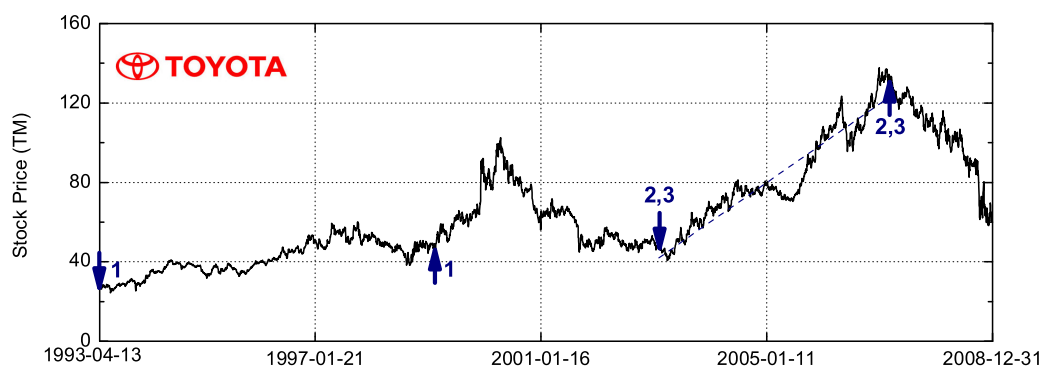


Анализ временных рядов

Курс лекций

А. Ю. Лоскутов

Физический факультет МГУ



Содержание

1	Введение. Рекомендуемая литература	4
2	Понятие временного ряда	8
2.1	Задачи анализа временных рядов	8
2.2	Описание процессов временными рядами	9
3	Наблюдаемые	12
3.1	Бесконечномерное и конечномерное описание	12
3.2	Дискретное представление	13
4	Методы обработки временных рядов	17
4.1	Статистические методы обработки временных рядов	17
4.2	Динамические методы обработки временных рядов	19

5	Размерность вложения	21
5.1	Функциональный метод определения размерности вложения . . .	21
5.2	Метод Грассбергера-Прокаччия	24
6	Информация и энтропия динамических систем	27
6.1	Понятие энтропии	27
6.2	Информация	28
7	Обобщенная размерность	33
7.1	Фрактальная, информационная и корреляционная размерности	33
8	Разделение шума и детерминированного сигнала. Оценка энтропии	37
8.1	Выделение детерминированных переменных	37
8.2	Оценка энтропии	38
9	Математические основы анализа временных рядов	40
9.1	Вложения	40
9.2	О рядах	42
10	Оптимизация выбора параметров	44
10.1	Оптимальный выбор m	45
10.2	Выбор времени задержки	46
11	Оценка длины ряда	47
11.1	Оценка из свойств корреляционного интеграла	47
12	Основные методы прогноза нерегулярных временных рядов	51
12.1	Сингулярный спектральный анализ (SSA)	51
12.2	Прогноз методом SSA	56
12.3	Локальная аппроксимация (LA)	60
12.4	Построение прогноза на один шаг вперед	60
13	Сравнение различных методов прогноза LA и прогноз на несколько шагов	64
13.1	Порядок аппроксимации	64
13.2	Сравнение методов	65
13.3	Прогноз на несколько шагов вперед	66
14	Малые возмущения хаотических динамических систем	70
14.1	Анализ систем с внешними возмущениями	71

15	Подавление хаоса и управление хаотическими системами	74
15.1	Подавление хаоса	74
15.2	Управление с обратной связью	76
16	Стратегии конкурентных отношений	78
16.1	Шумы	78
16.2	Разорение игрока	82
16.3	Неблагоприятные шансы на выигрыш	83
16.4	Основные выводы	84
17	Разладка и сегментация временных рядов	85
17.1	Разладка	85
17.2	Сегментация	88
17.3	Показатель Хёрста и показатель Гёльдера	89
18	Фрактальные свойства временных рядов	90
18.1	Вейвлет-преобразование и локальное Фурье-преобразование	95
19	Явление гомодинамичности	97
19.1	Линейные гомодинамические модели	99
19.2	Исходные гипотезы	99
19.3	Техника оценки параметров	101
20	Прогноз временных рядов естественного происхождения	103
20.1	Краткий обзор LA	103
20.2	Примеры прогнозов экономических показателей	107

Прежде чем искать причину какого-либо факта, необходимо как следует убедиться в том, что такой факт действительно имеет место.
Фонтенель, “Истории оракулов”, 1687.

1 Введение. Рекомендуемая литература

В настоящее время для изучения свойств сложных систем, в том числе и при экспериментальных исследованиях, широко используется подход, основанный на анализе сигналов, произведенных системой. Это очень актуально в тех случаях, когда математически описать изучаемый процесс практически невозможно, но в нашем распоряжении имеется некоторая характерная наблюдаемая величина. Поэтому анализ систем, особенно при экспериментальных исследованиях, часто реализуется посредством обработки регистрируемых сигналов. Например, в аритмологии в качестве такого сигнала используется электрокардиограмма, в сейсмологии — запись колебаний земной коры, в метеорологии — данные метеонаблюдений и т.п. Обычно такой сигнал называется наблюдаемой, а метод исследования — реконструкцией динамических систем. Этот раздел теории динамических систем называется *анализом временных рядов*.

Наблюдаемая — это последовательность значений некоторой переменной (или переменных), регистрируемых непрерывно или через некоторые промежутки времени. Часто вместо термина наблюдаемая используется понятие временной ряд. Ясно, что наличие только лишь временного ряда вместо полного решения уравнений сильно ограничивает наши знания об изучаемой системе. Это налагает большие ограничения на возможности метода реконструкции.

Скалярным временным рядом $\{x_i\}_{i=1}^N$ называется массив из N чисел, представляющих собой значения некоторой измеренной (наблюдаемой) динамической переменной $x(t)$ с некоторым постоянным шагом τ по времени, $t_i = t_0 + (i - 1)\tau$: $x_i = x(t_i)$, $i = 1, \dots, N$. В анализе временных рядов выделяются две основные задачи: задача идентификации и задача прогноза.

Задача идентификации при анализе наблюдаемых предполагает ответ на вопрос, каковы параметры системы, породившей данный временной ряд — размерность вложения, корреляционная размерность, энтропия и др. Размерность вложения — это минимальное число динамических переменных, однозначно описывающих наблюдаемый процесс. Корреляционная размерность является оценкой фрактальной размерности аттрактора системы и частным случаем обобщенной вероятностной размерности. Понятие энтропии связано с предсказуемостью значений ряда и всей системы.

Задача прогноза имеет целью по данным наблюдений предсказать будущие значения измеряемых характеристик изучаемого объекта, т.е. составить прогноз на некоторый отрезок времени вперед. Сейчас разработано и обосновано несколько различных методов прогноза. Однако все они подразделяются на два основных класса: локальные и глобальные. Такое деление проводится по области определения параметров аппроксимирующей функции, рекуррентно устанавливающей следующее значение временного ряда по нескольким предыдущим.

Исторически первыми были разработаны глобальные методы, в которых на основе статистического анализа предлагалось использовать авторегрессию, скользящее среднее и др. Позже в рамках нелинейной динамики были разработаны новые практические методики:

- сингулярный спектральный анализ (SSA), который является глобальным методом;
- локальная аппроксимация (LA);
- сочетание SSA–LA.

Исследование временных рядов базируется на идее, что удовлетворительную геометрическую картину странного аттрактора можно получить, если вместо переменных, входящих в исходную систему, использовать так называемые векторы задержек наблюдаемой $z_i = \{x_i, x_{i+1}, \dots, x_{i+m-1}\}$. Впервые данный подход к анализу временных рядов был математически обоснован в работе Ф.Такенса.

Таким образом, наиболее интригующим и заманчивым применением теории динамических систем является прогнозирование динамики порож-

даемых ими временных рядов. При этом предполагается, что *a priori* характеристики систем, которые порождают этот ряд, могут быть неизвестны.

Сейчас стало ясно, что теория игр теснейшим образом переплетена с теорией динамических систем, фрактальных множеств и нелинейной динамикой, поскольку большинство реальных временных рядов имеют самоподобную структуру. Эта особенность позволяет переосмыслить подходы к анализу временных рядов и иным (в основном, более успешным образом) подойти к их описанию. При этом выявляются различные стратегии прогноза, обосновывается невозможность использовать здравый (привычный) смысл в некоторых казалось бы очевидных ситуациях и т.п. Более того, если принять во внимание теорию управления хаотическими системами, то становится возможным на основе совершенно иных подходов, чем это принято в обычной теории, управлять динамической системой.

Таким образом, теоретические исследования, основанные на анализе временных рядов, могут дать мощный инструмент для понимания многих явлений, особенно когда имеющихся данных для построения модели может быть недостаточно.

Список литературы

- [1] А.Ю.Лоскутов, А.С.Михайлов. *Основы теории сложных систем.*— Москва, Регулярная и хаотич. динамика, 2007.
- [2] Gouesbet G, Meunier-Guttin-Cluzel S, Ménard O (eds.) *Chaos and Its Reconstruction* (New York: Nova Sci. Publ., 2003).
- [3] Безручко Б П, Смирнов Д А *Математическое моделирование и хаотические временные ряды* (Саратов: ГосУНЦ «Колледж», 2005)
- [4] В.С.Афраймович, А.М.Рейман. Размерность и энтропия в многомерных системах.— В сб. *Нелинейные волны. Динамика и эволюция.* Ред. А.В.Гапонов-Грехов, И.М.Рабинович.— М., Наука, 1989, с.238–262.
- [5] А.Ю.Лоскутов, О.Л.Котляров, И.А.Истомин, Д.И.Журавлев. Проблемы нелинейной динамики. III. Локальные методы прогнозирования временных рядов.— *Вестн. Моск. ун-та, сер. Физ.-астр.*, 2002, №6, с.3–21.
- [6] А.Ю.Лоскутов. Очарование хаоса.— *Успехи физ. наук*, 2010, т.180, № 12, с.1305–1329.

-
- [7] Г.Шустер. *Детерминированный хаос. Введение.*— М., Мир, 1988.
- [8] Г.Г.Малинецкий, А.Б.Потапов. *Современные проблемы нелинейной динамики.*— м.: УРСС, 2000.
- [9] А.Н.Ширяев. *Статистический последовательный анализ.*— М.: Наука, 1976.
- [10] A.Loskutov. Parametric perturbations and non-feedback controlling chaotic motion.— *Discrete and Cont. Dyn. Syst.* 2006, v.6, No5, p.1157–1174.

2 Понятие временного ряда

Многочисленные приложения анализа временных рядов составляет в настоящее время бурно развивающаяся область науки.

2.1 Задачи анализа временных рядов

Временной ряд — последовательность регистрируемого сигнала (наблюдаемая). Такой подход используется, когда нет возможности построить уравнения движения.

Примеры: солнечная активность, землетрясения, ЭЭГ, ЭКГ, биржевой курс и т.п.

Какие задачи здесь возникают?

Физика солнца:

- а) скрытые периодичности;
- б) прогноз активности.

Электрокардиограммы (ЭКГ):

- а) природа наблюдающихся аритмий;
- б) прогноз развития состояния.

Экономические ряды:

- а) задача сегментации;
- б) задача прогноза.

Химическая кинетика:

- а) анализ динамики;
- б) построение модели.

Можно ли прогнозировать динамику изучаемой системы по порождаемому временному ряду и на основе этого предложить модель?

Теория Такенса:

- 1) Определить, динамическая ли система (конечность степеней свободы системы, порождающей ряд) → размерность вложения.

- 2) Горизонт прогноза: энтропия временного ряда.
- 3) Точность прогноза.

2.2 Описание процессов временными рядами

Аттрактор — «притягатель» (см. предыдущий курс).

Странный аттрактор — не многообразие (отличен от конечного объединения подмногообразий). Отсюда название «странный».

Характеристики «странности»: те же, что для канторовых множеств.



Фрактальные множества в природе:

- медицина (капилляры, бронхи);
- осаждение ионов металла на затравочном образце;
- электрический разряд (фигура Лихтенберга);
- самоподобные явления на бирже.

Основной вывод: существует универсальность в описании многих явлений, хотя, конечно, при помощи фрактального подхода описать можно далеко не все.

Анализ рядов, порожденных системами, на основе спектров → “спад” сглаженных спектров. Тогда можно обнаружить аналогии с *шумами*:

шум белый	}	отличие этих процессов в кластеризации данных
шум розовый		
шум коричневый		
шум черный		

Это поясняет приведенный ниже рисунок. Хорошо видно, что наблюдаемые значения группируются. Иными словами, в белом шуме за определенный (небольшой) интервал времени имеется примерно одинаковое

количество положительных и отрицательных значений. Для коричневого процесса это не так: значения, если они положительные, некоторое время и останутся таковыми (*они кластеризуются*), т.е. не следует быстро ожидать, что они станут отрицательными. Вероятность этого для коричневого и черного шумов мала.

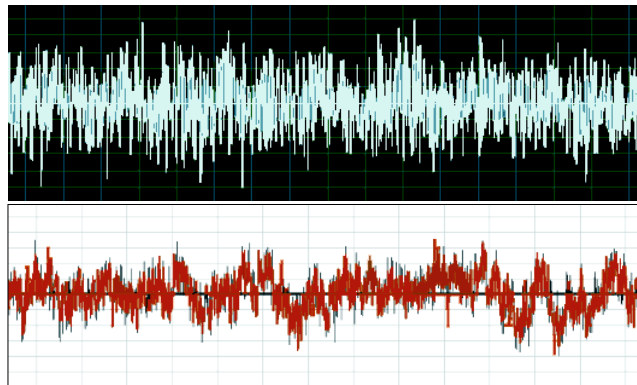


Рис. 1: Отличие белого и коричневого шумов.

Используя такой подход, можно объяснить некоторые тенденции, скажем, в экономике и на биржах. Например, попав в какой-либо из кластеров, практически невозможно выйти из него быстро \Rightarrow **после спада не следует ожидать быстрого подъема.**

В азартных играх: если капитал упал до нуля, то вероятность того, что он в течение короткого промежутка времени восстановится, практически нулевая.

{ Так объясняется интуиция опытных игроков: "плохой день \equiv пепёр".
{ Наоборот, выигрыши тоже кластеризуются: пёр пошёл.

Отсюда можно **строго обосновать** стратегии:

- жадного игрока (быстро получить прибыль);
- беспечного игрока (ставить почти все деньги);
- робкого игрока (играет по маленькой);
- опытного игрока (использует побочные сведения).

Кроме того, в случае действия группы фирм на рынке, можно обосновать:

- стратегию разорения конкурентов (это может быть не очень выгодно, т.к. останешься один);
- стабилизация прибыли варьированием инвестиций.

|| Это все существует в условиях *хаотического развития ситуации*.
|| Поэтому сначала уместно напомнить основные положения **теории хаоса**.

Формализация словосочетания “хаотическое поведение” представлено в курсе, который был прочитан в предыдущем семестре “Основы теории динамического хаоса”.

3 Наблюдаемые

Ранее мы, как правило, рассматривали *сосредоточенные* системы → т.е. системы с конечным числом степеней свободы. Изучение таких систем не представляет собой никаких **принципиальных** трудностей (однако вычислительных трудностей может быть достаточно). Можно найти показатели Ляпунова, энтропию, осуществить прогнозирование. → Уравнения известны (модели), законы взаимодействия заданы. Поэтому для адекватности описания остается определить только параметры.

3.1 Бесконечномерное и конечномерное описание

Совсем иное — *распределенные* системы. Их в природе большинство. типичный пример — имеется производная по пространству (уравнение Навье–Стокса, с диффузией). Таким образом, фазовое пространство является **бесконечномерным** (т.к. надо задать начальные условия в каждой точке пространства!) Каждая точка этого пространства отвечает определенному распределению величин, характеризующих систему: полю температуры, давления, плотности и т.п.

Как изучать такие системы? Конечно, существуют мощные численные методы. Но зачастую они оказываются некорректными, и проверить численно полученные результаты невозможно. Более того, как **корректно составить** такую модель так, чтобы она была адекватна рассматриваемой системе, ведь часто неизвестно даже число факторов, вовлеченных в систему? Поэтому здесь возникает два основных вопроса.

- Можно ли распределенные системы свести к конечномерным?
- Можно ли изучать системы, не прибегая к составлению моделей?

Большинство систем диссипативны: с течением времени все фазовые траектории стягиваются к некоторому подмножеству исходного бесконечномерного пространства. Как правило, это подмножество *конечномерно*. Таким образом, можно изучить динамику исходной системы, рассматри-

вая это подмножество. Раз это подмножество конечномерно, оно описывается каким-то конечным набором переменных.

3.2 Дискретное представление

Наименьшее число независимых переменных, однозначно определяющих установившееся движение исходной диссипативной распределенной системы, называют **размерность вложения** и обозначают d_e (от embedding).

Очевидно, подмножество, к которому стягиваются траектории системы — это **аттрактор**. Размерность вложения *аттрактора* — это минимальная размерность фазового пространства, в которое без самопересечений может быть помещено гладкое многообразие, целиком содержащее этот аттрактор.

Примеры.

1) Двумерный тор. Он вложим только в трехмерное пространство и пространство большей размерности.

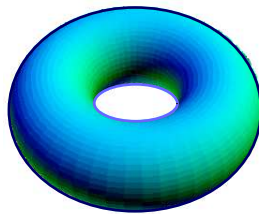


Рис. 2: Двумерный тор.

2) Двумерная сфера. Она вложима тоже только в трехмерное пространство и пространство большей размерности.

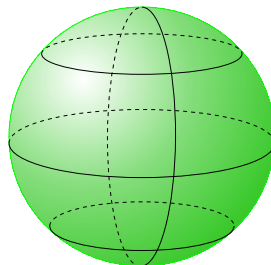


Рис. 3: Двумерная сфера.

3) Достаточно сложная пространственная кривая (см. рисунок) вложима опять только в трехмерное пространство и пространство большей размерности.

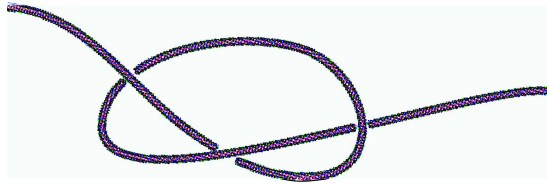


Рис. 4: Пространственная кривая.

- В общем случае справедливо следующее точное утверждение.

Лемма вложения Уитни: Любое гладкое многообразие размерности m вложимо в пространство размерности $d \geq 2m + 1$.

Отсюда сразу следует, что в пространствах размерности 4 и выше не существует узлов, кос и т.п.! Любая попытка завязать узел приведет к тому, что он не “затягивается”, а “распускается”. Странный аттрактор с фрактальной размерностью d_F всегда вложим в пространство размерности $K \geq 2[d_F] + 1$. В некоторых случаях $K = [d_F] + 1$, где $[d_F]$ — целая часть d_F .

Как установить размерность вложения? *Это очень важно, т.к. можно определить число главных факторов, влияющих на формирование динамики исследуемой системы!*

⇒ Здесь нужно использовать порождаемые системой временные ряды. Такой подход позволяет найти ответ и на второй основной вопрос об изучении системы без построения математических моделей.

Для определения размерности вложения как для системы уравнений в частных производных, так и для изучаемой системы, где *a priori* данных для моделирования мало, достаточно знать, как меняется единственная переменная в какой-либо точке. Изменение со временем такой типичной переменной называется (*скалярным*) *временным рядом*. Если измеряется несколько переменных, то ряд называется *векторным*.

Временные ряды естественным образом возникают в экспериментах

как *натурных* (тогда получается непрерывный ряд $y(t)$), так и *численных* (тогда естественным образом получим дискретный ряд). *Для многих приложений естественны дискретные ряды.*

Поясним качественно, почему можно для определения размерности вложения использовать скалярный ряд.

Будем отслеживать переменную (типичную) $x(t)$ через равные промежутки времени τ :

$$x(t) \equiv y_1, x(t + \tau) \equiv y_2, \dots, x(t + (n - 1)\tau) \equiv y_n, \dots$$

Допустим, что динамика исходной рассматриваемой системы описывается одним дифференциальным уравнением первого порядка. Тогда эта последовательность будет определяться единственным начальным значением y_1 , а все остальные $y_n, n > 1$, будут функционально зависеть от него:

$$y_2 = f(y_1), y_3 = f(f(y_1)), \dots$$

Это означает, что все значения y_n ложатся на определенную **линию** $y_{n+1} = f(y_n)$ на плоскости (y_{n+1}, y_n) .

Предположим теперь, что для описания системы требуется два дифференциальных уравнения первого порядка. Тогда точки y_n не будут ложиться на линию, а будут разбросаны по плоскости (y_{n+1}, y_n) . Теперь для задания последовательности y_n недостаточно только y_1 . Необходимо также y_2 . В этом случае $y_{n+1} = f(y_{n-1}, y_n)$.

В общем случае, если размерность равна k , то для k последовательных значений $y_n, y_{n-1}, \dots, y_{n-k+1}$, можно однозначно восстановить по ним следующее значение y_{n+1} . Иными словами, должна существовать некоторая функция f :

$$y_{n+1} = f(y_n, y_{n-1}, \dots, y_{n-k+1}).$$

Теперь возникают следующие естественные вопросы.

- Как практически определить размерность из временного ряда и установить конечномерность или бесконечномерность рассматриваемого явления?

- Можно ли из временного ряда идентифицировать (распознать) систему: отличить норму от патологий, не разрушая систему?
- Можно ли на основе анализа временного ряда решить задачу прогноза: предсказать будущие значения или будущее состояние анализируемой системы?

