial x^2} - \chi \frac{\partial f}{\partial x}. \end{aligned}$$

(3.17)

Введем обозначения для входящих в (3.17) интегралов:

$$\begin{aligned} e(x, t) f(x, t) &= \int v^2 F(x, v, t) dv, \\ a(x, t) f(x, t) &= \int W(x, v, t) F(x, v, t) dv. \end{aligned} \quad (3.18)$$

Если мы имеем дело с динамической системой, то статистико-механический смысл интегралов в (3.18) следующий: $e(x, t)$ есть удвоенная плотность средней кинетической энергии, а $a(x, t)$ представляет среднее локальное (по координате и времени) ускорение от внешнего поля. В результате из (3.16) и (3.17) получаем уравнение эволюции локальной скорости $u(x, t)$:

$$f \frac{\partial u}{\partial t} = u \left(\frac{\partial(uf)}{\partial x} - \frac{\lambda}{2} \frac{\partial^2 f}{\partial x^2} \right) - \frac{\partial(ef)}{\partial x} + af + \frac{\lambda}{2} \frac{\partial^2(uf)}{\partial x^2} - \chi \frac{\partial f}{\partial x}. \quad (3.19)$$

Уравнение (3.19) отличается от выводимых из уравнения Лиувилля уравнений (2.19) или (2.20). Из (3.19) следует, что изменение со временем средней макроскопической скорости $U(t)$ равно среднему ускорению, обусловленному внешними причинами, что согласуется с традиционными механическими представлениями:

$$\frac{dU(t)}{dt} = A(t) \equiv \int a(x, t) f(x, t) dx = \iint W(x, v, t) F(x, v, t) dx dv. \quad (3.20)$$

Заметим теперь, что в уравнение (3.19) входит величина $e(x, t)$, эволюция которой также должна быть определена. Используя тот же метод, что и при выводе (3.19), получаем из (3.18) и (3.13)

$$\begin{aligned} \frac{\partial(ef)}{\partial t} &= \int v^2 \frac{\partial F(x, v, t)}{\partial t} dv = \\ &= -\frac{\partial(e_3 f)}{\partial x} - 2\chi \frac{\partial(uf)}{\partial x} + \frac{\lambda}{2} \frac{\partial^2(ef)}{\partial x^2} + \mu + \beta f. \end{aligned} \quad (3.21)$$

Здесь введены две новые величины:

$$\begin{aligned} e_3(x, t) f(x, t) &= \int v^3 F(x, v, t) dv, \\ \beta(x, t) f(x, t) &= \int \left(2vW + v^2 \frac{\partial W}{\partial v} \right) F(x, v, t) dv. \end{aligned} \quad (3.22)$$

Таким образом, эволюция момента второго порядка выражается, кроме ранее введенных величин, через градиент момента третьего порядка и новый потоковый член. Если обозначить через $e_k(x, t)$ плотность момента k -го порядка двумерной ВПФР по скоростям

$$e_k(x, t) f(x, t) = \int v^k F(x, v, t) dv, \quad (3.23)$$

то уравнение эволюции этой величины имеет следующий вид:

$$\begin{aligned} \frac{\partial(e_k f)}{\partial t} &= \int v^k \frac{\partial F(x, v, t)}{\partial t} dv = \\ &= \int v^k \left(-v \frac{\partial F}{\partial x} - \frac{\partial(WF)}{\partial v} + \frac{\lambda}{2} \frac{\partial^2 F}{\partial x^2} + \chi \frac{\partial^2 F}{\partial x \partial v} + \frac{\mu}{2} \frac{\partial^2 F}{\partial v^2} \right) dv = \\ &= -\frac{\partial(e_{k+1} f)}{\partial x} - k\chi \frac{\partial(e_{k-1} f)}{\partial x} + \frac{\lambda}{2} \frac{\partial^2(e_k f)}{\partial x^2} + \\ &\quad + k(k-1) \frac{\mu}{2} e_{k-2} f + \beta_k f, \end{aligned} \quad (3.24)$$

где

$$\beta_k(x, t) f(x, t) = \int \left(kW + v \frac{\partial W}{\partial v} \right) v^{k-1} F(x, v, t) dv. \quad (3.25)$$

Во многих случаях можно приближенно считать, что локальные ускорения W малы, либо что локальное ускорение является однородной формой скорости порядка s . Тогда без учета W эволюция момента k -го порядка выражается через уже известные величины и момент $(k+1)$ -го порядка:

$$\begin{aligned} \frac{\partial(e_k f)}{\partial t} &= -\frac{\partial(e_{k+1} f)}{\partial x} + \frac{\lambda}{2} \frac{\partial^2(e_k f)}{\partial x^2} - k\chi \frac{\partial(e_{k-1} f)}{\partial x} + \\ &\quad + (k+s)ae_{k-1}f + k(k-1) \frac{\mu}{2} e_{k-2}f. \end{aligned} \quad (3.26)$$

Если на каком-либо порядке оборвать моментную систему, т.е. задать $e_{k+1}(x, t)$ независимо, то уравнение для момента k -го порядка будет представлять собой уравнение диффузии со сносом и источником.

Выясним теперь статистический смысл остальных коэффициентов, входящих в двумерное уравнение Фоккера-Планка (3.13). Рассмотрим сначала, как эволюционирует второй центральный выборочный момент одномерного распределения

$$\sigma^2(t) = \int (x - m(t))^2 f(x, t) dx.$$

В соответствии с (3.14) и (3.15) имеем

$$\begin{aligned} \frac{d\sigma^2}{dt} &= - \int (x - m(t))^2 \frac{\partial(uf)}{\partial x} dx + \frac{\lambda}{2} \int (x - m(t))^2 \frac{\partial^2 f}{\partial x^2} dx = \\ &= \lambda + 2 \int u x f dx - 2mU. \end{aligned}$$

Поскольку ковариация координаты и скорости равна

$$\begin{aligned} R(t) &= \text{cov}_{x,v}(t) = \\ &= \int x v F(x, v, t) dx dv - \int x F(x, v, t) dx dv \cdot \int v F(x, v, t) dx dv = \\ &= \int x u(x, t) f(x, t) dx - m(t)U(t) \equiv \text{cov}_{x,u}(t), \end{aligned}$$

то эволюция выборочной дисперсии в силу уравнения Фоккера-Планка (3.14) имеет вид

$$\frac{d\sigma^2}{dt} = \lambda + 2R. \quad (3.27)$$

Тем самым выявлен статистический смысл параметра $\lambda(t)$ в уравнении (3.13). Это есть разность между производной текущей выборочной дисперсии ряда и удвоенной выборочной ковариацией исходного ряда и ряда его первых разностей. Сам же коэффициент ковариации $R(t)$ эволюционирует следующим образом:

$$\begin{aligned} \frac{dR}{dt} &= - \frac{d(mU)}{dt} + \int x \frac{\partial(uf)}{\partial t} dx = \\ &= -U^2 - mA + \int x \left(- \frac{\partial(ef)}{\partial x} + af + \frac{\lambda}{2} \frac{\partial^2(uf)}{\partial x^2} - \chi \frac{\partial f}{\partial x} \right) dx = \\ &= -U^2 + \text{cov}_{x,a} + E + \chi. \end{aligned} \quad (3.28)$$

Здесь

$$E = E(t) = \int e(x, t) f(x, t) dx. \quad (3.29)$$

Следовательно, коэффициент $\chi(t)$ в уравнении (3.13) представляет собой разность между производной по времени ковариации координаты-

скорости и ковариацией координаты-ускорения, минус разность между средней энергией $E(t)$ и энергией среднего движения $U^2(t)$.

По аналогии с (3.27), эволюция выборочной дисперсии скорости имеет вид

$$\frac{d\sigma_v^2}{dt} = \mu + 2\text{cov}_{v,a}. \quad (3.30)$$

Таким образом, если задать левые части выражений (3.26-3.28) и (3.30) исходя из информации, известной к текущему моменту времени, то в рамках приближения двумерного уравнения Фоккера-Планка (3.13) его коэффициенты полностью определены.

Коэффициент асимметрии γ (или скос) одномерного распределения вычисляется по формуле

$$\gamma(t) = \frac{1}{\sigma^3(t)} \int (x - m(t))^3 f(x, t) dx. \quad (3.31)$$

Эволюция этого коэффициента в силу уравнения Фоккера-Планка (3.14) дается формулой

$$\frac{d\gamma}{dt} = -3\gamma \frac{\lambda + 2R}{2\sigma^2} + \frac{3}{\sigma^3} \int (x - m(t))^2 u(x, t) f(x, t) dx. \quad (3.32)$$

Аналогично можно рассмотреть и другие высшие центральные моменты одномерного выборочного распределения.

Заметим теперь, что из вывода уравнения Фоккера-Планка (3.13) из уравнения Чепмена-Колмогорова следует, матрица коэффициентов при вторых производных положительно определена. Смысл этих коэффициентов выяснен выше: ими определяется эволюция моментов и ковариаций двумерного распределения. С другой стороны, для выборочного распределения положительная определенность диффузионной матрицы отнюдь не очевидна, поскольку элементы этой матрицы определяются эмпирически по конкретной выборке временного ряда, для которой не известны моменты условной вероятности, требующиеся в уравнении Чепмена-Колмогорова. Иными словами, на практике матрица подбирается эмпирически так, чтобы согласовать эволюцию средних величин по выборке с их эволюцией в силу кинетического уравнения. Если же матрица будет отрицательно определена, не очень осмысленно решать в этом случае эволюционную задачу как диффузионную с мнимым среднеквадратичным отклонением. Таким образом, для корректности модели необходимо ис-

следовать главные миноры матрицы $D = \begin{vmatrix} \lambda(t) & \chi(t) \\ \chi(t) & \mu(t) \end{vmatrix}$. Заметим, одна-

ко, что для наших целей достаточно доказать неотрицательность только одного коэффициента $\lambda(t)$, поскольку в явном виде эволюционная модель моментов временного ряда имеет дело с одномерным уравнением Фоккера-Планка (3.14).

▼ **Теорема 3.1.** Если величина $\lambda(t)$ определена по некоторой выборке временного ряда в соответствии с уравнением (3.27)

$\lambda = \frac{d\sigma^2}{dt} - 2\text{cov}_{x,v}$, то уравнение (3.14) является уравнением Фоккера-Планка с неотрицательным коэффициентом диффузии.

Доказательство. Поскольку для временного ряда производная по времени понимается в разностном смысле, то скорость равна $v(t) = x(t+1) - x(t)$, а производная выборочной диспер-

сии по времени равна $\frac{d\sigma^2}{dt} = \sigma^2(t+1) - \sigma^2(t)$. Пусть длина выборки равна n . Тогда для выборочного среднего $\bar{x}(t)$ справедлива рекуррентная формула

$$\bar{x}(t+1) = \bar{x}(t) + \frac{x(t+1) - x(t-n+1)}{n}, \quad (3.33)$$

а для выборочной ковариации между двумя рядами имеем $\text{cov}_{xy}(t+1) =$

$$\begin{aligned} &= \text{cov}_{xy}(t) - \frac{1}{n^2} (y(t+1) - y(t-n+1)) \cdot (x(t+1) - x(t-n+1)) + \\ &+ \frac{1}{n} \{ y(t+1) \cdot (x(t+1) - \bar{x}(t)) - \bar{y}(t) \cdot (x(t+1) - x(t-n+1)) + \\ &+ \bar{x}(t) \cdot y(t-n+1) \}. \end{aligned} \quad (3.34)$$

Из (3.34) в частном случае $x = y$ получаем рекуррентную формулу для выборочной дисперсии:

$$\begin{aligned} \sigma_x^2(t+1) &= \sigma_x^2(t) - \frac{1}{n^2} (x(t+1) - x(t-n+1))^2 + \\ &+ \frac{1}{n} (x(t+1) - x(t-n+1)) (x(t+1) + x(t-n+1) - 2\bar{x}(t)). \end{aligned} \quad (3.35)$$

Ковариация же координаты и скорости равна

$$\begin{aligned} \text{cov}_{xv}(t) &= \frac{1}{n} \sum_{k=t-n+1}^t (x(k) - \bar{x}(t)) (v(k) - \bar{v}(t)) = \\ &= \frac{1}{n} \sum_{k=t-n+1}^t x(k) (x(k+1) - x(k)) - \frac{1}{n} \bar{x}(t) (x(t+1) - x(t-n+1)). \end{aligned} \quad (3.36)$$

В результате получаем, что

$$\begin{aligned} \lambda(t) &= \sigma^2(t+1) - \sigma^2(t) - 2\text{cov}_{x,v}(t) = \\ &= -\frac{1}{n^2} (x(t+1) - x(t-n+1))^2 + \\ &+ \frac{1}{n} (x^2(t+1) - x^2(t-n+1)) - \frac{2}{n} \sum_{k=t-n+1}^t x(k) (x(k+1) - x(k)) = \\ &= \frac{1}{n} \sum_{k=t-n+1}^t (x(k) - x(k+1))^2 - \frac{1}{n^2} (x(t+1) - x(t-n+1))^2 \geq 0. \end{aligned}$$

Последняя оценка в цепочке преобразований следует из того, что

$$\begin{aligned} \sum_{k=t-n+1}^t (x(k) - x(k+1))^2 &\geq \left(\sum_{k=t-n+1}^t (x(k) - x(k+1)) \right)^2 = \\ &= (x(t+1) - x(t-n+1))^2, \end{aligned}$$

и $1/n \geq 1/n^2$. Теорема 3.1 доказана. ▲

Аналогично показывается, что и коэффициент $\mu(t) \geq 0$. Поскольку же локально по времени симметрическая диффузионная матрица не зависит от x, v , ее можно привести к диагональному виду невырожденным линейным преобразованием. Для диагональной матрицы по доказанному

выше следует неотрицательность собственных значений, поэтому на каждом шаге по времени модель, использующая двумерное уравнение Фоккера-Планка (3.13), корректна.

Итак, возможны различные подходы к прогнозированию ВПФР: непосредственно по уравнению Лиувилля; по системе, включающей уравнение Лиувилля и уравнения эволюции моментов в пространстве скоростей; по системе, включающей уравнение Фоккера-Планка и уравнения эволюции моментов. В следующем параграфе будет рассмотрен общий подход к конструированию прогнозных моделей для ВПФР и самого временного ряда.

9.4. Кинетический подход к выводу моделей временных рядов

Как уже говорилось ранее, метод кинетических уравнений является основным инструментом для вывода моделей макроскопических процессов «из первых принципов», т.е. из микроскопического описания динамики отдельных частиц. Этот метод может быть распространен и на модели временных рядов.

Стандартные модели стационарных временных рядов представляют соотношения между некоторыми выборочными статистиками, которые перенесены затем на сами значения временного ряда. Например, регрессионные модели предполагают, что связь между переменными такая же, как и между их средними величинами по выборке, и т.п. В этой связи методически полезно рассмотреть статистические модели рядов как следствия вероятностных гипотез относительно их выборочных распределений.

Обратимся снова к уравнению Лиувилля (1.2) для одночастичной ВПФР. Из него прямым следствием получается уравнение эволюции (1.6) среднего значения первого выборочного момента: $\dot{m} = U(t)$. Если прогнозом значения ряда считать прогноз среднего значения соответствующего распределения, то уравнение $\dot{m} = U(t)$, понимаемое в конечно-разностном смысле, при заданной априори средней скорости $U(t)$ представляет модель выделения заданного временного тренда:

$$x(t+1) = x(t) + U(t). \quad (4.1)$$

В другом случае, если ковариация $\text{cov}(x, \dot{x})$ достаточно мала, допустимо рассмотреть модель, в которой ряд и его приращения независимы. Тогда можно считать, что двумерная ВПФР факторизуется: $F(x, \dot{x}, t) = f(x, t)\varphi(\dot{x}, t)$. В этом случае ЭЛС не зависит от x и совпа-

дает со средней скоростью U : $u(x, t) = U(t)$. Следовательно, трендовый прогноз (4.1) получается, в частности, при принятии гипотезы статистической независимости координат и скоростей. В случае, если $U(t) = U_0 \neq 0$, из (4.1) следует линейная по времени регрессионная модель.

Если априорные соображения относятся к локальной скорости $u(x, t)$, получаются модели динамических систем. Эти модели возникают из статистического описания в тех случаях, если прогнозируется не сам ряд, а скорость его изменения. Тогда вместо (4.1) естественно предложить модель временного ряда, в которой приращения определяются средней локальной скоростью:

$$\dot{x} = u(x, t) \Rightarrow x_{n+1} - x_n = u(x_n; n). \quad (4.2)$$

Например, считая, что $u(x, t)$ не зависит от времени, получаем одномерную динамическую систему, порождающую стационарный временной ряд:

$$x_{n+1} = x_n + u(x_n) \equiv g(x_n). \quad (4.3)$$

В частном случае, когда $u(x)$ линейна по x , причем $U = 0$, из (4.3) получается авторегрессионная модель первого порядка.

Аналогично можно рассмотреть ВПФР, зависящую и от большего числа производных, т.е. функцию вида $F(x, \dot{x}, \ddot{x}, \dots, t)$, но надо помнить об ограниченности объема выборки определенной величиной T . В силу конечности объема выборки невозможно эмпирически определить производную более высокого порядка, чем $T-1$. Более того, чтобы оперировать одинаковыми массивами данных по каждой из r производных, необходимо иметь исходный ряд в количестве по крайней мере $T+r$ элементов. Если это условие выполнено, то теми же соображениями можно получить многомерные аналоги вышеперечисленных прогнозных моделей.

Например, уравнение Лиувилля для многомерной ВПФР при обрыве цепочки на производной по времени r -го порядка имеет вид

$$\begin{aligned}
& \frac{\partial F_r(x, \dot{x}, \dots, x^{(r-1)}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F_r + \frac{\partial}{\partial \dot{x}} \ddot{x} F_r + \dots + \\
& \quad + \frac{\partial}{\partial x^{(r-1)}} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r = 0, \\
& w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) = \\
& \quad = \int x^{(r)} F_{r+1}(x, \dot{x}, \dots, x^r; t) dx^{(r)}.
\end{aligned}
\tag{4.4}$$

Замыкание цепочки эволюционных уравнений на функции распределения r -ой размерности состоит в том, что среднее ускорение следующего порядка полагается равным нулю: $W(x, \dot{x}, \dots, x^{(r-1)}, x^{(r)}; t) = 0$. В этом случае уравнение эволюции $w(x, \dot{x}, \dots, x^{(r-1)}; t)$ будет содержать только моменты плотности распределения F_{r+1} :

$$\begin{aligned}
& \frac{\partial}{\partial t} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) = \\
& = F_r \frac{\partial w}{\partial t} + w \frac{\partial F_r}{\partial t} = \int x^{(r)} \frac{\partial F_{r+1}(x, \dot{x}, \dots, x^r; t)}{\partial t} dx^{(r)}.
\end{aligned}$$

Используя (4.21) и записывая аналогичное ему уравнение эволюции для F_{r+1} с той лишь разницей, что в последнем следует положить $W(x, \dot{x}, \dots, x^{(r-1)}, x^{(r)}; t) = 0$, получаем

$$\begin{aligned}
& \frac{\partial}{\partial t} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) = \\
& = F_r \frac{\partial w}{\partial t} + w \frac{\partial F_r}{\partial t} = \int x^{(r)} \frac{\partial F_{r+1}(x, \dot{x}, \dots, x^r; t)}{\partial t} dx^{(r)} = \\
& = - \int x^{(r)} \left(\frac{\partial}{\partial x} \dot{x} F_{r+1} + \frac{\partial}{\partial \dot{x}} \ddot{x} F_{r+1} + \dots + \frac{\partial}{\partial x^{(r-1)}} x^{(r)} F_{r+1} \right) dx^{(r)},
\end{aligned}$$

откуда следует

$$\begin{aligned}
& \left(\frac{\partial w}{\partial t} + \frac{\partial \langle \dot{x} x^{(r)} \rangle}{\partial x} + \frac{\partial \langle \ddot{x} x^{(r)} \rangle}{\partial \dot{x}} + \dots + w \frac{\partial w}{\partial x^{(r-1)}} \right) F_r = \\
& = \left(w \dot{x} - \langle \dot{x} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial x} + \left(w \ddot{x} - \langle \ddot{x} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial \dot{x}} + \dots + \\
& + \left(w x^{(r-1)} - \langle x^{(r-1)} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial x^{(r-2)}} + \frac{\partial}{\partial x^{(r-1)}} \left(w^2 - \langle (x^{(r)})^2 \rangle \right) F_r.
\end{aligned}
\tag{4.5}$$

Что касается модели временного ряда, основанной на уравнении (4.4), то соответствующей модельной функцией здесь является $w(x, \dot{x}, \dots, x^{(r-1)}; t)$. Именно, r -мерная прогнозная модель имеет вид

$$x^{(r)}(t) = w(x, \dot{x}, \dots, x^{(r-1)}; t). \tag{4.6}$$

Заметим, что уравнение (4.4) формально получается в результате интегрирования цепочки уравнений для многомерных ВПФР по всем производным $x^{(k)}$ от порядка r включительно и до бесконечности. Если затем полученное уравнение проинтегрировать по первым $s < r-1$ переменным, то получающиеся в результате модели рядов относительно производной $x^{(r-1)}$ будут иметь порядок $r-s-1$. Эти модели составляют класс интегрированных моделей порядка s .

Можно также проинтегрировать (4.4) по выборочным аргументам, идущим не обязательно подряд, после чего полученное уравнение позволит связать между собой переменные, разделенные заданными лагами, что дает в итоге модели с распределенным запаздыванием. Если затем сделать предположения о статистической зависимости распределений — например, считать, что многомерные распределения переменных с лагами функционально зависят только от одной одномерной плотности распределения, то в результате получится некоторая самосогласованная модель. В класс самосогласованных моделей попадают, в частности, адаптивные модели, в которых прогноз значений ряда использует данные о точности прогноза в предыдущие моменты времени.

Перечисленные возможности применения кинетических уравнений показывают, что эти уравнения могут рассматриваться как инструмент для создания статистических моделей временных рядов с заданными свойствами, а также для отбора моделей, наиболее близко подходящих для описания тех или иных реальных ситуаций.

Дадим теперь схему практических действий, которые следует предпринять для получения прогнозной модели временного ряда, исходя из кинетического описания его ВПФР. Общая идея прогнозирования состоит в следующем. Сначала строится горизонтная статистика и определяется оптимальный объем выборки для прогнозирования ВПФР на τ шагов вперед с заданной точностью ε . Затем по имеющимся данным вычисляются эмпирические статистики: ЭЛС, второй и третий моменты, фигурирующие в уравнениях «гидродинамики», т.е. в уравнении эволюции средних степеней приращений элементов ряда. Если какие-либо из этих статистик оказалось возможным функционально смоделировать с требуемой точностью ε , то выбирается соответствующий способ замыкания моментных уравнений, после чего решается система этих уравнений, включая исходное кинетическое уравнение — Лиувилля, Фоккера-Планка или более общее, из описанных выше. Прогнозом временного ряда на следующем шаге по времени называется среднее значение построенного прогнозного распределения.

Продemonстрируем эту схему решения на примере эмпирического уравнения Лиувилля. Если нам известна скорость $u(x, t)$ в текущий момент времени, тем самым известна и функция распределения в будущий момент времени согласно формуле

$$u(i+1, t)f(i+1, t) = - \sum_{k=1}^i (f(k, t+1) - f(k, t)). \quad (4.7)$$

Предположим, что удалось замкнуть моментную систему, так что она оказалась рекуррентно разрешима. Тогда после численного нахождения $u(i, t)$ распределение $f(i, t+1)$ определяется по скоростной модели:

$$\begin{aligned} f(1, t+1) &= f(1, t) - u(2, t)f(2, t); \\ f(2, t+1) &= f(2, t) - u(3, t)f(3, t) + u(2, t)f(2, t); \dots \end{aligned} \quad (4.8)$$

Прогнозным значением $\hat{x}(t+1)$ будет считаться среднее значение по распределению $f(x, t+1)$:

$$\hat{x}(t+1) = \int x f(x, t+1) dx = \sum_i i f(i, t+1). \quad (4.9)$$

Динамическая система, порождающая этот временной ряд, приближенно находится из (4.7). Обозначая правую часть (4.7), деленную на $f(i+1, t)$, через $g(i, t)$, получаем

$$u(i, t) = g(i-1, t). \quad (4.10)$$

В формуле (4.10) скорость $u(i, t)$ трактуется как изменение значения самого временного ряда, т.е. $u(i, t) = x(t+1) - x(t)$. Если окажется, что правая часть приближенно может считаться независимой явно от времени t , то получающаяся динамическая система будет автономна, т.е. приращение значения случайной величины $x(t)$ (номера ячейки i) будет определяться только значением самой этой величины. В нестационарном случае правая часть (4.10) помимо $x(t)$ зависит также явно от времени:

$$x(t+1) = x(t) + g(x(t), t). \quad (4.11)$$

Для стационарного распределения лиувиллева скорость равна нулю, и прогноз дает одну и ту же величину, равную среднему значению по этому распределению.

Аналогичный подход будем применять и для уравнения Фоккера-Планка и согласованных с ним уравнений эволюции моментов. В зависимости от порядка момента двумерной ВПФР, на котором замыкается кинетическое уравнение для одномерной ВПФР, получается более или менее сложная динамическая система, моделирующая временной ряд. Однако, если учесть, что естественным шагом по времени является обратная частота наблюдения данных временного ряда, то для двумерного уравнения Фоккера-Планка эволюция моментов порядка выше первого не дает новой информации. В результате замыкание может быть проведено на втором уравнении системы, т.е. на уравнении эволюции для скорости.

Считаем, что параметр $\lambda(t)$ в уравнении Фоккера-Планка (3.14), оцениваемый по временному ряду, известен с предыдущего шага по времени: $\lambda(t) = \lambda(t-1)$. Тогда, если считать, что скорость $u(x, t)$ в момент t откуда-либо известна, например, по наивному прогнозу $u(x, t) = u(x, t-1)$, то модель замыкается на первом же уравнении, т.е. на уравнении (3.14). Сама ВПФР находится в этом случае автоматически по уравнению (4.7). С точки зрения прогнозирования временного ряда как среднего значения по текущему распределению, такая модель представляет наивный прогноз по скорости изменения среднего значения. При этом для прогноза ряда находить будущую ВПФР не нужно, так как для этих целей достаточно знания скорости изменения ВПФР $u(x, t)$, а эта скорость известна с предыдущего шага по времени в рамках наивного прогноза.

Если требуется спрогнозировать и саму эту скорость, то в описанную схему надо добавить уравнение (3.19) для эволюции $u(x, t)$, в котором

параметры $e(x, t)$, $a(x, t)$ и $\chi(t)$ считаются известными с предыдущего шага по времени, т.е. теперь уже в отношении этих величин используется концепция наивного прогноза. Тогда система уравнений, которые надо совместно решать, следующая:

$$\begin{aligned}\frac{\partial f}{\partial t} &= \frac{\lambda}{2} \frac{\partial^2 f}{\partial x^2} - \frac{\partial}{\partial x}(uf); \\ \frac{\partial(uf)}{\partial t} &= \frac{\lambda}{2} \frac{\partial^2(uf)}{\partial x^2} - \frac{\partial}{\partial x}(e + \chi)f + af.\end{aligned}\quad (4.12)$$

Может показаться, что получилась самосогласованная система диффузионных уравнений со сносом и источником. Однако на самом деле вторая строка решается независимо от первой. Дело в том, что начальные условия для плотности $f(x, t)$ и скорости $u(x, t)$ задаются в разные моменты времени: если задана $f_0(x) = f(x, t_0)$, то скорость задается в предыдущий дискретный момент времени $u_0(x) = u(x, t_0 - 1)$. В противном случае, если задать и $u(x, t_0)$, то в силу первого уравнения (4.12) сразу будет найдена $f(x, t_0 + 1)$, а второе уравнение превратится в тождество.