Этим, в основном, мы и будем заниматься. Мы изучим также их приложения (тактика и стратегия биржевых игроков и игроков в казино, разладка, разорение, крах здравого смысла при анализе торгов и др.).

4 Методы обработки временных рядов

Несмотря на то, что нелинейные системы могут значительно отличаться в конкретных проявлениях и деталях, существуют глубокие аналогии в их организации и функционировании. Это предопределило интерес к методам, которые развиваются в рамках теории динамических систем, как универсальному инструменту исследования объектов самой различной природы. Наиболее явно такой объединяющий подход проявляется при анализе временных рядов.

В настоящее время существует два качественно различных подхода к исследованию временных рядов:

- статистические
- динамические

К статистическим подходам относятся вероятностные модели. К динамическим — теория Такенса (Такенса-Мане). Современное представление о возможности описания наблюдаемых дает т.н. *эмбедология* (от англ. embedding — вложение), объединяющая элементы теории размерности, теории информации, топологии, дифференциальной динамики и теории динамических систем.

4.1 Статистические методы обработки временных рядов

Основа: имеется ряд y_n и шум — последовательность некоррелирующих и одинаково распределенных случайных величин ξ_i с нулевым средним. Тогда можно записать, что

$$y_n = F(y_{n-1}, \dots, y_{n-m}, \xi_n, \dots, \xi_{n-k}),$$

где k и m — некоторые конечные числа.

Основные модели здесь — модели типа ARMA (Auto Regression Moving Average — авторегрессия скользящего среднего):

$$y_i = a_0 + \sum_{j=1}^m a_j y_{i-j} + \sum_{j=0}^k b_j \xi_{i-j}.$$

Нахождение искоемых коэффициентов — это возможное решение задачи идентификации, а само соотношение для y_i можно использовать для прогноза по m предыдущим.

Это все линейные модели!

В качестве прогнозируемой величины обычно используют среднее значение:

$$\hat{y}_i = a_0 + \sum_{j=1}^m a_j y_{i-j}$$

и предшествующие значения известны точно. Следует подчеркнуть, что шум является неотъемлемой частью модели; в его отсутствие поведение абсолютно непохоже на исследуемый ряд.

Модели типа ARMA — это **линейные** модели. Часто строят **нелинейные статистические модели** (NARMA). Они делятся на

параметрические

непараметрические

Параметрические модели: функция $F(\vec{y}, \vec{\xi}, \vec{a})$ одна и та же для всех \vec{y} и $\vec{\xi}$. Несколько параметров $\vec{a} = \{a_1, \dots, a_k\}$ необходимо найти по временному ряду.

Непараметрические модели: набор $\{\vec{y}_k, \vec{\xi}_k\}$ в окрестности которого используется локальная аппроксимация.

Таким образом, при данных подходах часто используется интуиция, на основе которой *тем или иным уравнением ставятся в соответствие определенные ситуации, а параметры неизвестны.*

Следующая задача — выписать **детерминированные уравнения** для средних. *Это — типичная теория случайных процессов.*

Основная трудность — удаление тренда (средние не зануляются, когда есть тренд), получение статистических оценок, фильтрация шума. Если случайный процесс не стационарен — используются ARCH-модели.

В случаях, требующих быстрого обновления прогноза на основе вновь поступивших данных, используются адаптивные методы прогноза. К ним относится, например, метод экспоненциального сглаживания (метод Брауна). Следуя ему, каждому значению ряда в процессе идентификации модели присваивается весовой коэффициент, экспоненциально убывающий

со временем, отделяющим это значение от последнего известного значения ряда. Таким образом, самые “старые” значения ряда практически не влияют на результаты прогноза, тогда как последние известные величины имеют наибольший вес. Тем самым этот метод приближается к локальным методам, так как основной вклад при прогнозе дает лишь небольшая часть самых последних по времени значений ряда.

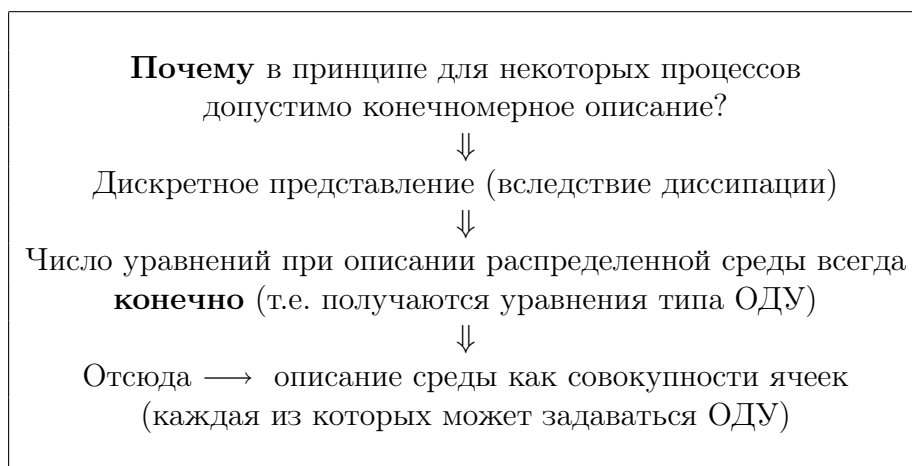
Методы авторегрессии разработаны наиболее тщательно и применяются, как правило, в прикладных задачах. Они реализованы практически во всех программных пакетах статистической обработки данных.

4.2 Динамические методы обработки временных рядов

В статистических методах обработки используется допущение, что *всегда имеется шум*. Это довольно сильное условие. Поэтому мы должны априори полагать, что изучаемый ряд — стохастический! А если это детерминированный хаос, допустим ли такой подход? Может в этом случае лучше работают другие методы?

Здесь надо очень хорошо себе представлять *отличие стохастичности от хаоса*.

Статистические методы не позволяют отличить конечномерный процесс от бесконечномерного, но это вполне можно сделать динамическими методами.



Для многих систем конечномерное описание вполне допустимо. Так, в *финансовых рядах* число спекулянтов всегда конечно, другое дело, как те или иные из них влияют на процесс (сильно, слабо). Дискретное (конечномерное) описание здесь подразумевает выявление числа крупных спекулянтов. Тогда присутствие остальных можно рассматривать как влияние некоторого шума (его *можно оценить*, выявив попутно скрытые динамические процессы, ответственные за крупных игроков). Отсюда — **смешанное описание: статистическое и динамическое**.

Первый вопрос, которым мы задаем, — определение возможности конечномерного описания, т.е. в приложении к конкретным рядам — вопрос о количестве факторов, влияющих на динамику.

Это можно сделать посредством нахождения *размерности вложения* — минимального числа динамических переменных, однозначно описывающих поведение исследуемой системы.

5 Размерность вложения

При анализе временных рядов главной задачей является реконструкция породившей этот ряд динамической системы. В соответствии с теорией Такенса–Мане приемлемое описание фазового пространства динамической системы можно получить, если взять вместо реальных переменных системы (которые могут быть неизвестны вообще!) k -мерные векторы задержек, составленные из значений ряда в последовательные моменты времени. При выполнении условия $k \geq 2d_e + 1$ (результат, восходящий к лемме вложения Уитни), где d_e — размерность вложения, возможно реконструировать фазовое пространство (пространство состояний) системы. При условии стационарности временного ряда на базе этой реконструкции строится *прогноз его дальнейшей динамики*.

На сегодняшний день наиболее часто используемым алгоритмом для оценки величины k является алгоритм Грассбергера–Прокаччия. Однако и он оказывается неэффективным при работе с короткими (до 10^4 точек) временными рядами. Имеются также и другие методы, среди которых наиболее приемлемый — это т.н. функциональный метод. Существуют также и другие возможности расчетов величины d_e .

Однако все они имеют свои недостатки: сложность реализации, большая длительность расчетов, неоднозначность либо сомнительность результатов. В связи с этим величина k , за исключением модельных примеров, в которых она достоверно известна, как правило, определяется эмпирически. Главный критерий в этом случае — выбор такого k , начиная с которого прекращается качественное изменение прогноза.

5.1 Функциональный метод определения размерности вложения

Построим через равные промежутки времени τ последовательность $y_1, y_2, \dots, y_n, y_{n+1}, \dots$ наблюдаемой величины $x(t)$ следующим образом: $x(t) \equiv y_1, x(t + \tau) \equiv y_2, \dots, x(t + (n - 1)\tau) \equiv y_n, \dots$. Выберем в этой последовательности, начиная с какого-то n , произвольную подпоследова-

тельность длины k и рассмотрим все возможные векторы

$$\vec{w}^{(n)} = \{y_{n-k+1}, y_{n-k+2}, \dots, y_{n-1}, y_n\},$$

Все они удовлетворяют соотношению $\vec{w}^{(n)} \in \mathbb{R}^k$.

Зафиксируем одну из последовательностей, начинающуюся с какого-то номера n_0 и обозначим $\vec{w}^{(0)}$. Рассмотрим расстояние между произвольным вектором $\vec{w}^{(n)}$ и этим $\vec{w}^{(0)}$:

$$\rho_k(n, n_0) = |\vec{w}^{(n)} - \vec{w}^{(0)}| = \left[\sum_{i=1}^k (y_{n-k+i} - y_{n_0-k+i})^2 \right]^{1/2},$$

а также величину

$$r(n, n_0) = |y_{n+1} - y_{n_0+1}|,$$

Построим для различных n зависимость $r(\rho)$, откладывая $\rho_k(n, n_0)$ и $r(n, n_0)$ по осям (см. рис.5):

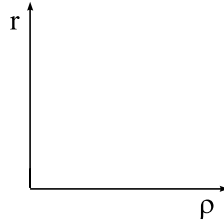


Рис. 5: К вопросу определения функциональной зависимости.

Вспомним, что **наличие функциональной зависимости** (что означает существование динамической системы, описывающей данный временной ряд) между y_{n+1} и предыдущими k величинами $y_n, y_{n-1}, \dots, y_{n-k+1}$ означает, что существует функция f :

$$y_{n+1} = f(\vec{w}^{(n)}).$$

При этом можно записать

$$r(n, n_0) = |f(\vec{w}^{(n)}) - f(\vec{w}^{(0)})|.$$

Расстояние $\rho_k(n, n_0) \rightarrow 0$, когда $\vec{w}^{(n)} \rightarrow \vec{w}^{(0)}$. Тогда $r(n, n_0)$ также должно стремиться к нулю.

Если же функциональная зависимость отсутствует при выбранном k , то вектор $\vec{w}^{(n)}$ не определяет однозначно значение y_{n+1} и, следовательно, $r(n, n_0)$ не будет стремиться к нулю при уменьшении $\rho_k(n, n_0)$.

Процедура нахождения функциональной зависимости:

- для различных n найти значения $\rho_k(n, n_0)$ и $r(n, n_0)$;
- отложить соответствующие точки на плоскости (r, ρ) ;
- последовательно соединить полученные точки.

Если при выбранном k функциональная зависимость достигнута, то в области малых ρ все участки ломаной расположатся вблизи начала координат. **Размерность вложения** d_e — это минимальное значение k , начиная с которого ломаная обладает этим свойством (см. рис.6, где $k = 3$):

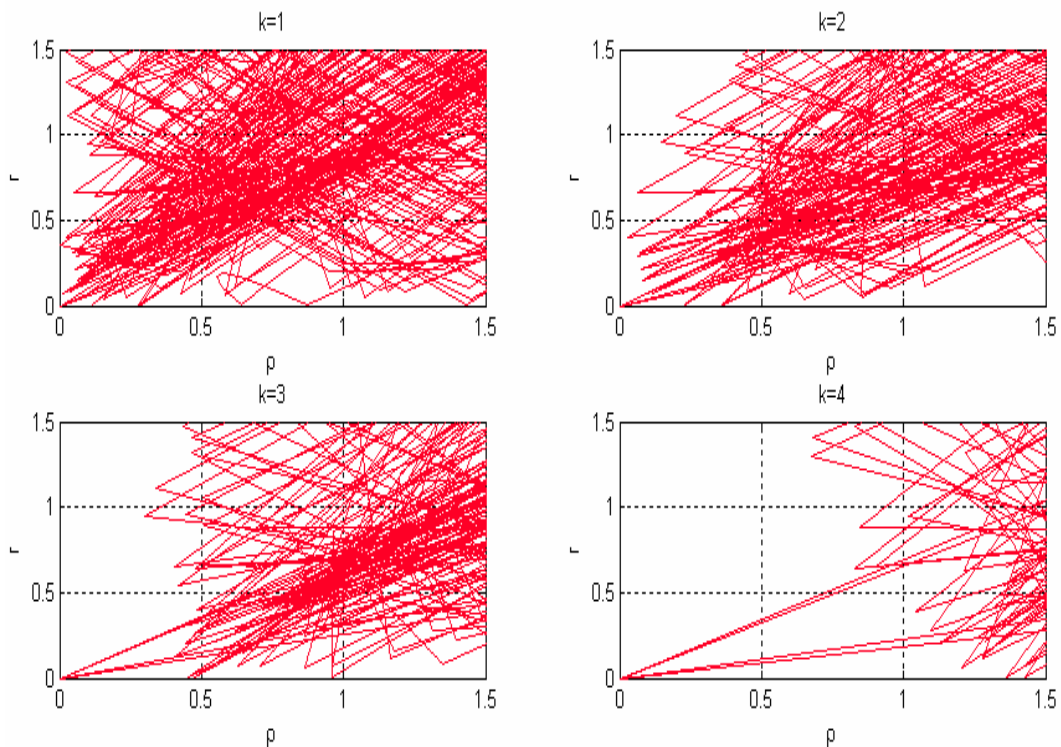
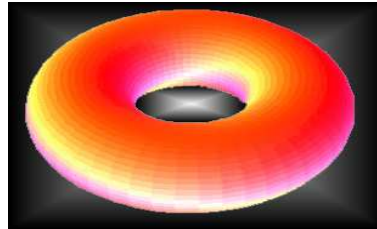


Рис. 6: Графическое определение размерности вложения.

Для двумерного тора Tor^2



зависимость $r(\rho)$ показана на рис.7:

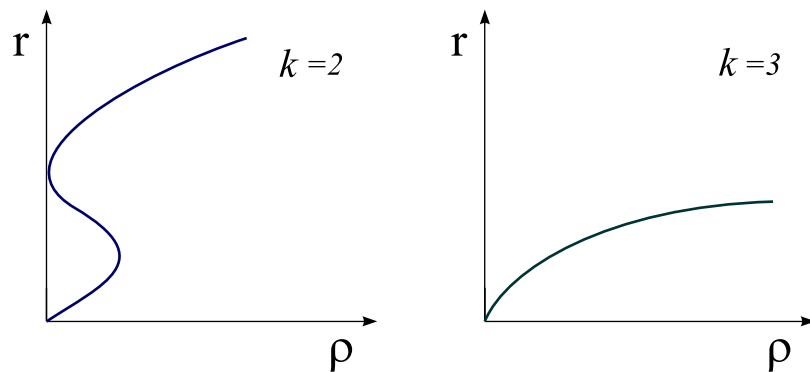


Рис. 7: Вложение двумерного тора.

Отсюда сразу следует, что размерность вложения двумерного тора равна трем.

5.2 Метод Грассбергера-Прокаччия

Пусть, как и прежде, имеется $y_1, y_2, \dots, y_n, y_{n+1}, \dots$. Выделим различные произвольные последовательности длины k и построим векторы $\vec{w}^{(n)}$. Теперь определим

$$\rho_k(n, n') = |\vec{w}^{(n)} - \vec{w}^{(n')}| = \sqrt{\sum_{i=1}^k (y_{n-k+i} - y_{n'-k+i})^2},$$

для любой пары $\vec{w}^{(n)}, \vec{w}^{(n')}$.

Возьмем некоторое значение l и вычислим *корреляционный интеграл* $C_k(l)$ как

$$C_k(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{n=1}^N \sum_{n'=1}^N \Theta[l - \rho_k(n, n')],$$

где N — это полное число элементов в последовательности y_n , $\Theta(z)$ — ступенчатая функция Хевисайда:

$$\Theta(z) = \begin{cases} \Theta(z) = 0, & z < 0, \\ \Theta(z) = 1, & z \geq 0. \end{cases}$$

Легко видеть, как интерпретировать корреляционный интеграл: это нормированное на N^2 количество пар точек, расстояние между которыми меньше l . Иными словами, вклад в $C_k(l)$ дают только те пары $\vec{w}^{(n)}$, $\vec{w}^{(n')}$ расстояние между которыми меньше или равно l .

Теперь можно ввести понятие *корреляционной размерности*:

$$d_c = \lim_{l \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\ln C(l)}{\ln l},$$

где l — это характерный размер ячейки фазового пространства. Аналитически или численно этот двойной предел оценить непросто. Однако геометрически выражение d_c определяется как тангенс угла наклона $\ln C(l)$ от $\ln l$. При этом, конечно, наклон графика будет зависеть от k .

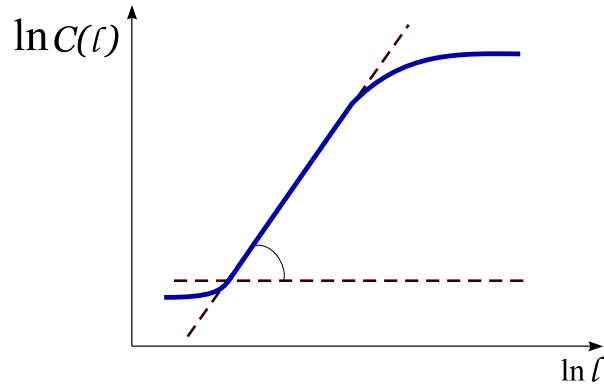


Рис. 8: Зависимость корреляционного интеграла от ширины ячейки в логарифмическом масштабе.

Если **конечномерное** представление данной системы существует, то с увеличением k тангенс будет стремиться к определенной **конечной** величине. Эта величина и будет d_e . Значение k , при котором наклон перестает

изменяться, дает размерность вложения d_c . Типичная зависимость $\ln C(l)$ от $\ln l$ представлена на рис.8:

Качественная интерпретация этого графика следующая.

$$\begin{cases} l \text{ велико} & \text{— в ячейку попадают все точки; “корреляции” } \mathbf{велики}. \\ l \text{ мало} & \text{— структура аттрактора } \mathbf{неразрешима}. \end{cases}$$

Таким образом, типичная картина при увеличении k такова:

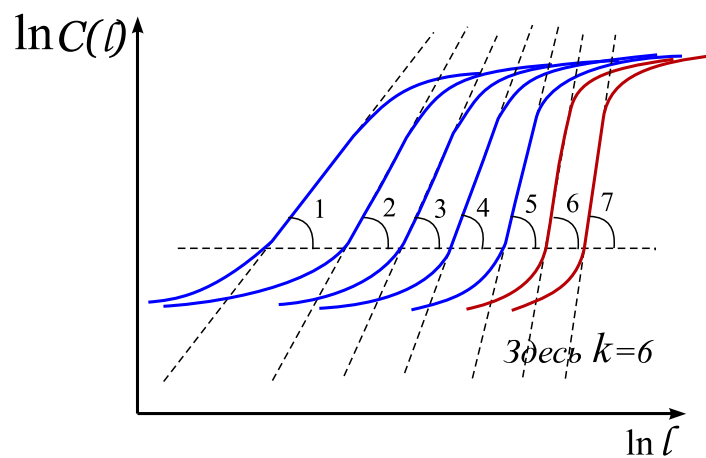


Рис. 9: Изменение наклона кривой $\ln C$ как функции $\ln l$ при увеличении размерности вложения k .

Замечание: значение d_c не превышает фрактальной размерности d_f .

- ЗАДАНИЕ для курсовой работы. Найти (скажем, в Интернете) любой **равномерный** ряд (например, ЭКГ или курс акций). Определить, сходится ли для него размерность вложения. Если да (а это может быть так для некоторых рядов), то тщательно определить эту величину.

Использовать подходы а) $r(\rho)$; б) d_c . Результаты сравнить. Выполнять можно группами по 2-3 человека.

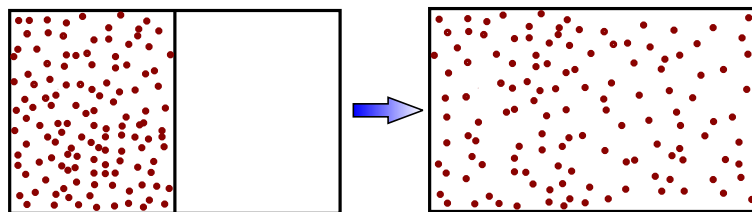
6 Информация и энтропия динамических систем

Основные вопросы при изучении временных рядов:

- время предсказуемости;
- какие инвариантные характеристики можно найти;
- отличить шум от хаоса.

6.1 Понятие энтропии

Ответы на них дает понятие **энтропии**. Сначала рассмотрим качественно это понятие. **Термодинамическая энтропия** S — мера беспорядка в системе. Пример, когда S растет:



Рост беспорядка связан с ростом нашего незнания о состоянии системы: до того, как была убрана перегородка, о расположении молекул мы знали больше.

$$S \sim - \sum_i P_i \ln P_i,$$

P_i — вероятности системы оказаться в состоянии i . S — это мера информации, необходимой для определения местонахождения системы в некотором состоянии.

Таким образом, беспорядок есть понятие из теории информации.

Пояснение.

Сначала задачка. *На раздаче в баре совершено убийство; яд находится в одном из 1200 стаканов. Сколько надо сделать проб для определения, в каком стакане яд?*

Решение этой задачи поможет понять материал данной лекции.

Пусть система имеет вид $\begin{array}{|c|c|} \hline & \bullet \\ \hline \end{array}$. Если положение шара заранее неизвестно и мы его узнаём, то тем самым мы увеличиваем суммарную информацию на 1 бит (по определению). Имея эту информацию, мы *экономим один вопрос*, необходимый для определения местоположения.

Для другого случая $\begin{array}{|c|c|} \hline & \bullet \\ \hline & \\ \hline \end{array}$ необходимо *два вопроса*, т.е. максимальное количество информации $I = 2$ бит. Это же можно записать как

$$I = \log_2 4,$$

от числа возможных состояний, которое равно 4.

Вообще, между максимальным количеством информации I и числом состояний N выполнимо соотношение:

$$I = \log_2 N.$$

Найдем **приращение информации**, когда мы узнаем результат статистических событий.

6.2 Информация

Рассмотрим пример с подбрасыванием монеты: пусть p_1 — вероятность "орла", p_2 — вероятность "решки".

Предположим, что $p_1 = p_2 = 1/2$ (не фальшивая монета). Информация I , полученная, если мы узнаем, что результат эксперимента — "орел": $I = 1$, т.к. имеется только *два* равновероятных состояния $\begin{array}{|c|c|} \hline & \\ \hline \end{array}$. Используя вероятности p_i , найдем:

$$I = - \left[\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right].$$

Обобщая, можно записать $I = - \sum_i p_i \log_2 p_i$. Если вероятности выпадения "орла" и "решки" разные (т.е. мы подбрасываем деформированную монету), то $p_1 \neq p_2 = 1 - p_1$. Однако соотношение для I , как легко видеть, сохраняется.

Вообще, если априори известно только, что $1, \dots, n$ событий (или состояний системы) имеют место с вероятностями p_i (так что $\sum_{i=1}^n p_i = 1$) и мы

узнаем путем измерения, что имеет место событие j (или система находится сейчас в некотором состоянии), то, многократно повторяя измерения, получим *среднее приращение информации*:

$$I = - \sum_i p_i \log_2 p_i.$$

*Результат
Шеннона*

Вернемся к КС-энтропии.

Поскольку беспорядок есть понятие теории информации, КС-энтропию K , показывающую, *насколько система хаотична*, тоже можно определить соотношением Шеннона: это **средняя скорость потери информации о состоянии динамической системы с течением времени**.

Пусть $\vec{x}(t) = \{x_1(t), \dots, x_d(t)\}$ — траектория с компонентами x_i , $i = 1, 2, \dots, k$. Разделим d -мерное фазовое пространство M на d -мерные ячейки l^d . Состояние этой системы будем измерять через временные интервалы τ . Пусть $P_{i_0 \dots i_n}$ — совместная вероятность того, что

$$\begin{aligned} \vec{x}|_{t=0}, & \text{ находится в ячейке } i_0; \\ \vec{x}|_{t=\tau}, & \text{ находится в ячейке } i_1; \\ \vec{x}|_{t=2\tau}, & \text{ находится в ячейке } i_2; \\ & \dots \dots \dots \dots \dots \dots \dots; \\ \vec{x}|_{t+n\tau}, & \text{ находится в ячейке } i_n. \end{aligned}$$

Тогда величина

$$K_n = - \sum_{i_0 \dots i_n} P_{i_0 \dots i_n} \ln P_{i_0 \dots i_n}$$

пропорциональна информации, необходимой для определения местоположения системы на заданной траектории $i_1^* \dots i_n^*$ с точностью до l .

Поэтому $K_{n+1} - K_n$ есть **дополнительная информация**, необходимая для **предсказания**, в какой ячейке i_{n+1}^* будет система, если известно, что она находилась в $i_1^* \dots i_n^*$.

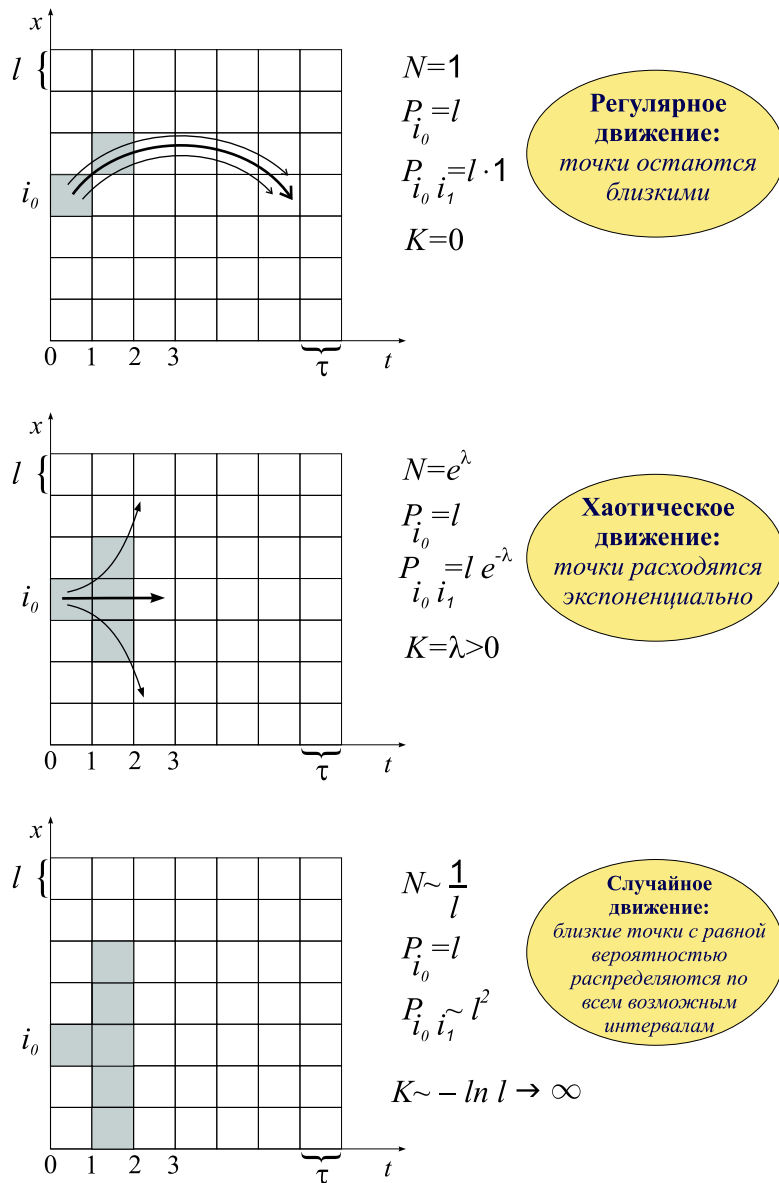
Иными словами, $K_{n+1} - K_n$ описывает потерю информации о системе на интервале $\{n, n + 1\}$. КС-энтропия K определяется как средняя скорость потери информации:

$$K = \lim_{\tau \rightarrow 0} \lim_{l \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{\tau N} \sum_{n=0}^{N-1} (K_{n+1} - K_n) =$$

$$= - \lim_{\tau \rightarrow 0} \lim_{l \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{\tau N} \sum_{i_0 \dots i_N} P_{i_0 \dots i_N} \ln P_{i_0 \dots i_N}.$$

Важно: порядок следования пределов. Предел $l \rightarrow 0$ определяет независимость энтропии от частного вида разбиения. Для отображений с дискретным шагом $\tau = const = 1$ предел $\tau \rightarrow 0$ опускается.

Рассмотрим K для регулярного, хаотичного и случайного (стохастического) движений:



$$N=1$$

$$P_{i_0} = l$$

$$P_{i_0 i_1} = l \cdot 1$$

$$K=0$$

Регулярное движение:

точки остаются близкими

$$N=e^\lambda$$

$$P_{i_0} = l$$

$$P_{i_0 i_1} = l e^{-\lambda}$$

$$K=\lambda > 0$$

Хаотическое движение:

точки расходятся экспоненциально

$$N \sim \frac{1}{l}$$

$$P_{i_0} = l$$

$$P_{i_0 i_1} \sim l^2$$

$$K \sim -\ln l \rightarrow \infty$$

Случайное движение:

близкие точки с равной вероятностью распределяются по всем возможным интервалам

Для одномерных отображений $K = \lambda$ — показатель Ляпунова. Для систем большей размерности информация о состоянии системы с хаосом теряется, т.к. имеет место экспоненциальное "разбегание" траекторий. Эта скорость потери определяется положительными показателями Ляпунова (их может быть несколько):

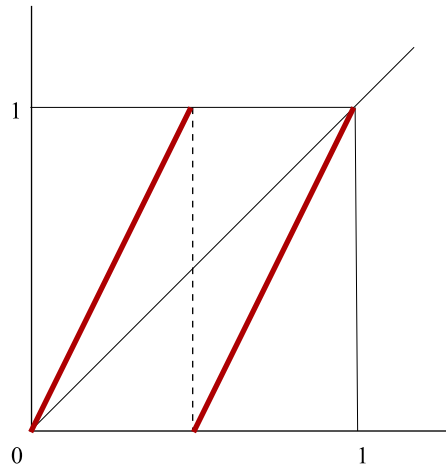
$$K = \int dx^d \rho(\vec{x}) \sum_i \lambda_i^+(\vec{x}),$$

где $\rho(\vec{x})$ — инвариантная плотность на аттракторе. Но для многих систем $\lambda \neq \lambda(\vec{x})$. Тогда интеграл равен 1 и

$$K = \sum_i \lambda_i^+.$$

Используя K , можно **достаточно точно** определить среднее время предсказуемости поведения системы.

Рассмотрим преобразование Бернулли (модель подбрасывания монеты): $x_{n+1} = 2x_n \bmod 1$. Его график представлен на следующем рисунке:



После n шагов интеграл l , где мы хотим определить положение функции, возрастает до размеров $L = le^{\lambda n}$, λ — показатель Ляпунова. Если на каком-то шаге $L > 1$, то невозможно определить местоположение функции на $[0, 1]$. Можно только сказать, что система с вероятностью

$$\rho_0(x) dx$$

находится в интервале $[x, x + dx] \subset [0, 1]$, где $\rho_0(x)$ — инвариантная мера (плотность) системы.

Иначе говоря, точное предсказание состояния возможно на интервале T_m , когда

$$le^{\lambda T_m} = 1 \Rightarrow T_m = \frac{1}{\lambda} \ln \frac{1}{l}.$$

При $t > T_m$ возможны только вероятностные (или статистические) прогнозы.

В общем случае

$$T_m \sim \frac{1}{K} \ln \frac{1}{l},$$

где l — точность задания точки. Отметим, что влияние этой точности l только логарифмическое (т.е. малое).

Подведем некоторые итоги.

Энтропия K дает возможность:

- сказать, хаотична ли система (когда K конечна) или случайна (когда $K \rightarrow \infty$);
- найти меру потери информации о состоянии динамической системы с течением времени.
- K можно определить и из анализа временных рядов.

7 Обобщенная размерность

Преобразуем траекторию $\vec{x}(t) = [x_1(t), \dots, x_d(t)]$ в последовательность \vec{y}_{i+n-1} , а d -мерное фазовое пространство разделим на ячейки l^d .

Тогда **вероятность попадания точки**, принадлежащей предельному множеству (т.е. аттрактору), в i -ю ячейку ($i = 1, 2, \dots, N(l)$), равно:

$$p_i = \lim_{N \rightarrow \infty} \frac{N_i}{N},$$

где N_i —число точек $\{\vec{y}\}$ в этой ячейке.

Вообще говоря, при таком подходе всегда предельное множество неоднозначно. Очевидно, эта неоднозначность не может быть снята только емкостью (фрактальной размерностью d_F).

7.1 Фрактальная, информационная и корреляционная размерности

Для выявления структуры предельного множества было предложено использовать множество размерностей D_q , $q = 0, 1, 2, \dots$, характеризующих *статистическую структуру* (т.е. некоторую степень неоднородности) множества:

$$D_q = \lim_{l \rightarrow 0} \frac{1}{q-1} \frac{\ln \sum_{i=1}^{N(l)} p_i^q}{\ln l}, \quad q = 0, 1, 2, \dots$$

При $q \rightarrow 0$ находим:

$$D_0 = - \lim_{l \rightarrow 0} \frac{\ln \sum_{i=1}^{N(l)} 1}{\ln l} = - \lim_{l \rightarrow 0} \frac{\ln N(l)}{\ln l} = \lim_{l \rightarrow 0} \frac{\ln N(l)}{\ln 1/l}.$$

Это совпадает с определением емкости, $D_0 = d_F$.

При $q \rightarrow 1$ имеем следующее. Введем $\kappa = q - 1$. Тогда легко получить:

$$\begin{aligned} D_1 &= \lim_{l \rightarrow 0} \lim_{\kappa \rightarrow 0} \frac{\ln \left[\sum_{i=1}^{N(l)} p_i e^{\kappa \ln p_i} \right]}{\kappa \ln l} = \lim_{l \rightarrow 0} \lim_{\kappa \rightarrow 0} \frac{\ln \left(1 + \kappa \sum_{i=1}^{N(l)} p_i \ln p_i \right)}{\kappa \ln l} = \\ &= \lim_{l \rightarrow 0} \frac{\sum_{i=1}^{N(l)} p_i \ln p_i}{\ln l}. \end{aligned}$$

Таким образом, величина D_1 показывает, как при $l \rightarrow 0$ возрастает количество информации об изучаемом предельном множестве. Поэтому D_1 называют **информационной размерностью**.

Для однородных предельных множеств все p_i равны, $p_i = 1/N(l)$. Значит в этом случае информационная размерность совпадает с емкостью:

$$D_1 = d_F.$$

При $q = 2$ получим

$$D_2 = \lim_{l \rightarrow 0} \frac{\ln \sum_{i=1}^{N(l)} p_i^2}{\ln l}.$$

Что представляет собой числитель?

Это вероятность того, что **две** точки находятся внутри ячейки со стороной l , т.е. **разделены расстоянием l** . Это позволяет вычислить D_2 через корреляционный интеграл, который, напомним, определяется как

$$C(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{ij} \Theta [l - |\vec{y}_i - \vec{y}_j|].$$

Поэтому можно сказать

$$\begin{aligned} \sum_i p_i^2 &= \text{вероятность того, что две точки лежат внутри } l^d; \\ &\simeq \text{вероятность того, что две точки разделены расстоянием,} \\ &\text{меньшим } l; \\ &= \lim_{N \rightarrow \infty} \frac{1}{N^2} \left\{ \text{число пар } (ij), \text{ для которых расстояние } |\vec{y}_i - \vec{y}_j| < l \right\} \end{aligned}$$



Таким образом, из временного ряда мы всегда определяем корреляционную размерность.

Замечание к определению объема и расстояния в *многомерном пространстве*.

1) Рассмотрим квадрат 4×4 и вписанные в него 4 равные окружности, а также окружность, вписанную в центральную часть (см. рис.10). Вычислим радиус малой окружности в центре: $r_2 = \sqrt{2} - 1$.

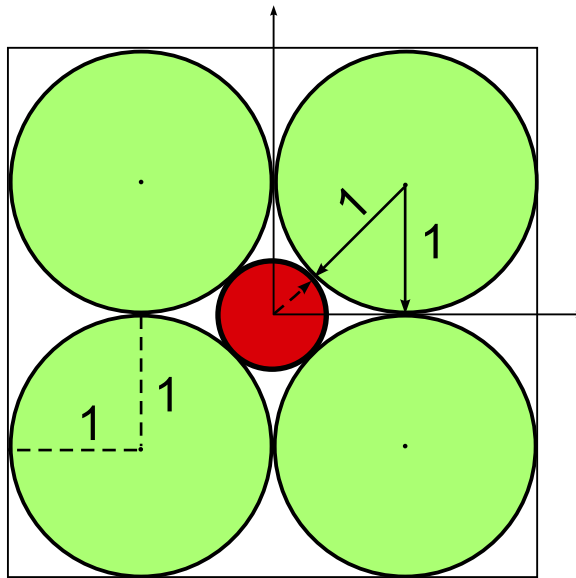


Рис. 10: Квадрат 4×4 и вписанные 4 единичные окружности.

2) Рассмотрим теперь куб и аналогично вписанные сферы. Тогда радиус малой сферы в центре составит $r_3 = \sqrt{3} - 1$.

3) И т.д. В десятимерном пространстве получим, что

$$r_{10} = \sqrt{10} - 1 \approx 2,16 > 2.$$

Таким образом, внутренняя сфера “простирается” за пределы внешнего куба!

Это связано с тем, что при увеличении размерности объем центральной сферы становится сколь угодно большим по сравнению с объемом куба, содержащим все единичные сферы (число которых 2^n). \square

Но вернемся к размерностям. Для $q = 3, 4, \dots$ размерности D_3, D_4 связаны соответственно с кубическими, квадратичными и т.д. корреляционными интегралами и дают дополнительную информацию о геометрической структуре предельного множества.

Справедливо неравенство $D_{q'} \leq D_q$ для $q' > q$. Например, всегда $D_0 \geq D_1$. Равенство имеет место только для **однородного распределения точек на аттракторе**.

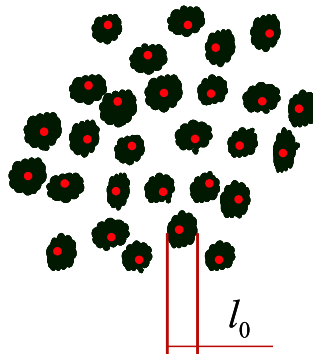
Определена также величина $D_q|_{q \rightarrow \infty}$. Следовательно, D_q всегда ограничена D_0 и D_∞ .

8 Разделение шума и детерминированного сигнала. Оценка энтропии

В этой лекции мы рассмотрим возможность селектирования детерминированного сигнала от случайных (шумовых) компонент. Это позволяет определить размерность вложения для временных рядов (и тем самым число эффективных степеней свободы) которые в реальной ситуации зашумлены.

8.1 Выделение детерминированных переменных

Допустим, что предельное множество вложено в d -мерное пространство и в систему вносится белый шум. Тогда каждая точка во временном ряду *будет окружена равномерно заполненным d -мерным облаком точек*. Радиус облака определяется амплитудой шума l_0 :



При $l \gg l_0$ эти “облака” можно рассматривать в корреляционном интеграле как точки, и наклон графика $\ln C(l)$ от $\ln l$ даст значение корреляционной размерности d_c . Если $l < l_0$, то бóльшая часть точек попадет внутрь заполненных d -мерных ячеек, и наклон становится равным d . Таким образом, получается картина, показанная на рис.11.

С увеличением размерности вложения картина будет меняться. В области малых l наклон растет, но при больших значениях l с ростом размерности вложения наклон перестает меняться (см.рис.12).

|| *Следовательно, можно сказать, сколько в зашумленном ряде основных степеней свободы. Этот результат используется для селекции шума.*

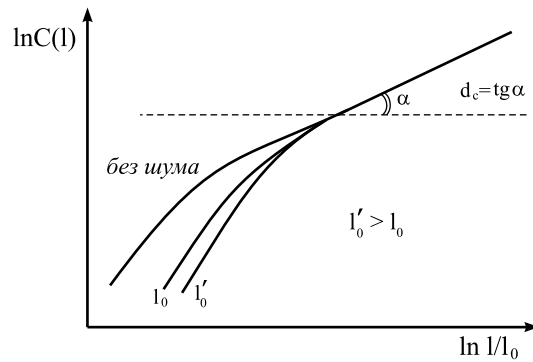


Рис. 11: Типичный вид зависимости корреляционного интеграла для неза шумленного временного ряда.

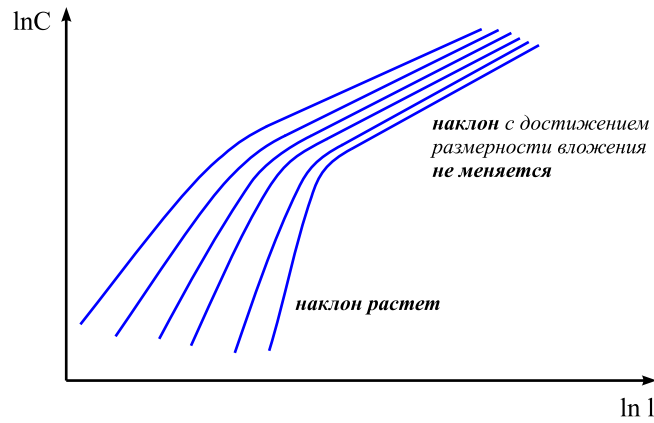


Рис. 12: Пример зависимости корреляционного интеграла для системы с шумом.

8.2 Оценка энтропии

$$K = \lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i_1 \dots i_n} P_{i_1 \dots i_n} \ln P_{i_1 \dots i_n}.$$

По аналогии с D_q обобщим это выражение (можно показать, что это допустимо):

$$K_q = - \lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \frac{1}{q-1} \ln \sum_{i_1 \dots i_n} P_{i_1 \dots i_n}^q.$$

При этом $K_1 = K$ и $K_{q'} \leq K_q$, $q < q'$. Рассмотрим $K_2 \leq K$:

$$K_2 = - \lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{i_1 \dots i_n} P_{i_1 \dots i_n}^2.$$

Обобщим корреляционный интеграл

$$C(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \left\{ \text{число пар } (ij), \text{ для которых } |\vec{y}_i - \vec{y}_j| < l \right\} \simeq \sum_i P_i^2$$

следующим образом:

$$C_n(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \left\{ \text{число пар } (ij) : \sqrt{\sum_{k=0}^{n-1} (\vec{y}_{i+k} - \vec{y}_{j+k})^2} < l \right\} \simeq \sum_{i_1 \dots i_n} P_{i_1 \dots i_n}^2$$

Поэтому, используя K_2 , найдем:

$$K_2 = \lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \ln \frac{C_n(l)}{C_{n+1}(l)} \leq K \rightarrow \lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln C_n(l).$$

C_n нетрудно определить из временного ряда. Условие $K_2 > 0$ говорит о существовании хаоса (см. рис.13).