На следующем уровне по порядку учитываемых моментов добавляется уравнение (3.21) эволюции для $e(x, t)$. Если бы мы не оборвали цепочку кинетических уравнений на двумерной ВПФР, а учли бы и уравнение эволюции для трехмерной ВПФР $\Phi(x, v, w, t)$, то потребовалось бы еще и уравнение эволюции локальной плотности ускорения $a(x, t)$. В модели, замыкаемой на двумерной ВПФР, ускорение определяется в рамках наивного прогноза. Поэтому $a(x, t)$ считается известной функцией с пред-предыдущего шага по времени. В результате имеем

$$\begin{aligned}\frac{\partial f}{\partial t} &= \frac{\lambda}{2} \frac{\partial^2 f}{\partial x^2} - \frac{\partial}{\partial x}(uf); \\ \frac{\partial(uf)}{\partial t} &= \frac{\lambda}{2} \frac{\partial^2(uf)}{\partial x^2} - \frac{\partial}{\partial x}(e + \chi)f + af; \\ \frac{\partial(e f)}{\partial t} &= \frac{\lambda}{2} \frac{\partial^2(e f)}{\partial x^2} - \frac{\partial}{\partial x}(e_3 + 2\chi u)f + \beta f + \mu.\end{aligned}\quad (4.13)$$

Аналогично можно получить замкнутую систему относительно выборочных моментов двумерного распределения более высокого порядка. Но на практике при прогнозировании выборочных распределений с шагом по времени, равным шагу появления значений временного ряда (например, цены закрытия какого-либо финансового инструмента с шагом 1 минута), достаточно использовать систему (4.12). Это связано с тем, что моменты высших порядков $e_k(x, t)$, $k \geq 2$ известны в тот же момент времени, что и скорость $u(x, t)$. В результате третье уравнение в системе (4.13) обращается в тождество.

Запишем дискретную схему для решения задачи (4.12) на шаг вперед. Шляпкой обозначим прогнозируемые величины. Используем явную разностную схему для эволюции по времени и стандартный шаблон левой разностной производной.

Для ВПФР:

$$\begin{aligned}\mathbf{f} : \hat{f}(x, t+1) &= f(x, t) + f(x, t)\hat{u}(x, t) - f(x+1, t)\hat{u}(x+1, t) + \\ &+ \frac{\lambda(t-1)}{2}(f(x+2, t) - 2f(x+1, t) + f(x, t));\end{aligned}\quad (4.14)$$

для ЭЛС:

$$\begin{aligned}
u: f(x, t) \hat{u}(x, t) = & f(x, t-1)u(x, t-1) + a(x, t-2)f(x, t-1) + \\
& + (e(x, t-1) + \chi(t-1))f(x, t-1) - \\
& - (e(x+1, t-1) + \chi(t-1))f(x+1, t-1) + \\
& + \frac{\lambda(t-1)}{2} (f(x+2, t-1)u(x+2, t-1) + f(x, t-1)u(x, t-1)) - \\
& - \lambda(t-1)f(x+1, t-1)u(x+1, t-1).
\end{aligned}
\tag{4.15}$$

Правая часть уравнения в (4.15) известна, поэтому прогноз скорости строится автоматически во всех ячейках гистограммы, представляющей ВПФР, в которых $f(x, t) > 0$. В тех ячейках, где $f(x, t) = 0$, скорость $\hat{u}(x, t)$ будем полагать равной нулю. Это не приведет к артефактам вычислений, так как в уравнении эволюции ВПФР скорость входит в комбинации с произведением на плотность распределения в той же ячейке. После нахождения $\hat{f}(x, t)$ отрицательные значения (если они есть) обнуляются, и проводится перенормировка построенного распределения.

Эта схема может быть продолжена для прогнозирования на произвольный горизонт вперед. При этом каждое найденное решение на следующем шаге присоединяется к скользящему окну выборки, по которой строится уравнение Фоккера-Планка. Начальные условия ставятся на момент начала прогнозирования, граничные условия — нулевые.

Альтернативным методом прогнозирования эмпирической скорости является вероятностное продолжение траектории динамической системы. Если по имеющимся данным к текущему моменту времени собрана информация о распределении параметра λ , соответствующую плотность обозначим через $h(\lambda)$, и построено распределение $p_x(u)$ скорости u в x -ой ячейке гистограммы, то уравнение системы (4.14) можно усреднить по этим распределениям. В результате на шаге, равном горизонту прогнозирования, получаем модель

$$\begin{aligned}
\hat{f}(x, t+1) = & f(x, t)(1 + \bar{u}(x)) - f(x+1, t)\bar{u}(x+1) + \\
& + \frac{\bar{\lambda}}{2} (f(x+2, t) - 2f(x+1, t) + f(x, t)); \\
\bar{u}(x) = & \int p_x(u) du, \quad \bar{\lambda} = \int \lambda h(\lambda) d\lambda.
\end{aligned}
\tag{4.16}$$

Возможна и другая модификация модели (4.14) - (4.15), когда аналогичное усреднение проводится во втором уравнении системы.

В результате получено одно уравнение на ВПФР, коэффициенты которого определены из статистических наблюдений за рядом в предыдущие моменты времени. Практически это более удобный подход с вычислительной точки зрения.

9.5. Эволюция внутри горизонта квазистационарности

Рассмотрим задачу прогнозирования ВПФР внутри границ τ - ε -стационарности при малых значениях ε , т.е. в предположении, что плотность распределения близка к стационарной в узком смысле. Будем искать решение цепочки уравнений Лиувилля для плотностей $f(x, t)$, $F(x, \dot{x}, t)$ и т.д. в виде формального ряда по степеням параметра ε . «Формальность» в данном случае означает, что вопросы сходимости такого ряда не обсуждаются, как не ставится и вопрос о существовании функции, аналитической по ε в окрестности нуля, разложение которой порождает указанный ряд. Итак, пусть

$$\begin{aligned}
f(x, t) = & f^{(0)}(x) + \varepsilon f^{(1)}(x, t) + \varepsilon^2 f^{(2)}(x, t) + \dots; \\
F(x, \dot{x}, t) = & F^{(0)}(x, \dot{x}) + \varepsilon F^{(1)}(x, \dot{x}, t) + \varepsilon^2 F^{(2)}(x, \dot{x}, t) + \dots.
\end{aligned}
\tag{5.1}$$

Аналогичные разложения записываются и для высших ВПФР. Здесь мы рассмотрим подробный вывод слабо нестационарных уравнений в случае, когда замыкание цепочки проводится на трехмерной плотности распределения.

Заметим, что, по предположению близости ВПФР к стационарным, распределения нулевого по ε приближения $f^{(0)}$, $F^{(0)}$, $F_3^{(0)}$ не зависят от времени. Это — квазистационарные распределения в рамках концепции «наивного прогноза». Поскольку же тогда

$$\frac{\partial f^{(0)}}{\partial t} = -\frac{\partial}{\partial x} (u f^{(0)}) = 0,$$

причем $f^{(0)}$ не равна тождественно нулю, то в стационарном случае ЭЛС равна нулю: $u(x, t) = 0$. Отсюда следует, что для

стационарных распределений первый момент по скорости двумерной ПФР должен быть равен нулю:

$$u_0(x) = \frac{1}{f^{(0)}(x)} \int \dot{x} F^{(0)}(x, \dot{x}) d\dot{x} = 0, \quad f^{(0)}(x) \neq 0. \quad (5.2)$$

Учтем теперь, что имеющаяся в начальный момент времени двумерная плотность $F^{(0)}(x, \dot{x})$ является не «истинной» стационарной плотностью совместного распределения координат и скоростей, а выборочной плотностью, т.е. не ПФР, а ВПФР. Поэтому интеграл в (5.2), вычисленный по эмпирической функции $F^{(0)}(x, \dot{x})$, не обязательно равен нулю. Отличия его от нуля должны быть пренебрежимо малыми (в рамках требований, предъявляемых к конкретной задаче), чтобы можно было применить разложения (5.2).

Подчеркнем обнаруженное методологическое различие между одномерной $f^{(0)}(x)$ и двумерной $F^{(0)}(x, \dot{x})$ плотностями. Если любая отрицательная эмпирическая плотность $f^{(0)}(x)$, удовлетворяющая условию нормировки на $[0; 1]$, может быть интерпретирована как распределение некоторой генеральной совокупности, то не любая совместная плотность распределения значений временного ряда и их приращений годится на такую роль, а лишь такая, которая удовлетворяет условию (5.2).

Следовательно, система эволюционных уравнений последовательных приближений по ε для ВПФР должна быть такой, чтобы частная производная по времени в уравнении Лиувилля сразу имела бы порядок ε . Эта ситуация аналогична разложению распределений в классической цепочке Боголюбова [6] по параметру неоднородности в окрестности положения термодинамического равновесия (см. Часть I), с той разницей, что в статистической механике нулевое приближение для плотностей распределений частиц по координатам и скоростям пространственно однородно, а в нашем случае нулевое приближение не зависит от времени.

Снабдим тогда среднюю скорость $u(x, t)$ в правой части уравнения Лиувилля параметром ε для удобства выделения порядка приближения:

$$u(x, t) = \varepsilon u_1(x, t) + \varepsilon^2 u_2(x, t) + \dots \quad (5.3)$$

По построению, функции u_1, u_2, \dots представляют собой уточнения к нулевому приближению, причем средняя скорость первого приближения определяется по исходной двумерной ВПФР, т.е. по эмпирическому распределению нулевого приближения:

$$\varepsilon u_1(x, t) = \varepsilon u_1(x) = \frac{1}{f^{(0)}(x)} \int \dot{x} F^{(0)}(x, \dot{x}) d\dot{x}. \quad (5.4)$$

Поскольку же

$$\begin{aligned} u(x, t) &= \frac{1}{f(x)} \int \dot{x} F(x, \dot{x}) d\dot{x} = \\ &= \frac{\int \dot{x} F(x, \dot{x}) d\dot{x}}{\int F(x, \dot{x}) d\dot{x}} = \frac{\int \dot{x} (F^{(0)}(x, \dot{x}) + \varepsilon F^{(1)}(x, \dot{x}, t) + \dots) d\dot{x}}{f^{(0)}(x) + \varepsilon f^{(1)}(x, t) + \dots}, \end{aligned}$$

то поправка второго порядка по ε к средней локальной скорости связана с поправками первого порядка к ВПФР:

$$\varepsilon^2 u_2(x, t) = \frac{\varepsilon}{f^{(0)}(x)} \int \dot{x} F^{(1)}(x, \dot{x}, t) d\dot{x} - \frac{\varepsilon^2 f^{(1)}(x, t)}{f^{(0)}(x)} u_1(x). \quad (5.5)$$

Подставим разложения (5.1) и (5.3) в уравнение Лиувилля (1.2). Приравнявая члены одинакового порядка по ε , получаем следующую систему последовательно разрешаемых уравнений для одномерной ВПФР:

$$\begin{aligned} \frac{\partial f^{(0)}}{\partial t} &= 0; \\ \frac{\partial f^{(1)}}{\partial t} &= -\frac{\partial}{\partial x} (f^{(0)} u_1); \\ \frac{\partial f^{(2)}}{\partial t} &= -\frac{\partial}{\partial x} (f^{(0)} u_2) - \frac{\partial}{\partial x} (f^{(1)} u_1); \dots \end{aligned} \quad (5.6)$$

Первое уравнение системы (5.6) определяет квазистационарное распределение, под которым в данном случае следует понимать выборочную плотность $f(x, t_0)$, построенную по выборке оптимального объема данных, известных к моменту времени t_0 .

В силу равенств (5.4) и (5.5) правая часть второго уравнения в (5.6) также является известной функцией координаты x , построенной по квазистационарному совместному распределению координат и скоростей в момент времени t_0 . Тогда это уравнение может быть проинтегрировано по времени, что дает

$$f^{(1)}(x, t) = -t \frac{\partial}{\partial x} (f^{(0)} u_1) + c_1(x),$$

где $c_1(x)$ есть некоторая произвольная функция координаты. Поскольку в начальный момент времени при $t = t_0$ распределение по построению было стационарным и определялось плотностью $f^{(0)}(x)$, то начальное условие должно быть таким: $f^{(1)}(x, t_0) = 0$. Это приводит к следующему выражению для $f^{(1)}(x, t)$:

$$f^{(1)}(x, t) = -(t - t_0) \frac{\partial}{\partial x} (f^{(0)} u_1). \quad (5.7)$$

Далее, чтобы найти из третьего уравнения системы (5.6) поправку второго порядка $f^{(2)}(x, t)$, следует определить $u_2(x, t)$, для чего, согласно (5.5), требуется построить первую поправку к двумерной квазистационарной ВПФР. Эта поправка $F^{(1)}(x, \dot{x}, t)$ строится таким же образом, как и поправка $f^{(1)}(x, t)$, с той лишь разницей, что для этой цели используется уравнение Лиувилля для двумерной ВПФР, которое при подстановке в него разложений двумерной и трехмерной ВПФР в виде (5.1) имеет вид:

$$\frac{\partial F^{(0)}}{\partial t} = 0; \quad \frac{\partial F^{(1)}}{\partial t} = -\frac{\partial}{\partial x} (F^{(0)} \dot{x}) - \frac{\partial}{\partial \dot{x}} (F^{(0)} w_1), \quad (5.8)$$

где величина

$$w_1(x, \dot{x}) = \frac{1}{F^{(0)}(x, \dot{x})} \int \ddot{x} F_3^{(0)}(x, \dot{x}, \ddot{x}) d\ddot{x} \quad (5.9)$$

является известной функцией. В результате получаем

$$F^{(1)}(x, \dot{x}, t) = -(t - t_0) \cdot \left(\frac{\partial}{\partial x} (F^{(0)} \dot{x}) + \frac{\partial}{\partial \dot{x}} (F^{(0)} w_1) \right). \quad (5.10)$$

Легко проверить, что при интегрировании (5.10) по \dot{x} получается поправка к одномерной ВПФР (5.7). Определим также соответствующий интеграл от приближения (5.9) первого порядка для ускорения:

$$W_1(x) f^{(0)}(x) = \int \ddot{x} F_3^{(0)}(x, \dot{x}, \ddot{x}) d\ddot{x}. \quad (5.11)$$

Из (5.10) - (5.11) и уравнения Лиувилля для $F(x, \dot{x}, t)$ получаем выражение для второго приближения к средней локальной скорости:

$$u_2(x, t) = (t - t_0) \cdot \left(W_1(x) - \frac{1}{f^{(0)}(x)} \frac{\partial}{\partial x} (e(x) f^{(0)}(x)) \right). \quad (5.12)$$

Подставляя найденную поправку (5.12) в третье уравнение системы (5.6) и интегрируя его по времени, получаем итоговое выражение для второго приближения к одномерной ВПФР:

$$f^{(2)}(x, t) = -\frac{(t - t_0)^2}{2} \frac{\partial}{\partial x} \left(f^{(0)} W_1 - u_1 \frac{\partial}{\partial x} f^{(0)} u_1 - \frac{\partial}{\partial x} f^{(0)} e \right). \quad (5.13)$$

Таким образом, если первая поправка представляет собой дивергенцию потока плотности вероятности, то вторая поправка имеет диффузионный вид. Выведенные формулы справедливы в приближении слабой нестационарности ВПФР внутри горизонта прогноза, т.е. при $0 < t - t_0 \leq \tau$. Как и в уравнениях цепочки Боголюбова, ВПФР высших приближений представляются в виде разложения по градиентам средних величин, играющих здесь роль соответствующих «гидродинамических» параметров.

Тем же методом можно получить и следующие приближения. Для построения k -го приближения к $f^{(0)}$ необходимо использовать уравнение Лиувилля для $F_k^{(0)}$, а также выражение для эволюции ускорения, определяемого функцией $F_{k+1}^{(0)}$. В результате на k -ом шаге величины

$f^{(k)}(x, t)$ будут представляться в виде частичной суммы порядка k степенного ряда по времени, коэффициенты которого определяются по известной ВПФР в начальный момент времени посредством дифференцирования по координате вплоть до k -го порядка.

Глава X. Маркированные точечные случайные процессы

10.1. Нестационарный пуассоновский поток событий

В этой главе изучаются двумерные ряды событий, когда каждое событие характеризуется двумя параметрами (координатами) — значением случайной величины и моментом времени (также случайным), когда это событие произошло. Такие процессы называются маркированными, маркировкой в данном случае является момент времени. Поток событий называется последовательность этих событий во времени. Интенсивность потока — это среднее число событий в единицу времени (например, для биржевого ряда это тик/мин).

Введем величину $p_k(t, t + \tau)$ как вероятность того, что за промежуток времени $\Delta_t(\tau) = [t; t + \tau]$ произойдет ровно k событий. Тогда интенсивность потока $\mu(t, t + \tau)$ на промежутке $\Delta_t(\tau)$ определяется формулой

$$\mu(t, t + \tau) = \frac{1}{\tau} \sum_{k=1}^{\infty} k p_k(t, t + \tau). \quad (1.1)$$

Если события независимы, то по теореме Хинчина [28] вероятность $p_k(t, t + \tau)$ для ординарного потока, когда в любой момент времени может произойти только одно событие, дается нестационарным распределением Пуассона

$$p_k(t, t + \tau) = \frac{(\Lambda(t, \tau))^k}{k!} \exp(-\Lambda(t, \tau)), \quad \Lambda(t, \tau) = \tau \mu(t, t + \tau). \quad (1.2)$$

Величина $\Lambda(t, \tau)$ называется параметром потока. В стационарном случае $\mu = \text{const}$ и $\Lambda(t, \tau) = \mu \tau$ — это среднее число событий за время τ . Следовательно, параметр потока равен среднему числу событий на промежутке $\Delta_t(\tau)$. Тогда в нестационарном случае величину $\Lambda(t, \tau)$ можно оценить непосредственно по наблюдениям. Для определенности будем считать единицу измерения минутой. Подчеркнем, что мы рассматриваем поток событий, принципиально не являющийся непрерывным во

времени, поэтому на практике обязательно приходится вводить некоторую единицу агрегирования данных, отличную от единицы измерения времени. Пусть, в частности, единица измерения — секунда, а единица агрегирования — минута. При этом мы часто будем писать производную функции распределения по времени, считая эту функцию, таким образом, дифференцируемой, и, следовательно, непрерывной. Эта идеализация будет относиться только к представлению о том, что единица агрегирования по времени выбрана достаточно малой по сравнению с масштабом времени наблюдения за процессом, и не будет использоваться при доказательствах тех или иных утверждений о свойствах функции распределения.

Нашей задачей является построение прогноза эмпирического распределения на горизонт τ и нахождение оценки точности такого прогноза. При этом ВПФР одномерного ряда последовательности собственно значений случайной величины принципиально строится по количественному промежутку данных, а не по временному. Однако следует заметить, что в любой момент времени t вся предыстория ряда нам известна, поэтому для любого промежутка времени в прошлом можно построить ВПФР по тому объему данных, который содержался в этом промежутке времени. Следовательно, опорным или стартовым распределением, известным в начальный момент времени t , всегда можно считать тиковую ВПФР $F_N(x, n)$, где n есть номер ближайшего слева события к моменту t . Для удобства, если это не приведет к искажению смысла утверждения, будем иногда вместо натурального аргумента n писать время t , подразумевая указанную упорядоченность номера события и момента времени. Тиковая ВПФР будет в этой главе обозначаться как $F_N(x, n)$, а ВПФР, построенная по тому количеству событий, которое произошло за промежуток времени $[t - T; t]$, если это количество событий отлично от нуля, обозначим $f_T(x, t)$. Событием, например, является отдельная торговая сделка на фондовом рынке.

Используя горизонтный анализ, для любого натурального k можно, по крайней мере, формально, определить такой оптимальный объем выборки $N(k)$, что для всех $i = 1, 2, \dots, k$ будет выполнено условие $\|F_{N(k)}(x, n) - \hat{F}_{N(k)}(x, n + i)\| \leq \varepsilon$. «Шляпкой» сверху, как и ранее, обозначена прогнозная ВПФР, вместо которой с точностью ε можно взять саму $F_{N(k)}(x, n)$, являющуюся на этом количестве шагов ε -стационарной. Теперь следует учесть, что число тиков k в промежутке

времени τ является случайной величиной, вероятность которой определяется формулой (1.2). Тогда прогнозной плотностью функции распределения $\hat{f}(x, t + \tau)$, реализующейся через время τ , естественно считать прогнозную тиковую ЭПФР, усредненную по возможному числу событий за этот промежуток времени:

$$\hat{f}_T(x, t + \tau) = \sum_{k=0}^{\infty} p_k(t, t + \tau) F_{N(k)}(x, n(t)). \quad (1.3)$$

Для простоты в (1.3) можно считать, что $N(k) = [2k / \varepsilon]$, хотя в общем случае $N(k)$ есть оптимальное значение, получаемое из анализа горизонтной статистики.

Очевидно, распределение (1.3) нормировано на единицу, поскольку при интегрировании правой части (1.3) по x и после перестановки операций суммирования и интегрирования остается сумма вероятностей $p_k(t, t + \tau)$, равная единице. Переставлять же сумму и интеграл можно в силу равномерной сходимости интеграла по параметру k , поскольку этот параметр выделяет объем выборки для построения ВПФР $F_{N(k)}(x, n)$ независимо от x .

Следующее утверждение показывает, что так построенная ВПФР $\hat{f}(x, t + \tau)$ является ε -стационарной при любом фактически состоявшемся числе событий за этот промежуток времени.

▼ **Теорема 1.1.** Пусть $F_{N(k)}(x, n)$ — ВПФР, построенная по выборке объема $N(k) = [2k / \varepsilon]$ в момент времени t . Тогда, каково бы ни было распределение вероятностей $p_k(t, t + \tau)$ при условии равномерной ограниченности параметра потока по обоим переменным, существует такое $s = s(\varepsilon)$, что ВПФР $\hat{f}(x, t + \tau)$, определяемая по формуле

$$\hat{f}(x, t + \tau) = \sum_{k=0}^s p_k(t, t + \tau) F_{N(k)}(x, n(t)), \quad (1.4)$$

отличается от фактически состоявшейся ВПФР, построенной по выборке $N(m)$, где m есть число фактических событий на промежутке $\Delta_t(\tau)$, не более чем на 2ε .

Доказательство. Рассмотрим следующую цепочку оценок:

$$\begin{aligned}
& \left\| F_{N(m)}(x, n) - \sum_{k=0}^{\infty} p_k(t, t + \tau) F_{N(k)}(x, n(t)) \right\| = \\
& = \left\| \sum_{k=0}^{\infty} p_k(t, t + \tau) (F_{N(m)}(x, n) - F_{N(k)}(x, n(t))) \right\| \leq \\
& \leq \sum_{k=0}^{\infty} p_k(t, t + \tau) \|F_{N(m)}(x, n) - F_{N(k)}(x, n(t))\| = \\
& = \sum_{k=0}^m p_k(t, t + \tau) \|F_{N(m)}(x, n) - F_{N(k)}(x, n(t))\| + \\
& + \sum_{k=m+1}^{\infty} p_k(t, t + \tau) \|F_{N(m)}(x, n) - F_{N(k)}(x, n(t))\|.
\end{aligned}$$

Отличие нормы во второй сумме в последнем выражении этой цепочки равенств не превосходит ε в силу определения ε -стационарности, а норма в первой сумме допускает очевидную оценку

$$\|F_{N(m)}(x, n) - F_{N(k)}(x, n(t))\| \leq \frac{2(m-k)}{N(m)} \leq \left(1 - \frac{k}{m}\right) \varepsilon.$$

В результате получаем

$$\begin{aligned}
& \left\| F_{N(m)}(x, n) - \sum_{k=0}^{\infty} p_k(t, t + \tau) F_{N(k)}(x, n(t)) \right\| \leq \\
& \leq \varepsilon \sum_{k=0}^m p_k(t, t + \tau) \left(1 - \frac{k}{m}\right) + \varepsilon \sum_{k=m+1}^{\infty} p_k(t, t + \tau) \leq \\
& \leq \varepsilon \sum_{k=0}^m p_k(t, t + \tau) + \varepsilon \sum_{k=m+1}^{\infty} p_k(t, t + \tau) = \varepsilon \sum_{k=0}^{\infty} p_k(t, t + \tau) = \varepsilon.
\end{aligned}$$

Этим доказана ε -стационарность прогнозной ВПФР (1.3). Заметим теперь, что на практике приходится иметь дело с конечным объемом данных, поэтому необходимо иметь представление о допустимой неточности, если вместо бесконечного ряда использовать конечную сумму (1.4). Тогда рассмотрим величину остатка

$$r_s = \left\| \sum_{k=s+1}^{\infty} p_k(t, t + \tau) F_{N(k)}(x, n(t)) \right\|.$$

Поскольку ВПФР $F_N(x, n)$ нормирована на единицу, то при равномерной ограниченности параметра потока по времени величиной λ получаем для остаточного члена ряда оценку

$$r_s \leq \sum_{k=s+1}^{\infty} p_k(t, t + \tau) = e^{-\lambda} \sum_{k=s+1}^{\infty} \frac{\lambda^k}{k!}.$$

Последняя сумма представляет собой выражение для остаточного члена разложения экспоненты в ряд Тэйлора в окрестности нуля, поэтому величина остатка может быть оценена по остаточному члену в форме Коши в виде $\frac{\lambda^{s+1}}{(s+1)!} e^{\lambda}$, что дает в итоге оценку

$$r_s \leq \frac{\lambda^{s+1}}{(s+1)!}. \quad (1.5)$$

На практике при конечном горизонте прогнозирования среднее число событий λ на этом промежутке времени конечно, поэтому предположение о равномерной ограниченности параметра потока оправданно. Поскольку предел выражения (1.5) при фиксированном λ и при $s \rightarrow \infty$ равен нулю, всегда можно найти такой конечный номер $s(\varepsilon)$, что $r_s < \varepsilon$ для всех $s > s(\varepsilon)$. В результате по правилу «двух ε » ($F_N(x, n)$ отличается от полной суммы не более чем на ε , а полная отличается от частичной не более чем на ε) получаем

$$\left\| F_{N(m)}(x, n) - \sum_{k=0}^s p_k(t, t + \tau) F_{N(k)}(x, n(t)) \right\| < 2\varepsilon.$$

Теорема 1.1 доказана. ▲

Заметим здесь, что если вместо модели (1.3) использовать в качестве прогнозной ВПФР $\hat{F}_{N(k)}(x, n + k)$ результат решения по кинетической модели, описанной в параграфе 1.5, то ошибка прогноза временной ЭПФР не ухудшится.

Таким образом, если параметр потока событий $\Lambda(t, \tau)$ известен, то всегда можно построить ε -стационарную ВПФР не только по прошествии любого числа тиков, но и через любой промежуток времени. В ре-

зультате задача прогнозирования ВПФР свелась к нахождению нестационарного параметра потока и подстановке его в эмпирическое уравнение Лиувилля, которое следует преобразовать в терминах временного промежутка.

10.2. Уравнение Лиувилля в терминах времени и номера события

Предположим, что параметр потока $\Lambda(t, \tau)$ из (1.2) известен. Тогда можно перейти к построению прогноза ВПФР $\hat{F}_{N(k)}(x, n+k)$.

По доказанной выше теореме 1.1, в качестве прогнозной плотности на k шагов вперед с ε -точностью можно взять исходную ВПФР $F_{N(k)}(x, n)$. После этого, если требуется построить прогноз не в тиковых терминах, а во времени, можно воспользоваться формулой (1.3). Для повышения точности прогноза следует построить эволюцию ВПФР на заданном горизонте.

В тиковых терминах уравнение эволюции ВПФР имеет вид (9.1.2), где указание на объем выборки для краткости опущено:

$$\begin{aligned} \frac{\partial F(x, n)}{\partial n} + \frac{\partial}{\partial x} u(x, n) F(x, n) &= 0, \\ u(x, n) &= \int \dot{x} F^{(2)}(x, \dot{x}, n) d\dot{x}, \end{aligned} \quad (2.1)$$

где эмпирическая скорость $u(x, n)$ в тиковых терминах выражается через совместную ВПФР $F^{(2)}(x, \dot{x}, n)$ значений ряда и его приращений в терминах номера события.

Чтобы построить аналогичное уравнение для ВПФР в терминах времени, надо вспомнить, что на самом деле наш случайный процесс не непрерывный, поэтому уравнение будет записано не для произвольного приращения по времени, а для единицы агрегирования, т.е. в нашем случае для шага по времени в 1 минуту. В этом смысле оно будет похоже на уравнение (2.1), которое на самом деле записано для шага в 1 тик.

Поскольку производная по времени выражается через производную по тикам согласно формуле

$$\frac{\partial F(x, n)}{\partial t} = \frac{\partial F(x, n)}{\partial n} \frac{dn}{dt}, \quad (2.2)$$

и эту последнюю надо понимать в агрегированном смысле, т.е. $dn/dt = n(t+1) - n(t)$, то

$$\frac{dn}{dt} = \Lambda(t, 1). \quad (2.3)$$

Сравнивая (2.2) с (2.1) замечаем, что уравнение Лиувилля можно переписать в виде эволюции по агрегированному моменту времени (соответствующую ВПФР обозначаем, как и в п.10.1, через $f(x, t)$):

$$\begin{aligned} \frac{\partial f(x, t)}{\partial t} &= -\Lambda(t, 1) \frac{\partial}{\partial x} u(x, t) f(x, t) = -\frac{\partial}{\partial x} v(x, t) f(x, t), \\ v(x, t) &= \Lambda(t, 1) u(x, t). \end{aligned}$$

(2.4)

Прогнозная модель для ВПФР $f(x, t)$ следует из (2.4) заменой $u(x, t)$ и $\Lambda(t, 1)$ на их прогнозные значения:

$$\hat{f}(x, t+1) = f(x, t) - \frac{\partial}{\partial x} \hat{\Lambda}(t, 1) \hat{u}(x, t) f(x, t). \quad (2.5)$$

Согласно теореме 1.1, объем выборки, по которому вычисляется скорость $\hat{u}(x, t)$, отвечает ε -стационарной ВПФР $F_{N(s)}(x, n)$, прогнозируемой на s шагов вперед, где s есть верхний предел суммирования в (1.4).

Таким образом, построено эволюционное уравнение для ВПФР неэквидистантного нестационарного временного ряда в терминах времени. Для прогноза самого ряда можно, например, положить

$$\hat{x}(t(n+k)) = \int \hat{F}_{N(k)}(x, n+k) dx. \quad (2.6)$$

Проблема замыкания уравнения (2.5) состоит в независимом оценивании эмпирической скорости $\hat{u}(x, t)$ и параметра потока $\hat{\Lambda}(t, 1)$ по прошлым данным. Например, оценивая ЭЛС по предыдущему тикку, получаем выражение в конечных разностях, когда промежуток изменения величины x разбит на конечное число равных промежутков. Обозначая их номера i , имеем

$$u(i+1, n-1) F(i+1, n-1) = \sum_{k=1}^i (F(k, n) - F(k, n-1)). \quad (2.7)$$

Поскольку ВПФР на данном шаге n известна, как и на предыдущем, то уравнение (2.7) рекуррентно разрешимо относительно $u(i, n-1)$. Прогнозом скорости тогда можно считать, например, «наивное» представление

$$\hat{u}(i, n) = u(i, n-1). \quad (2.8)$$

Выше в главе IX были рассмотрены и другие модели нахождения эмпирической скорости. Нашей целью в данном случае является не разработка прогнозной модели для скорости, а уточнение собственно уравнения Лиувилля и формализация вида эмпирической скорости в терминах времени.

Более сложной является задача прогнозирования параметра потока, чему посвящен следующий параграф.

10.3. Прогнозные модели интенсивности потока событий

В условиях естественной суточной периодичности, характерной для многих рядов, связанных с торговыми операциями, оценка параметра потока получается усреднением суточных профилей активности. Введем следующие обозначения.

Пусть $n_m(j)$ есть число событий в m -ю минуту j -ых суток. Тогда число событий за эти сутки есть

$$N(j) = \sum_{m=1}^{1440} n_m(j), \quad (3.1)$$

а суточный нормированный профиль интенсивности определяется по формуле

$$y_m(j) = \frac{n_m(j)}{N(j)}. \quad (3.2)$$

После этого среднее число событий (т.е. собственно параметр потока) на промежутке $\Delta_t(\tau)$ может быть введено следующим образом.

Первый вариант — построение средневзвешенного профиля суточной активности за период T суток. Для этого вводим средневзвешенную интенсивность $w(m)$ в m -ю минуту и среднее число тиков за сутки W :

$$w(m) = \frac{1}{T} \sum_{j=1}^T n_m(j), \quad W = \sum_{m=1}^{1440} w(m) = \frac{1}{T} \sum_{j=1}^T N(j). \quad (3.3)$$

Затем определяем средневзвешенный нормированный профиль активности:

$$y^{(w)}(m) = \frac{w(m)}{W}. \quad (3.4)$$

Пусть, таким образом, среднее число событий за сутки определено и равно W . Тогда среднее число тиков за промежуток τ минут с момента времени t (мин) есть

$$\Lambda^{(w)}(t, \tau) = \sum_{m=1}^{\tau} w(t+m) = W \sum_{m=1}^{\tau} y^{(w)}(t+m). \quad (3.5)$$

Поскольку профиль $y^{(w)}(m)$ в (3.5) нормирован на 1, то его можно рассматривать как плотность распределения интенсивности по минутам в сутках. Соответствующая функция распределения есть

$$Y^{(w)}(m) = \sum_{k=1}^m y^{(w)}(k), \quad (3.6)$$

так что

$$\Lambda^{(w)}(t, \tau) = W \cdot (Y^{(w)}(t+\tau) - Y^{(w)}(t)). \quad (3.7)$$

Тем самым определен параметр потока событий в (1.2).

Второй вариант — построение среднего профиля активности за период T суток и построение модели для оценки ожидаемого количества тиков W^* за сутки. Введем величину среднего числа событий в m -ю минуту суток:

$$y^*(m) = \frac{1}{T} \sum_{j=1}^T y_m(j) \quad \text{и} \quad Y^*(m) = \sum_{k=1}^m y^*(k), \quad (3.8)$$

после чего получаем оценку параметра потока в виде

$$\Lambda^*(t, \tau) = W^* \cdot (Y^*(t+\tau) - Y^*(t)). \quad (3.9)$$

Представление суточного параметра потока в виде (3.9) отличается от (3.7) тем, что в нем среднее число событий в сутках не связано с интенсивностью $w(m)$ в m -ю минуту, а является независимой оценкой числа тиков за сутки. Таким образом, подход (3.7) однопараметрический, в нем оценивается только одна величина — интенсивность потока тик/мин. Подход (3.9) двухпараметрический, в нем независимо оцениваются нор-

мированный профиль суточной активности и число тиков за сутки. Вторым вариантом представляется практически более полезным с той точки зрения, что он позволяет выделить две составляющих нестационарности при оценке среднего числа событий за некоторое количество суток T . Во-первых, сам профиль суточной активности может меняться от одних суток к другим. Кроме того, число событий за сутки не является постоянным, и следует определить либо модель изменения этой величины, либо определить такое количество суток T , усреднение по которым числа событий даст минимальную ошибку при использовании его в качестве прогноза в следующие сутки.

Третий вариант — нахождение не среднего профиля в течение суток, а средней интенсивности на каждом характерном промежутке суточной активности, что может дать некоторое уточнение прогноза этой активности.

Оптимизация объема усреднения для получения среднего параметра потока в любом из этих трех вариантов проводится путем минимизации отклонения среднего профиля от фактического профиля следующего дня (или, возможно, с некоторым лагом вперед). Фактический параметр потока в j -е сутки есть

$$\Lambda_j(t, \tau) = N(j) \sum_{m=1}^{\tau} y_{t+m}(j). \quad (3.10)$$

Введем τ -сглаженный профиль суточной активности

$$y_j^{smooth}(t, \tau) = \frac{1}{\tau} \sum_{m=1}^{\tau} y_{t+m}(j). \quad (3.11)$$

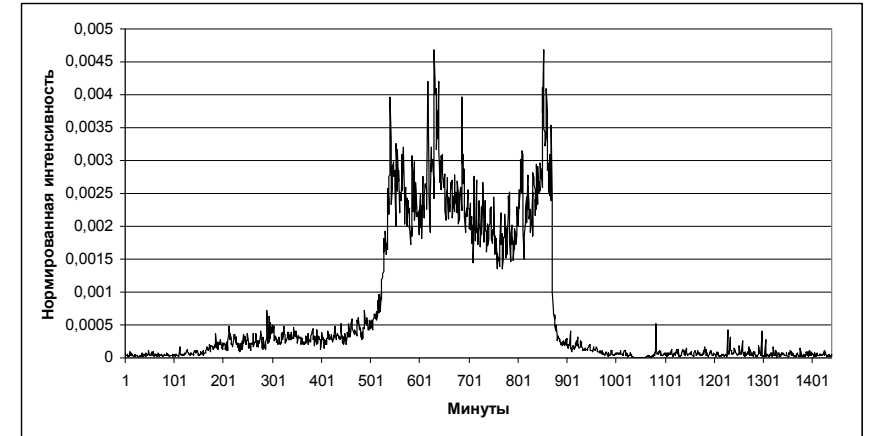


Рис. 10.3.1 — Средний профиль минутной интенсивности

Тогда

$$\Lambda_j(t, \tau) = N(j) \cdot \tau \cdot y_j^{smooth}(t, \tau). \quad (3.12)$$

Пример среднего профиля (3.2) суточной активности, агрегированной за 1 минуту, показан на рис. 10.3.1. В терминах уравнения Лиувилля (2.5) график на рис. 10.3.1 представляет собой величину $\Lambda(t, 1)$.

Пусть построен средний профиль $y(m; T, j)$ активности в m -ю минуту при усреднении по T суткам, отсчитанным назад от j -ых суток. Например, $y(m; T, j)$ определяется по формуле (3.2) или по формуле (3.8). Тогда и для этого профиля введем сглаженный по τ минутам профиль

$$y^{smooth}(t, \tau; T, j) = \frac{1}{\tau} \sum_{m=1}^{\tau} y(t+m; T, j), \quad (3.13)$$

после чего рассмотрим расстояние

$$\rho(\tau; T, j) = \sum_{t=1}^{1440} \left| y^{smooth}(t, \tau; T, j) - y_{j+1}^{smooth}(t, \tau) \right|. \quad (3.14)$$

Пусть имеются данные за S суток. Тогда вычисляется среднее отклонение

$$\bar{\rho}(\tau; T) = \frac{1}{S - T} \sum_{j=T}^{S-1} \rho(\tau; T, j) \quad (3.15)$$

и находится оптимальный объем для сбора статистики по суточной активности

$$T_{opt}(\tau) = \arg \min_T \bar{\rho}(\tau; T). \quad (3.16)$$

Чтобы замкнуть любой из этих методов, необходимо построить модель изменения суточной активности как таковой, т.е. оценить среднее число событий в сутках. Для этого надо построить модель эволюции распределения эквидистантного временного ряда, каковым является число событий в сутках, на один шаг (т.е. на сутки или на другой конкретный горизонт) вперед.

Пример числа событий за сутки для трех фьючерсных контрактов на нефть WTI, следующих один за другим непрерывно по оси времени, показано на рис. 10.3.2.

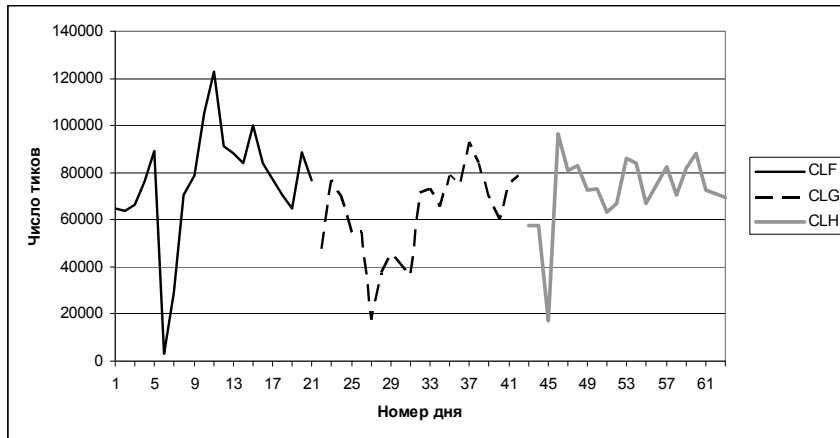


Рис. 10.3.2 — Количество событий за сутки

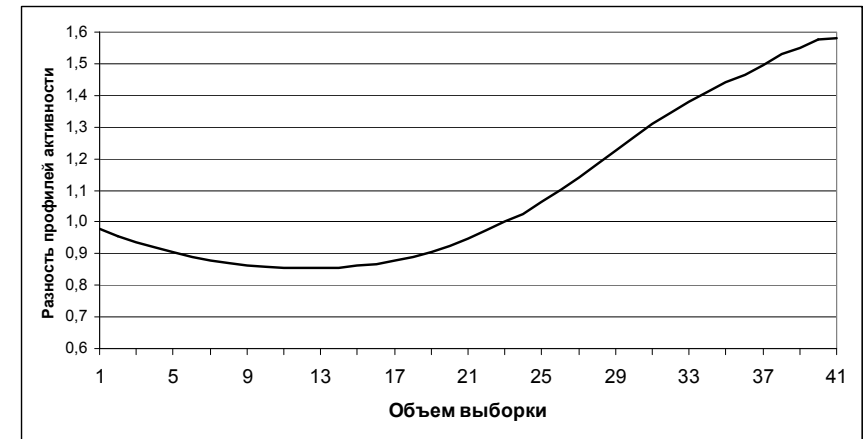


Рис. 10.3.3 — Расстояние между средневзвешенным и текущим профилями активности

Если моделирование осуществляется с помощью средневзвешенного профиля суточной активности, то количество суток для этого усреднения находится из условия минимального среднего расстояния между двумя профилями активности — средним и фактическим суточным, следующим за промежутком усреднения. График зависимости этого расстояния от промежутка усреднения в сутках приведен на рис. 10.3.3. Из него видно, что оптимальное количество суток для нахождения среднего суточного параметра потока в терминах агрегированной интенсивности событий при анализе данных за три контракта, динамика интенсивности которых показана выше, оказалось равным 13 суткам.

Однако, что касается минимального расхождения тиковых профилей активности (среднего и фактического не нормированных, а в натуральных величинах чисел событий), то наилучшим вариантом оказывается профиль, средневзвешенный по объему всего 2 суток. Это сравнение показано на рис. 10.3.4. Соответствующие графики получены усреднением отдельных отклонений за сутки по промежутку времени за три рассматриваемых контракта.



Рис. 10.3.4 — Расстояние между средними и текущим профилями активности

Но существенно более точный прогноз интенсивности событий получается при разбиении суток на периоды активности. Из рис. 10.3.1 видно, что в течение суток явно выделяются три характерных периода: период «сна», т.е. очень малой активности длится с 0 до примерно 150 первых минут и с 1000 до 1444 (конец суток); «оживление» в период от 150 до 500 минут и аналогичное «засыпание» с 900 до 1000; период активности длится от 500 до 900 минут от начала суток.

Выяснилось, что число событий в период «сна» не влияет на число событий в другое время суток, тогда как периоды оживления и активности связаны между собой с положительной корреляционной связью на уровне 0,7. В частности, среднее число событий в минуту в период оживления равно $\bar{\lambda}_2 = 23$, а среднеквадратичное отклонение этой величины равно $\sigma_2 = 13$. В период активности эти показатели равны соответственно $\bar{\lambda}_3 = 170$, $\sigma_3 = 55$. Достоверность регрессионной связи между средней минутной интенсивностью событий в период активности и в период оживления составила величину 0,8 согласно эмпирической формуле регрессии

$$\lambda_3 = \bar{\lambda}_3 + 3(\lambda_2 - \bar{\lambda}_2). \quad (3.17)$$

Далее, согласно общей концепции идентификации текущего фрагмента распределения, следует создать библиотеку эталонов, т.е. эталонных

распределений, типичных для того или иного поведения исходного временного тикового ряда. Методика кластеризации эталонов формулируется в следующем параграфе.

10.4. Кластеризация паттернов маркированных временных рядов

По имеющимся историческим данным экспертно выделяются фрагменты временного ряда, интересные для исследователя. На основе этой нарезки составляется набор эталонов, как описано в п. 7.3 по формулам (7.3.7-7.3.8).

Протестируем метод разбиения на кластеры на примере тиковых рядов фьючерсных контрактов на нефть WTI. Анализ изменения распределений временного ряда с суточной периодичностью в трех суточных фазах — сна, оживления и активности (см. рис. 10.3.1) — показали, что в каждой из этих фаз поведение ряда в статистическом смысле различно. Опишем характерные черты каждой из этих фаз.

В фазе сна преобладают фрагменты с высоким уровнем нестационарности, часто совпадающие по длительности со всем этим участком. Эти фрагменты образуют три кластера при критериальном уровне разделения между ними, равном $\rho_c = 0,14$. Отметим, что это значение примерно совпадает с согласованным уровнем нестационарности тикового ряда приростов цен (см. п. 5.4). Эталонные распределения кластеров в фазе сна показаны на диаграмме рис. 10.4.1.

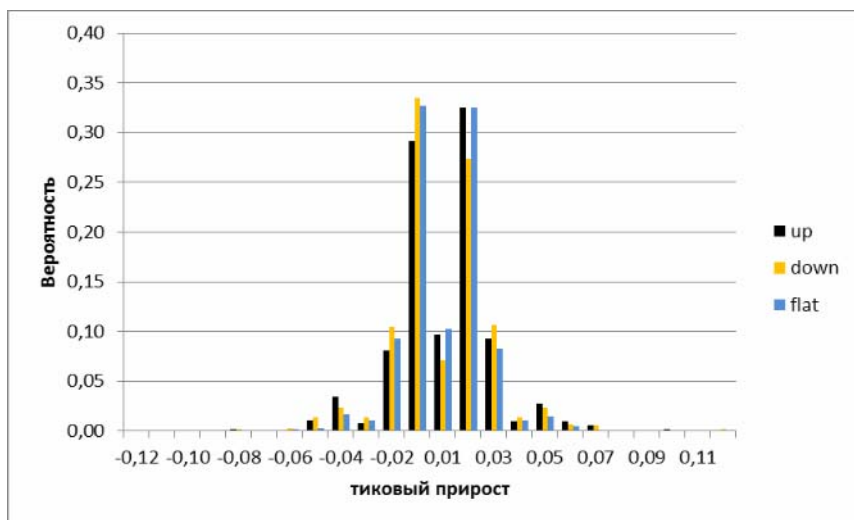


Рис. 10.4.1 — Эталонные ПФР кластеров в фазе сна

Несмотря на то, что диаграммы на рис. 10.4.1 похожи по форме, как и вообще распределения приростов значений рассматриваемых рядов по выборке любого объема и в любой момент времени, они позволяют различить «тонкую структуру» разного типа поведения исходного ряда. Кластер «up» отвечает ситуации возрастающего тренда, кластер «down» — напротив, убывающего, а кластер «flat» — «боковому» движению цены, когда левая часть соответствующего эталона идет близко к эталону «down», а правая — к эталону «up». Расстояния между эталонными распределениями следующие:

$$\rho_{12} = 0,20, \quad \rho_{13} = 0,15, \quad \rho_{23} = 0,16.$$

При этом среднее расстояние между элементами одного кластера равно примерно 0,08, а максимальное расстояние не превосходит, как указывалось, критерияльного уровня $\rho_c = 0,14$. В рассмотренном примере любые два элемента из разных кластеров с вероятностью 0,95 находятся на расстоянии большем, чем расстояние между каждым из этих элементов и любым элементом «его» кластера.

Фаза *оживления* на рис. 10.3.1 оказалась специфической: в ней нет разделения на кластеры, распределение этой фазы в целом очень близко к распределению «бокового» эталона в фазе сна. Расстояние между этими двумя распределениями всего 0,05.

В активной фазе кластеризация оказалась многоуровневой. Критерий разделения фрагментов ряда по близости их ВПФР на «свои» и «чужие» был выбран, исходя из распределения попарных расстояний между всеми найденными фрагментами, отвечающими трем типовым ситуациям. Как и все распределения расстояний между ВПФР, оно унимодально с «тяжелым хвостом». Оптимальный критический уровень разделения расстояний оказался равным $\rho_c = 0,08$. При этом в каждом типовом кластере «up», «down» и «flat» появились свои под-кластеры.

В кластере up-тренда исходного ряда выделяются три под-кластера, отвечающие разному уровню активности: относительно слабой, средней и высокой. В первый под-кластер up1 попали фрагменты, отвечающие интенсивности потока в диапазоне 10-20 тик/мин, второй (up2) содержит фрагменты с интенсивностью в диапазоне 100-200 тик/мин, а третий (up3) отвечает высокой интенсивности порядка 1000-1500 тик/мин. Эти эталоны приведены на рис. 10.4.2.

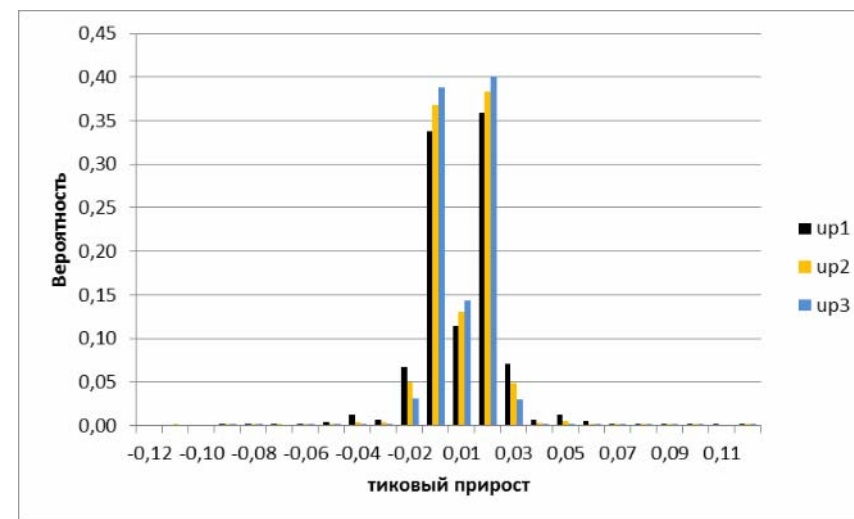


Рис. 10.4.2 — Эталонные ПФР в up-кластере

Средние расстояния между ВПФР фрагментов внутри одного кластера составили примерно 0,04. При этом расстояния между самими эталонами оказались равными

$$\rho_{12}^{up} = 0,14, \quad \rho_{13}^{up} = 0,24, \quad \rho_{23}^{up} = 0,10.$$

Сам up-кластер характеризуется тем, что среди двух наиболее вероятных типовых приростов положительный прирост более вероятен, чем отрицательный. В down-кластере, наоборот, левая ветвь гистограмм эталонных распределений с отрицательными приростами выше, чем правая (рис. 10.4.3).

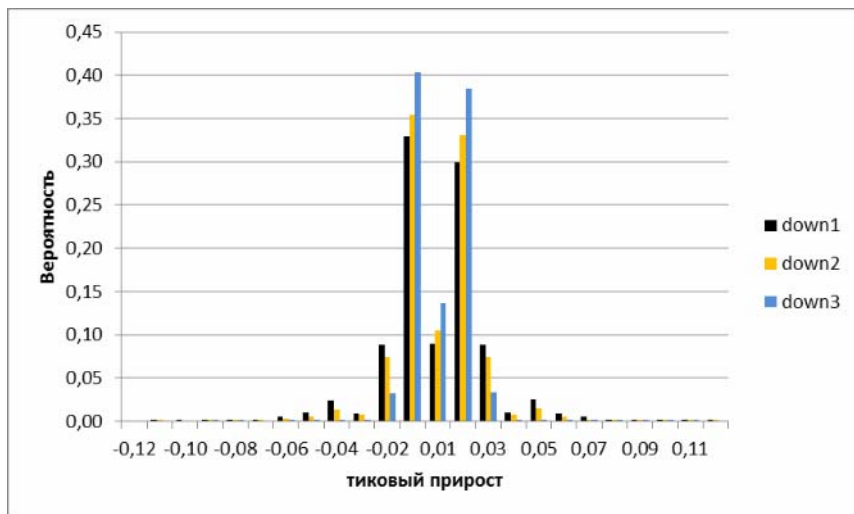


Рис. 10.4.3 — Эталонные ПФР в down-кластере

Средние расстояния между ВПФР фрагментов внутри одного down-кластера составили около 0,04. При этом расстояния между самими эталонами оказались равными

$$\rho_{12}^{down} = 0,17, \quad \rho_{13}^{down} = 0,31, \quad \rho_{23}^{down} = 0,27.$$

В принципе, возможны и рваные приросты вверх или вниз, когда, например, имеется большое количество малых отрицательных приростов и малое количество резких положительных скачков, но такие фазы сравнительно редки и отвечают нестандартной ситуации.

Flat-кластер в активной фазе характеризуется высокой хаотичностью, в связи с чем фрагменты, выделяемые в этом кластере, объединяются с меньшей достоверностью. В целом, под-кластеры не выделяются, а весь боковой тренд совпадает по форме с аналогичным распределением в фазе оживления. Различаются эти графики всего лишь большей или меньшей интенсивностью.

Подчеркнем, что число эталонов в каждом типовом кластере определяется не какими-то внутренне присущими изучаемому ряду особенностями, а лишь теми фрагментами, которые оказались в данном временном промежутке. В другом промежутке число под-кластеров может оказаться и другим, в соответствии с тем, какая интенсивность событий будет на этом промежутке. В принципе, если бы период наблюдений был бы достаточно большим, можно было бы построить типовую поверхность каждого из трех кластеров (up, down, flat), второй координатой на которой служила бы интенсивность событий. Чем выше интенсивность, тем больше вероятность наиболее вероятных значений приростов. В среднем наблюдается следующая скейлинговая зависимость между максимальной частотой q положительных или отрицательных приростов в зависимости от параметра потока $\lambda = \Lambda(t,1)$:

$$\frac{q}{q_0} = 1 + 0,02 \cdot \lg \frac{\lambda}{\lambda_0}. \quad (4.1)$$

Методика анализа достоверности построенной кластеризации состоит в следующем. Для каждого кластера a строится функция распределения $g_a^+(\rho)$ отклонений $\rho_{i_a,a}$ «его» фрагментов, а также распределение $g_a^-(\rho)$ отклонений $\rho_{k_b,a}$ фрагментов, отнесенных в другие кластеры, от средней ПФР \bar{f}_a данного кластера. Минимальное значение ρ , при котором $g_a^+(\rho) = 1$, обозначим ρ_a^+ , а максимальное значение ρ , при котором $g_a^-(\rho) = 0$, обозначим ρ_a^- . Смысл введенных величин в том, что все ПФР фрагментов типа a находятся на расстоянии не более ρ_a^+ от средней ПФР, и аналогично все ПФР других типов находятся от нее на расстоянии не менее ρ_a^- . Величина $1 - g_a^+(\rho_a^-)$ есть вероятность ошибочно признать за тип « a » другой фрагмент (ошибка второго рода), а величина $g_a^-(\rho_a^+)$ есть вероятность ошибочно отвергнуть тип « a », посчитав его за другой (ошибка первого рода).