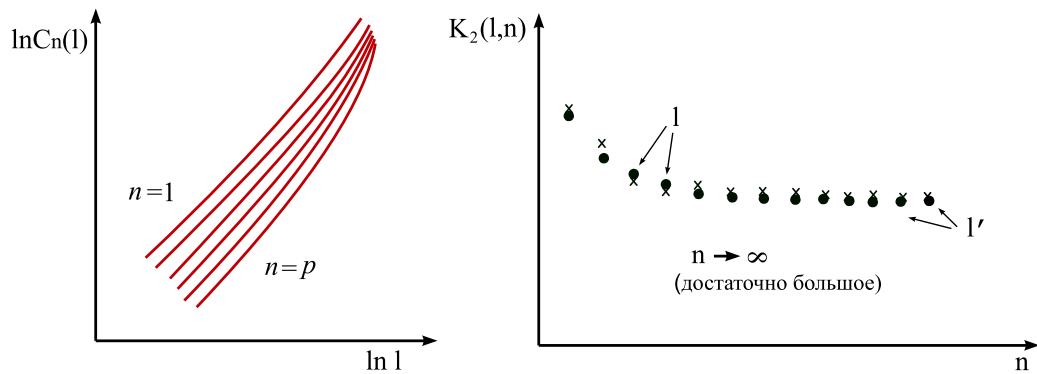


Рис. 13: Расчет корреляционной энтропии K_2 .

9 Математические основы анализа временных рядов

Вопрос, связанный с обоснованием применения описанной теории является довольно сложным. Например, некоторые ее положения остаются до настоящего времени строго не обоснованными.

В данной лекции приводятся некоторые элементы строгой теории, которые могут помочь понять, почему идеи авторегрессии оказываются верными.

9.1 Вложения

Напомним, что подмногообразием пространства M называют всякое подмножество \mathcal{M} в пространстве M' ($M' \subset M$) такое, что локально оно выглядит как «кусочек» пространства M' и имеет в каждой точке единственную касательную гиперплоскость, т.е., как говорят, \mathcal{M} вложено в M гладко.

Локально K -мерное многообразие \mathcal{M}^k — это гладкая K -мерная поверхность, которую можно параметризовать k евклидовыми координатами.

Глобально: сфера, сфера с ручкой, с двумя и т.п. ручками (см.рис.14):

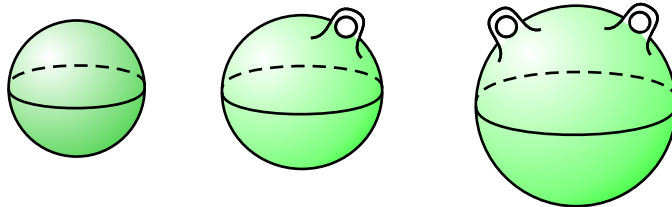


Рис. 14: Сфера, сфера с ручкой и сфера с двумя ручками.

Если в n -мерном евклидовом пространстве \mathbb{R}^k многообразие \mathcal{M}^k реализуется в виде поверхности S^k **без самопересечений**, то S^k вложено в \mathbb{R}^k гладко.

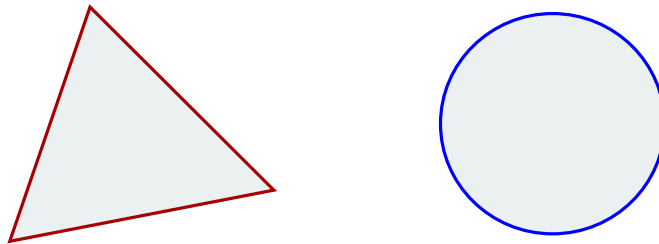
Вложение: дифференциальная функция f , определенная на \mathcal{M}^k , для которой отображение $\mathcal{M}^k \rightarrow S^k$ является взаимно однозначным и существует f^{-1} , отображающая S^k обратно в \mathcal{M}^k , $S^k = f(\mathcal{M}^k)$, $\mathcal{M}^k = f^{-1}(S^k)$.

Выбирая f и n , можно получить различные *представления* одного и того же многообразия. Если все f и f^{-1} дифференцируемы, то говорят,

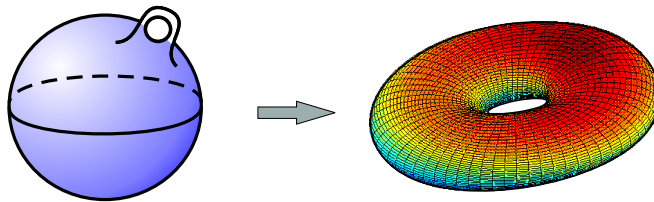
что эти представления *диффеоморфны*.

Примеры.

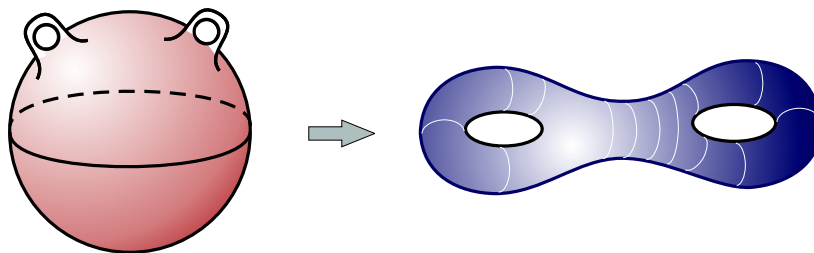
1) Треугольник и окружность не диффеоморфны, но топологически эквивалентны:



2) Сфера с ручкой и тор топологически эквивалентны и диффеоморфны:



3) Сфера с двумя ручками и крендель топологически эквивалентны и диффеоморфны:



Пусть на многообразии \mathcal{M}^k (или на S^k , которое диффеоморфно ему) определена векторная функция необходимой гладкости, отображающая \mathcal{M}^k в \mathbb{R}^m . Что в общем случае получится в \mathbb{R}^m ?

|| Это один из основных вопросов,
|| которые нужно решить при анализе временных рядов.

9.2 О рядах

Пусть задана динамическая система (поток или отображение) с фазовым пространством M , $\vec{F}^t : M \rightarrow M$.

Предположим, что $\vec{x}(t)$ — состояние динамической системы. Тогда, вспоминая

$$\vec{w}_i = \{y_i, y_{i+1}, \dots, y_{i+m-1}\},$$

найдем

$$y_i = h(\vec{x}(t_i)) = h(\vec{F}^{t_i}(\vec{x}_0)),$$

t_i — i -й фиксированный момент времени.

Пусть шаг дискретизации τ . Тогда

$$\begin{aligned} \vec{x}(t_{i+1}) &= \vec{F}^\tau(\vec{x}(t_i)), \\ \vec{x}(t_{i+2}) &= \vec{F}^{2\tau}(\vec{x}(t_i)), \\ &\dots\dots\dots, \\ \vec{x}(t_{i+m-1}) &= \vec{F}^{(m-1)\tau}(\vec{x}(t_i)). \end{aligned}$$

\Downarrow

$$\begin{aligned} y_i &= h(\vec{x}(t_i)) \equiv \Phi_0(\vec{x}(t_i)), \\ y_{i+1} &= h(\vec{x}(t_{i+1})) = h(\vec{F}^\tau(\vec{x}(t_i))) \equiv \Phi_1(\vec{x}(t_i)), \\ y_{i+2} &= h(\vec{x}(t_{i+2})) = h(\vec{F}^{2\tau}(\vec{x}(t_i))) \equiv \Phi_2(\vec{x}(t_i)), \\ &\dots\dots\dots; \\ y_{i+m-1} &= h(\vec{x}(t_{i+m-1})) = h(\vec{F}^{(m-1)\tau}(\vec{x}(t_i))) \equiv \Phi_{m-1}(\vec{x}(t_i)). \end{aligned}$$

Таким образом, существует *связь* между компонентами вектора \vec{w}_i с *одним и тем же* состоянием динамической системы $\vec{x}(t_i)$. Значит, найдется векторная функция Λ , которая отображает векторы $\vec{x}(t_i) \in \mathcal{M}^d \subset M$ в точки m -мерного евклидова пространства \mathbb{R}^m :

$$\vec{w}_i = \Lambda(\vec{x}(t_i)), \quad \vec{x}(t_i) \in \mathcal{M}^d,$$

$\vec{w}_i \in \mathbb{R}^m$, \mathcal{M}^d — многообразие.

Типичное свойство отображения Λ — при $m \geq 2d + 1$ оно дает **вложение** \mathcal{M}^d в \mathbb{R}^m . Образ многообразия \mathcal{M}^d в \mathbb{R}^m обозначим S^d , т.е. $S^d = \Lambda(\mathcal{M}^d)$.

Основное следствие:

На S^d можно определить динамическую систему $\vec{x}(t_i) = \Lambda^{-1}(\vec{w}_i)$, $\vec{x}(t_{i+1}) = \vec{F}^\tau(\vec{w}_i)$, $\vec{w}_{i+1} = \Lambda(\vec{x}(t_{i+1})) = \Lambda(\vec{F}^\tau(\Lambda^{-1}(\vec{w}_i))) \equiv \Psi(\vec{w}_i)$, $\vec{w}_i \in S^d$, $\Psi : S^d \rightarrow S^d$, но вне S^d функция Ψ не определена.

Таким образом, мы имеем две динамические системы:

$$\Phi : M^d \rightarrow M^d$$

и

$$\Psi : S^d \rightarrow S^d.$$

Их связь: $\vec{w} = \Lambda(\vec{x})$.

Основной вывод: инвариантные характеристики исходной системы и полученные из временного ряда должны совпадать, т.е. D_i , K_i , λ_i одинаковы \Rightarrow *Эти характеристики можно определить из эксперимента, не зная всех динамических переменных системы.*

Итак, понятно, почему можно говорить об идентичных свойствах наблюдаемой и исходной системы.

Какие могут возникать трудности из определения τ (времени задержки) и m (размерности вложения)?



Эта задача называется **задачей оптимального выбора параметров**.

10 Оптимизация выбора параметров

Пусть снова имеется N точек, которые являются значениями наблюдаемой, характеризующей *одну и ту же* динамическую систему. Тогда векторы \vec{w} дадут $(N - m)$ точек на $S^d \in \mathbb{R}^m$.

Т.к. число $N < \infty$, то существует характерное расстояние l между точкой и ее ближайшим соседом. Это дает **ограничение на исследование на меньших масштабах** для длинного временного ряда.

Строгих критериев выбора m и τ нет. Но есть эмпирические правила.

Предположим, что, как и прежде, имеется временной ряд $x(t) \equiv y_1$, $x(t + \tau) \equiv y_2, \dots, x(t + (n - 1)\tau) \equiv y_n, \dots$. Обозначим n -мерные векторы, получаемые из элементов временного ряда, как

$$\vec{w}_i = \{y_i, y_{i+1}, \dots, y_{i+n-1}\}.$$

В корреляционном интеграле используется обычное евклидово расстояние

$$\|\vec{w}^i - \vec{w}^j\| = \sqrt{\sum_{k=1}^m (w_k^i - w_k^j)^2}.$$

Очевидно, с ростом m все расстояния монотонно возрастают. При этом среднеквадратичное расстояние

$$\langle \|\vec{w}^i - \vec{w}^j\|^2 \rangle \sim m.$$

Поэтому одни и те же участки в $C(l)$ будут относиться к разным значениям величины l .



сдвиг масштабов

|| Это может препятствовать использовать корреляционный интеграл для расчетов некоторых характеристик.

Следовательно, если $C(l)$ вычисляется для нескольких m , то необходимо использовать **коррекцию метрики**, т.е. нормировать расстояние на m :

$$\|\vec{w}^i - \vec{w}^j\|_{2C} = \sqrt{\sum_{k=1}^m \frac{(w_k^i - w_k^j)^2}{m}}.$$

Обычная метрика *называется* L_2 . Эта последняя *называется* L_{2C} (C — от англ. "correction").

Иногда используют две другие метрики:

$$\|\vec{w}^i - \vec{w}^j\|_1 = \sum_{k=1}^m |w_k^i - w_k^j|,$$

которая нуждается в коррекции,

$$\|\vec{w}^i - \vec{w}^j\|_{1C} = \sum_{k=1}^m \frac{|w_k^i - w_k^j|}{m},$$

и

$$\|\vec{w}^i - \vec{w}^j\|_\infty = \sup_k |w_k^i - w_k^j|,$$

которая в коррекции не нуждается.

Использование: приведенные способы вычисления расстояния (две последних) существенно упрощают алгоритмы расчета.

10.1 Оптимальный выбор m

Поиск ложных соседей: пары векторов \vec{w}_i и \vec{w}_j , которые оказались *близкими*, но их образы *далеки*.

Пусть $\vec{w}_i^{(m)}$, $\vec{w}_j^{(m)}$ — два соседа в пространстве размерности $m + 1$. Если это *истинно близкие соседи*, то, в основном, они будут близки в пространстве \mathbb{R}^{m+1} . Но ложные соседи не будут близки в этом пространстве. Как правило, они превращаются в истинные с ростом m .

Итак, пары, для которых

$$\|\vec{w}_i^{(m)} - \vec{w}_j^{(m)}\| \text{ малó,}$$

а $\|\vec{w}_i^{(m+1)} - \vec{w}_j^{(m+1)}\|$ не малó, называются **ложными соседями**. Если увеличивать m и оценивать их число, то при достижении нужной (правильной) размерности это количество резко уменьшается.

Это понятно из соотношения, что при достижении размерности вложения

$$\vec{w}_{i+1} = \Psi(\vec{w}_i)$$

и для близких \vec{w}_i , \vec{w}_j близки \vec{w}_{i+1} , \vec{w}_{j+1} .

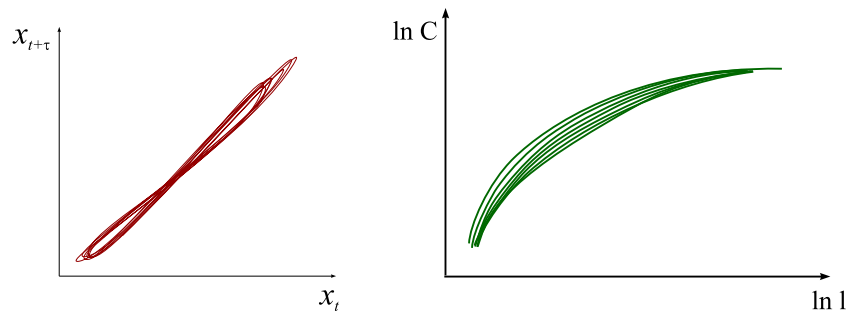
Из этого же ясно, что ложные соседи все же могут остаться: *в хаосе траектории экспоненциально быстро расходятся*.

10.2 Выбор времени задержки

Простейший способ — выбрать τ близким к первому нулю автокорреляционной функции ряда:

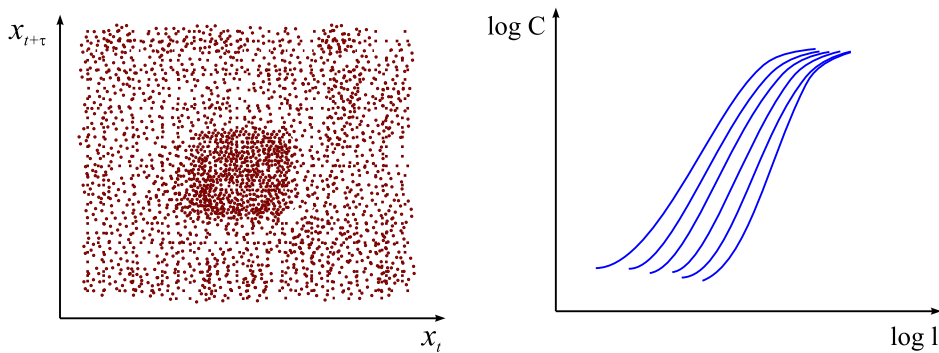
$$g(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t + \tau) dt .$$

Важно: при *малых* τ недостаток информации:



Т.е. имеет место вытягивание вдоль линии \rightarrow существует множество ложных соседей.

При *больших* τ наблюдается следующая картина:



Временной масштаб должен определяться величиной $1/K$. Но как найти K ? \rightarrow Может быть замкнутый круг!

11 Оценка длины ряда

Итак, знаем, как выбрать τ (по первому нулю автокорреляционной функции) и m (по истинным соседям). Рассмотрим, как можно оценить необходимую длину временного ряда. При анализе реальных рядов это является, как правило, главным параметром.

Пусть идеальная ситуация: однородное распределение, размерность множества, на котором сосредоточены точки, равна d . Сколько таких точек необходимо взять?

Для обнаружения скейлинга надо изменить масштаб хотя бы в 10 раз. Если бы рассматриваемое множество было d -мерным кубом со стороной a , то необходимо, чтобы расстояние до ближайшей точки $\sim 0,1a$. Таким образом, на одну точку приходится объем $(0,1a)^d$. Тогда для заполнения куба число точек

$$N \sim \frac{a^d}{(0,1a)^d} = 10^d.$$

Следовательно, если $N < 10^5$, то множества с $d > 5$ невозможно корректно обработать.

В общем случае возникают *трудности, если*

$$d < \lg N.$$

11.1 Оценка из свойств корреляционного интеграла

Пусть график $\lg C(\lg l)$ — идеальный отрезок прямой с наклоном d_C (см. рис.15):

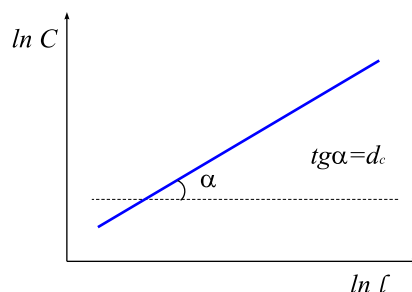


Рис. 15: Наклон $\lg C(\lg l)$ в идеальном случае.

Очевидно, максимальное значение $C_{max} = 1$, а минимальное $C_{min} \simeq 1/N^2$, т.к.

$$C(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{n=1}^N \sum_{n'=1}^N \Theta[l - \rho(n, n')].$$

Это нормированное на N^2 количество пар точек, расстояние между которыми меньше l . Тогда \log изменения масштаба, соответствующего этому линейному участку, составит

$$\lg \rho = \frac{\lg C_{max} - \lg C_{min}}{d} \simeq 2 \frac{\lg N}{d}.$$

Мы приняли, что $\rho \sim 10$. Это дает

$$1 \simeq 2 \frac{\lg N}{d}.$$

Или $d < 2 \lg N$. Для $N = 10^5$ получим $d < 10$.

Это все для **идеального** случая, и **правильная асимптотика** получится при

$$N > 10^d.$$

.....

В реальных ситуациях, однако,

$$N > N_0 10^d,$$

где N_0 — некоторое число, большее 1. Оно возникает вследствие нелинейности и неоднородных свойств аттрактора. Поэтому при $l > l_{fr}$ (fr. от fractal) график $C(l)$ утрачивает линейность. Чем выше l_{fr} , тем больше число N_0 .

Оценим N_0 . Пусть \vec{x}_1 и \vec{x}_2 — достаточно близкие точки на аттракторе. Тогда

$$\left. \begin{aligned} \hat{\vec{x}}_1 &= \varphi(\vec{x}_1, t), \\ \hat{\vec{x}}_2 &= \varphi(\vec{x}_2, t). \end{aligned} \right\} \text{эволюция.}$$

Поэтому вектор $\vec{y} = \vec{x}_2 - \vec{x}_1$ отобразится как

$$\hat{\vec{y}} = \varphi(\vec{x}_1 + y, t) - \varphi(\vec{x}_1, t) \equiv g(\vec{x}_1, \vec{y}, t).$$

Предположим, что $\|\vec{y}\|$ мало настолько, что отображение $g(\vec{x}_1, \vec{y}, t)$ линейно на малых временах t . Тогда масштабу, где заканчивается линейная часть C отвечает "утрата линейности" отображения g .

Что это означает? Если система с хаосом, то отображение $g(\vec{x}_1, \vec{y}, t)$ должно быть в среднем растягивающим, а длина вектора \vec{y} должна расти как $\sim e^{\lambda t}$, где λ — показатель Ляпунова.

На больших масштабах **нелинейные члены становятся существенными и отношение нельзя считать линейным**. Таким образом,

$$\hat{\vec{y}} = g(\vec{x}, \vec{y}, t) = \varphi(\vec{x} + \vec{y}, t) - \varphi(\vec{x}, t) = D\varphi(\vec{x}, t)\vec{y} + \frac{1}{2}D^2\varphi(\vec{x}, t)(\vec{y}, \vec{y}) + \dots$$

Предел нелинейности можно выбрать, когда величины линейного и квадратного членов совпадают:

$$\|D\varphi(\vec{x}, t)\vec{y}\| \simeq \|D^2\varphi(\vec{x}, t)(\vec{y}, \vec{y})\| \approx l_{fr}.$$

Используем оценки:

$$\|D\varphi(\vec{x}, t)\vec{y}\| \simeq \|D\varphi(\vec{x}, t)\| \|\vec{y}\|,$$

$$\|D^2\varphi(\vec{x}, t)(\vec{y}, \vec{y})\| \simeq \|D^2\varphi(\vec{x}, t)\| \|y\|^2,$$

Тогда

$$\|y\| \simeq \frac{\|D\varphi(\vec{x}, t)\|}{\|D^2\varphi(\vec{x}, t)\|};$$

$$l_{fr}(\vec{x}, t) \simeq \frac{\|D\varphi(\vec{x}, t)\|^2}{\|D^2\varphi(\vec{x}, t)\|}.$$

Но эта величина характеризует лишь **одно значение t** и только **одну точку \vec{x}** .