При тестировании идентификации кластерной принадлежности был рассмотрен конкретный контракт CLN, следующий за контрактами CLF и CLG, которые были использованы для построения эталонных ПФР. Для него с помощью функционала-индикатора разладки были найдены 86

фрагментов. Кластерной принадлежности удовлетворили 74 из них, а остальные не были идентифицированы как элементы какого-либо кластера. Из этих 74 фрагментов правильно (т.е. в свой кластер) было отнесено 65 фрагментов, так что ошибка идентификации составила примерно 0,12. Это показывает достаточно высокую точность предложенного метода идентификации текущего фрагмента временного ряда. Отметим также, что эта идентификация осуществлялась только по близости распределений в норме $L1$. Если к условию близости добавить второй параметр — а именно, относительное расположение максимумов распределений (правый находится выше левого или наоборот), то ошибка идентификации фрагмента составит всего 0,02.

10.5. Зависимость выборочных статистик от времени

Пусть $\bar{b}_N(n)$ есть среднее значение величины $b(x_n)$, вычисленное по выборке длины N , отсчитанной назад от события с номером n :

$$\bar{b}_N(n) = \frac{1}{N} \sum_{k=n-N+1}^n b(x_k). \quad (5.1)$$

Поставим вопрос о том, как пересчитать выборочное среднее в терминах времени, учитывая, что временной ряд $\{x_n\}$ имеет маркировку $t(n)$. Параметр потока $\Lambda(t, \tau)$ считается известным для всех моментов времени t и величин агрегирования τ . Для определенности будем считать, что внутри минимального промежутка агрегирования интенсивность потока является постоянной величиной, равной

$$\mu(t, t + \tau) = \frac{1}{[\tau] + 1} \Lambda(t, [\tau] + 1). \quad (5.2)$$

Это дополнение позволяет определять временные средние на любых промежутках времени, не меньших, конечно, чем единица измерения времени.

По-прежнему считаем, что вероятность наступления k событий за время τ дается нестационарным распределением Пуассона (1.2):

$$p_k(t, t + \tau) = \frac{(\Lambda(t, \tau))^k}{k!} \exp(-\Lambda(t, \tau)). \quad (5.3)$$

Временной средней $\tilde{b}_T(t)$ на промежутке длины T , отсчитанном назад от момента времени $t(n)$, будем называть выборочную среднюю, взвешенную по нестационарному распределению (5.3):

$$\tilde{b}_T(t) = \frac{1}{1 - p_0(t - T, t)} \sum_{k=1}^{\infty} p_k(t - T, t) \bar{b}_k(n(t)). \quad (5.4)$$

Поскольку вероятность того, что на промежутке времени длины T состоялось ровно $k \geq 1$ событий, равна $p_k(t - T, t)$, то этой же вероятности равна вероятность того, что временное среднее величины $b(x_n)$ равно $\bar{b}_k(t(n))$. Следовательно, среднее по времени значение величины $b(x_n)$, определяемое по выборке ненулевой длины, вычисляется по формуле (5.4).

В частности, если обозначить $a(x_n) = x_n$, то в соответствии с (5.1), (5.4), среднее выборочное значение маркированного ряда на промежутке времени $[t - T; t]$ выражается через среднее значение $a_N(n(t))$, построенное по выборке длины N в момент t , по формуле

$$\begin{aligned} \tilde{a}_T(t) &= \frac{\sum_{k=1}^{\infty} p_k(t - T, t) \bar{a}_k(n(t))}{1 - p_0(t - T, t)} = \\ &= \frac{\sum_{k=1}^{\infty} p_k(t - T, t) \frac{1}{k} \sum_{j=n(t)-k+1}^{n(t)} x_j}{1 - p_0(t - T, t)}. \end{aligned} \quad (5.5)$$

Если же функция $b(x_n)$ зависит еще и от $n(t)$, например, через посредство средних величин вида (5.5), то временное среднее для таких функций определяется неоднозначно. Например, рассмотрим в качестве функции $b(x_n)$ выборочную дисперсию. С одной стороны, в соответствии с (5.4), временная выборочная дисперсия $\tilde{\sigma}^2(t, T)$ в момент времени t по промежутку длительности T , содержащему не менее двух событий, определяется через событийную дисперсию $\sigma^2(n(t), k)$ формулой

$$\tilde{\sigma}^2(t, T) = \frac{\sum_{k=2}^{\infty} p_k(t-T, t) \sigma^2(n(t), k)}{1 - p_0(t-T, t) - p_1(t-T, t)}. \quad (5.6)$$

С другой стороны, дисперсию по временному промежутку можно определить по временной плотности распределения (1.3), что дает

$$\hat{\sigma}^2(t, T) = \frac{\sum_{k=2}^{\infty} p_k(t-T, t) \bar{b}_k(n(t))}{1 - p_0(t-T, t) - p_1(t-T, t)} - \left[\frac{\sum_{k=2}^{\infty} p_k(t-T, t) \bar{a}_k(n(t))}{1 - p_0(t-T, t) - p_1(t-T, t)} \right]^2, \quad (5.7)$$

где $a(x_n) = x_n$, $b(x_n) = x_n^2$. В случае эквидистантного временного ряда оба выражения — (5.6) и (5.7) — совпадают с обычным выражением для выборочной дисперсии. Какое из этих двух выражений более адекватно для описания свойств ряда, решается опытным путем из сравнения с фактической дисперсией ряда на выбранном временном отрезке. Практически важной является оценка

$$\hat{\sigma}^2(t, T) \geq \tilde{\sigma}^2(t, T), \quad (5.8)$$

которая позволяет оценить снизу временную выборочную дисперсию нестационарного маркированного случайного процесса через выборочное среднее значение шаговых выборочных дисперсий.

Рассмотрим теперь задачу о пересчете автокорреляционной функции как функции длины выборки и длины лага к функции от временного промежутка и сдвига по времени. Пусть $A_N(n, l)$ есть выборочная АКФ исследуемого ряда с лагом l событий в момент времени $t(n)$ по выборке длины N :

$$A_N(n, l) = \frac{(N-l) \sum_{k=n-N+1}^{n-l} x_k x_{k+l} - \sum_{k=n-N+1}^{n-l} x_k \sum_{j=n-N+1}^{n-l} x_{j+l}}{\sqrt{(N-l) \sum_{k=n-N+1}^{n-l} (x_k)^2 - \left(\sum_{k=n-N+1}^{n-l} x_k \right)^2} \sqrt{(N-l) \sum_{k=n-N+1}^{n-l} (x_{k+l})^2 - \left(\sum_{k=n-N+1}^{n-l} x_{k+l} \right)^2}}. \quad (5.9)$$

Этой функции надо сопоставить временную АКФ $\tilde{A}_T(t, \tau)$, построенную в момент времени t по временному промежутку T с лагом по времени τ . Обозначим $c(x_n, x_k) = x_n x_k$. В терминах (5.1) выражение (5.9) переписывается в виде

$$A_N(n, l) = \frac{\bar{c}_{N-l}(n-l, n) - \bar{a}_{N-l}(n-l) \bar{a}_{N-l}(n)}{\sqrt{\bar{b}_{N-l}(n-l) - (\bar{a}_{N-l}(n-l))^2} \sqrt{\bar{b}_{N-l}(n) - (\bar{a}_{N-l}(n))^2}}. \quad (5.10)$$

Для получения временной АКФ сравниваются две выборки: одна заключена в промежутке времени $[t-T; t]$, а другая — в промежутке $[t-T+\tau; t+\tau]$. При этом в обоих промежутках должно быть одинаковое число событий, как и в промежутках длительности τ : $[t-T; t-T+\tau]$ и $[t; t+\tau]$. Как и для дисперсии, можно рассмотреть два способа определения временной АКФ. С одной стороны, вероятность того, что временная АКФ $\tilde{A}_T(t, \tau)$ равна событийной АКФ $A_N(n, l)$, есть

$$p_l(t, t+\tau) p_l(t-T, t-T+\tau) p_{N-l}(t-T+\tau, t).$$

В результате получаем, что если в указанных промежутках находится попарно равное, но произвольное число событий, то $\tilde{A}_T(t, \tau)$ следует определить формулой

$$\tilde{A}_T(t, \tau) = \frac{\sum_{k=2}^{\infty} \sum_{l=1}^{k-1} p_l(t, t+\tau) p_l(t-T, t-T+\tau) p_k(t-T+\tau, t) A_k(n(t), l)}{\sum_{k=2}^{\infty} \sum_{l=1}^{k-1} p_l(t, t+\tau) p_l(t-T, t-T+\tau) p_k(t-T+\tau, t)}. \quad (5.11)$$

С другой стороны, временную АКФ можно определить и как функцию от средневзвешенных временных средних в соответствии с формулой (5.9) по аналогии с определением выборочной временной дисперсии (5.7). Тогда имеем

$$\tilde{A}_T(t, \tau) = \frac{\tilde{c}_{T-\tau}(t-\tau, t) - \tilde{a}_{T-\tau}(t-\tau)\tilde{a}_{T-\tau}(t)}{\sqrt{\sigma_{T-\tau}^2(t-\tau) - (\tilde{a}_{T-\tau}(t-\tau))^2} \sqrt{\sigma_{T-\tau}^2(t) - (\tilde{a}_{T-\tau}(t))^2}}. \quad (5.12)$$

Какой из вариантов более корректен, выясняется на практике при построении скользящих индикаторов нестационарности как функций не длины выборки, а длительности временного промежутка. Обе возможности следует учитывать при изучении разрядки как функции времени.

Глава XI. Прикладные задачи нестационарной статистики

11.1. Слабо меняющиеся временные ряды

Существует большое количество временных рядов, относительное изменение соседних величин в которых мало, но которые, тем не менее, не являются стационарными ни в узком, ни в широком смыслах. Среди классических регрессионных моделей наилучшим прогнозом в смысле среднеквадратичной ошибки для таких рядов оказывается наивный, в котором предполагается, что следующее значение ряда равно предыдущему. Характерными рядами в этом смысле являются валютные ряды, т.е. ряды, образованные обменным курсом валют. Принципиально важной задачей является создание прогнозной модели ряда, которая имеет ошибку, меньшую ошибки наивного прогноза.

Рассмотрим применение гидродинамической модели (9.4.16) для прогнозирования временного ряда, образованного минутными данными закрытия $p(t)$ валютного курса EUR/USD (для краткости на рисунках ниже приведен фрагмент анализа данных за первый квартал 2011 г.). Обозначим

$$M = \max_t (p(t+1) - p(t)), \quad m = \min_t (p(t+1) - p(t)) \quad (1.1)$$

и введем нормированные приросты курса закрытия

$$x(t) = \frac{p(t+1) - p(t) - m}{M - m}. \quad (1.2)$$

Временной ряд $x(t)$ представлен на рис. 11.1.1.

Автокорреляция с лагом от 1 до 1440 (сутки) для этого ряда не превосходит по модулю величину 0,04, так что величины приростов исходного ряда $p(t)$ можно с этой точностью считать статистически независимыми. Минимальный прирост курса в исходных единицах в этом фрагменте составил величину - 0,01, а максимальный 0,004, максимум вероятности пришелся на промежуток от 0,0005 до 0,001. Среднеквадратичное относительное отклонение минутных приростов курса EUR/USD составило $3,57 \cdot 10^{-6}$, а среднеквадратичное отклонение нормированных вели-

чин $x(t)$ по равно примерно 0,018. Заметим, что такая же точность $\varepsilon = 0,018$ гарантируется при прогнозировании изменения ВПФР ряда $x(t)$ по выборке длины $n = 800$ на $\tau = 7$ шагов (т.е. минут) вперед, согласно оценке $\tau \leq n\varepsilon/2$.

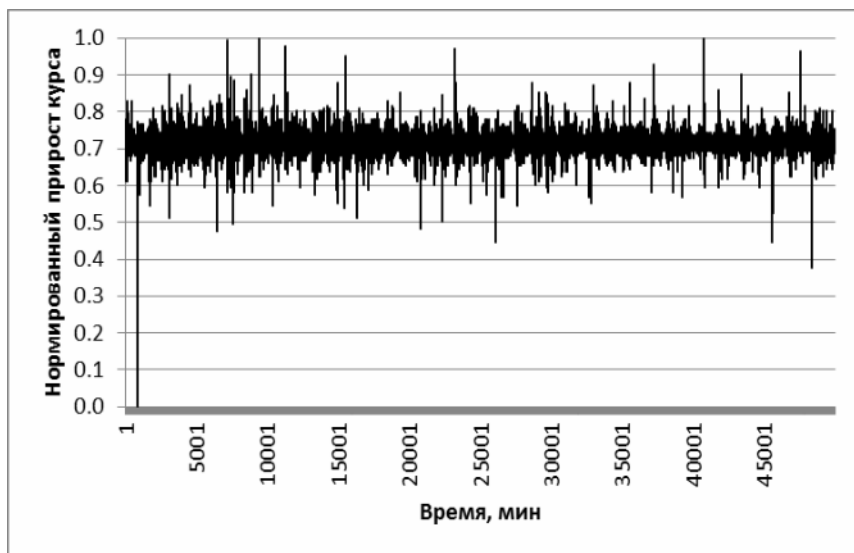


Рис. 11.1.1 — Фрагмент ряда $x(t)$ нормированных приростов валютного курса

На рис. 11.1.2 показаны две ВПФР, построенные за два соседних месяца.

Из рис. 11.1.2 может показаться, что распределение приростов за довольно большой промежуток времени практически не меняется. На самом деле изменения хоть и малы, но значительны по сравнению со стационарной ситуацией. Стационарный уровень значимости для этих распределений равен $\varepsilon^* = 0,0075$. Это означает, что если бы распределение приростов было стационарным, характерное расстояние между двумя гистограммами было бы равно $2\varepsilon^* = 0,015$, т.е. вероятность того, что расстояние между ними больше, чем 0,015, была бы равна 0,015. В нашем же примере расстояние между двумя встык-выборками равно 0,134, т.е. поч-

ти на порядок больше. Это показывает, сколь тонкий эффект требуется смоделировать при анализе данного нестационарного ряда.



Рис. 11.1.2 — ВПФР двух встык-выборок ряда $x(t)$ длиной в месяц

Построим плотность распределения $g_n(\rho)$ расстояний между встык-выборками длины n в норме $L1$ при равномерном разбиении гистограммы на оптимальное число $k(n)$ промежутков и найдем долю $\beta(n)$ расстояний, меньших, чем точность определения самих этих расстояний:

$$\int_0^{2\alpha(n)} g_n(\rho) d\rho = \beta(n). \quad (1.3)$$

Величина $\beta(n)$ представляет собой вероятность того, что распределение стационарно с согласованной точностью $\alpha(n)$. Если оказалось, что $\beta \leq \alpha$, то эффект нестационарности отображается корректно и можно применять эволюционную модель ВПФР. Если же $\beta > \alpha$, то в статистически значимой доле случаев наилучший прогноз ВПФР — наивный, и

лишь в доле $1 - \beta$ случаев кинетический прогноз даст хороший результат.

На рис. 11.1.3 показаны распределения расстояний между встык-выборками в зависимости от числа классовых интервалов. В целом при увеличении числа интервалов расстояние между выборками возрастает.

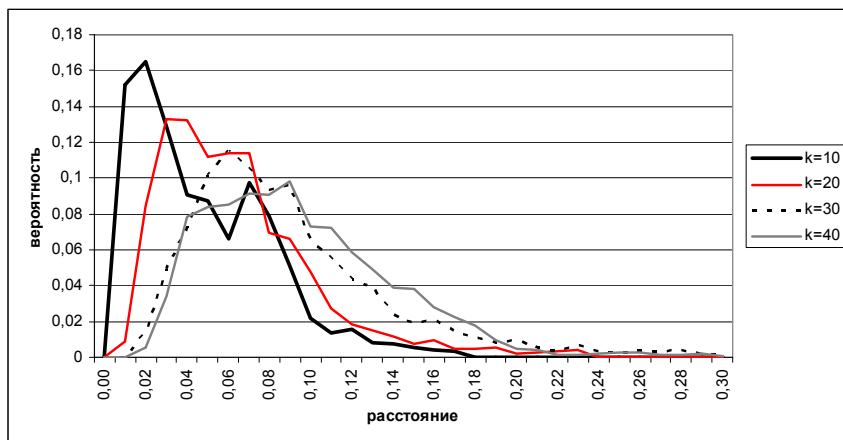


Рис. 11.1.3 — Распределение расстояний между встык-выборками в зависимости от числа k классовых интервалов

Анализ согласованного уровня значимости показал, что при длинах выборок порядка 1000-2000 точек наилучшим прогнозом ВПФР будет наивный, поскольку почти все расстояния между ВПФР укладываются в точность, обусловленную чисто статистическим разбросом данных. Для выборок же меньших длин $n = 500 \div 600$ точек нестационарность достоверно проявляется примерно в 70-80% случаев, когда и следует применять кинетическую модель прогнозирования.

Приведем результаты прогнозирования ВПФР и временного ряда по модели (9.4.16). Заметим, что прогнозировать ВПФР на 1 шаг в терминах частоты появления значений временного ряда не очень осмысленно, т.к. изменение ВПФР за этот промежуток не превзойдет малой величины $2/n$. Такая точность не может быть достигнута в модели эволюции моментов ВПФР, поэтому на малый горизонт прогнозирования наилучшее приближение — квазистационарное, т.е. наивный прогноз ВПФР. Применение модели представляет интерес при прогнозировании на более дли-

тельный горизонт, например, равный длине окна выборки для случая $n = 600$.

Точность прогнозирования на τ шагов вперед включает в себя несколько аспектов. Во-первых, следует определить точность прогнозирования ВПФР в определенной норме, в данном случае в норме $L1$. Эта величина определяется по формуле

$$\bar{\varepsilon} = \frac{1}{N} \sum_{t=1}^N \|f(x, t + \tau) - f(x, t)\|, \quad (1.4)$$

где N есть число экспериментов по прогнозированию ВПФР.

Во-вторых, если целью моделирования было получение прогноза значения ряда через определенный промежуток времени, то интерес представляет как относительная среднеквадратичная точность отклонения прогноза $\hat{x}(t; \tau)$ от факта $x(t)$

$$\bar{\delta}(\tau) = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{x}(t; \tau) - x(t))^2}, \quad (1.5)$$

так и точность τ -шаговой траектории

$$\delta = \sqrt{\frac{1}{N\tau} \sum_{t=1}^N \sum_{k=1}^{\tau} (\hat{x}(t; k) - x(t))^2}, \quad (1.6)$$

а также скорость нарастания ошибки с увеличением горизонта прогнозирования

$$d(k) = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{x}(t; k) - x(t))^2}, \quad k = 1, 2, \dots, \tau. \quad (1.7)$$

Наилучшая точность прогнозирования ВПФР по формуле (1.4) оказалась равной 0,03. При этом следует подчеркнуть, что характерное расстояние между ВПФР встык-выборок в 600 точек равно 0,24. Это означает, что наивный прогноз распределения в квазистационарном приближении на порядок хуже, чем прогнозы по кинетическим моделям, так что они представляют собой не просто математические упражнения по статистической механике, но вполне адекватны для описания нестационарной эволюции распределений временных рядов. Следует также сказать, что стандартные прогнозы приростов курса по стационарным регрессионным и авторегрессионным моделям любого порядка на оптимальных объемах

выборки дают ошибку, приблизительно совпадающую (до второго знака) с ошибкой наивного прогноза, т.е. они не являются эффективными.

11.2. Ряды с высоким уровнем нестационарности

В этом разделе разработанная методика применяется для прогнозирования рядов с высоким уровнем нестационарности. Типовым представителем такого ряда является ряд цен на рынках ценных бумаг. Эти рынки отличаются от валютных, сырьевых или энергетических более высокой волатильностью, и, одновременно, более тесной корреляционной связью первого порядка. Это приводит к тому, что без пред-подготовки ряда точность низкоразмерных гидродинамических моделей прогнозирования, описанных в главе VII, невелика: их ошибка тогда оказывается на уровне наивного прогноза.

Рассмотрим, например, ряд 5-минутных цен закрытия акция General Electric (GE) и построим прогноз ВПФР на 24 шага (т.е. на 2 часа) вперед с помощью уравнения Лиувилля с точностью 0,05. Оптимальный объем выборки для прогнозирования на 24 шага вперед с точностью не хуже 0,05 находится из анализа горизонтной статистики, как это описано в главе VI, и равен в данном случае 880 (меньше, чем $2\tau/\varepsilon = 960$).

В результате оказалось, что ВПФР для цен GE была спрогнозирована с ошибкой 0,04. Если прогнозным значением считать среднее по ВПФР, то оказалось, что прогноз ряда, основанный на кинетическом прогнозе ВПФР, в среднем имеет относительную среднеквадратичную ошибку 0,22, что примерно равно среднеквадратичному отклонению самого ряда. Это весьма большая ошибка, но ее происхождение вполне естественно, поскольку ряд не был пред-подготовлен, а выборочное распределение имеет дисперсию, равную выборочной дисперсии ряда.

Если же не использовать кинетический прогноз, а рассмотреть некоторый стандартный метод прогноза ряда, например, модель АР невысокого порядка (второго или третьего), примененную к первым разностям исходного ряда, то для каждой такой модели можно эмпирически установить наиболее оптимальный объем выборки. Применим, например, АР модель второго порядка для того же прогноза цен на акции GE. Используем скользящую АКФ, определяемую по выборке объема N , которая начинается в момент $j+1$ и имеет лаг n :

$$A_N(j, n) = \frac{(N-n) \sum_{k=j+1}^{j+N-n} x_k x_{k+n} - \sum_{k=j+1}^{j+N-n} x_k \sum_{l=j+1}^{j+N-n} x_{l+n}}{\sqrt{(N-n) \sum_{k=j+1}^{j+N-n} (x_k)^2 - \left(\sum_{k=j+1}^{j+N-n} x_k \right)^2} \sqrt{(N-n) \sum_{k=j+1}^{j+N-n} (x_{k+n})^2 - \left(\sum_{k=j+1}^{j+N-n} x_{k+n} \right)^2}}. \quad (2.1)$$

Для модели второго порядка потребуются величины $A_N(j, 1)$ и $A_N(j, 2)$. Тогда прогноз на 1 шаг вперед определяется формулой $\hat{y}_{i+1}(N) =$

$$= \frac{A_N(i-N, 1)(1 - A_N(i-N, 2))y_i(N) + (A_N(i-N, 2) - (A_N(i-N, 1))^2)y_{i-1}(N)}{1 - (A_N(i-N, 1))^2},$$

$$y_i(N) = x_i - m_1(i-N, N) \equiv x_i - \frac{1}{N} \sum_{l=i-N+1}^i x_l.$$

(2.2)

Обозначая для краткости правую часть (2.2) через $\varphi_N(i)$, получаем прогнозное значение исходного ряда:

$$\hat{x}_{i+1}(N) = \frac{Nm_1(i-N, N) - x_{i-N+1} + N\varphi_N(i)}{N-1}. \quad (2.3)$$

Прогноз на большее число шагов строится по такой же формуле, в которой прогнозное значение, найденное на предыдущем шаге, рассматривается как фактическое.

Среднеквадратичное отклонение прогноза от факта по серии из T последовательных данных определяется формулой

$$\sigma(N) = \sqrt{\frac{1}{T} \sum_{i=1}^T (\hat{x}_i(N) - x_i)^2}, \quad (2.4)$$

а относительное отклонение —

$$\delta(N) = \sqrt{\frac{1}{T} \sum_{i=1}^T (\hat{x}_i(N)/x_i - 1)^2}. \quad (2.5)$$

На рис. 11.2.1 приведены результаты расчетов нормированных значений рядов цен по кинетической модели уравнения Лиувилля, по авторегрессионной модели второго порядка, для которой была проведена прямая оптимизация объема выборки для получения минимальной ошибки прогноза, и фактические значения этого ряда.

Ошибка прогноза по авторегрессионной модели второго порядка составила всего 0,06. Это показывает, что ряды без пред-подготовки прогнозируются кинетическим методом с ошибкой, определяемой дисперсией ряда, а более точный прогноз может быть достигнут подходящим стандартным методом, для которого, однако, следует определить оптимальный объем выборки.

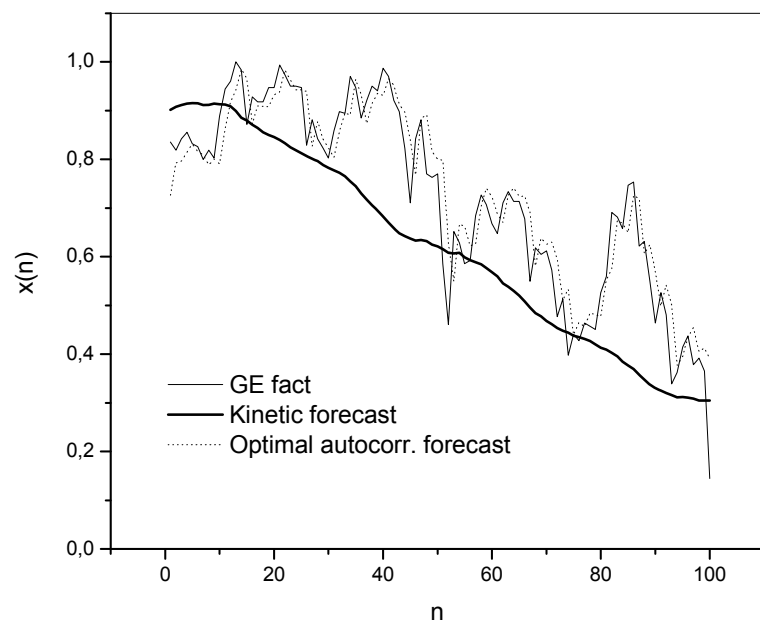


Рис. 11.2.1 — Сравнение прогнозов цен на акции компании GE

Метод минимизации ошибки прогноза ряда как функции объема выборки для прогнозирования на заданный горизонт состоит в построении скользящих средних с произвольным окном выборки в произвольный же момент времени. Рассмотрим ряд из первых разностей $x(t) = c(t+1) - c(t)$ исходных цен. Для этого массива составляется

массив скользящих средних и дисперсий в момент времени t по объему выборки j :

$$a(t, j) = \frac{1}{j} \sum_{i=t-j+1}^t x_i, \quad D(t, j) = \frac{1}{j} \sum_{i=t-j+1}^t (x_i - a(t, j))^2. \quad (2.6)$$

Затем проводится авторегрессионная подготовка ряда $x(t)$ по модели заданного порядка на множестве данных, выделяемых парой $\{t, j\}$. Вычисляется набор АКФ $A(t, j)$ с соответствующими лагами по выборке объема j в момент t и составляется массив аппроксимаций $y_k(t, j)$ в соответствии с уравнением, например, (2.2). Остатки ряда после этой обработки обозначаем через $z_k(t, j)$:

$$z_k(t, j) = p(k) - y_k(t, j); \quad z_{t-j+1} \equiv 0. \quad (2.7)$$

Эти остатки, в свою очередь, можно аппроксимировать регрессионной зависимостью на время. Двумерный массив коэффициентов регрессии определяется формулой

$$b(t, j) = \frac{1}{j\sigma^2(t, j)} \sum_{i=t-j+1}^t (z_i(t, j) - \bar{z}(t, j))(i - \theta(t, j)), \quad (2.8)$$

где

$$\bar{z}(t, j) = \frac{1}{j} \sum_{i=t-j+1}^t z_i(t, j), \quad \theta(t, j) = \frac{1}{j} \sum_{i=t-j+1}^t i, \\ \sigma^2(t, j) = \frac{1}{j} \sum_{i=t-j+1}^t (i - \theta(t, j))^2,$$

а сама аппроксимация имеет вид

$$w_k(t, j) = \bar{z}(t, j) + b(t, j)(k - \bar{z}(t, j)), \quad k = t - j + 1, \dots, t. \quad (2.9)$$

Аппроксимация (2.9), продолженная в будущее за пределы выборки, является регрессионным прогнозом. Если τ — горизонт прогноза, то прогнозные значения величин (2.9) определяются формулой

$$v_k(t, j) = \bar{z}(t, j) + b(t, j)(t + k - \bar{z}(t, j)), \quad k = 1, 2, \dots, \tau. \quad (2.10)$$

По регрессионному прогнозу (2.10) восстанавливается прогнозный массив исходных величин. Для этого строятся массивы

$$\begin{aligned} u_n(t, j) &= a(t, j) + A(t, j)(u_{n-1}(t, j) - a(t, j)), \\ u_0(t, j) &= y_i(t, j) \end{aligned} \quad (2.11)$$

и

$$\xi_n(t, j) = v_n(t, j) + u_n(t, j), \quad n = 1, 2, \dots, \tau. \quad (2.12)$$

Собственно прогнозные значения $c_n(t, j)$ исходного ряда цен по выборке объема j , взятой в момент t , на n -ом шаге прогноза строятся по алгоритму

$$c_n(t, j) = c_{n-1}(t, j) + \xi_n(t, j), \quad c_0(t, j) = c(t). \quad (2.13)$$

Для анализа ошибки прогноза строятся массивы относительных локальных отклонений прогнозных значений от фактических,

$$d(t, j; n) = \left(\frac{c_n(t, j)}{c(t + n)} - 1 \right)^2, \quad (2.14)$$

а также средних отклонений на всем горизонте прогноза

$$\delta(t, j; \tau) = \sqrt{\frac{1}{\tau} \sum_{n=1}^{\tau} d(t, j; n)}. \quad (2.15)$$

Первая из ошибок — величина (2.14) — позволит определить тот горизонт прогноза, на котором локальное отклонение не превосходит заданной величины, а вторая — (2.15) — тот объем выборки, свой для каждого горизонта, при котором средняя ошибка прогноза минимальна.