Для аттрактора в целом: поведение средних величин при достаточно больших t , т.е.

$$l_{fr} = \lim_{t \rightarrow \infty} \exp \left\langle \ln \frac{\|D\varphi(\vec{x}, t)\|^2}{\|D^2\varphi(\vec{x}, t)\|} \right\rangle_{\vec{x}}.$$

Здесь логарифм выбран потому, что влияние малых членов *не будет после усреднения пренебрежимо малым*.

Таким образом, мы получаем основной **сценарий обработки** рядов:

- выбрать длинный ряд; он должен быть как можно ровнее, без сильных трендов;
- если есть видимые изменения (падения, рост), обрезать ряд до значения, где это происходит;
- оценить число точек N ;
- проверить сходимость $C(l)$.
- обработать ряд, т.е. найти m , d_c , d_e , K_2 и др.;
- построить зависимость $\rho(r)$, найти значение d_e и сверить полученное значение с результатом, полученным методом корреляционного интеграла. *Результаты, строго говоря, могут и не совпадать!*

12 Основные методы прогноза нерегулярных временных рядов

В зависимости от области определения параметров аппроксимирующей функции выделяют две основные группы методов:

- **глобальные** — параметры аппроксимирующей функции идентифицируются посредством использования всех известных значений ряда. Основное приложение — получение глобальных характеристик системы. К ним относятся: авторегрессионные модели, сингулярный спектральный анализ.
- **локальные** методы основаны на принципе локальной аппроксимации (LA). Применяется в целях прогнозирования динамики ряда.

Преимущества локальных методов прогноза *нерегулярных* (хаотических, квазипериодических) рядов:

- применение не требует априорной информации о системе, породившей временной ряд;
- нет необходимости в построении специфической модели динамики исследуемого ряда;
- использования для прогнозирования наиболее близких к *стартовой точке* (стартовому вектору) значений ряда;
- использование меньшего количества исходных данных;
- горизонт прогноза ограничивается не возможностями метода, а особенностями динамики ряда.

12.1 Сингулярный спектральный анализ (SSA)

Метод **сингулярного спектрального анализа (SSA)** используется для определения основных составляющих временного ряда и подавления шума. Метод SSA позволяет:

- различать составляющие временного ряда, полученные из последовательности значений какой-либо величины, взятой через равные промежутки времени;
- находить заранее неизвестные периодичности ряда;
- сглаживать исходные данные на основе отобранных составляющих;
- наилучшим образом выделять компоненту с заранее известным периодом;
- предсказывать дальнейшее поведение наблюдаемой зависимости.

В основе SSA лежит построение **множества векторов задержек**:

$$\mathbf{x}_t = (x_t \ x_{t-1} \ \cdots \ x_{t-p+1})^T, \quad t = p, p+1, \dots, N.$$

Метод задержек устанавливает переход от исходного одномерного (скалярного) временного ряда к многомерному (векторному) представлению. Каждый многомерный вектор образуется из некоторого числа p следующих друг за другом значений исходного временного ряда:

$$\mathbf{X}_{p \times (N-p+1)} = \begin{pmatrix} \vdots & \vdots & & \\ & x_{p+1} & & \\ \left[\begin{array}{c} x_p \\ \vdots \\ x_2 \\ x_1 \end{array} \right] & \left[\begin{array}{c} x_{p+1} \\ \vdots \\ x_3 \\ x_2 \\ x_1 \end{array} \right] & \cdots & \left[\begin{array}{c} x_N \\ \vdots \\ x_{N-p+2} \\ x_{N-p+1} \\ x_{N-p} \\ \vdots \end{array} \right] \\ \downarrow & \downarrow & & \end{pmatrix}$$

Здесь каждая квадратная скобка — вектор в p -мерном пространстве задержек; последовательность таких векторов задает **матрицу задержек** $\mathbf{X}_{p \times (N-p+1)}$, где N — число элементов исходного ряда.

Особенностью SSA является обработка матрицы \mathbf{X} по алгоритму, похожему на метод главных компонент. Суть метода **главных компонент** состоит в снижении размерности исходного пространства факторов (задержек) с помощью ортогонального линейного преобразования. Полученные таким образом новые переменные и называют *главными компонентами*. Применение этого метода позволяет *сгладить исходный ряд, снизить уровень случайных возмущений, повысить отношение сигнал/шум*.

Способы прогнозирования на базе SSA следующие: (а) “Гусеница” и (б) метод авторегрессии, который применяется по отдельности к каждой из выбранных компонент разложения. Ниже представлены основные этапы применения SSA к конкретному ряду $(x_i)_{i=1}^N$.

1 Преобразование одномерного ряда в многомерный.

Для такого преобразования необходимо взять некоторое число задержек $\tau \leq [(N + 1)/2]$:

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_\tau & \dots & x_n \\ x_2 & x_3 & x_4 & \dots & x_{\tau+1} & \dots & x_{n+1} \\ x_3 & x_4 & x_5 & \dots & x_{\tau+2} & \dots & x_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_\tau & x_{\tau+1} & x_{\tau+2} & \dots & x_{2\tau-1} & \dots & x_N \end{pmatrix}$$

Построенная матрица X является прямоугольной, но в предельном случае, т.е. при $\tau = (N + 1)/2$ и нечетном N , она вырождается в квадратную.

2 Построение для матрицы X соответствующей ковариационной матрицы:

$$C = \frac{1}{n} X X^T.$$

3 Определение собственных значений и собственных векторов матрицы C .

Для этого ее необходимо разложить по собственным векторам: $C = V \Lambda V^T$, где введены:

диагональная матрица собственных чисел:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_\tau \end{pmatrix}$$

и ортогональная матрица собственных векторов матрицы C :

$$V = (V^1, V^2, \dots, V^\tau) = \begin{pmatrix} v_1^1 & v_1^2 & \dots & v_1^\tau \\ v_2^1 & v_2^2 & \dots & v_2^\tau \\ \vdots & \vdots & \ddots & \vdots \\ v_\tau^1 & v_\tau^2 & \dots & v_\tau^\tau \end{pmatrix}$$

Ясно, что $\Lambda = V^T C V$, $\det C = \prod_{i=1}^{\tau} \lambda_i$ и $\sum_{i=1}^{\tau} \lambda_i = \tau$ (последнее равенство справедливо только в случае предварительно нормированных строк матрицы X).

4 Переход к главным компонентам

Матрицу собственных векторов V обычно представляют как матрицу перехода к **главным компонентам** $Y = V^T X = (Y_1, Y_2, \dots, Y_{\tau})$ исходного ряда, где величины $Y_i, i = 1, 2, \dots, \tau$, — суть строки длины n . При этом собственные значения $\lambda_1, \lambda_2, \dots, \lambda_{\tau}$ можно рассматривать как вклад главных компонент $Y_1, Y_2, \dots, Y_{\tau}$ в общее информационное содержание временного ряда $(x_i)_{i=1}^N$. Тогда по полученным главным компонентам можно полностью восстановить исходную матрицу

$$X = (V^1, V^2, \dots, V^{\tau}) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{\tau} \end{pmatrix} = \sum_{i=1}^{\tau} V^i Y_i,$$

а по ней в свою очередь можно восстановить временной ряд $(x_i)_{i=1}^N$.

Преобразования $Y_i = (V^i)^T X, i = 1, 2, \dots, \tau, Y_i[l] = \sum_{q=1}^{\tau} v_q^i x_q^l, l = 1, 2, \dots, n$, являются *линейными фильтрами*.

Каждый i -й собственный вектор включает в себя τ компонент, т.е. $V^i = (v_1^i, v_2^i, \dots, v_{\tau}^i)^T$. Построим зависимость значений компонент $v_k^i, k = 1, 2, \dots, \tau$, от их номера: $v^i = v^i(k)$. Тогда, используя свойство ортогональности собственных векторов, дальнейший анализ последовательности $(x_i)_{i=1}^N$ можно провести, изучая диаграммы, построенные по аналогии с *фигурами Лиссажу*. Именно по осям попарно откладываются компоненты v_k^i, v_k^j . Если построенные диаграммы оказываются близкими по форме к окружности, то функции $v^i = v^i(k), v^j = v^j(k)$ будут подобны периодическим с близкими амплитудами и со сдвигом фазы около четверти периода.

Таким образом, для некоторых пар собственных векторов V^i, V^j можно вычислить величину, имеющую смысл периода. Следовательно-

но, графический анализ дает представление о частотах компонент, формирующих исходный временной ряд $(x_i)_{i=1}^N$.

5 SSA–сглаживание

Предположим, что из τ компонент для дальнейшего анализа оставлены лишь первые r . Тогда для восстановления исходной матрицы X используются первые r собственных векторов V^i . В таком случае

$$\tilde{X} = (V^1, V^2, \dots, V^\tau) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_\tau \end{pmatrix} = \sum_{i=1}^r V^i Y_i,$$

где \tilde{X} — восстановленная матрица, имеющая n столбцов и τ строк. Теперь исходный временной ряд, воссозданный из этой матрицы, определится как

$$\tilde{x}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^s \tilde{x}_{i,s-i+1}, & 1 \leq s \leq \tau, \\ \frac{1}{\tau} \sum_{i=1}^{\tau} \tilde{x}_{i,s-i+1}, & \tau \leq s \leq n, \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{x}_{i+s-n,n-i+1}, & n \leq s \leq N. \end{cases}$$

Данный способ получения последовательности $(\tilde{x}_i)_{i=1}^N$ называется **SSA-сглаживанием** исходного временного ряда $(x_i)_{i=1}^N$ по первым r компонентам из τ .

12.2 Прогноз методом SSA

6 Прогнозирование исходной временной последовательности

На этом этапе строится ряд $(x_i)_{i=1}^{N+p}$. При этом прогноз на p точек вперед осуществляется как применение p раз операции прогноза на одну точку.

Базовая идея нахождения значения x_{N+1} состоит в следующем. Пусть имеется набор значений x_1, x_2, \dots, x_N . Теперь построим выборку в виде матрицы X . В качестве базиса поверхности, содержащей эту выборку, можно взять отобранные ранее собственные векторы V^1, V^2, \dots, v^r матрицы C .

Запишем параметрическое уравнение этой поверхности как $S(P) = \sum_{i=1}^r p_i V^i$, где каждому значению вектора $S(P)$, представляющего собой столбец высотой τ , соответствует набор из r параметров p_i . Тогда k -му ($k = 1, 2, \dots, n$) столбцу исходной матрицы X соответствует свой набор значений параметров $P^k = (p_1^k, p_2^k, \dots, p_r^k)$ и, следовательно,

$$X^1 = S(P^1), X^2 = S(P^2), \dots, X^n = S(P^n).$$

Для предсказания значения x_{N+1} необходимо построить $(n+1)$ -й столбец X^{n+1} матрицы X , которому соответствует некоторое значение параметров $P^{n+1} = (p_1^{n+1}, p_2^{n+1}, \dots, p_r^{n+1})$. Этот набор параметров можно найти из соотношения $S(P) = \sum_{i=1}^r p_i V^i$, исходя лишь из данных $(x_i)_{i=1}^N$. При этом прогнозируемый столбец будет записан следующим образом: $X^{N+1} = S(P^{n+1})$.

Введем следующие обозначения:

$$V_* = \begin{pmatrix} v_1^1 & v_1^2 & \dots & v_1^r \\ v_2^1 & v_2^2 & \dots & v_2^r \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau-1}^1 & v_{\tau-1}^2 & \dots & v_{\tau-1}^r \end{pmatrix},$$

$$\tilde{p} = \begin{pmatrix} \tilde{p}_1^{n+1} \\ \tilde{p}_2^{n+1} \\ \vdots \\ \tilde{p}_r^{n+1} \end{pmatrix}, Q = \begin{pmatrix} x_{N-\tau+2} \\ x_{N-\tau+3} \\ \vdots \\ x_N \end{pmatrix},$$

$$V_\tau = (v_\tau^1, v_\tau^2, \dots, v_\tau^r).$$

Параметры $(p_1^{n+1}, p_2^{n+1}, \dots, p_r^{n+1})$ можно определить из системы уравнений $V_* \tilde{P} = Q$ относительно \tilde{P} . Таким образом, окончательное выражение для прогнозируемого значения будет выглядеть как

$$x_{N+1} = \frac{V_\tau V_*^T Q}{1 - V_\tau V_\tau^T}.$$

Для предсказания следующих значений в простейшем случае требуется лишь изменить соответствующим образом матрицу Q и вновь умножить ее на величину $V_\tau V_*^T / (1 - V_\tau V_\tau^T)$. Однако дополнительно при этом можно для каждой следующей точки повторять частично или полностью весь алгоритм метода SSA. В этом случае будут также меняться матрицы V_τ и V_* .

7 Выбор количества задержек τ для построения многомерной выборки X .

Как и в случае с отбором главных компонент, выбор значения τ существенным образом зависит от исследуемой проблемы.

Пусть задача состоит в сглаживании ряда методом SSA, т.е. в восстановлении ряда с использованием известных периодичностей. Тогда выделение главной компоненты есть фильтрация ряда с переходной функцией фильтра в виде собственного вектора этой главной компоненты. Чем больше величина τ , тем больше параллельных фильтров, уже полоса пропускания каждого из них и лучше сглаживание ряда.

Если требуется определить неизвестные (скрытые) периодичности в наблюдаемой последовательности, то следует сначала взять как можно большее значение τ . Затем, после отбрасывания близких к нулю собственных чисел, величину задержки необходимо сократить.

Предположим, что необходимо выделить одну известную периодическую компоненту. В этом случае следует выбрать значение τ , равное искомому периоду.

Наконец, пусть задача заключается в продолжении изучаемого ряда на заданную величину (т.е. в предсказании эволюции наблюдаемого процесса). Тогда следует взять максимально допустимое значение τ и затем подбирать величину r .

Примеры использования SSA–“Гусеница” для прогнозирования временных рядов.

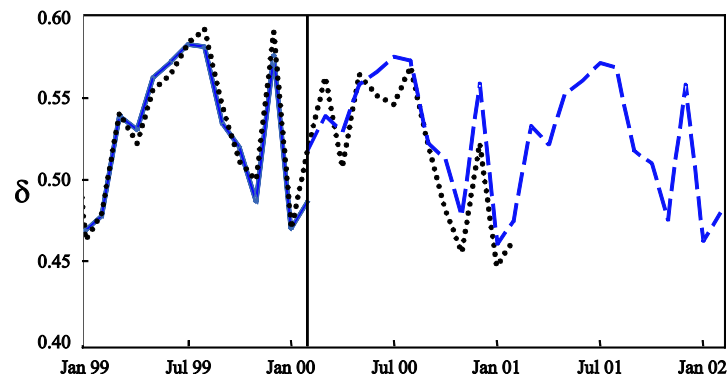


Рис. 16: Фактическая динамика, реконструкция и прогноз отношения товарных запасов к продажам товаров длительного пользования в США.

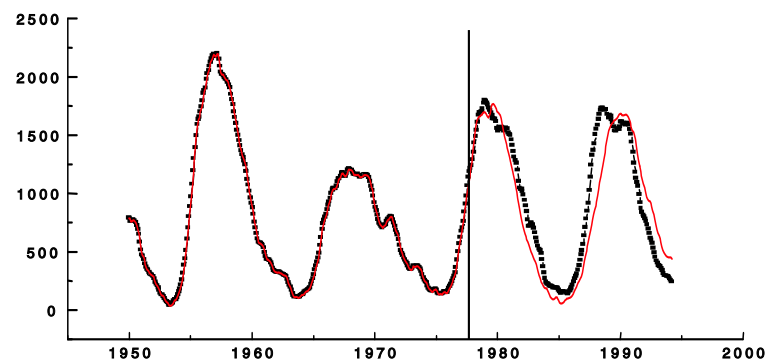


Рис. 17: Прогноз по методу SSA–“Гусеница” и реальные значения чисел Вольфа, характеризующие солнечную активность. Вертикальная линия отделяет начало прогноза.

Первый пример демонстрирует возможности метода для прогноза временного ряда отношения товарных запасов к продажам товаров длительного пользования в США. Результат прогноза иллюстрирует рис.16.

Второй пример связан с прогнозированием солнечной активности, характеризуемой значениями чисел Вольфа. Прогноз на 18 лет, т.е. примерно на полтора 11-летних солнечных цикла оказался весьма точным (см. рис.17).

12.3 Локальная аппроксимация (LA)

В соответствии с теорией Такенса–Мане приемлемое описание фазового пространства динамической системы можно получить, если взять вместо реальных переменных системы p -мерные векторы задержек из значений ряда в последовательные моменты времени. При выполнении условия $p \geq 2d + 1$, где d — размерность вложения, возможно реконструировать фазовое пространство (пространство состояний) системы. При условии стационарности временного ряда на базе этой реконструкции строится прогноз его дальнейшей динамики.

Способы определения величины p :

— Алгоритм Грассбергера–Прокаччия. *Недостаток метода*: неэффективен при работе с короткими (до 10^4 точек) временными рядами.

— Другие методы тоже имеют свои *недостатки*: сложность реализации, большая длительность расчетов, неоднозначность либо сомнительность результатов.

В связи с этим величина p , за исключением модельных примеров, в которых она достоверно известна, как правило, определяется **эмпирически**. *Главный критерий* в этом случае — выбор такого p , начиная с которого прекращается качественное изменение прогноза.

12.4 Построение прогноза на один шаг вперед

1. Построение матрицы задержек и выбор локального представления

Преобразуем скалярный временной ряд, содержащий N значений наблюдаемой, в *матрицу задержек*:

$$\{x_1, x_2, \dots, x_N\} \rightarrow \mathbf{X}_{p \times (N-p+1)} = \begin{pmatrix} x_p & x_{p+1} & \cdots & x_N \\ \vdots & \vdots & \ddots & \vdots \\ x_2 & x_3 & \cdots & x_{N-p+2} \\ x_1 & x_2 & \cdots & x_{N-p+1} \end{pmatrix}.$$

Размерность матрицы \mathbf{X} определяется количеством задержек p .

После построения матрицы задержек выбирается *вид локального представления*, т.е. вид функции, связывающей следующие значения ряда с

предыдущими:

$$x_{t+1} = f(\mathbf{x}_t, \mathbf{a}),$$

где \mathbf{a} — вектор параметров представления.

Наиболее распространенный вариант — это линейная аппроксимация первого (LA1) порядка:

$$x_{t+1} = a_0 + \mathbf{x}_t^T \mathbf{a}.$$

Кроме него, используются еще два варианта: линейная аппроксимация нулевого (LA0) порядка

$$x_{t+1} = a_0$$

линейная аппроксимация второго (LA2) порядка

$$x_{t+1} = a_0 + \mathbf{x}_t^T \mathbf{a} + \mathbf{x}_t^T \mathbf{B} \mathbf{x}_t,$$

где $B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{12} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1p} & b_{2p} & \cdots & b_{pp} \end{pmatrix}$ — матрица дополнительных параметров представления.

2. Выбор ближайших соседей — выделение локальной подобласти фазового пространства

Будем считать **соседними** к стартовому вектору \mathbf{x}_{N-p+1} векторы, удовлетворяющие условию:

$$\{\mathbf{x}_s\} : \sum_{s \in \omega_{\Xi}} \|\mathbf{x}_{N-p+1} - \mathbf{x}_s\| \rightarrow \min_{\omega_{N-p+1}}, \quad (1)$$

где $s \in \omega_{\Xi} \subset \omega_{N-p+1}$, $\omega_{N-p+1} \equiv \{1, \dots, N - p + 1\}$, Ξ — количество соседей, ω_{Ξ} — набор номеров векторов-соседей. Норма обычно берется евклидова, хотя возможно использование и других норм. Выбор соседей автоматически определяет локальную подобласть, в которой параметры представления полагаются неизменными.

Допустим также и обратный вариант: выбрав некоторую окрестность, т.е. задавшись некоторым критерием вида

$$\|\mathbf{x}_{N-p+1} - \mathbf{x}_s\| \leq \varepsilon, \quad (2)$$

можно также определить соседей. Однако этот критерий практически не применяется, так как выбор величины ε является в достаточной мере произвольным, тогда как из-за необходимости оценки параметров на количество соседей налагается существенное ограничение снизу: $\Xi \geq p + 1$. В реальных расчетах стараются придерживаться соотношения $\Xi \cong 3(p + 1)$, чтобы получать более устойчивые оценки параметров. Таким образом, если выбрать условие (2) и подобрать величину ε , обеспечивающую нужное число соседей, все сведется к условию (1).

Данный этап алгоритма — *главный отличительный признак метода и самый важный его этап*. Выбор “хороших” соседей определяет качество прогноза.

3. Оценка параметров модели и построение прогноза на один шаг вперед

Выбором типа линейной аппроксимации и размерности реконструкции p устанавливается количество неизвестных параметров, которые требуется оценить. При этом круг используемых данных ограничивается набором соседей стартового вектора.