Проанализируем сначала величину (2.14). Для нее в каждый момент времени t для выборки объема j находится наименьший номер $n(t, j, \varepsilon)$ шага прогноза, при котором величина $\sqrt{d(t, j; n)}$ впервые превзойдет заданное значение ε . Для временного ряда величин $n(t, j, \varepsilon)$ находятся их среднее значение $\bar{n}(j, \varepsilon)$ и дисперсия $\sigma^2(j, \varepsilon)$ по имеющемуся массиву данных, доступных для тестирования модели. После этого находится

величина $K^*(\varepsilon)$, дающая минимум относительному среднеквадратичному отклонению по объему выборки:

$$K^*(\varepsilon) = \arg \min_j \frac{\sigma(j, \varepsilon)}{\bar{n}(j, \varepsilon)}. \quad (2.16)$$

Этот объем считается оптимальным для прогнозирования с ошибкой, локально не превосходящей ε , на горизонт $\bar{n}(K^*, \varepsilon)$.

Далее анализируется ошибка (2.15). Для каждой пары значений $\{t, \tau\}$ находится такой объем выборки J , который дает минимум ошибке $\delta(t, j; \tau)$:

$$J(t, \tau) = \arg \min_j \delta(t, j; \tau). \quad (2.17)$$

По имеющемуся объему данных определяются средние по времени величины $\bar{J}(\tau)$ и $\sigma_J^2(\tau)$ для объемов (2.17), а также и среднее значение соответствующих минимумов $\bar{\delta}(\tau)$ ошибки $\delta(t, J(t, \tau); \tau)$.

После этого вычисляется чувствительность $\bar{\lambda}(\tau)$ минимума ошибки к объему данных. Для этого на отрезке $\Delta_J(t) = [J(t, \tau) - \sigma_J(\tau), J(t, \tau) + \sigma_J(\tau)]$ численно определяется логарифмическая производная ошибки по объему выборки

$$\lambda(t, \tau) = \frac{\bar{J}(\tau)}{\bar{\delta}(\tau)} \cdot \frac{1}{2\sigma_J(\tau)} \cdot \left(\max_{\Delta_J(t)} \delta(t, j; \tau) - \min_{\Delta_J(t)} \delta(t, j; \tau) \right) \quad (2.18)$$

и находится ее среднее по времени значение $\bar{\lambda}(\tau)$.

На основе найденной чувствительности определяется ошибка, которую можно допустить при прогнозировании ряда, используя объем выборки $\bar{J}(\tau)$. Для этого вычисляем дисперсию минимальной ошибки

$$\sigma_{\delta}^2(\tau) = \left\langle (\delta(t, J(t, \tau); \tau) - \bar{\delta}(\tau))^2 \right\rangle \quad (2.19)$$

и коэффициент корреляции $\rho(\tau)$ между ошибкой $\delta(t, J(t, \tau); \tau)$ и объемом выборки $J(t, \tau)$:

$$\rho(\tau) = \frac{\langle (\delta(t, J(t, \tau); \tau) - \bar{\delta}(\tau)) (J(t, \tau) - \bar{J}(\tau)) \rangle}{\sigma_J(\tau) \sigma_\delta(\tau)}. \quad (2.20)$$

Тогда дисперсия ошибки прогноза оценивается по формуле

$$D_\delta^2(\tau) = \sigma_\delta^2(\tau) \cdot \left(1 + \left(1 - \rho^2(\tau) \right) \bar{\lambda}^2(\tau) \frac{\sigma_J^2(\tau)}{\bar{J}^2(\tau)} \right), \quad (2.21)$$

а сама ошибка прогноза оценивается величиной

$$\Delta(\tau) = \bar{\delta}(\tau) + D_\delta(\tau). \quad (2.22)$$

Минимум по τ выражения (2.22) даст наиболее оптимальный горизонт прогнозирования по выбранной модели.

Итак, в этом параграфе был рассмотрен пример оптимизации прогноза временного ряда по конкретной авторегрессионной модели. Такая оптимизация полезна в том случае, если прогноз по кинетической модели не дал желаемой точности из-за высокой зависимости между членами ряда. Конечно, оптимизация объема выборки для применения конкретной прогнозной модели эффективна лишь в том случае, если в итоге получается ошибка, меньшая выборочной дисперсии ряда на том же объеме данных. В противном случае необходимо сделать пред-подготовку ряда по описанной выше методике, а остатки прогнозировать по кинетической модели. Такой комбинированный прогноз будет давать в этом случае более высокую точность, чем каждый из этих методов в отдельности.

В целом относительно прогнозов нестационарных рядов можно сделать следующие выводы. Во-первых, если на рассматриваемом промежутке времени локальная ЭЛС изменяется незначительно, то кинетический прогноз ВПФР имеет высокую точность, как для валютного курса. Во-вторых, если ЭЛС сильно нестационарна, как для акций GE, то кинетический прогноз даст ошибку, по порядку величины совпадающую с ошибкой наивного прогноза, и тогда данный метод для соответствующего временного ряда не эффективен. В этом случае могли бы быть полезны гидродинамические модели, но, как уже говорилось, они имеют значение при анализе ВПФР большой размерности, т.е. требуют больших объемов выборки. Для малых объемов выборки более эффективна адаптивная регрессионная модель.

11.3. Ряды с сезонной периодичностью

Во многих прикладных задачах возникают ряды с так называемой естественной или назначенной периодичностью. Это, например, временные

ряды среднесуточной температуры в определенной географической точке Земли, почасовой график артериального давления человека, потребление воды или электроэнергии в жилом доме в зависимости от времени суток и т.п. Прогнозирование таких рядов, с одной стороны, кажется простым, поскольку для прогноза можно использовать детерминистическую модель среднего графика изучаемой величины за период, а отклонения от него считать белым шумом. С другой стороны, горизонтный анализ ряда этих отклонений показывает, что они образованы некоторой динамической системой, поэтому существует прогнозная модель, имеющая более высокую точность, чем наивный прогноз.

Рассмотрим временной ряд, образованный почасовыми ценами на рынке электроэнергии по европейской ценовой зоне России. На рис. 11.3.1. приведен для примера типичный почасовой график цены за одну неделю с понедельника по воскресенье.

Отметим, что относительное среднеквадратичное отклонение часовых данных за сутки меньше, чем за неделю, а то, в свою очередь, меньше, чем за месяц. Характерные величины соответственно равны 17%, 19% и 21%. Следовательно, если в качестве базового профиля взять суточный график (24 часовых значения), усредненный по дням за неделю или за месяц, то ошибка аппроксимации данных с его помощью на месячном интервале будет больше, чем при использовании средненедельного (168 часовых значений) профиля.

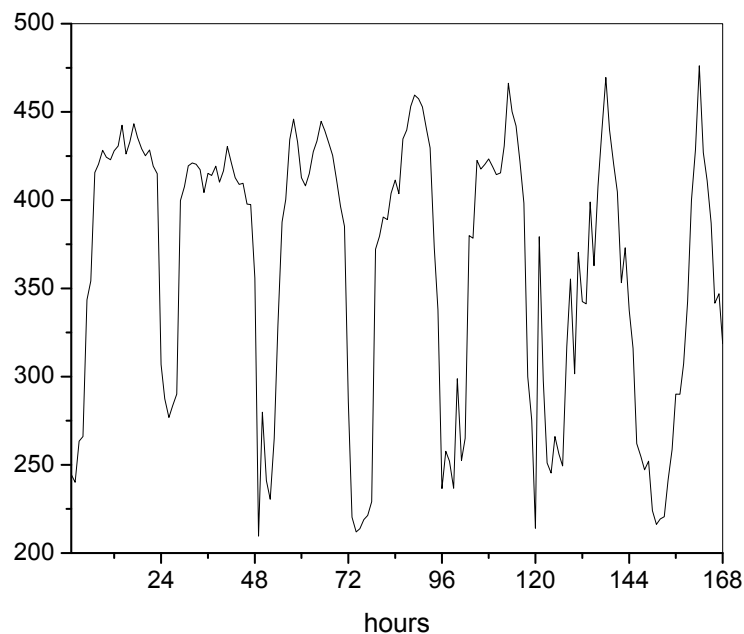


Рис. 11.3.1 — Пример почасового графика цены электроэнергии за неделю, руб./МВт*ч

Из рис. 11.3.1 видно, что недельный график имеет суточную периодичность с тремя достаточно хорошо выраженными интервалами для рабочего дня (от 0 до 9 часов, от 10 до 17 часов и от 18 до 23 часов) и двумя для выходного (от 5 до 16 часов и от 17 до 4 часов следующих суток). Однако напрашивающееся разбиение временного ряда на подмножества, разделенные хронологически (например, только выходные или только рабочие дни), затрудняет их изучение как статистического набора последовательных случайных событий, что увеличивает ошибку прогноза.

Среднемесячный же профиль (4 полных недели или 672 значения) неудобно использовать из-за того, что динамика продаж часто имеет особенность скачкообразного изменения от месяца к месяцу, поэтому годовая дисперсия среднемесячных профилей еще больше, чем месячная дисперсия недельных, так что относительное среднеквадратичное отклонение часовых данных за год составляет приблизительно 30%.

Проведем теперь пред-обработку ряда, состоящую в исключении из него главной периодической составляющей, усредненной по четырем неделям. Вычтем затем из каждого из четырех недельных графиков найденный средний профиль. В результате получаем ряд остатков, который имеет приблизительно в два раза меньший размах по сравнению с исходным. Фрагмент этого ряда представлен на рис. 11.3.2.

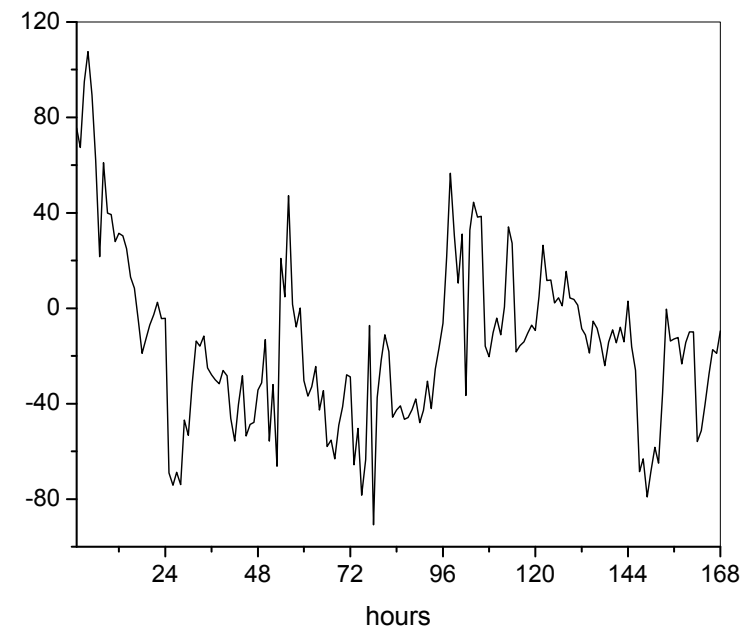


Рис. 11.3.2 — Остаток ценового ряда после вычитания среднего профиля

Относительная среднеквадратичная ошибка аппроксимации месячных данных соответствующим средненедельным профилем составила в рассматриваемом примере 11%. Таким образом, одно лишь выделение наилучшего среднего профиля в условиях назначенной периодичности позволяет прогнозировать ряд с относительной ошибкой порядка 0,1.

Поскольку интерес представляет прогнозирование не на один-два часа, а на сутки вперед, то суточный график естественно рассматривать как

24-мерную единицу, обозначаемую \mathbf{p}_n , где n нумерует суточные графики. Семь таких упорядоченных единиц образуют недельный блок $\mathbf{w}_n(j) = (\mathbf{p}_{7n-6+j}, \mathbf{p}_{7n-5+j}, \dots, \mathbf{p}_{7n+j})$. Индекс j нумерует скользящий начальный день недельного блока. Далее рассматриваются данные в количестве не менее одного недельного блока. Пусть таких блоков L . Тогда для каждого l от 1 до L может быть построен средненедельный профиль $\mathbf{W}_l(k, j)$, получаемый усреднением l последовательно идущих блоков:

$$\mathbf{W}_l(k, j) = \frac{1}{l} \sum_{i=0}^{l-1} \mathbf{w}_{k+i}(j). \quad (3.1)$$

Затем найденный профиль $\mathbf{W}_l(k, j)$ вычитается из всех рассмотренных l недельных блоков, после чего полученные данные разворачиваются в последовательность ряда остатков $r(t-168l+1), r(t-168l+2), \dots, r(t)$, которая насчитывает $168l$ значений. Первая разность (т.е. скорость) определена по этой выборке в моменты времени $t-1, t-2, \dots, t-168l+2$, т.е. данные приращений образуют набор из $168l-2$ значений. С добавлением следующего вектора суточных данных эта операция повторяется.

Полученный ряд первых разностей отображается на отрезок $[0; 1]$, на который с некоторым запасом нормируется размах выборки. Нормировка с запасом требуется для того, чтобы распределения последующих выборок сравнивались в одинаковых масштабах. Этот ряд, обозначаемый далее $x(t)$, и является объектом дальнейшего анализа.

Для рассматриваемого ряда остатков ошибка суточной аппроксимации по линейной регрессии на время в среднем по имеющемуся объему данных составила 5,7%. Однако для прогноза такая модель не очень хороша, поскольку параметр регрессии не постоянен для разных суток. В результате регрессионный прогноз приводит к ошибке в 11,5%, которая приблизительно равна среднеквадратичному отклонению от средненедельного графика, а последнее находится вовсе без всякого прогноза. Следовательно, необходимо прогнозировать сам параметр регрессии. Это можно сделать, например, следующим образом, формируя адаптивную регрессионную модель.

Обозначим a_n значение параметра регрессии a при аппроксимации данных за день n , а \bar{x}_n среднесуточное значение ряда. Именно разброс в значениях этих параметров является причиной увеличения ошибки прогноза на следующие сутки по сравнению с аппроксимацией суточных данных.

Автокорреляция между величинами a_n заметно отлична от нуля для единичного лага $A_a(1) = -0,35$, а для лагов, больших единицы, оказалась пренебрежимо малой, поэтому для параметра регрессии можно применить авторегрессионную модель первого порядка. Оценка среднего значения близка к нулю, так что модель авторегрессии первого порядка для оценки a_n имеет вид

$$\hat{a}_{n+1} = A_a(1)a_n. \quad (3.2)$$

Аналогично следует подправить и среднее суточное значение \bar{x}_n , которое также весьма значительно различается для разных суток. Коэффициент авторегрессии первого порядка для величин \bar{x}_n равен $A_x(1) = 0,33$, а остальные лаги вносят пренебрежимо малые поправки. Среднемесячное суточное значение равно нулю, поскольку мы изучаем центрированный ряд остатков. Тогда модель для прогноза \bar{x}_n имеет вид

$$\hat{\bar{x}}_{n+1} = A_x(1)\bar{x}_n. \quad (3.3)$$

Объединяя модели (3.2) и (3.3) с регрессионным прогнозом остатков x_n , получаем уточненный прогноз исходных величин $p(t)$ на сутки вперед, который имеет ошибку 8,9%. Таким образом, применение комбинированных моделей позволяет улучшить качество прогноза по сравнению с наивным прогнозом.

Дальнейшее снижение ошибки прогноза в рамках выбранной модели связано с оптимизацией объема выборки, т.е. в определении оптимальной длины усреднения при получении средненедельного часового профиля. Для решения этой задачи следует изучить горизонтную статистику рассматриваемого ряда, а также рассмотреть зависимость выборочной дисперсии от объема выборки.

Фрагмент горизонтного ряда для ряда остатков $x(t)$ показан на рис. 11.3.3 для сдвига $\tau = 24$ и фиксированного расстояния между ВПФР $\varepsilon = 0,05$. По оси абсцисс отложено текущее время, а по оси ординат — объем

выборки, достаточный в данный момент для прогнозирования на 24 шага вперед с указанной точностью. Напомним, что значения этого ряда мажорируются величиной $2\tau / \varepsilon = 960$.

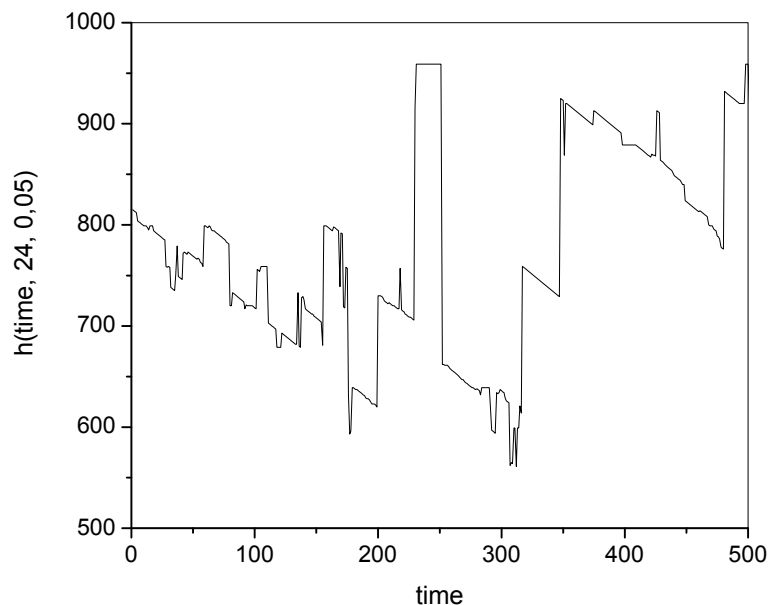


Рис. 11.3.3 — Горизонтный ряд для ряда остатков $x(t)$ при сдвиге на 24 шага для уровня стационарности 0,05

В приведенном примере наблюдается значительный разброс значений горизонтного ряда. Достижение точной верхней грани в 960 точек свидетельствует о максимальной неопределенности в поведении системы, а скачкообразный спад — о проявлении внутренней периодичности, т.е. об усилении детерминистских факторов. Плато означает сохранение сформировавшихся тенденций, скачкообразный рост — локальное исчезновение квазипериодической структуры, а эпизоды с кусочно-линейным трендом означают постепенное нарастание хаоса, когда каждое новое значение исходного ряда не совпадает с последним в выборке, увеличивая, таким образом, минимальный объем на 1.

Среднее значение данного горизонтного ряда составляет 802 точки, а ширина «полосы», в которой находятся значения горизонтного ряда, т.е. минимально достаточные объемы выборок, составляет 35 % от максимума

в 960 точек. Чем шире эта полоса, тем более значимо влияние детерминистских факторов, которых не удалось исключить на этапе пред-обработки ряда. Квантиль порядка 0,95 для распределения на рис. 11.3.3 равен 942. Это тот объем выборки, который в среднем наиболее оптимален для прогнозирования на 1 сутки вперед с ошибкой в определении ВПФР на уровне 0,05. Прогноз ВПФР по уравнению Лиувилля представлен на рис. 11.3.4. Его ошибка на сутки вперед составила 0,042, а собственно ряда цен — 0,065, что почти в 2 раза лучше наивного прогноза.



Рис. 11.3.4 — Фактическая и прогнозная ВПФР для ряда остатков $x(t)$ для объема выборки 942 и горизонта прогноза 24 шага

Еще более точный прогноз дает метод идентификации динамической системы по оптимальному для прогнозирования объему данных, т.е. по 942 значениям. Оказалось, что ряд остатков может быть смоделирован трехмерной ДС с квадратичной нелинейностью в правой части, т.е. в чем-то ряд остатков подобен логистической модели и ее обобщениям. Ошибка прогноза на сутки вперед составила 0,043, что меньше, чем ошибка регрессионной аппроксимации ряда остатков на время.

Отметим, что моделью вида системы Лоренца наиболее точно прогнозируется также и ряд среднесуточных температур, но в этом случае модель адекватна и по физическим соображениям. Интересно, что эта физическая модель может быть обнаружена чисто статистическими методами идентификации ДС по носителю совместного распределения. Таким

образом, сочетание горизонтного анализа с нелинейной авторегрессией в данном случае привело к наиболее точному прогнозу.

11.4. Анализ устойчивости распределения значений индикатора

Моделирование ансамбля траекторий нестационарного временного ряда с помощью кинетических уравнений позволяет по-новому подойти к проблеме анализа эффективности срабатывания того или иного индикатора. В качестве анализируемых величин могут выступать: индикатор разладки, индикатор эффективности торговой стратегии (кривая кумулятивной доходности), индикатор оптимального объема выборки при анализе корреляций, и т.п. Методика численного моделирования траекторий ряда описывается в следующей главе в п. 12.1.

Анализ индикатора на устойчивость относительно нестационарного временного ряда имеет два аспекта. Во-первых, следует выяснить, каково его выборочное распределение по выборкам достаточно больших объемов. Если это распределение стационарно в рамках приближенного представления о его ПФР через гистограмму ВПФР, то построенный индикатор определяет параметры выборки для нахождения скользящих индикаторов текущего состояния ряда. Если же распределение изучаемого индикатора нестационарно, то он является скользящим индикатором, и в этом случае возникает вторая задача: определить, при каких условиях на исходный временной ряд его срабатывание наиболее безошибочно. Первая задача была подробно разобрана ранее в п. 7.5 на примере показателя Херста. Здесь будет рассмотрена вторая задача — анализ статистики скользящего индикатора.

Пусть $q(t)$ есть значение анализируемого индикатора в момент времени t , а $p(t)$ есть значение временного ряда, который является «инструментом», на котором работает данный индикатор. Ряд первых разностей инструмента обозначим $x(t)$. Пусть также T есть длительность периода времени, в течение которого анализируется индикатор. Обычно этот период связан с внешними условиями по отношению к самой статистической задаче. Этими условиями определяется функционирование системы, порождающей данный временной ряд. Для простоты будем считать оба ряда $q(t)$ и $x(t)$ эквидистантными с единичным шагом.

Обозначим $f(x, t)$ есть ВПФР ряда приростов инструмента, а $g(q, t)$ есть ВПФР ряда значений индикатора. Обе ВПФР построены по

выборкам одинакового объема T , причем представляющие их гистограммы также имеют одинаковое равномерное разбиение, которое является оптимальным для одной из этих двух случайных величин — именно, для той, для которой число разбиений меньше.

Сначала следует провести проверку того, насколько сильно влияет на нестационарное поведение распределения индикатора нестационарность ряда приростов инструмента. Для этого запишем уравнение Лиувилля для обоих распределений:

$$\frac{\partial f}{\partial t} + \frac{\partial(uf)}{\partial x} = 0; \quad \frac{\partial g}{\partial t} + \frac{\partial(vg)}{\partial q} = 0.$$

Скорости $u(x, t)$ и $v(q, t)$ определяются по формуле (9.1.3) из сравнения двух встык-выборок длины T .

Затем вычисляется корреляция $corr_{u,v}(t)$ с нулевым лагом между скоростями как между двумя наборами из упорядоченного набора чисел в соответствии с разбиением гистограмм. Абсолютная величина корреляции между скоростями покажет, насколько сильно изменение распределения индикатора связано с изменением распределения приростов инструмента.

Сгенерируем затем достаточное количество (не менее T) модельных траекторий ряда $p(t)$, которые исходят из заданного состояния ряда $p(t - T)$ и получены из решения уравнения Лиувилля для эволюции ВПФР приростов $x(t)$. На каждой из этих траекторий протестируем действие индикатора и вычислим аналогичные модельные значения величины $corr_{u,v}(t)$. Если разброс корреляций небольшой, то найденное значение характеризует постоянное влияние инструмента на индикатор. В этом случае индикатор следует считать устойчивым относительно инструмента. Если же разброс большой, это свидетельствует о неудачном выборе индикатора, ибо он должен срабатывать на изменение распределения инструмента, которое на модельных траекториях ансамбля квазистационарно.

Следующей задачей является установление пределов применимости индикатора к уровню волатильности инструмента. Для этого эмпирическая скорость изменения ВПФР приростов инструмента модифицируется «волевым путем», изменяясь в определенных участках гистограммы в $z(x, t)$ раз, т.е. вместо скорости $u(x, t)$ рассматривается скорость $u(x, t)z(x, t)$. Функция модификации $z(x, t)$ выбирается в соответствии с целями анализа. В результате генерируется множество траекторий

$p(t)$, которые по сравнению с исходной траекторией имеют измененные показатели волатильности. Для классификации типов нестационарного поведения ряда $p(t)$ полезно рассматривать два типа волатильности: первый тип отвечает микроструктуре ряда, а второй тип характеризует глобальное поведение траектории на промежутке времени T . В качестве характеристики волатильности первого типа берется относительный среднеквадратичный прирост $\bar{\delta}$ величины $\delta(t) = x(t)/p(t)$, а в качестве волатильности второго типа — число пересечений траекторией $p(t)$ коридора своей средней линии на промежутке времени T , ширина коридора принимается равной $\bar{\delta}$. Пересечением считается полный выход траектории за противоположную границу коридора.

Может оказаться так, что для определенных сочетаний микро- и макроволатильностей ряда инструмента его влияние на распределение индикатора кластеризуется по величинам корреляций скоростей изменения распределений. Тогда имеет смысл рассмотреть изменение эффективности срабатывания индикатора при сдвиге распределения приростов инструмента из одного кластера в другой. Для этого выбираются два распределения из разных кластеров и вычисляется соответствующая ЭЛС в предположении, что переход между кластерами происходит за время T . Затем вводится функция эффективности срабатывания индикатора $e(q, x)$, понимаемая в простейшем случае как доля правильных срабатываний. Возможна модификация этой функции в виде премий за правильное срабатывание и штрафов за ложное срабатывание. Затем строится ВПФР $\varphi(e, t)$ значений функции эффективности для переходов между разными кластерами. Если распределение $\varphi(e, t)$ квазистационарно с точностью, отвечающей разбиению гистограммы, то индикатор считается устойчивым по эффективности относительно вариации инструмента. В противном случае индикатор неустойчив по вариации инструмента, и следует выяснить причину изменения распределения $\varphi(e, t)$. Как и выше, для этого вычисляется корреляция между скоростью изменения ВПФР при переходе из кластера в кластер и между скоростью изменения распределения $\varphi(e, t)$. Высокая корреляция свидетельствует о влиянии инструмента на эффективность индикатора, а низкая — о его внутренней хаотичности, что является негативной характеристикой индикатора.

Мерой устойчивости индикатора по отношению к вариации инструмента является отношение

$$S_e = \bar{e} / \sigma_e, \quad (4.1)$$

которое вычисляется по ансамблю модельных траекторий, имитирующих переход распределения прироста инструмента из одного кластера в другой. Чем больше величина S_e (в алгебраическом смысле, а не по модулю), тем более устойчиво работает индикатор.

11.5. Метод функций распределения в математической лингвистике

Существует большое количество статистических задач в области анализа текстов, которые формулируются в терминах функции распределения текста по символам или группам символов. В этом разделе будут рассмотрены методы, применяемые для анализа литературных текстов [29], т.е. текстов, написанных профессиональными писателями. Эти методы используют три типа распределений в зависимости от поставленной задачи. Распределение текста по буквам применяется для идентификации автора неизвестного текста в библиотеке известных авторских эталонов. Распределение букв по частоте встречаемости или ранговое распределение характеризует алфавит по избытку или недостатку используемых символов. Распределение расстояний между одинаковыми буквами применяется для распознавания однородных фрагментов текста. Рассмотрим последовательно эти три задачи.

Задача идентификации автора неизвестного текста состоит в следующем. Имеется библиотека, содержащая тексты, представленные в виде ПФР (однобуквенных или многобуквенных, т.е. n -грамм) для A известных авторов. Пусть K_a — имеющееся количество текстов a -го автора, и $N_{i,a}$ — количество букв в i -ом тексте этого автора, $i = 1, 2, \dots, K_a$. Длина каждого из текстов достаточна для проведения статистического анализа в соответствии с оценками на точность эмпирического распределения (см. п. 5.2). Обозначим $f_{i,a}(j)$ n -ПФР соответствующего текста,

где аргумент j меняется от 1 до $\alpha(n) = 32^n$, где 32 — число букв русского алфавита при условии отождествления «е» и «ё», что часто встречается в электронной форме текстов. Пробелы, знаки препинания и прочие символы не учитываются. Для каждого автора определим его средневзве-

шенную ПФР по тому же правилу, что и эталонные ПФР в задаче о кластеризации фрагментов временных рядов (п. 10.4):

$$F_a(j) = \frac{1}{N_a} \sum_{i=1}^{K_a} f_{i,a}(j) N_{i,a}, \quad N_a = \sum_{i=1}^{K_a} N_{i,a}. \quad (5.1)$$

Введем «библиотечную норму» ρ_{ik} как расстояние между ПФР текстов i и k в норме суммируемых функций:

$$\rho_{ik} = \|f_i - f_k\| = \sum_{j=1}^{\alpha(n)} |f_i(j) - f_k(j)|. \quad (5.2)$$

Для каждого автора a построим плотность функции распределения $g_a^+(\rho)$ отклонений $\rho_{i,a}$ «его» текстов, а также распределение $g_a^-(\rho)$ отклонений $\rho_{k,b,a}$ «чужих» произведений от его средней ПФР F_a . Обозначим $G_a^\pm(\rho)$ соответствующие интегральные функции распределения.

Минимальное значение ρ , при котором $G_a^+(\rho) = 1$, обозначим ρ_a^+ , а максимальное значение ρ , при котором $G_a^-(\rho) = 0$, обозначим ρ_a^- . Смысл введенных величин в том, что все ПФР текстов автора a находятся на расстоянии не более ρ_a^+ от его средней ПФР (5.1), и аналогично все

ПФР других авторов находятся от нее на расстоянии не менее ρ_a^- . Величина $1 - G_a^+(\rho_a^-)$ есть вероятность ошибочно признать за произведение автора « a » чужой текст (ошибка второго рода), а величина $G_a^-(\rho_a^+)$ есть вероятность ошибочно отвергнуть произведение автора « a », посчитав его за чужое (ошибка первого рода). Назовем расстоянием разделения авторов такое значение $\hat{\rho}$, для которого ошибка идентификации автора текста минимальна:

$$\hat{\rho} = \arg \min (1 - G^+(\rho) + G^-(\rho)) = \arg \max (G^+(\rho) - G^-(\rho)). \quad (5.3)$$

Эта величина может служить верхним уровнем для кластеризации текстов.

На рис. 11.5.1 приведены плотности $g^\pm(\rho)$ распределений расстояний для 3-ПФР по достаточно большой совокупности авторов (30 авторов, 300 текстов).

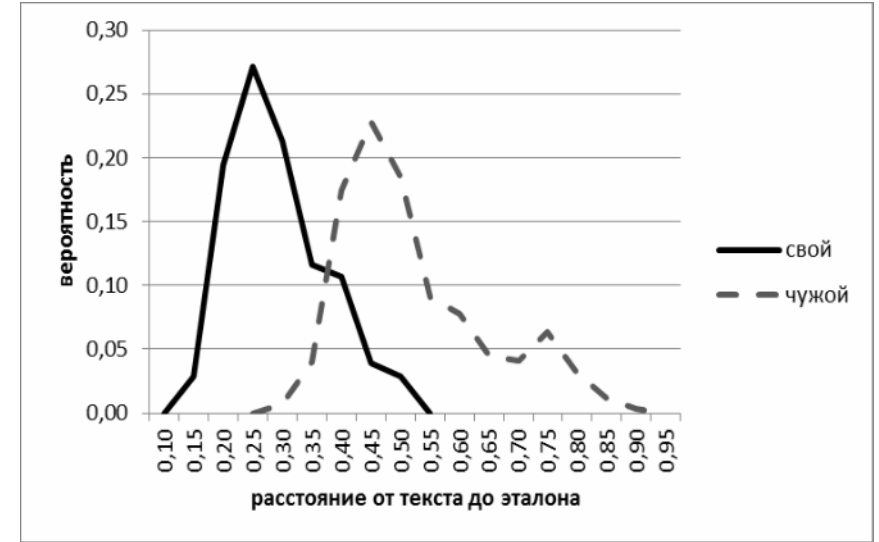


Рис. 11.5.1 — Распределение расстояний между 3-ПФР текста и авторского эталона

Видно, что расстояние от текста до эталона «своего» автора имеет заметно более узкое распределение, чем от текста до «чужого» эталона. Это обстоятельство позволяет весьма точно идентифицировать автора в библиотеке эталонов. Пусть имеется текст «0» неизвестного автора, который надо идентифицировать внутри данной библиотеки. Автором текста «0» считается тот из авторов « a », для которого норма $\rho_a^0 = \|f_0 - F_a\|$ разности между ПФР $f_0(j)$ текста «0» и средней авторской ПФР $F_a(j)$ минимальна:

$$\rho_a^0 = \|f_0 - F_a\|, \quad a^0 = \arg \min_a \rho_a^0. \quad (5.4)$$

Некоторые характеристики распределений рис.11.5.1 приведены в табл. 11.5.1.

Табл. 11.5.1. Характеристики распределений расстояний между текстом и эталоном

Показатель	1-ПФР	2-ПФР	3-ПФР
Среднее значение l_s (свой текст)	0,04	0,13	0,27
Среднее значение l_d (чужой текст)	0,08	0,23	0,50
Стандартное отклонение σ_s (свой текст)	0,02	0,04	0,08
Стандартное отклонение σ_d (чужой текст)	0,03	0,07	0,12
Расстояние разделения $\hat{\rho}$	0,05	0,19	0,35
$\rho +$	0,11	0,30	0,55
$\rho -$	0,03	0,10	0,25
Вероятность ошибки I рода: $G^-(\rho +)$	0,47	0,65	0,72
Вероятность ошибки II рода: $1 - G^+(\rho -)$	0,72	0,57	0,50
Общая площадь под распределениями	0,38	0,35	0,22
Интервал 0,90 нормировки (свой текст)	0,03-0,06	0,08-0,16	0,20-0,40

Из табл. 11.5.1 можно вывести статистические аргументы в пользу эффективности метода 3-ПФР по сравнению с распределениями другой размерности. Во-первых, ошибка второго рода, являющаяся в задаче идентификации ключевой, минимальна для 3-ПФР. Во-вторых, для распределений расстояний от текстов до «своих-чужих» эталонов площадь перекрытия графиков также минимальна для 3-ПФР. В-третьих, если вычислить разность средних расстояний от текста до «чужого» и до «своего» авторов $l_d - l_s$ и сравнить ее с суммой соответствующих дисперсий

$\sigma_d + \sigma_s$, то для 1,2,-ПФР отношение $\frac{l_d - l_s}{\sigma_d + \sigma_s}$ равно 0,9, а для 3-ПФР

оно равно 1,1, т.е. различающая способность авторских эталонов здесь наибольшая. Интересно отметить, что и для 4-ПФР отношение $\frac{l_d - l_s}{\sigma_d + \sigma_s}$

также равно 0,9. Поэтому увеличивать размерность ПФР для повышения точности анализа имеет смысл лишь до определенных пределов. Обе последовательности характерных расстояний между текстами — «своими» и «чужими» — возрастают с увеличением размерности ПФР, но они также и ограничены сверху одним и тем же числом, ибо эти расстояния заведомо не превосходят двойки. Следовательно, точность, с которой могут быть различены тексты разных авторов, при увеличении размерности ПФР начнет уменьшаться, что потребует для анализа очень больших объемов текстов. Это наблюдается уже на 4-ПФР.

Описанный метод оказался весьма точным в определении автора неизвестного текста: ошибка идентификации на множестве «30 авторов, 300 текстов» по близости ПФР в норме $L1$ составила: для 1-ПФР 15%, для 2-ПФР 5%, для 3-ПФР 0%, для 4-ПФР 7%. Метод сравнения с эталоном 3-ПФР является наиболее точным из всех существующих статистических способов идентификации автора.

Рассмотрим теперь ранговое распределение букв в русском и некоторых европейских языках. В отличие от алфавитного упорядочения ранговое распределение устойчиво независимо от автора, произведения или жанра. При этом в разных текстах на одном и том же порядковом месте могут оказаться разные буквы, однако форма распределения от этого не меняется. Отсюда можно сделать вывод, что ранговое распределение букв характеризует язык как таковой.

В среднем по большой группе текстов на первом месте по частоте использования в русской художественной литературе стоит буква «о», на предпоследнем — буква «ф», а на последнем — твердый знак. Остальные места не имеют четкой привязки к буквам, хотя можно выделить группы примерно постоянного состава: первые 7 букв, составляющие 0,52 нормировки ПФР, последние 6 букв, дающие суммарный вклад в нормировку на незначительном уровне 0,01, и середину (19 букв). Однако следует подчеркнуть, что нельзя провести условную черту, навсегда отделяющую одни буквы от других. У любой буквы в центральной части алфавита «персональное» место по частоте встречаемости флуктуирует на 2-6 позиций.

На рис. 11.5.2 приведена эмпирическая кривая рангового распределения по 1000 текстов разных авторов и аппроксимирующая ее модельная функциональная кривая.

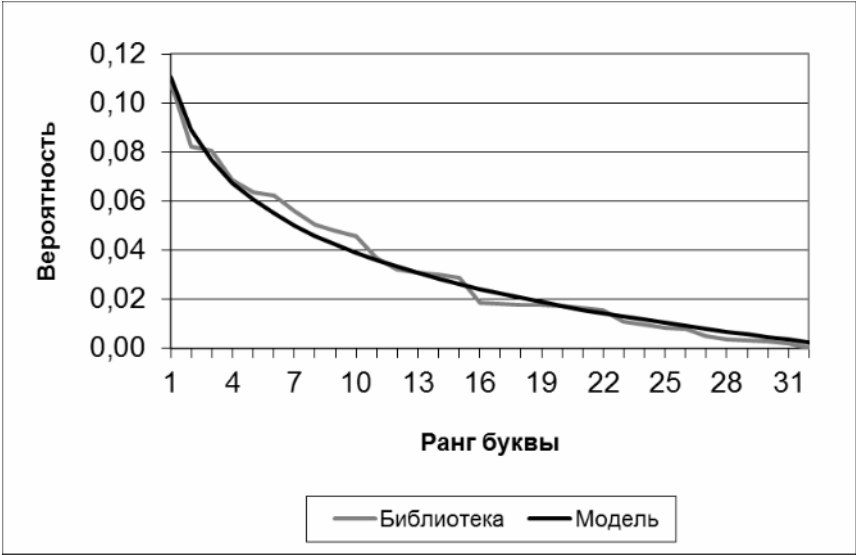


Рис. 11.5.2 — Модельное и эмпирическое распределения букв по частоте встречаемости

Оказалось, что ранговое распределение частот букв в тексте с высокой точностью аппроксимируется логарифмической функцией:

$$f(k) = \frac{1}{n} \left(1 + \frac{1}{n+o} \ln \frac{n!}{k^n} \right). \quad (5.5)$$

Распределение (5.5) нормировано на единицу. В нем k — есть ранг букв по убыванию, n — количество букв в алфавите (в данном случае 32), $f(k)$ — частота встречаемости буквы, а параметр o выбирается из требования наилучшей детерминации эмпирической зависимости.

Численный анализ показал, что тексты в прозе на современном русском языке имеют наилучшую аппроксимацию в классе распределений (5.5) при значении параметра $o = o_{rus} = 0$. Детерминация формулы (5.5) составила 0,99, среднеквадратичная ошибка аппроксимации эмпири-

ческого распределения равна 0,003, а расстояние между эмпирической и модельной кривыми в норме $L1$ равно 0,09.

Интересно отметить, что если перевести русскоязычные тексты в транслит, т.е. записать их латиницей с использованием 23 букв, то наилучшая аппроксимация частотного упорядочения для таких текстов дается формулой (5.5) при значении параметра $o_{translit} = 9$. Ошибка этой аппроксимации в норме $L1$ составила 0,07 при детерминации 0,99. Отсюда вытекает интерпретация параметра o в формуле (5.5) как оценки избыточности или недостаточности алфавита. Можно предложить, что звуковая структура языка наиболее адекватно отражена в алфавите, число символов в котором есть $n + o$, поэтому o далее принимается равным целому числу.

Аналогичная картина наблюдается и для текстов, написанных на европейских языках, т.е. тогда, когда слова составляются из букв. Поскольку в некоторых алфавитах (чешский, польский, французский, немецкий и др.) используются диакритические знаки и диграфы, то следует определиться, считать ли их за самостоятельные символы или объединить со сходными по звучанию буквами. В ряде случаев при таком объединении удавалось добиться лучшей модельной аппроксимации. Результаты анализа для текстов на распространенных европейских языках приведены в табл. 11.5.2.

Отметим, что с точки зрения интерпретации параметра o различия в языках индоевропейской семьи проявляются только в использовании огласовки. Согласных букв в кириллическом и латинском алфавитах одинаковое количество — по 20. Если перенормировать частоты использования букв только на согласные, то оказывается, что с достоверностью 0,96 распределения букв в текстах на всех языках аппроксимируются формулой (5.5) с параметрами $n = 20$, $o = 0$.

Табл. 11.5.2. Характеристики ранговых распределений букв в текстах

Язык	Число букв в алфавите	Параметр o в (5.5)	Детерминация	Ошибка в норме $L1$
Старославянский, XI век	43	-6	0,97	0,11
Русский, XIX век	37	0	0,98	0,10

Русский, XXI век	32 (без ё)	0	0,99	0,09
Русский/транслит	23	+9	0,99	0,07
Болгарский	30	-4	0,98	0,09
Польский	32 с диакритич. знаками	+3	0,98	0,09
Чешский	42 с диакритич. знаками	-6	0,98	0,10
Чешский	30=26 ос- новных+4 диакритич. знаков	0	0,99	0,08
Шведский	29	-4	0,98	0,09
Немецкий	30 с диакритич. знаками	-7	0,97	0,09
Английский	26	0	0,96	0,10
Итальянский	26	-4	0,97	0,08
Испанский	27	-4	0,99	0,07
Французский	42 с диакритич. знаками	-9	0,96	0,11
Тексты без глас- ных на любом языке	20	0	0,96	0,08

Вернемся к текстам на русском языке и рассмотрим еще один аспект применения буквенных распределений, а именно — распределения расстояний между одинаковыми буквами. Будем обозначать ПФР-Б $B_i(l)$ вероятность того, что расстояние между двумя последовательными буквами «i» равно числу l , которое определяется как разность между порядковыми номерами букв в тексте, считая от начала. Оказалось, что распределения для всех букв, кроме трех малоиспользуемых («щ», «э», «ъ») имеют похожий вид, т.е. являются характеристикой языка, на котором написаны произведения. На рис. 11.5.3-11.5.5 приведены типичные распределения расстояний между одинаковыми буквами на объемах текстов

порядка 500 тыс. знаков. Заметим, что если построить фазовую диаграмму в координатах «частота f буквы — максимум L расстояния между буквами», то окажется, что с достаточно хорошей достоверностью 0,92 укладываются в эмпирическую зависимость (коэффициенты даны в округленном виде)

$$\ln L \approx 0,2 \exp\left(\frac{2}{f}\right). \quad (5.6)$$



Рис. 11.5.3 — Распределение промежутков между буквами «а»

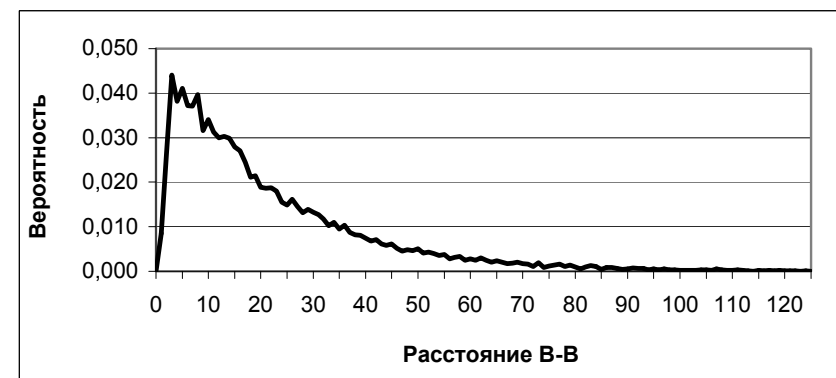


Рис. 11.5.4 — Распределение промежутков между буквами «в»

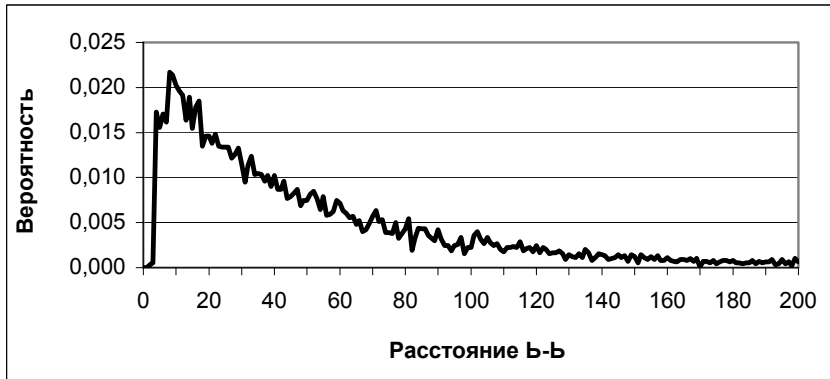


Рис. 11.5.5 — Распределение промежутков между буквами «б»

Поскольку среднее расстояние \bar{l}_i между буквами « i » точно равно обратной частоте соответствующей 1-ПФР, т.е. $\bar{l}_i = 1/f(i)$, то оценка носителя ПФР-Б имеет вид

$$\ln L_i = 0,2 \exp(2\bar{l}_i). \quad (5.7)$$

Точки глобального максимума эмпирических ПФР-Б оказались примерно с такой же достоверностью (0,91) лежащими на аппроксимирующей кривой

$$l_{i,\max} = -2,6 \ln f(i) - 3,8. \quad (5.8)$$

В результате ПФР-Б с ошибкой порядка 0,1 моделируются однопараметрическим гамма-распределением, параметр которого — частота встречаемости данной буквы:

$$B_i(l) = \frac{(\lambda_i l)^{\nu_i}}{\Gamma(\nu_i + 1)} e^{-\lambda_i l}, \quad (5.9)$$

где параметры ν_i и λ_i отвечают соответственно за максимум ПФР-Б и за скорость ее убывания при больших l . Эти параметры считаем зависящими только от частоты $f(i)$. Из условия $\bar{l}_i = 1/f(i)$ следует, что параметры ν_i и λ_i связаны соотношением

$$\bar{l}_i = \int_0^{\infty} l B_i(l) dl = \frac{\nu_i + 1}{\lambda_i} = \frac{1}{f(i)},$$

откуда получаем

$$\nu_i = \frac{\lambda_i}{f(i)} - 1. \quad (5.10)$$

Единственный максимум распределения (5.9) находится в точке

$$l_{i,\max} = \frac{\nu_i}{\lambda_i} = \frac{1}{f(i)} - \frac{1}{\lambda_i}. \quad (5.11)$$

Определив для всех букв « i » точки максимума $l_{i,\max}$ из аппроксимации (5.8), можно из (5.11) определить параметр λ_i , а затем из (5.10) и параметр ν_i . Например, для буквы «о» ($f = 0,11$) получается $\lambda = 0,140$, а для буквы «т» ($f = 0,06$) $\lambda = 0,076$. Параметр распределения ν для всех букв оказался одинаковым, равным примерно $\nu = 0,27$. Ошибка аппроксимации эмпирических ПФР-Б распределениями вида (5.9) в норме L_1 составила 0,11.

Интересно отметить, что ряды расстояний между буквами имеют весьма малую корреляцию и автокорреляцию, не превосходящую по модулю 0,1, а чаще всего имеющую величину 0,02. Это означает, что текст как последовательность символов приближенно представляет собой n -мерный белый шум, где n — число букв в алфавите. Если построить горизонтную статистику (см. главу VI) для каждой последовательности расстояний между буквами, то окажется, что для всех букв горизонтные ряды имеют приблизительно одинаковые распределения, характерные для распределений горизонтных рядов, порожденных хаотическими динамическими системами (п. 6.3). Соответствующий пример приведен на рис. 11.5.6.

Следовательно, можно предположить, что существует алгоритмическое механическое правило по составлению символов в информационную цепочку, т.е. в текст из слов. Определить эту ДС было бы весьма любопытно, но в вычислительном плане сказываются ограничения по размерности, препятствующие провести анализ с высокой точностью.

Из практического применения ПФР-Б отметим возможность проверки текста на однородность с помощью горизонтного ряда последовательно-

сти расстояний между одинаковыми (и достаточно часто употребляемыми) буквами. Какие именно это буквы, значения не имеет, так как поведение горизонтной статистики для них одинаково.



Рис. 11.5.6 — Распределение горизонтного ряда для ПФР-Б, $\tau = 10$, $\varepsilon = 0,05$

Относительно малые значения горизонтного ряда свидетельствуют о том, что в этом месте сработали внутренние взаимосвязи между элементами исходного временного ряда, а большие значения, напротив, показывают, что последовательность случайных величин независима. Несмотря на то, что нормированные ПФР-Б для разных авторов практически совпадают, т.е. все авторы пишут с одинаковым гамма-распределением, отражающим только общезыковые свойства, в местах склейки текстов разных авторов наблюдается разладка в виде резких всплесков горизонтного ряда.

Были рассмотрены склеенные в один текст фрагменты двадцати произведений разных писателей так, чтобы ряд расстояний между буквами для каждого из авторов содержал бы не менее 1000 данных. Параметр стационарности ε горизонтных распределений менялся от 0,01 до 0,1, а лаг сдвига τ — от 1 до 20. Наилучшие индикативные свойства были обнаружены для пары $\tau = 10$, $\varepsilon = 0,05$.

Оказалось, что места склейки текстов разных авторов могут быть опознаны с точностью $\pm \tau$ следующим образом. Сами по себе всплески горизонтного ряда могут быть весьма большой амплитуды и для одного писателя, но поведение горизонтного ряда в окрестности такого пика имеет

определенную специфику. Именно, максимумы горизонтного ряда идут сериями по 2-5 значений подряд с последующим квазилинейным убыванием. В области же перехода от одного текста к другому наблюдался провал большой амплитуды (порядка 80-90) и мгновенный скачок вверх такой же большой амплитуды. Например, типичная последовательность значений горизонтного ряда в этой переходной области имеет вид ..., 387, 295, 382, ... и т.п. Для одного автора амплитуда таких скачков «вниз-вверх» не превосходит 60 с эмпирической вероятностью 0,95. Это означает, что лишь с вероятностью 0,05 двойной скачок горизонтного ряда с амплитудой выше 60 может произойти в тексте одного писателя, а скачок с амплитудой выше 80 — с вероятностью менее 0,01. В результате в текстах, написанных авторским тандемом, можно указать места, где авторы менялись, с точностью до полстраницы.

Косвенная проверка достоверности этого метода может быть осуществлена в тех случаях, если соавторы имеют собственные произведения. Тогда с помощью описанного индикатора разладки в виде «двойного скачка» можно разрезать текст на фрагменты, написанные, по предположению, разными авторами, а затем, сгруппировав эти фрагменты, с помощью авторских эталонов определить их принадлежность. Так, примеры произведений братьев Стругацких, Ильфа и Петрова, Перумова и Лукьяненко и ряда других авторов показали (см. [29]) непротиворечивость идентификации разладки с вероятностью ошибки 0,05, т.е. метод дает в среднем одну ошибку на 20 фрагментов.

Глава XII. Вычислительные аспекты нестационарной статистики

12.1. Генерация нестационарного временного ряда

Во многих задачах прикладного статистического анализа существует необходимость тестирования тех или иных индикаторов локального поведения временного ряда с целью оценки вероятности их правильного срабатывания. Как правило, индикаторы представляют собой функционалы от фрагментов траектории случайного процесса. Примеры индикаторов: отношение числа положительных приростов значений временного ряда к числу отрицательных за определенный период времени; угловой коэффициент прямой регрессии для выборки заданной длины; расстояние между выборочными плотностями функции распределения ряда в той или иной норме, дисперсия накопленного размаха за определенный период и т.п.

Чтобы оценить эмпирическую условную вероятность того, что определенный интервал значений индикатора отвечает ожидаемому исследователем поведению ряда в настоящем или будущем, нужно иметь много реализаций изучаемого процесса, тогда как в наличии имеется лишь одна фактически наблюденная траектория. На практике берется фрагмент траектории, который представляется достаточно большим, и на нем собирается требуемая статистика по индикатору: число ошибочных срабатываний, число ошибочных несрабатываний и число правильных срабатываний. Однако, если временной ряд нестационарный, то, например, оптимизация длины выборки для получения минимальной ошибки индикатора по некоторому фрагменту прошлой траектории не имеет особого смысла, поскольку оптимум искался для конкретного фрагмента траектории случайного процесса. При другой последовательности тех же самых значений временного ряда возможен иной результат оптимизации. Возникает задача тестирования индикатора на устойчивость относительно различных реализаций случайного процесса, имеющего близкие (с точки зрения исследователя — квазистационарные) выборочные распределения. Для этого требуется сгенерировать пучок возможных траекторий временного ряда, исходящий из заданного текущего состояния, и проверить на нем устойчивость срабатывания индикатора.

Если процесс стационарный, то набор его траекторий может быть получен, исходя из следующих соображений.