Параметры модели (вектор \mathbf{a}) оцениваются *методом наименьших квадратов (МНК)*. Это самый распространенный и наиболее эффективный метод. Оценка по МНК для вектора \mathbf{a} , обозначим ее как $\hat{\mathbf{a}}$, находится из условия:

$$\hat{\mathbf{a}} : \sum_{\omega_s} (x_{s+1} - f(\mathbf{x}_s, \hat{\mathbf{a}}))^2 \rightarrow \min_{\hat{\mathbf{a}}}.$$

Однако в большинстве случаев применение МНК в чистом виде невозможно из-за *вырожденности матрицы факторов*, что является следствием взаимной близости соседей. Поэтому обычно применяется так называемое *сингулярное разложение* (SVD — Singular Value Decomposition). При этом не всегда учитывается, что оценки, даваемые этим методом, в общем случае смещенные, сильно зависящие от машинной точности и выбора минимального значащего сингулярного числа. Следовательно, при использовании SVD важно всегда контролировать устойчивость получаемых результатов.

Оценив значения параметров аппроксимации, нетрудно построить про-

гноз следующего значения ряда (стартовый вектор, обозначим индексом L):

$$\hat{x}_{L+1} = f(\mathbf{x}_L, \hat{\mathbf{a}}).$$

Таким образом, пройдя три описанных шага алгоритма LA, можно построить прогноз одного нового значения ряда.

Примером использования данного алгоритма может служить прогноз температуры воздуха. Один из возможных способов построения прогноза температуры на следующий день состоит в том, чтобы, найдя в какой из предшествующих дней температура была максимально близкой к сегодняшней (случай $p = 1$), взять в качестве прогноза температуры на завтра ее величину в следующий за найденным день. Аппроксимация более высокого порядка позволяет учитывать влияние на прогноз отклонений “сегодняшней” температуры от температуры в день, наиболее похожий на сегодняшний.

13 Сравнение различных методов прогноза LA и прогноз на несколько шагов

Сравним основные способы прогноза, использующие принцип локальной аппроксимации. Сравнение проводится на временных рядах, полученных для системы Лоренца

$$\begin{aligned}\dot{x} &= \sigma(y - x), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= zy - bz\end{aligned}$$

при $\sigma = 5$, $r = 15$, $b = 1$.

13.1 Порядок аппроксимации

При ограниченном количестве данных важным фактором становится **минимально необходимое количество точек** для оценки параметров. Оно *зависит от порядка аппроксимации*: для нулевого порядка достаточно одной точки, так как есть всего один неизвестный параметр a_0 . При линейной аппроксимации LA1 неизвестных параметров $p+1$. **Роль порядка аппроксимации возрастает с уменьшением длины исходного ряда. Для очень длинных рядов роль порядка аппроксимации значительно снижается.**

Для получения надежных оценок количество соседей должно быть достаточно большим. Существует эмпирическое правило, в соответствии с которым для успешного применения МНК соседей должно быть как *минимум втрое больше*, чем число оцениваемых параметров. Таким образом, для LA1 в каждой локальной подобласти желательно иметь не менее $3(p+1)$ известных векторов.

Порядок аппроксимации можно повышать и дальше, однако при этом резко возрастает количество необходимых соседей и увеличивается риск выхода из области справедливости **гипотезы о постоянстве коэффициентов**, т.е. происходит глобализация метода с потерей преимуществ локальной аппроксимации. Поэтому локальную аппроксимацию выше второго порядка (LA2) обычно не используют.

LA2 отличается от LA1 добавлением в модель парных произведений

координат векторов-соседей:

$$x_{t+1} = a_0 + \sum_{i=1}^p a_i x_{t-p+i} + \sum_{j=1}^p \sum_{k \geq j}^p b_{jk} x_{t-p+j} x_{t-p+k}.$$

Минимальное количество соседей здесь рано $p + 1 + p(p + 1)/2$, и их число растет как квадрат размерности реконструкции. Поэтому в число соседей могут попасть векторы, которые очень непохожи на стартовый.

Таким образом, главная причина неэффективности локальной аппроксимации старших порядков — одновременное использование большого числа векторов задержек.

13.2 Сравнение методов

Результат использования большого числа соседей (часть которых с трудом можно так назвать) показан на рис.18.

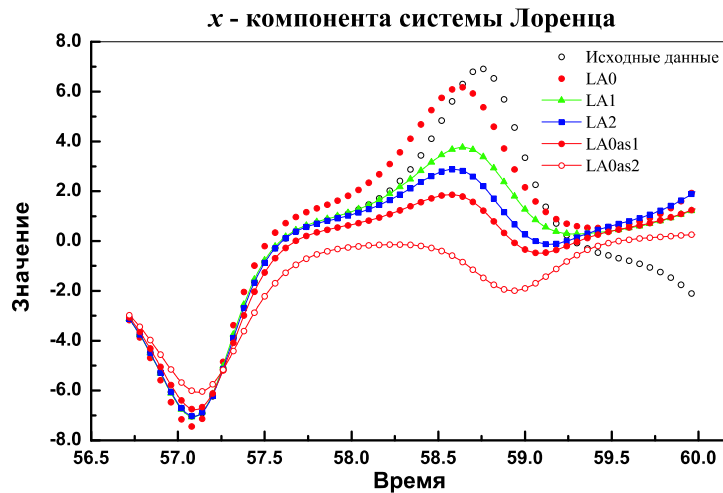


Рис. 18: Фактическая динамика и результат прогнозирования x -компоненты системы Лоренца методом LA нулевого, первого и второго порядков.

Сгенерированный ряд длиной 3600 точек разбивался на две части, первая из которых использовалась для оценки параметров аппроксимации, а по второй (последние 164 точки) проверялось качество прогноза. На рисунке показана только вторая часть исходного ряда и результаты прогнозов, полученные при разных порядках аппроксимации. Параметр

p принят равным пяти. Число соседей взято втрое больше минимально необходимого, т.е. для LA0 — 3 соседа, LA1 — $3(5 + 1) = 18$ соседей, LA2 — 63. Видно, что на коротких интервалах времени LA0 уступает остальным в точности, а LA1 и LA2 дают близкие результаты. Естественно, по одному примеру нельзя судить о том, какой вариант LA1 или LA2 лучше, но важно другое — роль числа соседей.

Для сравнения на этом же рисунке приведены прогнозы методом LA0 с тем же числом соседей, как для LA1 (линия LA0as1) и LA2 (линия LA0as2). Легко видеть, что с ростом количества соседей происходит “сглаживание” прогноза LA0, вызванное увеличением различий между векторами, попавшими в число соседей стартового вектора, а следовательно, и их прогнозами.

Таким образом, локальная аппроксимация первого порядка имеет по сравнению с аппроксимацией второго порядка два существенных **преимущества**. Во-первых, *она требует меньше данных, что важно для коротких рядов*. Во-вторых, *получить более точное приближение легче в малой области, чем в большой, так как в первом случае различия между величинами менее выражены*.

13.3 Прогноз на несколько шагов вперед

Способы, построения прогноза динамики ряда на несколько шагов: **итеративный, итеративный вариант с пересчетом и прямой**.

1) Итеративный способ прогноза

Пусть требуется построить прогноз на T шагов вперед. Идея итеративного способа состоит в *последовательном построении прогноза на один шаг с добавлением его результата к исходным данным*. Иными словами, сначала делается прогноз на один шаг, а затем на основе полученного значения строится новый вектор в пространстве задержек:

$$\hat{\mathbf{x}}_{L+1} = \begin{pmatrix} \hat{x}_{L+1} \\ x_L \\ \vdots \\ x_{L-p+2} \end{pmatrix}.$$

Этот вектор подставляется в правую часть соотношения $\hat{x}_{L+1} = f(\mathbf{x}_L, \hat{\mathbf{a}})$ и делается прогноз еще на один шаг. И так далее. При этом параметры $\hat{\mathbf{a}}$ выбираются те же, что и на первом шаге. Таким образом, каждое следующее прогнозируемое значение выражается через предыдущее (предыдущие). Отсюда название способа — итеративный.

Основное преимущество данного способа — простота и легкость прогнозирования следующих значений, так как оценка параметров модели проводится всего однажды — на первом шаге. **Основной недостаток** — предпосылка, что соседями всех строящихся новых векторов задержек являются те же векторы, которые были соседями стартового.

2) Итеративный способ с пересчетом параметров

Итеративный способ с пересчетом предполагает пересчет параметров $\hat{\mathbf{a}}$. Чтобы построить прогноз на два шага, после получения прогноза на один шаг и определения нового вектора в пространстве задержек, который будет теперь считаться стартовым, необходимо полностью повторить процедуру прогнозирования на один шаг и так далее.

Главный недостаток — это высокая вероятность ошибочного выбора соседей. Поскольку выбор соседей в этом варианте проводится на каждом шаге, то также на каждом шаге возрастает ошибка прогноза, связанная с вероятностью неправильного выбора. Вследствие этого, особенно при работе со стохастическими рядами, высока вероятность перехода на смежную траекторию. Контроль правильности выбора “нужных” соседей затруднителен, и поэтому применение рассматриваемого способа также может не обеспечивать желаемой надежности прогноза.

3) Прямой прогноз

При прямом способе прогноза на \mathcal{T} шагов вперед стартовый вектор и все его соседи остаются неизменными, а прогноз строится независимо для всех $t \in [1, \dots, \mathcal{T}]$:

$$\hat{x}_{L+t} = f(\mathbf{x}_L, \hat{\mathbf{a}}^{(t)}) .$$

Таким образом, для выбранных векторов-соседей строится зависимость значения ряда через t шагов от координат исходного вектора, оцениваются параметры этой зависимости, т.е. коэффициенты $\hat{\mathbf{a}}^{(t)}$ и с помощью них

затем строится прогноз на t шагов для стартового вектора. Коэффициенты аппроксимации для всех t оцениваются независимо.

При прямом прогнозе отсутствуют недостатки двух первых способов — гипотеза о неизменности соседей здесь не вызывает сомнения, а накопления ошибки не происходит. Однако этот вариант более требователен к длине ряда, так как для прогноза на t шагов необходимы соседи, после которых имеется еще как минимум t известных векторов.

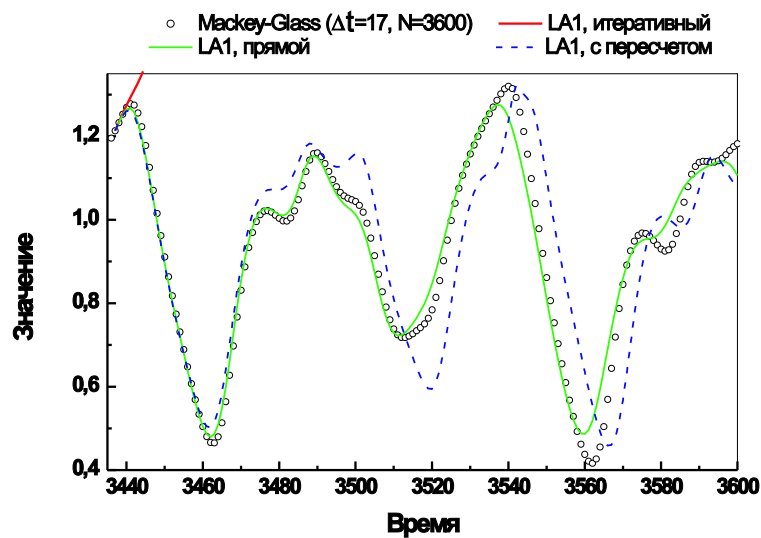


Рис. 19: Итеративный (iterat), прямой (direct) и прямой с пересчетом (recalc) прогнозы методом LA1.

Пример. Рассмотрим уравнение Маккея–Гласса (MG), которое используется как модель процесса регенерации кровяных телец:

$$\frac{dx}{dt} = \frac{ax(t - \Delta t)}{1 + x^c(t - \Delta t)} - bx(t),$$

где $a = 0,2$; $b = 0,1$; $c = 10$, $\Delta t = 17$. Как известно, при данных значениях параметров и времени задержки это уравнение проявляет хаотические свойства.

Прогноза для уравнения MG на $\mathcal{T} = 164$ шага тремя рассмотренными способами иллюстрирует рис.19. Видно, что при итеративном способе прогнозируемые значения, начиная с четвертого шага экспоненциально быстро отклоняются от истинных (для $\mathcal{T} = 164$ это отклонение уже порядка 10^{15}).

Прогноз с пересчетом давал приемлемые результаты дольше, но приблизительно с сорокового шага он начал существенно уступать прямому прогнозу.

На основе этого примера можно предположить, что **итеративный вариант будет лучшим при очень малых величинах T** , когда справедливо предположение о том, что предсказанный вектор остается соседом стартового. В остальных случаях, вероятно, **прямой прогноз даст более достоверный результат.**

14 Малые возмущения хаотических динамических систем

Хаос, как мы уже убедились, — достаточно общее свойство. Из анализа временных рядов мы теперь знаем, что можно оценить глубину хаоса, горизонт прогноза, инвариантные характеристики и др. Однако **можно ли как-то повлиять на хаотическое поведение (например, исходы торгов), используя только малое воздействие**? Скажем, играя на бирже "по маленькой"? Это не так уж и фантастически звучит, если вспомнить, что хаотические системы экспоненциально чувствительны к сколь угодно *малым* (но конечным) возмущениям.

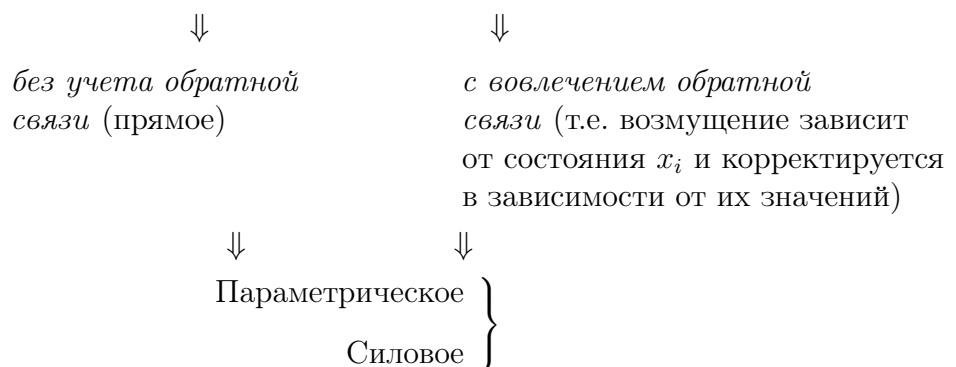
Однако до того, как корректно подойти к решению этой задачи, необходимо установить следующее:

- определить, *детерминированный ли это хаос (!)* (если это случайный процесс, то задача некорректна);
- отделить шум;
- найти размерность вложения и др. характеристики.

В обозначенной задаче речь идет о таком воздействии, когда оно приводило бы к нужной нам (т.е. заданной заранее) динамике системы. Трудностей только две:

- а) существуют ли такие возмущения;
- б) если существуют, как их найти.

Внешние возмущения



Если при таком воздействии динамика системы выходит на нужный нам режим, то

в первом случае — это называется $\begin{cases} \text{подавлением хаоса;} \\ \text{или контролем без обратной связи;} \end{cases}$

во втором случае — *управлением (контролем) с обратной связью.*

Прежде всего, поскольку мы собираемся изучить влияние внешних воздействий на хаотичные системы, рассмотрим, какие бывают возмущения, к чему они могут привести и формализуем задачу.

14.1 Анализ систем с внешними возмущениями

Рассмотрим динамическую систему:

$$\dot{x} = v(x, a), \quad x = \{x_1, \dots, x_n\}, \quad v = \{v_1, \dots, v_n\}.$$

Задача управления может быть сформулирована следующим образом: найти внешнее воздействие u , при котором фазовый поток $F^t(x, u)$ стремится к выбранному подмножеству $A \supset M$, т.е.

$$\lim_{t \rightarrow \infty} F^t(x, u) = A.$$

При введении возмущения u подмножество A может превратиться в аттрактор, но может остаться и неустойчивым множеством. В последнем случае u так модифицирует поток $F^t(x, u)$, что фазовая траектория подходит к A и остается под действием u в некоторой окрестности.

Если u реализуется мультипликативным образом, то говорят о **параметрическом** уравнении:

$$\dot{x} = v'(x, a),$$

$v' = v(x, a_0 + a_1(t))$, где $a_1(t)$ — T -периодическая функция времени. При введении обратной связи имеет место учет текущего состояния системы:

$$v'(x, a) = v(x, a(x(t))).$$

Тогда параметр a изменяется специальным образом.

Иногда мультипликативное возмущение невозможно. Тогда фазовый поток $F^t(x, u)$ разлагается на две составляющие:

поток $F^t(u)$, который инициируется только возмущением u ;

и поток $F^t(x)$, который порождается *невозмущенной* системой.

Следовательно, здесь реализуется *силовая составляющая*:

$$v'(x, a) = v(x, a) + g(t)$$

Параметрическое возмущение в прикладных задачах имеет **преимущества**:

1. Это просто изменение ресурсов системы.
2. Часто динамика определяется при $x_i > 0$. Тогда силовое воздействие может привести к вырождению системы.
3. **Но** силовое воздействие, как правило, приводит к нужному результату, т.к. можно попросту "задавить" внутреннюю динамику.

Таким образом, мы видим сильное разветвление проблемы стабилизации.

Что будет *в самом общем виде*? Пусть

$$T_a : x \mapsto f(x, a),$$

$$a \in A, \quad f = \{f_1, \dots, f_n\}, \quad x = \{x_1, \dots, x_n\}.$$

Возмущение:

$$G : A \rightarrow A,$$

$$G : a \mapsto g(a), \quad a \in A.$$

Тогда

$$\bar{T}_a : \begin{cases} x \mapsto f(x, a) \\ a \mapsto g(a), x \in M, a \in A. \end{cases}$$

Пусть возмущение *периодическое* с периодом τ :

$$a_{i+1} = g(a_i), \quad i = 1, 2, \dots, \tau - 1,$$

$$a_1 = g(a_\tau), \quad a_i \neq a_j, \quad a_i \in A_c.$$

В этом случае возмущенное отображение запишется как

$$\bar{T} : \begin{cases} T_{a_1} : x \mapsto f(x, a_1) \equiv f_1, \\ T_{a_2} : x \mapsto f(x, a_2) \equiv f_2, \\ \dots\dots\dots, \\ T_{a_\tau} : x \mapsto f(x, a_\tau) \equiv f_\tau. \end{cases}$$

Введем в рассмотрение τ функций:

$$\left. \begin{aligned} F_1 &= f_\tau(f_{\tau-1}(\dots f_2(f_1(x))))), \\ F_2 &= f_1(f_\tau(f_{\tau-1}(\dots f_3(f_2(x))))), \\ \dots\dots\dots, \\ F_\tau &= f_{\tau-1}(f_{\tau-2}(\dots f_1(f_\tau(x)))). \end{aligned} \right\} \text{последовательные перестановки}$$

Тогда исходное отношение с периодическим возмущением будет

$$\begin{aligned} T_1 : x &\mapsto F_1(x, a_1, \dots, a_\tau), \\ T_2 : x &\mapsto F_2(x, a_1, \dots, a_\tau), \\ \dots\dots\dots, \\ T_\tau : x &\mapsto F_\tau(x, a_1, \dots, a_\tau), \end{aligned}$$

для которых $x_1 = f_1(x_0), x_2 = f_2(x_1), \dots, x_{\tau-1} = f_{\tau-1}(x_{\tau-2})$.

Как связаны T_i ? \Rightarrow Только через начальные условия. Что можно сказать о T_i , циклах и т.п.?

Утверждение 1. Если $T_k, 1 \leq k \leq \tau$, имеет цикл периода t и $f_k(x) \equiv f(x, a_k)$ непрерывны, то $T_p, p = k + 1 \pmod{\tau}$, также имеет цикл того же периода t . Если цикл отображения T_k устойчив, то цикл отображения T_p тоже устойчив.

Утверждение 2. Период P любого цикла периодически возмущаемого отображения равен $P = \tau k, k \in \mathbf{Z}$ (положительное целое).

Таким образом, если есть модель реальной системы и мы вводим τ -периодическое возмущение, то исследование можно упростить. Вместо исходного неавтономного отображения достаточно рассмотреть одно из автономных T_1, T_2, \dots, T_τ .

15 Подавление хаоса и управление хаотическими системами

Пока речь шла только о каком-то периодическом воздействии. А как изменяется *динамика* отображения? Возможно ли управление?

Введем $A_c \subset A$ — множество такое, что если $a \in A_c$, то отображение будет обладать хаотическим поведением (порождать аperiodические последовательности).

Можно ли стабилизировать динамику в общем случае? Если нет, то в каком случае это можно сделать?

15.1 Подавление хаоса

Здесь речь пойдет о внешних возмущениях, производимых на систему (ее параметры), при которых происходит вырождение хаоса и переход системы на регулярный режим.

Наиболее общий результат:

|| Существуют динамические системы, для которых найдется подмножество $A_d \subset A_c$ такое, что если $a_i \in A_d$, то возмущенные системы будут обладать устойчивыми циклами конечных периодов.

Это означает, что внешние возмущения приводят в определенных динамических системах к стабилизации динамики. Однако в такой формулировке результат мало конструктивен.

Если система эффективно описывается одномерным отображением, то *наиболее подходящий для приложений результат следующий:*

|| Для полимодальных одномерных отображений найдутся такие параметрические возмущения, что любой выбранный заранее цикл можно стабилизировать.