Рассмотрим равномерно распределенную на конечном отрезке $[a; b]$ случайную величину ξ . Она имеет ПФР

$$f_U(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases} \quad (1.1)$$

Распределение (1.1) бывает удобно задать в виде, когда явно выделен центр μ промежутка длины $\omega = |b - a|$:

$$f_U(x; \mu, \omega) = \begin{cases} \frac{1}{\omega}, & \mu - \frac{\omega}{2} \leq x \leq \mu + \frac{\omega}{2}, \\ 0, & x \notin [\mu - \omega/2, \mu + \omega/2]. \end{cases} \quad (1.2)$$

Среднее значение этого распределения равно μ , а дисперсия равна $\omega^2/12$.

Стандартизованным равномерным распределением является распределение на отрезке $[0; 1]$, ПФР которого в терминах (1.2) есть $f_U(x; 1/2, 1)$. Распределение (1.2) переводится в $f_U(y; 1/2, 1)$ посредством линейного преобразования $y = \frac{1}{2} + \frac{x - \mu}{\omega}$.

Равномерное распределение используется для того, чтобы определить, является ли выборка $\{x_1, x_2, \dots, x_n\}$ выборкой из распределения с некоторой предполагаемой функцией распределения $F(x)$. В основе различных статистических критериев согласия с гипотезой о виде распределения $F(x)$ лежит следующее утверждение. Пусть ξ есть случайная величина с непрерывной ФР $F(x)$. Тогда случайная величина $\eta = F(\xi)$ имеет ПФР $f_U(y; 1/2, 1)$. Действительно, поскольку

$$P(F(x) \leq y) = \begin{cases} 1, & y > 1 \\ y, & 0 < y \leq 1 \\ 0, & y \leq 0 \end{cases}$$

то плотность этой функции распределения является $f_U(y; 1/2, 1)$.

На основе этого утверждения строятся алгоритмы генерации временного ряда $\{x_k\}$ с заданной функцией распределения $F(x)$. Для этого генерируется произвольная последовательность чисел $\{y_k\}$, равномерно распределенных на $[0; 1]$, после чего по формуле

$$y_k = F(x_k), \quad x_k = F^{-1}(y_k) \quad (1.3)$$

находятся элементы ряда $\{x_k\}$. Обращение функции распределения в (1.3) возможно в силу ее строгой монотонности.

На практике условие непрерывности ФР формально нарушается, потому что наблюдаемые значения получены с конечной точностью измерения, но это не является принципиальным ограничением. Процесс считается непрерывным с той точностью, с которой проводятся измерения. Поэтому, несмотря на то, что гистограмма, дающая оценку плотности вероятности, формально отвечает некоторой дискретной случайной величине, функция распределения может быть сделана непрерывной, если внутри ячейки классового интервала гистограммы постулировать некоторое — например, равномерное — распределение.

Что касается генерации равномерно распределенного ряда на $[0; 1]$, то одним из наиболее употребляемых способов является алгоритм Лемера. Он состоит в том, что строится последовательность псевдослучайных чисел в соответствии с формулой

$$x_{n+1} = (ax_n + b) \bmod m, \quad m > x_0, m > a, m > b.$$

Длина периода такой последовательности равна m в том и только в том случае, если выполнены условия:

- 1) b, m взаимно просты;
- 2) если некоторое число c есть делитель m , то число $a - 1$ должно быть кратно c ;
- 3) если m кратно 4, то число $a - 1$ должно быть кратно 4.

Числа x_n / m равномерно распределены на $[0; 1]$. Псевдослучайность этих чисел обусловлена неизбежным заикливанием рекуррентной последовательности при численной реализации модели на компьютере с ограниченной длиной машинного слова.

Построим на основе этих утверждений численную модель нестационарного временного ряда в соответствии с определенными требованиями, накладываемыми на его ВПФР. Сложность анализа и, следовательно, моделирования нестационарного временного ряда с заданными статистическими свойствами его распределения состоит в том, что на практике в каждый момент времени наблюдается только одно значение из предполагаемой генеральной совокупности текущего распределения вероятностей. Поэтому с увеличением периода наблюдения выборочное распределение, построенное по увеличивающемуся количеству данных, не сходится к генеральной совокупности, из которой эти данные получены. Следовательно, анализу доступны только выборочные распределения, и задача анализа состоит не в установлении вида генеральной совокупности, что невозможно, а в нахождении эмпирических правил изменения ВПФР на определенном промежутке времени. Эти правила и представляют собой те свойства, которые могут быть приписаны выборочному распределению и протестированы на различных реализациях случайного процесса с близкими распределениями.

Из практических соображений удобно считать, что наша случайная величина равномерно ограничена по времени, так что все ее значения принадлежат отрезку $[0; 1]$. Пусть $x(t)$ есть значение случайной величины в дискретный момент времени t , где шаг по времени считается единичным, а $f_T(x, t)$ есть ВПФР выборки длины T с окончанием в момент времени t , т.е. выборки фрагмента ряда $\{x(t - T + 1), \dots, x(t)\}$. По ВПФР $f_T(x, t)$ можно построить непрерывную функцию распределения $F_T(x, t)$, считая распределение вероятностей в отдельном классовом интервале равномерным, равным частоте попадания в этот интервал наблюдаемых значений, деленной на длину интервала. Тогда

$$F_T(x, t) = \int_0^x f_T(y, t) dy. \quad (1.4)$$

Чтобы имитировать процесс, близкий к реальным наблюдениям, например, за динамикой цен на бирже на какой-либо финансовый инструмент (цена акции, индекс набора ценных бумаг, курс валюты и т.п.), предлагается следующая схема действий. На первом этапе по имеющимся историческим данным строятся выборочные распределения приростов $x(t)$ цен на этот инструмент за тот промежуток времени T , который представляет интерес. Например, строятся распределения приростов цен за

крытия в определенном временном интервале для какого-нибудь инструмента за два соседних месяца по скользящей выборке длиной в месяц. Тем самым в каждый момент времени t определена ВПФР $f_T(x, t)$, допустим, 5-минутных приростов цен, и соответствующая ей $F_T(x, t)$ согласно (1.4). Здесь длина T выражена в единицах 5-минутных интервалов. Затем генерируется стационарный равномерно распределенный на $[0; 1]$ ряд чисел $\{y_k\}$ длиной T . Пусть t_0 есть начальный момент времени, в который известна функция распределения за первый месяц наблюдения. Тогда в последующие моменты времени одна из возможных траекторий случайного процесса, для которого ВПФР меняется от $f_T(x, t_0)$ до $f_T(x, t_0 + T)$, строится по формуле обращения соответствующей локальной по времени функции распределения, движущейся в скользящем окне длины T :

$$y_k = F_T(x_k, t_0 + k). \quad (1.5)$$

Подчеркнем, что, согласно (1.5), в каждый момент времени t из распределения $F_T(x, t)$ генерируется только одно значение ряда. Сама же $F_T(x, t)$ выступает в этот момент времени как генеральная совокупность. Тем самым имитируется процесс наблюдения за динамикой нестационарного временного ряда.

Задавая различные равномерно распределенные ряды $\{y_k\}$, можно получить пучок траекторий, ассоциированных с двумя ВПФР: $f_T(x, t)$ и $f_T(x, t + T)$, согласно наблюдаемой эволюции этих распределений. Каждая из этих траекторий порождает на отрезке $[t_0; t_0 + T]$ ВПФР $\tilde{f}_T(\{y\}; x, t_0 + T)$, отличную, вообще говоря, от наблюдаемой $f_T(x, t_0 + T)$. По совокупности сгенерированных траекторий можно оценить, насколько значимо отклонение фактических распределений в норме непрерывных функций:

$$\rho(F, F') = \|F - F'\| = \sup_x |F(x, t) - F'(x, t')|. \quad (1.6)$$

На рис. 12.1.1 приведен пучок траекторий, сгенерированных описанным методом для ряда приростов цен закрытия 5-минутных интервалов для нефтяных фьючерсов WTI по выборке длиной в 1 месяц, контракты

2010 года. Жирной линией выделена фактически состоявшаяся траектория. Для удобства вычислений цены были нормированы на отрезок $[0; 1]$, так что приросты этих нормированных цен лежат в отрезке $[-1; 2]$. В начальный момент времени t_0 , отмеченный на рис. 12.1.1 левой вертикальной пунктирной линией, из точки $x(t_0)$ выпускается пучок траекторий в соответствии с распределением, которое известно к моменту t_0 по выборке длины T , как описано выше. Затем в момент $t_0 + 1$ берется следующее распределение по выборке такой же длины, но сдвинутой на один шаг вперед, и так далее.

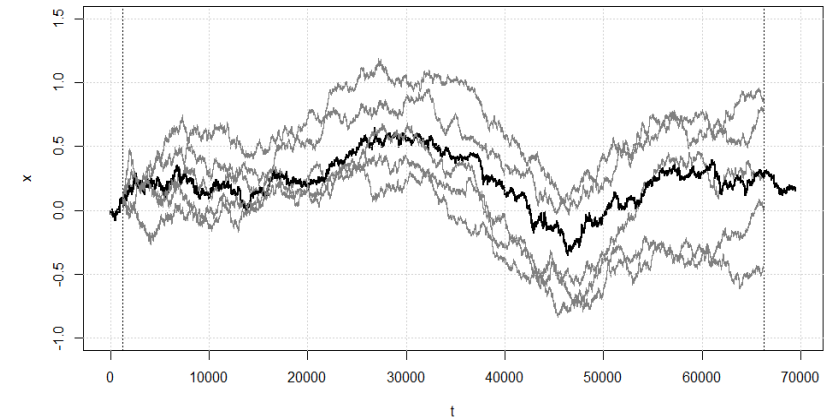


Рис. 12.1.1 — Пучок траекторий нестационарного процесса

Этот пример показывает, как моделируются траектории нестационарного случайного процесса, реализации которого образуют ВПФР, близкую к заданной. Для него оказалось, что согласованное расстояние между сгенерированными распределениями по выборкам длины T на промежутке $[t_0, t_0 + T]$, равно $\rho^* = 0,07$, тогда как согласованное расстояние между фактическими распределениями, разделенными временным промежутком длины T , равно $\rho_{fact} = 0,50$. Близкой к величине ρ^* оказалось и согласованное расстояние между сгенерированными выбор-

ками и фактически наблюдаемой ВПФР $f_T(x, t_0 + T)$: оно равно $\rho_{fact}^* = 0,06$. Заметим, что если бы ряд был стационарным, то согласованное расстояние между встык-выборками в соответствии с критерием согласования (5.1.9), примененным к функции Колмогорова, было бы равно 0,02, а не 0,50, как в данном примере. Поэтому можно считать, что эксперимент генерации нестационарного ряда в заданном классе распределений проведен корректно: сгенерированные траектории имеют уровень стационарности, значимо меньший, чем расстояние между встык-выборками, и в то же время больший, чем для стационарного процесса. Это служит указанием на то, что отклонения между фактическими распределениями, разделенными временным промежутком T , скорее всего, вызваны именно нестационарностью, а не статистическими флуктуациями.

Таким образом, можно построить численный алгоритм моделирования нестационарного временного ряда с определенными непараметрическими свойствами его ВПФР. Эти свойства выражаются в том, что определено некоторое эмпирическое кинетическое уравнение — Лиувилля, Фоккера-Планка или иное модельное уравнение, по которому предположительно эволюционирует ВПФР рассматриваемого ряда. Пусть, в частности, эволюция определяется уравнением Лиувилля

$$\frac{\partial f}{\partial t} + \frac{\partial(uf)}{\partial x} = 0.$$

Выбираем равномерное разбиение гистограммы на классовые интервалы и считаем шаг по времени единичным. Тогда, опуская далее у ВПФР указание на длину выборки и обозначая через $f(i, t)$ значение ВПФР в i -ом классовом интервале, получаем для ЭЛС выражение

$$u(i+1, t) = \frac{u(i, t)f(i, t) + f(i, t) - f(i, t+1)}{f(i+1, t)}.$$

Обозначим

$$F(i, t) = \sum_{k=1}^i f(k, t).$$

Тогда

$$u(1, t) = 0; \quad u(i+1, t) = \frac{F(i, t) - F(i, t+1)}{f(i+1, t)}. \quad (1.7)$$

Единичным шагом естественно считать шаг по времени, с которым наблюдаются значения временного ряда. Зафиксировав длину выборки T в терминах этих единичных шагов, можно построить изменение ВПФР при смещении по времени на длину окна выборки. Из формулы (1.7) следует, что скорость $U(i+1, t)$ изменения ВПФР при смещении на T шагов равна умноженной на T средневзвешенной скорости смещения на 1 шаг:

$$\begin{aligned} U(i+1, t) &= \frac{F(i, t) - F(i, t+T)}{f(i+1, t)} = \\ &= \frac{1}{f(i+1, t)} \sum_{k=0}^{T-1} u(i+1, t+k) f(i+1, t+k). \end{aligned} \quad (1.8)$$

Формула (1.8) показывает, как можно построить траекторию временного ряда с заданным профилем скорости изменения ВПФР в классе \mathcal{E} -стационарных распределений, т.е. таких, для которых при сдвиге на ширину окна выборки выполняется условие

$$\sum_{i=1}^n |f(i, t+T) - f(i, t)| \leq \varepsilon, \quad (1.9)$$

где n есть число классовых интервалов гистограммы, а ε представляет максимум среди наблюдаемых попарных расстояний между распределениями по непересекающимся выборкам длины T . Например, чтобы условие (1.9) было заведомо выполнено, можно положить скорость изменения ВПФР на шаге длины 1 равной

$$u(i+1, t+k) = \frac{U(i+1, t)}{(1+\varepsilon)T}, \quad k = 0, 1, \dots, T-1. \quad (1.10)$$

Если требуется смоделировать изменение вероятности срабатывания индикатора в определенной области значений случайной величины, то для этого можно провести «модуляцию» скорости в (1.10) с помощью произвольной подходящей функции $\lambda(i, t)$, т.е. рассматривать вместо $u(i, t+k)$ величину $v(i, t+k) = u(i, t+k)\lambda(i, t+k)$. Величина $\lambda(i, t)$ должна быть такой, чтобы на следующем шаге по времени расчетная вероятность, получаемая из (1.7) по формуле $f(i, t+1) = f(i, t)(1+v(i, t)) - f(i+1, t)v(i+1, t)$, $v = \lambda u$. (1.11)

была бы неотрицательной.

Тем самым на всем интервале времени длиной T можно смоделировать промежуточные распределения, являющиеся решениями модельного уравнения Лиувилля. Более того, если есть модель изменения эмпирической скорости, то можно построить и пучок прогнозных траекторий будущего случайного нестационарного процесса.

Задачу генерации траектории временного ряда можно решить и другим способом, основанным на ассоциировании с уравнением Лиувилля соответствующей динамической системы. Как следует из результатов главы IX, эмпирическая скорость, фигурирующая в уравнении Лиувилля, может быть естественным образом использована для нахождения среднего значения приращений ряда, по которому и было построено распределение:

$$\bar{x}(t) = \frac{1}{T} \sum_{i=t-T+1}^t x(i); \quad \frac{d\bar{x}(t)}{dt} = \int_0^1 u_T(x, t) f_T(x, t) dx. \quad (1.12)$$

Из (1.12) видно, что $u_T(x, t)$ можно трактовать как некоторую микроскопическую скорость движения «центра тяжести» временного ряда. Рассмотрим неавтономную динамическую систему, порожденную эмпирической скоростью $u_T(x, t)$:

$$\frac{dz}{dt} \equiv z(t+1) - z(t) = u_T(z(t), t). \quad (1.13)$$

Переменная $z(t)$ введена для того, чтобы отличать динамически сгенерированные значения от фактически наблюдаемых $x(t)$. Функция $u_T(x, t)$ может быть взята, например, из уравнения (1.10) или из его модификации в духе того, как это сделано в (1.11). При различных начальных условиях динамическая система (1.13) дает набор траекторий, среднее значение ряда по которым эволюционирует по заданному закону (1.12). Следовательно, выборочное распределение ряда $z(t)$ будет принадлежать к классу распределений средних выборочных значений исходного ряда $x(t)$.

Динамика среднего значения ряда и вероятность найти это значение в определенном промежутке представляют большую практическую важность при анализе ценовых рядов. Однако классические методы интервальных оценок для среднего или дисперсии основаны на стационарных критериях и не позволяют делать оценки на табулированном уровне зна-

чимости, если ряд не стационарный. В таком случае можно сгенерировать набор траекторий среднего значения, используя динамическую систему (1.13). Разумеется, траекторию среднего значения можно получить и скользящим усреднением по траектории случайного процесса, сгенерированной в соответствии с представлениями об эволюции ВПФР. Но если ставится задача только об отыскании коридора изменения среднего значения, то нет необходимости генерировать семейство случайных траекторий. Достаточно знать только эмпирическую скорость, удовлетворяющую уравнению Лиувилля.

Усреднив полученные траектории движения среднего значения и/или функционалы на них по начальному распределению $f_T(x, t_0)$, можно более корректно определить коридор изменения скользящих средних и аналогичных характеристик. Пусть, например, $\varphi(z, x, t_0 + T)$ есть ВПФР ряда, сгенерированного по системе (1.13) для начального значения x при $t = t_0$. Тогда средняя ВПФР определяется формулой

$$\bar{\varphi}(z, t_0 + T) = \int_0^1 \varphi(z, x, t_0 + T) f_T(x, t_0) dx. \quad (1.14)$$

Если отдельные траектории $z(t)$ близки в смысле уровня квазистационарности, то их отличие от средней траектории также не будет превосходить ε . Разбегание траекторий, т.е. изменение ширины коридора, в котором следует искать среднее значение исходного ряда $x(t)$, определяется показателем Ляпунова динамической системы (1.13). Напомним (см. главу III), что если известна динамически-инвариантная мера $\varphi_{inv}(z)$, то показатель Ляпунова для динамической системы, задаваемой рекуррентно формулой $z(n+1) = g(z(n))$, определяется как

$$\Lambda = \int_0^1 \varphi_{inv}(z) \ln |dg/dz| dz. \quad (1.15)$$

Тогда с точностью ε в качестве инвариантной меры можно взять $\varphi_{inv}(z) = \bar{\varphi}(z, t_0 + T)$. Если изменения скоростей $u_T(x, t)$ при переходе между классовыми интервалами малы по сравнению с единицей, находим из (1.13) и (1.15), что

$$\Lambda \approx \int_0^1 \bar{\varphi}(z, t_0 + T) \frac{\partial u_T(z, t_0)}{\partial z} dz. \quad (1.16)$$

На рис. 12.1.2 на фоне фактически состоявшегося ряда $x(t)$ приведены траектории его средних выборочных значений, полученные описанным здесь динамическим методом. Показатель Ляпунова для них имеет величину $\Lambda = 0,002$.

Из (1.16) следует, что при конечной точности в измерении среднего значения близкие траектории через конечное время разойдутся на расстояние порядка размера носителя распределения. Для рассматриваемого примера характерным промежутком времени является 2000 шагов. Тем самым динамическая система (1.13), несмотря на детерминированность, имеет конечный горизонт прогнозирования.

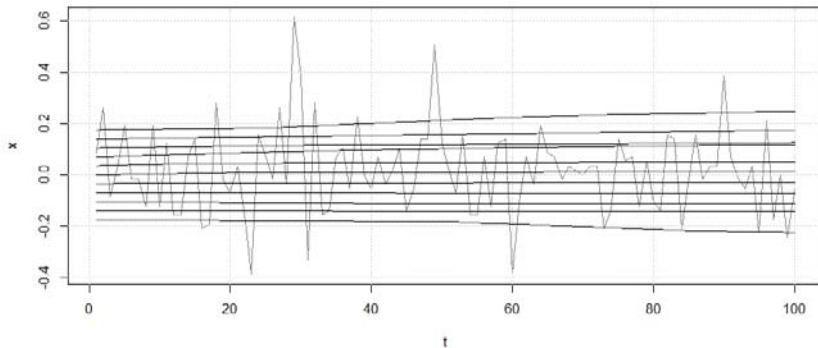


Рис. 12.1.2 — Пучок траекторий среднего выборочного значения нестационарного временного ряда

Итак, описан метод генерации траекторий нестационарных временных рядов, имитирующих временной ряд, возникающий на практике в процессе наблюдения случайной величины, распределение которой не является стационарным. Моделирование ряда основано на решении эмпирического кинетического уравнения, что позволяет задавать скользящее среднее, дисперсию, размах и аналогичные величины изменяющимися во времени в соответствии с требованиями статистического эксперимента, причем выборочные распределения остаются в классе распределений,

характерных для данного ряда. Подчеркнем, что класс распределений выделяется непараметрическими методами.

Метод позволяет тестировать индикаторы-предикторы изменения какого-либо свойства временного ряда в широком диапазоне изменения его средних выборочных характеристик. На практике для такого тестирования используются исторические данные за достаточно большой промежуток времени, но надо помнить, что при этом берется единственная реализация случайного процесса. Можно предположить, что такой промежуток действительно содержит все нужные для теста ситуации, и одной прогонки по длинной истории будет достаточно. По этому поводу следует заметить, что тест на прошлых данных почти никогда не подтверждается в будущем с требуемой точностью. Это связано в первую очередь с тем, что в будущем реализуется произвольная траектория из ансамбля возможных путей, а в прошлом есть только одна. Этой одной траектории достаточно, если только процесс эргодический. Поскольку это последнее требование не выполняется (во всяком случае на финансовых рынках), то тестирования в прошлом всегда недостаточно.

Разумеется, описанный в работе метод не моделирует сам финансовый инструмент, поскольку, возможно, в действительной траектории содержатся скрытые статистические зависимости, тогда как мы моделируем процесс с независимыми приращениями. Основанием для такой модели служит то обстоятельство, что большинство ценовых рядов таковы, что их автокорреляция близка к нулю. Это, конечно, не доказывает статистическую независимость приращений, но может служить удобным приближением. К преимуществам метода следует отнести то, что он позволяет провести стресс-тест на работоспособность индикатора в пределах, контролируемых исследователем. Исторический же ряд данных не предоставляет таких возможностей. Кроме того, для квалифицированного тестирования ряд прошлых данных требует предварительного выявления интересных ситуаций, кластеризации их, определения ошибок при кластеризации, что весьма трудоемко и не дает полного представления об имеющихся локальных паттернах ряда. Таким образом, численный код, генерирующий по фрагменту траектории нестационарного ряда ансамбль его нестационарных же реализаций, представляет практическую важность.

12.2. Численное моделирование фрактального блуждания

Наряду с моделированием «обычных» нестационарных случайных процессов большую важность представляют методы генерации процессов переноса во фрактальных структурах сложной конфигурации. Эти про-

цессы связаны с уравнениями в производных дробного порядка и стохастическими процессами Леви, которые используются в приложениях для описания так называемой «странной кинетики», когда скорость расплывания диффузионного пакета идет быстрее или, наоборот, медленнее, чем в случае классического случайного блуждания. В связи с этим задача построения случайного блуждания на множествах с фрактальной структурой весьма актуальна. В большинстве случаев для этой цели предлагается рассмотрение случайных блужданий на графах, порождаемых соответствующим фрактальным множеством.

В этом разделе описывается более простой, но вместе с тем весьма эффективный метод, опирающийся на использование случайных последовательностей символов над некоторым алфавитом. Рассмотрение будет проведено для обобщенного множества Кантора C_λ . Развитый подход может быть расширен и на другие фрактальные множества.

Простейшим способом определения обобщенного множества Кантора C_λ является следующая геометрическая итерационная процедура: из замкнутого отрезка $[0; 1]$ исключается открытый интервал длины λ , составляющий среднюю часть исходного отрезка; затем аналогичная операция применяется к каждому из оставшихся замкнутых отрезков. Таким образом, длина каждого из 2^{n-1} удаляемых на n -ой итерации интервалов равна β^{n-1} , где $\beta = (1 - \lambda)/2$, так что на n -ом шаге удаляются интервалы совокупной длины l_n , равной

$$l_n = \lambda(2\beta)^{n-1}, \quad \beta = \frac{1-\lambda}{2}. \quad (2.1)$$

Суммарная длина этих интервалов, как легко видеть, равна единице, поэтому мера Лебега множества C_λ равна нулю. Размерность Хаусдорфа такого множества определяется выражением $-\log_\beta 2$, поскольку C_λ является объединением двух подобных ему множеств с масштабным коэффициентом β и может принимать любые значения между 0 и 1. Точки обобщенного множества Кантора могут быть представлены в виде

$$x = (1 - \beta) \sum_{k=1}^{\infty} b_k \beta^k, \quad b_k \in \{0, 1\}. \quad (2.2)$$

Одним из наиболее известных примеров множеств вида C_λ является тернарное множество Кантора с параметрами $\lambda = \beta = 1/3$, которое может быть описано в явном виде как

$$C_{1/3} = [0; 1] \setminus \bigcup_{m=1}^{\infty} \bigcup_{k=0}^{3^{m-1}-1} \left[\frac{3k+1}{3^m}; \frac{3k+2}{3^m} \right]. \quad (2.3)$$

Точки тернарного множества Кантора, как следует из (2.3), при записи в троичной системе счисления не содержат цифры 1. При этом следует учитывать, что любое троично-рациональное число (знаменатель которого является степенью 3) имеет несколько представлений: например, число $1/3$ может быть представлено как $0,1(0)_3$ и $0,0(2)_3$. Множество Смита-Вольтерра-Кантора, называемое также \mathcal{E} -множеством Кантора, имеющее структуру, весьма близкую к C_λ , может быть получено с помощью аналогичной итерационной процедуры с той лишь разницей, что на n -ой итерации из каждого замкнутого множества исключается открытая средняя часть длины 2^{-2n} . Мера Лебега получающегося множества в этом случае положительна и равна $1/2$, а размерность Хаусдорфа равна 1.

При построении множества Кантора можно также отказаться от требования симметрии, исключив, например, из отрезка $[0; 1]$ интервал $]1/4; 1/2[$, после чего аналогичное правило применяется к каждому из оставшихся замкнутых множеств.

Представление (2.2) устанавливает соответствие между точками обобщенного множества Кантора и бинарными последовательностями, что позволяет использовать последние для построения случайного блуждания на таком множестве. Будем рассматривать случайный процесс c_t на канторовом множестве с дискретным временем $t \in N$, представляемый в виде

$$c_t = (1 - \beta) \sum_{k=0}^{\infty} \xi_t^{(k)} \beta^k, \quad (2.4)$$

где $\xi_t^{(k)}$ — дискретные случайные величины с распределением Бернулли (принимающие лишь два значения — 0 и 1).

Один из простейших вариантов случайного блуждания можно получить, считая, что $\xi_t^{(k)}$ при любом фиксированном k образуют однородную цепь Маркова, и при любом фиксированном t случайные величины $\xi_t^{(k_1)}$ и $\xi_t^{(k_2)}$ независимы. Не ограничивая общности, будем считать, что рассматриваемый случайный процесс выходит из точки 0, т.е. $\forall k \ P(\xi_1^{(k)} = 0) = 1$. Будем также считать, что вероятности переходов не зависят от k . Вероятности состояний обозначим следующим образом:

$$P(\xi_t^{(k)} = 0) = u_t, \quad P(\xi_t^{(k)} = 1) = w_t, \quad u_t + w_t = 1. \quad (2.5)$$

Для вероятностей u_t и w_t получаем систему линейных разностных уравнений

$$\begin{pmatrix} u_{t+1} \\ w_{t+1} \end{pmatrix} = \Pi^T \begin{pmatrix} u_t \\ w_t \end{pmatrix}, \quad \Pi = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}. \quad (2.6)$$

Если $p_{00} = p_{11} = 1$, то решением, очевидно, будет $u_t = 1, w_t = 0$ и $c_t \equiv 0$. В противном случае решение системы имеет вид

$$\begin{aligned} u_t &= \frac{p_{01}}{p_{01} + p_{10}} + \frac{p_{11}}{p_{01} + p_{10}} (\text{Tr} \Pi - 1)^t, \\ w_t &= \frac{p_{01}}{p_{01} + p_{10}} - \frac{p_{11}}{p_{01} + p_{10}} (\text{Tr} \Pi - 1)^t. \end{aligned} \quad (2.7)$$

Таким образом, если $\text{Tr} \Pi = 1$, то процесс c_t является строго стационарным.