↓

Полимодальность — это определенное свойство функции, порождающей одномерное отображение. Полимодальная функция имеет максимумы и минимумы, а между ними она монотонно возрастающая или монотонно убывающая (см.рис.20).

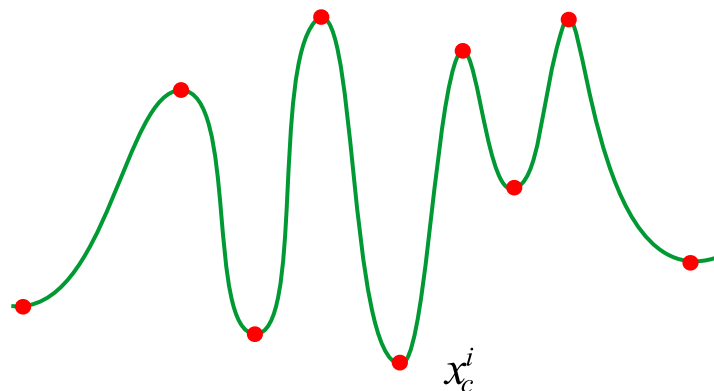


Рис. 20: Полимодалная функция.

Следовательно, если моделируемая система эффективно описывается (!!!) одномерным преобразованием (сводится к одномерному отображению), которое является полимодалным, то практически всегда можно найти возмущения, приводящие к стабилизации **нужной нам** динамики системы.

Для временных рядов картина следующая (рис.21):

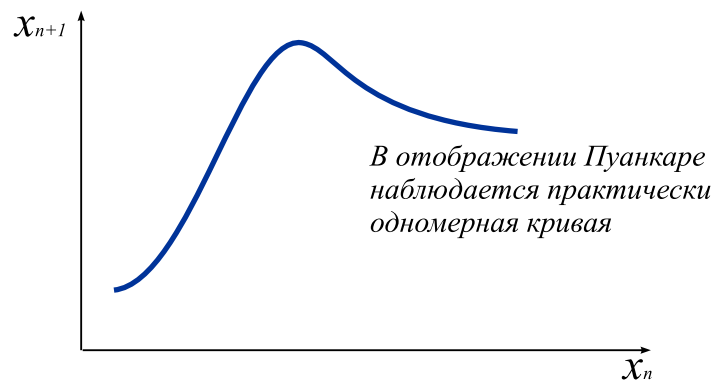


Рис. 21: При сильной диссипации в отображении Пуанкаре наблюдается практически одномерная кривая

Тогда воздействие задается **периодическими изменениями** параметра $(a_1, a_2, \dots, a_\tau)(a_1, a_2, \dots, a_\tau)(\dots)$. Но среди всех a_i должен быть такой, который соответствует критическому значению $f(x_c, a^*)$.

В этом случае будет стабилизированное (устойчивое) периодическое поведение. Почему это так?

У нас p точек, через которые проходит стабилизированный цикл. Среди этих точек существует x_c . Устойчивость цикла определяется из соотношения

$$|f'(x_1)f'(x_2)\dots f'(x_c)| < 1.$$

Но $f'(x_c) = 0$. Поэтому данное соотношение будет выполнено.

А что в случае систем, которые не сводятся к одномерным отображениям и не попадают в класс, для которых существует $A_d \subset A_c$? \Rightarrow

- искать параметры методом перебора (успех здесь часто возможен);
- ввести обратную связь.

15.2 Управление с обратной связью

Геометрическая идея состоит в следующем. Пусть необходимо стабилизировать подходящий цикл (он есть, т.к. в хаосе неустойчивые циклы всюду плотны). В отображении Пуанкаре (см.рис.22) точка p_1 , соответствующая циклу, сначала стремится к x^* , но затем, повинаясь W^u , уходит от x^* . Поэтому надо выбрать возмущение так, чтобы $p_1 \rightarrow W^s$. При этом, естественно, x^* передвинется в точку x_1^* .

Подбирая каждый раз возмущение u так, чтобы точка p_1 находилась вблизи W^s , можно добиться того, что система будет находиться вблизи точки $.$ Это не превращение в аттрактор (как в прежнем случае), а *удержание* p_1 в окрестности точки x^* посредством воздействий.

Этот способ реализован и использовался для управления системами самого различного происхождения. Но он *требует* обратной связи и поэтому его *труднее реализовать*. Поэтому первый способ (просто периодическое воздействие) гораздо более предпочтителен.

Формализация описанного подхода может быть упрощенно представлена следующим образом.

Пусть система задана отображением

$$x_{n+1} = f(x_n, \alpha), f = \{f_1, \dots, f_k\}.$$

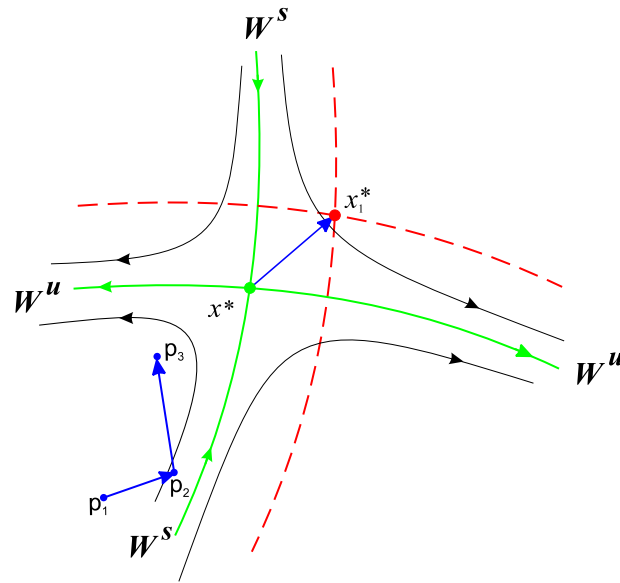


Рис. 22: Стабилизация седлового цикла, вложенного в хаотический аттрактор.

Неподвижная точка $x^* = f(x^*, \alpha_0)$. Вблизи неподвижной точки:

$$x_{n+1} - x^* = \hat{A}(x_n - x^*) + \hat{B}(\alpha - \alpha_0),$$

$$\hat{A} = \left. \frac{\partial f}{\partial x} \right|_{x=x^*; \alpha=\alpha_0}, \quad \hat{B} = \left. \frac{\partial f}{\partial \alpha} \right|_{x=x^*; \alpha=\alpha_0}.$$

В линейном приближении изменение параметра α :

$$\alpha_n - \alpha_0 = -\hat{L}^T(x_n - x^*).$$

Здесь \hat{L} — k -мерный вектор-столбец. Его выбор определяет закон управления, T — транспонирование. Поэтому

$$\delta x_{n+1} = (\hat{A} - \hat{B}\hat{L}^T)\delta x_n,$$

$\delta x_n = x_n - x^*$. Поэтому x^* стабилизируется, если определить \hat{L}^T так чтобы матрица $(\hat{A} - \hat{B}\hat{L}^T)$ имела собственные значения по модулю меньше единицы.

16 Стратегии конкурентных отношений

Одно из наиболее заманчивых приложений теории вероятностей и теории динамических систем — это анализ стратегии азартных игр и биржевых спекуляций. И хотя широко распространено мнение, что ничего здесь сделать нельзя, казино и биржевые спекулянты достаточно хорошо “защищены от вторжения” современных научных методов, данная лекция несколько опровергает его.

Рассмотрим сначала такое явление как шум.

16.1 Шумы

Часто временной ряд характеризуется спектром мощности (квадрат амплитуды преобразования Фурье) $S(\omega)$ относительно частотных интервалов ω . В общем случае такая зависимость может быть аппроксимирована соотношением

$$S(\omega) \approx c \omega^{-\beta},$$

где β — некоторый показатель “спада” функции S . Взяв логарифм, получим:

$$\ln S(\omega) = \text{const} - \beta \ln \omega.$$

Поэтому

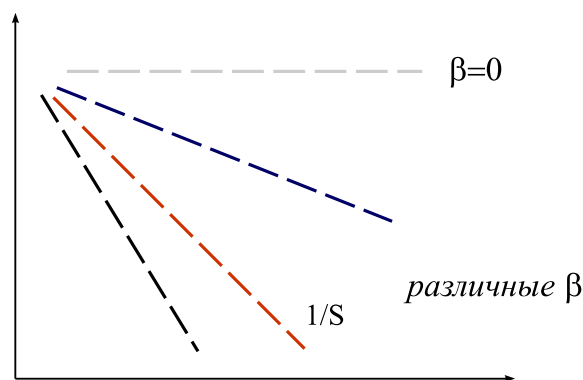


Рис. 23: Наклон сглаженных спектров в зависимости от показателя β .

Если $\beta = 0$, то шум называется **белым**: спектр мощности не зависит от частоты, т.е. он в среднем **не** спадает → *постоянный спектр*.

При $\beta = 2$ спектр будет пропорционален

$$1/\omega^2$$

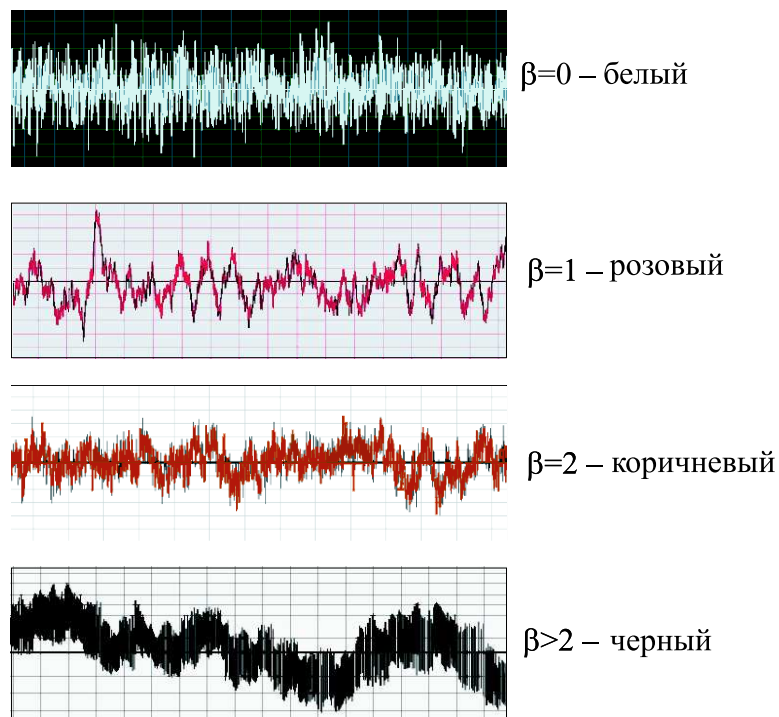
в широком диапазоне частот. Такой шум можно получить, если *проинтегрировать белый шум один раз по времени*. Тогда получим нечто вроде проекции броуновского движения на одно пространственное измерение. Этот шум называется **коричневым**.

Между белым и коричневым шумами располагается **розовый** шум со спектром $1/\omega$. Если $\beta > 2$, то спектр пропорционален

$$1/\omega^\beta,$$

и мы получаем так называемый **черный** шум.

Качественно временные реализации, соответствующие различным β , отличаются так называемой "кластеризацией" выбросов (см.рис.16.1)



Характерной **особенностью** процессов, спектры которых близки к коричневому или черному шумам — если *что-то случилось однажды, то в*

ближайшем будущем не следует ожидать качественного изменения ситуации. Применительно к биржевым спекуляциям, азартным играм и экономическим рядам это явление называется "кластеризация" или "скопление".

⇓

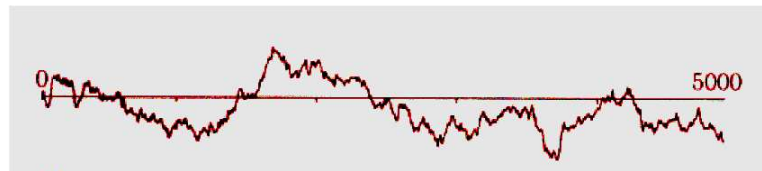
Снизившийся курс не сразу поднимается! В этом, в частности, заключается инертность рынка (если процессы удовлетворяют $\beta = 2, 3, \dots$).

Таким образом, если построить спектр $S(\omega)$ и его спад будет аппроксимирован как

$$\sim \frac{1}{\omega^\beta}, \quad \beta \geq 2,$$

то можно дать определенный прогноз развития процесса и даже выявить некоторые закономерности.

Рассмотрим *честную* азартную игру, в которой вероятность выиграть $p = 0,5$. Пусть текущий капитал $K(t)$ игрока изменяется в ходе игры как показано на рисунке:



Это типичная реализация со спектром $1/\omega^\beta, \beta \geq 2$. Видно, что если капитал упал до 0, то *вероятность того, что он в течение короткого промежутка времени восстановится, практически нулевая*

⇓

{ Интуиция опытных игроков: "плохой день \equiv непёр".
 { Наоборот, выигрыши тоже кластеризуются: "пёр пошёл".

⇓

Это очевидным образом объясняется процессами, где шум $\sim 1/\omega^\beta, \beta \geq 2$.

Данное свойство интересно рассмотреть с точки зрения **самоподобия**: если процесс сжать с постоянным коэффициентом подобия α , то соответствующий спектр растянется в $1/\alpha$ раз. Однако изменение масштаба частот в любое постоянное число раз **не меняет частотной зависимости: свою форму она сохраняет**. Это легко показать акустически. Запишем процесс с $1/\omega^\beta$ шумом на ленту с одной скоростью, а воспроизведем с другой, увеличенной в α раз скоростью. Тогда *звук не повысится*. Это есть явление самоподобия.

Известный пример коричневого шума ($\beta = 2$) — это флуктуации капитала игрока в казино.

↓

Допустим, что при каждом броске монеты вероятность выиграть 1\$ равна p , а проиграть 1\$, следовательно, $1 - p$. Какой стратегии надо придерживаться?

Ответ: если шансы на выигрыш не превышают 50\$, то лучше не играть вообще. Это будет ясно ниже, когда мы рассмотрим процесс разорения игрока.

Но бывает $p > 0,5$. Когда? Например, можно получить дополнительную информацию о статистике рулетки. Тогда *надо ставить*, но *не все деньги*. Какие?

Максимизация прироста дохода будет при ставке $(2p - 1)$ от капитала K . Например, при $p = 0,6$ надо ставить $0,2K$, т.е. 20% от своего капитала. Действительно, информационная емкость

$$C(p) = 1 - H(p),$$

где $H(p) = -(p \ln p + (1 - p) \ln(1 - p))$ — энтропия процесса. При $p = 0,6$ получим $H(p) = 0,97$ и $2^C = 1,02$ (что получается от взятия \log_2).

Таким образом, хорошо информированный игрок может ожидать прибыль в среднем 2% на ставку. Такая стратегия игры называется **стратегией опытного игрока**.

Робкий игрок, ставящий только, скажем, 5% K , будет выигрывать в среднем 1% на ставку. Легко понять, что после 200 ставок его прибыль

составит в среднем $0,1$ выигрыша опытного игрока. Ставки ниже $(2p - 1)$ называются **стратегией робкого игрока**.

Жадный игрок (жаждущий быстрее получить деньги), как правило, ставит половину текущего капитала. Он всегда проигрывает $\approx 3,5\%$ своих при каждом броске монетки. Спустя ~ 200 ставок он проиграет практически все свои деньги (или чужие), т.е. $99,9\%$!

Беспечный игрок, который ставит почти все имеющиеся у него деньги, полностью разорится после, в среднем, двух таких ставок.

16.2 Разорение игрока

Пусть $\begin{cases} \text{с вероятностью } p \text{ выпадает "орёл"} \rightarrow \text{выигрыш } 1\$; \\ \text{с вероятностью } q = 1 - p \text{ выпадает "решка"} \rightarrow \text{проигрыш } 1\$ \end{cases}$.

Обычно капитал игрока K есть броуновский процесс. После первого броска капитал K изменится:

$$\underbrace{(K + 1)}_{\text{с вероятностью } p} \quad \text{либо} \quad \underbrace{(K - 1)}_{\text{с вероятностью } q}$$

Пусть q_K — вероятность окончательного разорения. Тогда

$$q_K = p q_{K+1} + q q_{K-1}, \quad 0 < K < B,$$

где B — капитал банка, $q_0 = 1$ и $q_B = 0$.

Не вдаваясь в детали, решение этого уравнения можно записать как

$$q_K = \frac{(q/p)^B - (q/p)^K}{(q/p)^B - 1}, \quad p \neq q.$$

Если $p = q$, можно раскрыть предел $p \rightarrow 0,5$:

$$q_K = 1 - \frac{K}{B}.$$

Таким образом, при **очень продолжительной игре** разорение небогатого ($K \ll B$) игрока почти неизбежно: $q_k \approx 1$!

Следствие: малые предприятия в конце концов разоряются, если на рынке присутствуют очень крупные фирмы.

|| **Выход:** не работать там, где $p = q$.
|| Но тогда, как правило, будет $p < 0,5$, т.е. $q > p$. *Как действовать?*

16.3 Неблагоприятные шансы на выигрыш

Начальный капитал игрока равен K , капитал банка равен B . Вероятность при каждом испытании $p < 0,5$. Можно ли рискнуть (или работать на таком рынке)? Если да, то как?

Мы знаем, что при ставке в 1\$ вероятность достичь желанной цели

$$p_K = 1 - q_K = \frac{r^K - 1}{r^B - 1},$$

где

$$r = \frac{1-p}{p} > 1,$$

(p_K — вероятность "сорвать" банк).

Если рассмотреть рулетку, то там в колесе будет "зеро" (одно зеленое поле → европейское казино) или "два зера" (два зеленых поля → американское казино). Тогда

$$p = \frac{18}{37}, \quad r = \frac{19}{18}.$$

Мы начинаем игру с $K = 1000\$$. В банке $B = 10000\$$. Ставка в 1\$ дает вероятность сорвать банк $10^{-21}!!$ **Нечего и пробовать!**

При ставке в 10\$ $p_K \approx 10^{-21}$. Тоже шансы ничтожны. Но при ставке в 100\$ получим $p_K \approx 0,0025$. Теперь, в принципе, можно рискнуть!

Почему так выходит?

При малых ставках прежде, чем выиграть намеченную сумму, необходимо сыграть очень много партий. Но всякий раз, делая ставку, приходится "испытывать судьбу".

↓

При неблагоприятных шансах на выигрыш оптимальная стратегия состоит в том, чтобы *поставить все имеющиеся деньги K* .

Конкретно: пока $x = K/B < 1/2$, ставить нужно весь капитал. Если $x > 1/2$, чтобы выиграть B долларов, необходимо каждый раз ставить $(B - xB)\$$. Тогда в случае проигрыша останется

$$K - (B - K) = 2K - B.$$

Описанная стратегия называется **стратегией разорения конкурентов**.

16.4 Основные выводы

1) *Партнерские отношения иерархических систем обычно антагонистичны*, т.е. каждый стремится некоторым образом управлять ситуацией и противопоставить оппоненту свое поведение, не позволяющее строить никаких прогнозов. Такое отношение обычно называют игрой, а наука, которая их изучает, — *теорией игр*. Под игрой, таким образом, понимается *алгоритм принятия решений в условиях конфликта и неопределенности (относительно мотивов и ходов партнеров)*.

2) *Динамические системы тоже могут порождать цветные шумы*. Поэтому, в добавление к описанным явлениям полезно использовать весь арсенал известных в теории динамических систем методов.

3) Естественно, результаты, представленные в настоящей лекции, можно понимать и более широко. В частности, допустимо в качестве банка B (казино) рассмотреть некоторую фирму, а в качестве игрока K — держателя (или держателей) акций. Тогда, если условия рынка удовлетворяют заданным, то приведенные стратегии остаются верными. Таким образом, в более широком смысле *описанные подходы можно интерпретировать как стратегии разорения фирм*.

4) Основной вывод, который вытекает из приведенных результатов, следующий: *даже при не очень благоприятных шансах на успех вполне возможно найти стратегию, при которой игрок не только не остается в проигрыше, но даже выигрывает!* Для непосвященных но опытных игроков это обычно называется интуицией. Но, как видно, в данном случае такое “шестое” чувство нетрудно формализовать и выработать вполне разумный подход к анализу ситуации.

17 Разладка и сегментация временных рядов

Часто при наблюдении за какой-то переменной можно заметить в некоторые моменты времени ее резкое изменение.

17.1 Разладка

Традиционным при исследовании временных рядов является предположение о том, что *статистические свойства наблюдаемых постоянны во времени или медленно изменяются*. Вместе с тем, для многих практических целей *обнаружение резкого изменения свойств наблюдаемого ряда, происходящего в неизвестный заранее момент времени* (это явление иногда называют **разладкой**), является важной задачей.

Например, в эконометрике при управлении производством существуют задачи, решаемые методами обнаружения разладки. При выявлении момента разладки *требуется оценить момент появления изменений как можно точнее*. Алгоритмы, предназначенные для решения такого типа задач, называются **апостериорными**.

Появление новых теоретических результатов почти не отражается на методах построения конкретных систем контроля производства, обработки сигналов, диагностики и т.п. Это объясняется трудностями, возникающими при использовании оптимальных алгоритмов. Дело в том, что предположения теоретических работ о независимости последовательности наблюдений, точном знании параметров модели и свойств наблюдаемого сигнала или параметров и т.п., хотя и обеспечивают чистоту математического анализа, но при этом сильно ограничивают класс решаемых задач и препятствуют широкому практическому использованию полученных результатов.

Впервые задача апостериорного обнаружения разладки была поставлена и решена Е.С.Пейджем. Предполагалось, что

$\{x_1^N\}$ — независимая одномерная случайная последовательность.