Статистические характеристики случайной последовательности $\xi_t^{(k)}$ могут быть легко найдены:

$$M \xi_t^{(k)} = w_t, \quad D \xi_t^{(k)} = u_t w_t. \quad (2.8)$$

Совместное распределение $\xi_t^{(k)}$ и $\xi_s^{(k)}$ при фиксированных t и s таких, что $|t - s| = 1$, имеет вид

$$\begin{aligned} P(\xi_t^{(k)} = 0, \xi_{t+1}^{(k)} = 0) &= u_t p_{00}, \quad P(\xi_t^{(k)} = 0, \xi_{t+1}^{(k)} = 1) = u_t p_{01}; \\ P(\xi_t^{(k)} = 1, \xi_{t+1}^{(k)} = 0) &= w_t p_{10}, \quad P(\xi_t^{(k)} = 1, \xi_{t+1}^{(k)} = 1) = w_t p_{11}. \end{aligned} \quad (2.9)$$

Из (5.9) следует, что

$$M \xi_t^{(k)} \xi_{t+1}^{(k)} = w_t p_{11}. \quad (2.10)$$

Математическое ожидание и дисперсия процесса c_t даются следующими выражениями:

$$\begin{aligned} E c_t &= (1 - \beta) \sum_{k=0}^{\infty} E \xi_t^{(k)} \beta^k = w_t, \\ D c_t &= (1 - \beta)^2 \sum_{k=0}^{\infty} D \xi_t^{(k)} \beta^{2k} = u_t w_t \frac{1 - \beta}{1 + \beta}. \end{aligned} \quad (2.11)$$

Выражение для автоковариационной функции процесса имеет вид

$$K(t, s) = \begin{cases} p_{11} w_{\min\{t, s\}} \frac{1 - \beta}{1 + \beta} - w_t w_s, & |t - s| = 1; \\ u_t w_t \frac{1 - \beta}{1 + \beta}, & |t - s| = 0; \\ 0, & |t - s| > 1. \end{cases} \quad (2.12)$$

Для того чтобы получить траектории случайного процесса c_t , в сумме (2.2) необходимо удержать лишь конечное число n слагаемых. Разумеется, построенная таким образом траектория будет лишь приближением к истинной траектории c_t ; тем не менее, точностью аппроксимации можно управлять, задавая достаточно большим n .

Естественным обобщением описанной конструкции является случайное блуждание, где $\xi_t^{(k)}$ образуют неоднородную цепь Маркова, т.е. вероятности переходов зависят от времени t . Для простоты будем по-прежнему считать, что переходные вероятности не зависят от k . Но даже и в этом случае получить выражения вероятностей состояний при произвольном t , подобные (5.7), в общем случае весьма затруднительно.

Среди всевозможных видов функциональной зависимости переходных вероятностей от времени интерес представляют периодические $\theta(t, s; L)$ и «триггерные» $\tau(t, s; L)$ зависимости, которые могут быть описаны, например, следующими семействами функций:

$$\begin{aligned}\theta(t, s; L) &= \frac{1}{2} \left(\sin \left(\frac{2\pi}{L} (t-s) \right) + 1 \right); \\ \tau(t, s; L) &= \frac{1}{2} \left(\operatorname{erf} \left(\frac{t-s}{L\sqrt{2}} \right) + 1 \right).\end{aligned}\quad (2.13)$$

В первом случае траектория процесса обладает выраженной периодичностью, когда области «притяжения» траектории к 0 или 1 перемежаются с областями резких осцилляций. Во втором случае после серии случайных колебаний траектория в итоге «устанавливается» на некотором постоянном уровне.

В рассмотренную выше конструкцию случайного процесса можно также включить зависимость между случайными величинами $\xi_t^{(k)}$ при различных k . Например, значение величины $\xi_t^{(k)}$ может быть связано не только с ее предыдущим значением $\xi_{t-1}^{(k)}$, но и с предыдущими значениями $\xi_{t-1}^{(k\pm 1)}$.

При фиксированном t последовательность $\xi_t^{(k)}$ образует цепь Маркова (в общем случае неоднородную), имеющую с учетом всех возможных контекстов k -ого символа в бинарной последовательности восемь состояний. Чтобы правила перехода применялись одинаково ко всем символам бинарной последовательности, можно считать, что последовательности $\xi_t^{(l)}$ предшествует последовательность $\xi_{t-1}^{(k)}$.

Как уже отмечалось выше, существует сюръективное отображение множества бесконечных бинарных последовательностей на обобщенное множество Кантора; при этом элементы этих последовательностей являются коэффициентами в представлении (5.4). Однако возможна другая интерпретация бинарной последовательности, которая также позволяет описать множество Кантора. Рассмотрим отображения

$$F_0(x) = \beta x, \quad F_1(x) = F_0(x) + 1 - \beta = 1 - \beta(1-x). \quad (2.14)$$

Очевидно, они являются сжимающими в метрическом полном пространстве \mathbf{R} . Множество Кантора является единственным решением уравнения

$$C_\lambda = F_0(C_\lambda) \cup F_1(C_\lambda) \quad (2.15)$$

или, в эквивалентной формулировке

$$C_\lambda = \bigcap_{n=0}^{\infty} I_n, \quad I_n = F_0(I_{n-1}) \cup F_1(I_{n-1}), \quad I_0 = [0; 1]. \quad (2.16)$$

Следовательно, если задана некоторая бинарная последовательность $b_k \Big|_1^n$ и величина $x \in C_\lambda$, то результат n -кратного применения отображения (2.14) также принадлежит C_λ :

$$H_n(x) \in C_\lambda, \quad H_n = F_{b_n} \circ F_{b_{n-1}} \circ \dots \circ F_{b_1}.$$

Таким образом, случайное блуждание на C_λ можно описать в терминах случайных бинарных последовательностей, определяющих порядок применения отображений (2.14) к произвольной точке $x \in C_\lambda$ (например, к точке $x = 0$). Построение обобщенного множества Кантора с помощью системы сжимающих отображений является примером использования метода IFS (iterated function system), который, с некоторыми вариациями, применяется для создания различных фрактальных структур.

Рассмотренные стационарные и нестационарные модели блужданий по «разреженным» множествам полезны тем, что они представляют собой набор эталонов, свойства которых известны, и на которых следует тестировать различные методы идентификации эмпирических случайных процессов, относительно которых может быть выдвинута гипотеза о реализации фрактальной диффузии.

12.3. Возникновение искусственной корреляции

При численном моделировании временного ряда как последовательности независимых случайных событий важно соблюсти некоррелированность выборок, если указанная независимость лежала в основе создания модели. В более общем случае, если требовалось смоделировать ряд с определенной автокорреляционной связью, то такая связь должна быть соблюдена в процессе реализации численного алгоритма.

Проблема возникает из-за того, что фактически при численном моделировании используются пусть и очень большие, но ограниченные целые числа, а не бесконечные десятичные дроби, представляющие действительные числа. Поэтому может возникнуть ситуация, когда некоторое преобразование независимых случайных величин оставляет их независимыми в области действительных чисел, но создает зависимость в целых числах. Например, пусть ξ и η — пара независимых случайных величин. Тогда преобразование $\xi' = \xi + \eta$, $\eta' = \xi - \eta$ оставляет новую пару величин независимой, но на множестве целых чисел сумма и разность двух чисел либо обе четны, либо обе нечетны, поэтому новые случайные величины оказываются зависимыми, хотя корреляция между ними остается равной нулю. Описанная ситуация относится к стационарному случаю. Если же ряды нестационарны, то корреляционный эффект может иметь искусственное происхождение в том смысле, что зависимость между рядами носит не причинно-следственный характер, а обусловлена неизвестным третьим фактором, влияющим на оба ряда.

Поскольку коэффициент корреляции (или автокорреляции) не является линейным функционалом от распределения случайной величины, то сходимость выборочного коэффициента корреляции к генеральному при увеличении объема выборки отличается от сходимости выборочных моментов.

Если ряды стационарны, то относительно коэффициента корреляции чаще всего рассматриваются две основных задачи: проверка гипотезы о равенстве нулю генерального коэффициента корреляции ρ , а также определение доверительного интервала, накрывающего коэффициент корреляции на заданном уровне значимости.

Для первой задачи используется статистика

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad (3.1)$$

имеющая распределение Стьюдента с $n-2$ степенями свободы, где r есть выборочный коэффициент корреляции, а n — длина выборки. Тогда генеральный коэффициент корреляции отличен от нуля на уровне значимости α , если

$|t| > t_{1-\alpha}(n-2)$. Заметим, что вывод о неравенстве нулю генерального коэффициента корреляции делается на весьма малых уровнях значимости уже для выборок сравнительно ма-

лых длин, поскольку для такого вывода достаточно, чтобы выборочный коэффициент корреляции лишь незначительно был бы отличен от нуля:

$$r^2 > \frac{1}{1 + \frac{n-2}{t_{1-\alpha}^2(n-2)}} \approx \frac{1}{1 - \frac{2n}{\pi} \ln(1 - (1-2\alpha)^2)}. \quad (3.2)$$

Для решения второй задачи используется тот факт, что случайная величина

$$z = \frac{1}{2} \ln \frac{1+r}{1-r} \quad (3.3)$$

распределена асимптотически нормально со средним μ_z и дисперсией σ_z^2 , равными

$$\mu_z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}, \quad \sigma_z^2 = \frac{1}{\sqrt{n-3}}. \quad (3.4)$$

Тогда доверительный интервал для величины μ_z на уровне значимости α есть

$$z - \frac{u_{1-\alpha}}{\sqrt{n-3}} \leq \mu_z \leq z + \frac{u_{1-\alpha}}{\sqrt{n-3}}, \quad (3.5)$$

где $u_{1-\alpha}$ есть соответствующий квантиль нормального распределения.

Использовать стандартные методы оценки значимости коэффициента корреляции в нестационарном случае нецелесообразно. Как уже говорилось в п. 8.4, это связано с тем, что выборочный коэффициент корреляции сильно и немонотонно зависит как от времени, так и от объема выборки. Для тестирования модельных временных рядов на адекватность условиям задачи следует воспользоваться индексом нестационарности (см. п. 5.5). Если на интересующих исследователя объемах выборки этот индекс заметно больше единицы, то условие нестационарности доминирует над возможной сгенерированной корреляцией. Если же индекс меньше единицы, то сгенерированный временной ряд имеет черты стационарного ряда с зависимыми приращениями, и такой алгоритм численной генерации не может считаться корректным.

12.4. Возникновение отрицательных вероятностей

При численном решении уравнения Лиувилля для ВПФР может появиться такой вычислительный артефакт, как отрицательные вероятности в отдельных классовых интервалах. При этом нормировка новой ВПФР, полученной в результате такого решения, по-прежнему остается равной единице, что сразу следует из нулевых граничных условий для потока вероятности. Действительно, в соответствии с разностной схемой

$$\hat{f}(x, t+1) = f(x, t) + f(x, t)\hat{u}(x, t) - f(x+1, t)\hat{u}(x+1, t)$$

нормировка прогнозного значения ВПФР на следующем временном слое равна

$$\begin{aligned} Z(t+1) &= \sum_{x=1}^n \hat{f}(x, t+1) = \\ &= \sum_{x=1}^n f(x, t) + \sum_{x=1}^n f(x, t)\hat{u}(x, t) - \sum_{x=1}^{n-1} f(x+1, t)\hat{u}(x+1, t) = \quad (4.1) \\ &= Z(t) + f(1, t)\hat{u}(1, t) - f(n, t)\hat{u}(n, t) = 1 \end{aligned}$$

в силу того, что $Z(t) = \sum_{x=1}^n f(x, t) = 1$, а значение произведения

$f\hat{u}$ на границе области равно нулю.

Рассмотрим интегральную функцию распределения

$$F(x, t) = \sum_{k=1}^x f(k, t).$$

Ее приращение в дискретной схеме решения уравнения Лиувилля равно

$$\begin{aligned} \Delta F|_{t+1} &= F(x+1, t+1) - F(x, t+1) = \\ &= F(x+1, t) - F(x, t) + f(x+1, t)\hat{u}(x+1, t) - \\ &\quad - f(x+2, t)\hat{u}(x+2, t) = \Delta F|_t - \Delta(f\hat{u}). \end{aligned} \quad (4.2)$$

Таким образом, если ФР в момент времени t была монотонна, т.е. $\Delta F|_t > 0$, то в следующий момент времени это неравенство может и перестать выполняться.

Следует заметить, что решение уравнения Лиувилля в дифференциальной форме неотрицательно, поскольку может быть записано в виде действия оператора потока $S(t, t_0) = \exp(-(t - t_0)\hat{L})$, где \hat{L} есть дифференциальный оператор по фазовым переменным в уравнении Лиувилля, на ПФР, заданную в момент времени t_0 . Поэтому, если ПФР была неотрицательна в начальный момент времени, то она будет таковой и в последующие моменты.

Немонотонность интегральной функции распределения при численных расчетах может возникнуть по трем причинам.

Во-первых, дискретная схема какого бы то ни было конечного порядка точности не равномерно во всей области изменения фазовых переменных и времени аппроксимирует положительную экспоненту оператора потока. Полином же с чередующимися по знаку коэффициентами может, вообще говоря, принимать и отрицательные значения.

Во-вторых, может оказаться, что для данного конкретного ряда ВПФР чувствительна к численной схеме расчета. Это означает, что, например, изменение мелкости разбиения интервала значений временного ряда в допустимых пределах, определяемых точностью построения ВПФР, с одной стороны, и точностью измерения значений ряда $x(t)$, с другой, приведет к исчезновению эффекта. Однако напомним, что мелкость шага разбиения ограничивается снизу точностью измерения случайных величин и точностью, предъявляемой к ВПФР, поэтому уменьшение шага разбиения можно проводить только до определенных пределов.

В-третьих, эмпирическая скорость $u(x, t)$ корректно согласуется с ВПФР только постфактум, а в соответствии с (4.2) любое ее прогнозное значение может привести к качественно неверным результатам.

Поскольку решение ищется с точностью порядка $O(\varepsilon)$, то важно, насколько велики эти отрицательные вероятности. Если сумма абсолютных величин отрицательных вероятностей оказалась величиной порядка $O(\varepsilon)$ (на практике это проявляется в том, что указанная сумма не превосходит ε^2), то модель можно считать адекватной, а соответствующие вероятности, например, занулить.

Вычислив затем интеграл от такой скорректированной функции, следует провести ее перенормировку, разделив, например, каждое значение ВПФР на полученное значение интеграла. Разумеется, коррекцию можно проводить и другими способами, уменьшив не все значения распределения, а только те, в которых был получен наибольший рост, или те, кото-

рые составляют ε -квантиль в окрестности отрицательной вероятности и т.п.

Если же отрицательные вероятности оказались велики — порядка ε или больше, то используемая модель не является удовлетворительной, и требуется либо увеличение размерности ВПФР, либо вообще использование другого, не кинетического подхода.

Предположим, что отрицательные расчетные вероятности оказались достаточно малыми, так что их удаление можно считать действием, имеющим порядок $o(\varepsilon)$. Возникает вопрос: ошибку какой величины можно ожидать для прогнозной ВПФР, если модель прогнозирования использует эмпирическое уравнение Лиувилля?

Рассмотрим прогноз на один шаг по времени и оценим величину ошибки $\int |\hat{f}(x, t+1) - f(x, t+1)| dx$, где $\hat{f}(x, t+1)$ определяется по уравнению (9.1.2), а $f(x, t+1)$ есть фактическое распределение в следующий момент времени. Поскольку объем выборки таков, что $\int |f(x, t) - f(x, t+1)| dx \leq \varepsilon$, то

$$\begin{aligned} & \int |\hat{f}(x, t+1) - f(x, t+1)| dx = \\ & = \int \left| f(x, t+1) - f(x, t) + \frac{\partial}{\partial x} \hat{u}(x, t) f(x, t) \right| dx \leq \\ & \leq \int |f(x, t+1) - f(x, t)| dx + \int \left| \frac{\partial(\hat{u}f)}{\partial x} \right| dx \leq \varepsilon + \int \left| \frac{\partial(\hat{u}f)}{\partial x} \right| dx. \end{aligned} \quad (4.3)$$

Следовательно, любой другой прогноз, кроме наивного, заведомо имеет верхнюю грань оценки своей ошибки больше, чем требуемая точность ε . В отдельных случаях может оказаться, что

$$\int \left| \frac{\partial(\hat{u}f)}{\partial x} \right| dx = o(\varepsilon). \quad (4.4)$$

Это и есть условие применимости кинетической модели прогнозирования. Заметим, что, поскольку это поправка к верхней грани оценки ошибки прогноза, то вполне возможно, что отдельные прогнозы или их серия будут иметь ошибку, заметно меньшую, чем эта грань. Именно для таких случаев кинетическая модель является эффективной.

12.5. Численные схемы решения уравнения переноса

Численное решение уравнения переноса относительно ВПФР требует аккуратности в выборе разностной схемы, так как решение задачи Коши для одномерного уравнения Лиувилля $\partial f / \partial t + \partial(uf) / \partial x = 0$, $f|_{t=0} = f_0(x)$ даже при постоянной скорости u заметно отличается от точного решения $f(x, t) = f_0(x - ut)$.

Пусть τ есть шаг по временной сетке, а h — по пространственной, причем обе сетки для удобства равномерны. Число $r = u\tau/h$ называется числом Куранта, его величина определяет устойчивость схемы [30].

Приведем несколько типовых разностных шаблонов для решения уравнения переноса с постоянной скоростью. Традиционно обозначим сеточные величины, индексируя сверху слой по времени, а снизу — номер пространственного узла. Положим

$$f = f_i^j; \quad \hat{f} = f_i^{j+1}; \quad f_{\pm 1} = f_{i\pm 1}^j; \quad \hat{f}_{\pm 1} = f_{i\pm 1}^{j+1}. \quad (5.1)$$

Физически естественная явная схема с право-разностным шаблоном аппроксимации производных

$$\hat{f} = f(1+r) - rf_{+1} \quad (5.2)$$

имеет первый порядок аппроксимации по обоим переменным и является безусловно неустойчивой. Явная же схема с левой разностью

$$\hat{f} = f(1-r) + rf_{-1} \quad (5.3)$$

устойчива при $r \leq 1$. Смешанная схема

$$\hat{f} = f(1 + (1-2\alpha)r) + r(\alpha f_{-1} - (1-\alpha)f_{+1}), \quad \alpha \in [0; 1] \quad (5.4)$$

при всех $\alpha \neq 1$, в том числе и при $\alpha = 1/2$ (центральная разность) неустойчива.

Так называемая схема Лакса

$$\hat{f} = \frac{f_{+1} + f_{-1}}{2} - r \frac{f_{+1} - f_{-1}}{2} \quad (5.5)$$

устойчива при $r \leq 1$. Схема Лакса-Вендроффа

$$\hat{f} = f - \frac{r}{2}(2rf + (1-r)f_{+1} - (1+r)f_{-1}) \quad (5.6)$$

дает аппроксимацию второго порядка точности по обеим переменным и устойчива при $r \leq 1$.

Из неявных схем часто используется схема с левой разностью

$$\hat{f} = \frac{1}{1+r} (f + r\hat{f}_{-1}), \quad (5.7)$$

схема Кранк-Николсона

$$\hat{f} = f - \frac{r}{4} (f_{+1} - f_{-1} + \hat{f}_{+1} - \hat{f}_{-1}), \quad (5.8)$$

а также схема Бабенко

$$\hat{f}_{+1} = f - \frac{1-r}{1+r} (\hat{f} - f_{+1}). \quad (5.9)$$

Обе последних схемы устойчивы и имеют второй порядок аппроксимации по времени и пространству.

Более детально разностные схемы применительно к решению уравнения переноса описаны в [30].

Довольно распространено мнение, что уравнение $\partial f / \partial t + \partial(uf) / \partial x = 0$ является «точным», а его численное решение по разностной схеме — «приближенным», причем именно последнему присущи дефекты: ошибки округления, неустойчивость при стремлении к нулю шага по какой-либо переменной, сходимость не к точному решению. В действительности оба подхода являются приближенными. Дифференциальная форма уравнения выражает поведение разностного уравнения в главной части асимптотики при стремлении к нулю шага дискретизации, тогда как на практике этот шаг часто бывает ограничен естественными причинами. Например, в задаче прогнозирования ВПФР временного ряда шаг по времени равен интервалу между наблюдениями значений ряда, его не имеет смысла делать меньше и тем более стремиться к нулю. Пространственный шаг определяется оптимальным разбиением гистограммы, и его также нельзя устремлять к нулю, ибо это приведет к росту ошибки в определении эмпирических вероятностей.

В результате получается, что исследователь совершает двойную и даже тройную ошибку. Во-первых, при выводе уравнения изменения ПФР обычно стартуют с некоторой схемы в конечных разностях, исходя из физических соображений в виде законов сохранения. Величина этих разностей не обсуждается, а просто считается достаточно малой, чтобы в соответствующей ячейке изменение аппроксимируемой функции можно было бы считать пренебрежимо малым. Само это предположение является

слишком смелым, ибо именно оно приводит в итоге к нарастанию ошибки округления. Во-вторых, от уравнения в конечных разностях переходят к его асимптотической форме в виде дифференциального уравнения, что принципиально меняет свойства решения. В-третьих, чтобы численно решить это — на самом деле, неправильное с точки зрения задачи в ее исходной постановке — дифференциальное уравнение, вновь переходят к дискретизации, но уже не физической, а удобной в плане корректности вычислительных процедур.

Для ВПФР всех этих ошибок не возникает, так как шаги по времени и пространству определены однозначно самой изучаемой системой.

Заключение

В данной монографии изложен новый подход к анализу нестационарных временных рядов, использующий кинетические уравнения для прогноза выборочной плотности функции распределения временного ряда. Тем самым показано, что методы статистической механики с помощью нового понятийного аппарата могут быть перенесены и на немеханические системы. В частности, оказалось возможным проследить аналогию между цепочкой кинетических уравнений статистической механики и следующих из нее уравнений гидродинамики сплошной среды в приближении локально-равновесных распределений молекул с цепочкой уравнений эволюции выборочных плотностей функций распределения временного ряда и его выборочных моментов в квазистационарном приближении.

Цепочка кинетических уравнений, связывающих между собой эволюцию выборочных распределений разных размерностей, является инструментом для статистически обоснованного вывода различных моделей временных рядов. Кроме известных моделей (регрессионных, авторегрессионных, адаптивных, с распределенными лагами и т.п.) в рамках кинетического подхода удастся построить новый класс моделей гидродинамического типа, когда ВПФР зависит от времени через несколько первых своих моментов.

Для определения уровня нестационарности выборки была введена и исследована специальная статистика, названная горизонтным рядом. Теоретически выведенные свойства плотности распределения горизонтного ряда позволили построить методику оценки уровня стационарности выборки, проверить качество предподготовки ряда для исключения выраженного временного тренда, обнаружить наличие дальних нелинейных корреляций.

Важное значение на практике имеют индикаторы нестационарности, в основе которых лежит статистика согласованного уровня значимости. Эти индикаторы позволяют корректно подойти к решению нестационарной задачи о разладке.

Автор надеется, что изложенная в книге теория и практические примеры будут полезны как специалистам в области математической статистики, так и студентам и аспирантам, занимающимся изучением этой области знаний.

Литература

1. Аносов Д.В., Арансон С.Х., Бронштейн И.У., Гринес В.З. Гладкие динамические системы. – Итоги науки и техники. Серия: Современные проблемы математики. Фундаментальные направления, т.1. – М.: Наука, 1985. – 244 с.
2. Дубровин Б.А., Новиков С.П., Фоменко А.Т. Современная геометрия. – М.: Наука, 1986. – 759 с.
3. Орлов Ю.Н. Основы квантования вырожденных динамических систем. – М.: МФТИ, 2004. – 236 с.
4. Козлов В.В. Тепловое равновесие по Гиббсу и Пуанкаре. – Москва-Ижевск: Институт компьютерных исследований, 2002.
5. Боголюбов Н.Н. Проблемы динамической теории в статистической физике. – М.: -Л.: ГИТТЛ, 1946.
6. Боголюбов Н.Н. Уравнения гидродинамики в статистической механике. – Сб. трудов Института математики АН УССР, 1948. Т.10. № 41.
7. Козлов В.В., Смолянов О.Г. Бесконечномерные уравнения Лиувилля относительно мер. // Доклады РАН. – 2010. – Т. 432. - №1. С. 28-32.
8. Ландау Л.Д., Лифшиц Е.М. Теоретическая физика, Т.Х: Физическая кинетика. М.: Наука, 1979.
9. Власов А.А. Статистические функции распределения. – М.: Наука, 1966. – 355 с.
10. Гайда Р.П. Квазирелятивистские системы взаимодействующих частиц. ЭЧАЯ. 1982. Т. 13. № 2.
11. Darwin C.G. Phil. Mag. 1920. V.39. P.357.
12. Фок В.А. Теория пространства, времени и тяготения. М.: ГИФМЛ, 1961.
13. Гантмахер Ф.Р. Теория матриц. М.: Наука, 1966.
14. Абрамовиц М., Стиган И.М. Справочник по специальным функциям. – М.: Наука, 1979.
15. Тихонов В.И., Миронов М.А. Марковские процессы. – М.: Советское радио, 1977. – 488 с.
16. Gaida R.P., Tretyak V.I. Single-time form of Fokker-Plank relativistic dynamics. // Acta Physica Polonica. V. B11. P.509.
17. Аникин В.М., Голубенцев А.Ф. Аналитические модели детерминированного хаоса. – М.: Физматлит, 2007. – 326 с.

18. Самко С.Г., Килбас А.А., Маричев О.И. Интегралы и производные дробного порядка и некоторые их приложения. – Минск: Наука и техника, 1987.
19. Нигматуллин Р.Р. Дробный интеграл и его физическая интерпретация. // ТМФ. – 1992. – Т. 90. – № 3. – С. 354-368.
20. Zaslavsky G.M. Chaos, fractional kinetics and anomalous transport // Elsevier, Physics Report 371. – 2002. – P. 461-580.
21. Уилкс С. Математическая статистика. – М.: Наука, 1967. – 632 с.
22. Феллер В. Введение в теорию вероятностей и ее приложения. Т.1. – М.: Мир, 1964. – 498 с.
23. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2006. – 816 с.
24. Новицкий П.В., Зograф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, 1991. – 304 с.
25. Орлов Ю.Н., Осминин К.П. Нестационарные временные ряды: методы прогнозирования с примерами анализа финансовых и сырьевых рынков. – М.: Книжный дом «ЛИБРОКОМ», 2011. – 384 с.
26. Hurst H., Black R., Simaika Y. Long-term Storage. An Experimental Study. – London: Constable, 1965. – 184 p.
27. Вовк В.С., Новиков А.И., Глаголев А.И., Орлов Ю.Н., Бычков В.К., Удалов В.А. Мировое производство, перевозки и рынки СПГ: методы прогнозного моделирования. – М.: ООО «Газпром-ЭКСПО», 2009. – 345 с.
28. Хинчин А.Я. Работы по математической теории массового обслуживания. – М.: Физматлит, 1963. – 236 с.
29. Орлов Ю.Н., Осминин К.П. Методы статистического анализа литературных текстов. – М.: Эдиториал УРСС/Книжный дом «ЛИБРОКОМ», 2012. – 312 с.
30. Калиткин Н.Н. Численные методы. – М.: Наука, 1978. – 512 с.