При этом *плотность $\omega(x_t/\Theta)$ распределения наблюдений x_t* имела

$$\begin{cases} \text{при } t = 1, \dots, t_{0-1} \text{ скалярный параметр } \Theta = \Theta_1; \\ \text{а при } t = t_0, \dots, N \text{ параметр } \Theta = \Theta_2. \end{cases}$$

Значения Θ_1 и Θ_2 предполагались *известными заранее*, а все значения $t = 1, \dots, N$ — *равновероятными*.

Идея, описанная Пейджем, заключалась в проверке $N + 1$ гипотез H_1, H_2, \dots, H_{N+1} , состоящих в том, что $t_0 = 1, 2, \dots, N + 1$.

Например, гипотеза H_1 означает, что плотность распределения x_1, \dots, x_N суть $\omega(x/\Theta_2)$, а гипотеза H_{N+1} — что x_1, \dots, x_N имеют плотность распределения $\omega(x/\Theta_1)$.

В этих условия *вероятность ошибочной классификации будет минимальной*, если гипотеза H_k принимается при выполнении следующих условий:

$$l_k(x) \geq l_j(x) \text{ для любого } j \text{ такого, что } j \neq k,$$

$$l_m(x) = \sum_{i=1}^{m-1} \ln \omega(x_i/\Theta_1) + \sum_{i=m}^N \ln \omega(x_i/\Theta_2),$$

где $l_m(x)$ — **функция правдоподобия**, соответствующая гипотезе H_m .

Так как значения Θ_1 и Θ_2 заранее известны, *получение оценок методом максимального правдоподобия эквивалентно использованию схемы кумулятивных сумм*.

Тогда оценка момента разладки получается, как

$$t'_0 = \inf \{k : S_{k-1} \geq S_j, j = 2, \dots, N\},$$

где

$$S_k = \sum_{i=1}^k [\ln \omega(x_i/\Theta_1) - \ln \omega(x_i/\Theta_2)],$$

т.е. за оценку момента разладки t'_0 принимается момент времени k , соответствующий *максимальному значению S_k кумулятивной суммы, плюс единица*.

Первоначально рассматривались лишь независимые последовательности, но на практике такое предположение часто оказывается неверным. Поэтому большой интерес представляют работы по обнаружению разладки зависимых последовательностей. В одной из первых таких работ предполагалось, что случайная последовательность задается нестационарной

моделью проинтегрированного скользящего среднего. Критерий проверки факта изменения среднего уровня в некоторой точке основывался на t -распределении Стьюдента.

Однако в общем случае, когда разладка заключается в изменении не только среднего, но и коэффициентов корреляции последовательности, апостериорное обнаружение наталкивается на серьезные вычислительные трудности. Для их преодоления оказалось целесообразным применить параметрическое моделирование случайных последовательностей.

Основные выводы:

- *процедура апостериорной оценки разладки будет тем эффективнее, чем точнее определен диапазон возможных значений момента разладки, что особенно важно для случая наличия нескольких разладок на одной выборке;*
- *апостериорное оценивание момента разладки методом максимального правдоподобия статистически эффективнее, чем методом кумулятивных сумм;*
- *объем вычислений при оценке момента разладки методом максимального правдоподобия быстро растет с увеличением длины выборки, особенно для моделей авторегрессии-скользящего среднего и нескольких моментов разладки, существенно превосходя вычислительные затраты, необходимые для реализации алгоритма кумулятивных сумм.*

Таким образом, становится понятна взаимосвязь апостериорных и других методов обнаружения разладки:

1) Если нужно не только обнаружить, но и точно оценить момент разладки, то сначала нужно получить первичную оценку момента разладки, а апостериорный алгоритм должен ее уточнить после обработки участка реализации.

2) За первичным анализом временных рядов следует этап идентификации моделей временных рядов.

3) Этап идентификации моделей должен заканчиваться выбором типа модели, его параметризацией и оценкой диапазонов изменения коэффициентов.

17.2 Сегментация

При техническом анализе временных рядов их будущие значения либо их параметров (например, дисперсии) прогнозируются на основе только прошлых значений. В экономических приложениях правомочность такого подхода основывается на гипотезе эффективного рынка, согласно которой все агенты на рынке рациональны, на рынке нет арбитража и асимметрии информации.

При выполнении данной гипотезы существующие значения ряда уже аккумулировали всю доступную информацию, а значит, чтобы спрогнозировать завтрашнюю величину нам достаточно в модель заложить только предыдущие величины. Добавление в модель других временных рядов не должно улучшить качества прогноза, поскольку не добавляет новой информации в модель.

К техническому анализу можно отнести довольно широкий класс методов, которые разделяются на две основные группы: традиционный технический подход и развивающийся в последнее время анализ на основе теории нелинейных динамических систем. В базовых моделях временных рядов предполагается, что некоторая величина в данный момент времени определяется значением этой характеристики в предыдущие моменты времени и некоторой ошибкой (в общем случае суммой ошибок реализовавшейся в данный момент времени и ошибками, реализовавшимися в предыдущие моменты времени).

Как правило, используются линейные модели, т.е. предполагается, что конкретное значение является линейной комбинацией предыдущих и ошибки. Используя статистические методы, такие как метод наименьших квадратов или метод максимального правдоподобия, можно так подобрать коэффициенты в этих моделях, чтобы получить наилучшую подгонку имеющихся данных. На основании модели с подобранными коэффициентами

строются прогнозы будущих значений временного ряда. При этом предполагается, что в будущем модель не изменится. Однако в динамике ряда может произойти резкая смена и тогда такие модели перестают работать.

Хороший пример данного явления — экономический кризис или резкое изменение ритмов в кардиологии. Модели, хорошо работавшие в “спокойное время”, в этих случаях оказываются несостоятельными.

Выделение периодов с разной динамикой является очень важной задачей. Эта задача называется *сегментацией*. При этом разделяют задачи *априорной* и *апостериорной* сегментации. Под априорной сегментацией понимают задачу *прогноза момента разладки*. Под апостериорной сегментацией понимают задачу разделения имеющегося временного ряда на сегменты с разной динамикой.

17.3 Показатель Хёрста и показатель Гёльдера

Для решения задач сегментации в последнее время все чаще прибегают к инструментарию теории динамических систем. Данный инструментарий успешно применяется в медицине, геофизике, астрофизике и обработке сигналов и теории финансов.

Существует несколько подходов к задаче сегментации временных рядов. Один из них — метод нормированного размаха, или $\frac{R}{S}$. Он позволяет вычислить показатель Хёрста, меру хаотичности ряда. Значениям показателя Хёрста, близким к 0.5 соответствуют стохастические ряды. Показатель Хёрста, близкий к 1, говорит о том, что рассматриваемый временной ряд порожден некоторой хаотической системой. Можно предположить, что моменту разладки предшествует изменение показателя Хёрста. Из этого следует способ сегментации временного ряда на основе данного показателя.

Однако наиболее перспективным методом сегментации является показатель Гёльдера — локальная мера регулярности функции. Показателю Гёльдера, превышающему единицу, соответствуют гладкие функции. Если же данный показатель меньше единицы, то функция является сингулярной в данной точке.

Оказывается, что типичный временной ряд, порожденный некоторой природной сложной системой, функционирующей в нормальном (не кризисном) режиме, нерегулярен. Иными словами, в нормальном состоянии показатель Гёльдера не должен превышать единицу. Если же он достигает достаточно больших значений, то это свидетельствует о неправильном функционировании системы.

Данная закономерность используется, например, при анализе показателей жизнедеятельности человека (ЭКГ, ЭЭГ) с целью выявления сбоев. Исходная гипотеза состоит в том, что для временных рядов должна выполняться похожая зависимость. Резким падениям должен соответствовать рост показателя Гёльдера.

18 Фрактальные свойства ременных рядов

Начнем с краткого обзора параметров, часто применяемых для описания фрактальных свойств множеств, функций и временных рядов.

Часто используемые фрактальные характеристики

Многие экспериментальные сигналы обладают фрактальной статистикой, анализ которой может быть произведен с помощью **метода нормированного размаха**, предложенного Мандельбротом и Уоллисом на основе наблюдений Херста.

Наблюдаемый нормированный размах есть отношение

$$R/S,$$

где R — размах, т.е. разница между максимальным значением и минимальным значением наблюдаемой, S — стандартное отклонение. Эта величина очень хорошо описывается эмпирическим соотношением

$$(\tau/2)^H,$$

где H — показатель Херста, τ — период измерения.

При некоторых условиях запись измерений представляет кривую размерности

$$D_f = 2 - H$$

т.е. экспериментальные сигналы, получаемые при наблюдении многих природных явлений, имеют *фрактальную* зависимость от времени.

Поэтому было высказано следующее предположение: возможно, что содержательное истолкование допускает **сегментация на основе фрактальных характеристик ряда**. Рассмотрим ряд таких характеристик.

Функция $Y = f(X)$ называется **фрактальной**, если ее график представляет собой фрактальное множество или если существует значение функции Y_0 , которое является образом фрактального множества точек, т.е. если некоторый ее уровень имеет фрактальную структуру.

"Фрактал - это структура, состоящая из частей, которые в каком-то смысле подобны целому". Там же дано и другое определение:

"Фрактал - это множество, размерность Хаусдорфа-Безиковича которого строго больше его топологической размерности".

Второе определение слишком ограничительно; первое же содержит существенный отличительный признак: фрактал выглядит одинаково, в каком бы масштабе мы его не наблюдали.

Размерность Хаусдорфа-Безиковича $\dim_H(E)$ компактного множества E определяется при помощи **внешней хаусдорфовой меры**

$$M_d(E) = \liminf_{\varepsilon \rightarrow 0} \sum_C \text{diam}(S_i)^d,$$

где — покрытие E замкнутыми множествами S_i диаметра, не превосходящего ε .

Можно доказать, что существует константа α_0 такая, что при $d < \alpha_0$ $M_d(E) = \infty$, а при $d > \alpha_0$ $M_d(E) = 0$ (значение M_d при $d = \alpha_0$ может быть как конечным, так и бесконечным).

Величина α_0 и есть, по определению, размерность Хаусдорфа-Безиковича.

Множества, для которых эта размерность является дробным числом, называются *фрактальными множествами (фракталами) в узком смысле*. Множество называется *фрактальным (фракталом) в широком смысле*, если его топологическая размерность меньше размерности Хаусдорфа-Безиковича, т.е. $\alpha_0 > \dim E$.

С исследованием распределения физических или других величин на

геометрическом *носителе* связаны **мультифрактальные меры**. Неформально говоря, мультифрактальность меры μ означает наличие существенного диапазона изменения показателя локального скейлинга на фрактальном множестве — носителе μ (наименьшем замкнутом подмножестве, на котором сосредоточена мера).

Показателем локального скейлинга в точке x называется число $\alpha(x)$ такое, что

$$\mu(B_r(x)) \approx r^{\alpha(x)},$$

где $B_r(x)$ — окрестность радиуса r точки x .

Пусть X_α — объединение точек с локальным скейлингом α , $f(\alpha) = \dim H(X_\alpha)$.

Мера *мультифрактальна*, если функция $f(\alpha)$ отлична от нуля на целом сегменте значений α , и *монофрактальна*, если $f(\alpha)$ отлична от нуля в единственной точке.

Для функций параметром, аналогичным локальному скейлингу, является **локальный показатель Гёльдера**.

Основные понятия, лежащие в основе того, что теперь принято называть мультифракталами, были первоначально введены Мандельбротом при обсуждении турбулентности, и впоследствии распространены на многие другие ситуации. Особый интерес к мультифракталам начался с работ Грассбергера и др.. Анализ экспериментальных результатов и введение функции $f(\alpha)$ позволили достичь великолепного согласия между простой теоретической моделью и наблюдениями. Эти работы продемонстрировали полезность использования мультифракталов при описании экспериментальных данных. Аналогичный подход развивали в своих работах Бенсимон и др., Холси и др., Глайзер и др. Идея представления носителя меры в виде объединения своих подмножеств, соответствующих различным показателям локального скейлинга открыла новый инструмент для применения фрактальной геометрии к физическим системам. Каждое слагаемое в объединении фрактально и имеет свою фрактальную размерность. Максимальное значение фрактальной размерности этих подмножеств равно фрактальной размерности носителя меры P .

Фрактальные модели рядов: фрактальные интерполяционные функции

Рассмотрим **фрактальные интерполяционные функции (ФИФ)**.

Для того, чтобы задать ФИФ, достаточно задать произвольным образом три набора чисел:

1. **узлы интерполяции** $x_0 < x_1 < \dots < x_n$
2. **значения в узлах** y_0, y_1, \dots, y_n
3. **масштабные параметры** $c_0, c_1, \dots, c_n, 0 < |c_i| < 1$

Рассмотрим n аффинных отображений на плоскости, определенных посредством этих данных:

$$\Phi_i \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_i & 0 \\ b_i & c_i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} p_i \\ q_i \end{pmatrix},$$

где

$$a_i = \frac{x_i - x_{i-1}}{x_n - x_0}, \quad b_i = \frac{(y_i - y_{i-1}) - c_i(y_n - y_0)}{x_n - x_0},$$

$$p_i = \frac{x_n x_{i-1} - x_0 x_i}{x_n - x_0}, \quad q_i = \frac{(x_n y_{i-1}) - c_i(x_n y_0 - x_0 y_n)}{x_n - x_0}.$$

Введем обозначения:

$G = \text{graph}(f)$ на интервале $[X_0, X_N]$ и $G_i = \text{graph}(f)$ на интервале $[X_{i-1}, X_i]$.

Для набора отображений $\{\Phi_i\}$ существует единственная функция $f(x)$, непрерывная на интервале $[X_0, X_N]$ такая, что:

$$f(x_i) = y_i, \quad i = 0, \dots, n.$$

$$F_i(G) = G_i, \quad i = 1, \dots, n.$$

Первое из условий означает, что функция $f(x)$ интерполирует данные y_i в точках x_i , а *второе* — что график $f(x)$ является самоподобным множеством (ограничения этого графика на отрезки есть линейно уменьшенные копии графика в целом).

Другой формой записи системы уравнений (5) является следующая:

$$f(a_i x + p_i) = c_i f(x) + b_i x + q_i, \quad i = 1, \dots, n.$$

Из этого рекуррентного соотношения видно, что ФИФ нетрудно вычислить в точках всюду плотного подмножества отрезка определения.

Преимущества использования ФИФ: ФИФ полезны нам как тестовый материал, потому что они:

1. В общем случае фрактальны (т.е. размерность Хаусдорфа-Безиковича графика больше единицы),

2. Для них легко вычисляется нужный нам параметр, аналогичный показателю локального скейлинга для мер - показатель Гёльдера.

Гёльдеровский показатель непрерывной функции $f(x)$ в точке x_0 есть величина

$$h(x_0) = \sup\{\alpha : |f(x_0) - f(x)| \leq |x_0 - x|^\alpha, x \in O(x_0)\},$$

где $O(x_0)$ — окрестность точки x_0 .

Точка является **сингулярной точкой** $f(x)$, если показатель Гёльдера в ней меньше единицы: $h(x) < 1$.

Величина $h(x)$ характеризует "**степень сингулярности**" функции в данной точке интервала — чем меньше $h(x)$, тем сильнее сингулярность.

Показатель Гёльдера на интервале равен, по определению, наименьшему из показателей Гёльдера в точках интервала.

Показатель Гёльдера тесно связан с показателем Херста: например, для траекторий так называемого **фрактального броуновского движения** показатель Гёльдера постоянен во всех точках и равен показателю Херста.

Для широкого класса ФИФ показатель Гёльдера находится в диапазоне (h_{min}, h_{max}) , где

$$h_{min} = \min_i \{\log(c_i) / \log(a_i)\}, \quad h_{max} = \max_i \{\log(c_i) / \log(a_i)\}.$$

Варьируя параметры a_i, c_i можно генерировать ряды с известным распределением сингулярностей, в частности, моно- или мультифрактальные. Это позволяет тестировать методы сегментации, использующие показатель Гёльдера.

18.1 Вейвлет-преобразование и локальное Фурье-преобразование

Очень часто наиболее важная информация содержится в беспорядочных и хаотических на первый взгляд сигналах. Для многих типов сигналов — от электрокардиограмм до сигналов радара — интересующая нас информация содержится главным образом в пиках, т.е. весьма кратковременных явлениях. До недавних пор *преобразование Фурье* было главным математическим средством для анализа пространственного распределения сингулярности. Однако оно не приспособлено для выявления пространственного распределения сингулярностей. Это явилось главным мотивом для изучения *вейвлет-преобразования*. Последнее, в силу своей многомасштабной структуры, может быть использовано как "математический микроскоп", дающий понимание запутанного строения фракталов.

Вейвлет-преобразование

Рассмотрим комплекснозначную функцию

$$\psi(x) \in L^2(\mathbb{R}).$$

Она называется **вейвлетом**, если Фурье-преобразование $\hat{\psi}(w)$ удовлетворяет следующему соотношению:

$$\int_{-\infty}^0 \frac{|\hat{\psi}(w)|^2}{|w|} dw = \int_0^{+\infty} \frac{|\hat{\psi}(w)|^2}{w} dw = C_\psi < +\infty.$$

На практике это сводится к простому требованию:

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0.$$

Обозначим s -растяжение $\psi(x)$:

$$\psi_s(x) = \frac{1}{s} \psi\left(\frac{x}{s}\right).$$

Вейвлет-преобразование функции $f(x) \in L^2(\mathbb{R})$ — это функция двух переменных, которая задается выражением:

$$W_\psi f(x, s) = f * \psi_s(-x) = \frac{1}{s} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{x-t}{s}\right) dt.$$

Фурье-преобразование $Wf(x, s)$ по имеет вид:

$$\left(\widehat{Wf}\right)(w, s) = \hat{f}(w)\hat{\psi}(sw).$$

Из этих выражений ясно, что вейвлет-преобразование есть совокупность версий сигнала $f(x)$, профильтрованных фильтром $\psi_s(x)$ при различных s .

Равенство нулю среднего значения вейвлета означает, что (при выборе гладкого и хорошо локализованного в частотной области вейвлета) все эти фильтры локализованы в различных, хотя и пересекающихся областях частот.

Можно доказать, что *вейвлет-преобразование обратимо*, и $f(x)$ можно восстановить по формуле

$$f(x) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(t, s) \bar{\psi}_s(t-x) dt \frac{ds}{s}.$$

Замечательным свойством вейвлет-преобразования является его способность характеризовать локальную "степень сингулярности" функции, которая измеряется локальным показателем Гельдера.

Введем понятие **скелета максимумов** вейвлет-преобразования.

Точка x_0, s_0 называется точкой **максимума модуля** функции $Wf(x, s)$, если:

$$|Wf(x, s_0)| < |Wf(x_0, s_0)|$$

для в левой или правой ближайшей окрестности точки x_0 , и

$$|Wf(x, s_0)| \leq |Wf(x_0, s_0)|$$

для в противоположной окрестности точки x_0 , соответственно.

Линия максимумов — это любая кривая в (x, s) — пространстве, состоящая из точек максимума модуля. **Скелет максимумов** — это набор всех линий максимумов функции $Wf(x, s)$.

Рецепт локализации изолированных сингулярных точек и оценки показателя Гёльдера задается двумя утверждениями:

- Если в области $x \in (a, b)$, $0 < s < s_0$ нет линий максимума $Wf(x, s)$, то функция $f(x)$ регулярна на интервале (a, b) .
- Гёльдеровские сингулярности $f(x)$ являются пределами "локализуемых" последовательностей вида (x_n, s_n) , принадлежащих скелету максимумов, при $s \rightarrow 0$. Величину показателя можно оценить, исходя из условия $|Wf(x, s)| \leq A(s^h + |x - x_0|^h)$, которое обязано выполняться при достаточно малых s в точках локализующей последовательности. На практике это означает, что h можно оценить по наклону $\log - \log$ графика зависимости $|Wf(x, s)|$ от s вдоль линии максимумов, идущей в сингулярную точку.

Локальное преобразование Фурье

Классическим в теории функций средством для измерения показателя Гельдера функции $f(x)$ является оценка скорости убывания на бесконечности ее преобразования Фурье $\hat{f}(w)$. Можно доказать, что функция $f(x)$ ограничена и имеет показатель Гельдера h , если она удовлетворяет условию

$$\int_{-\infty}^{+\infty} |\hat{f}(w)| (1 + |w|^h) dw < +\infty.$$

Однако это дает лишь **глобальный** (т.е. наименьший из локальных) **показатель Гельдера**. Для того, чтобы получить представление об изменении *локального показателя Гельдера*, можно использовать **локальное Фурье-преобразование**, т.е. Фурье-преобразование исходной функции, умноженной на некоторое **окно** — гладкую функцию, локализованную в окрестности интересующей нас точки. В этом случае скорость убывания преобразования Фурье будет отвечать *гельдеровской регулярности функции* в окрестности данной точки.

19 Явление гомодинамичности

Допустим, что в результате наблюдений (например, торгов) была получена последовательность $\{y_m\}$.

Ряд $\hat{y} = \{y_m\}_{m=0}^{\infty}$ называется **детерминированно порожденным**, если выполнены следующие условия:

а) существует динамическая система $\{T\}$ с конечномерным фазовым пространством $M = \{x\}$, точка x_0 и липшиц-непрерывная функция ϕ такие, что $\phi(T^m(x_0)) = y_m$ для всех $m = 0, 1, 2, \dots$;

б) $\text{dist}(T^t x, T^t x') \leq \text{const} \cdot e^{\lambda t} \text{dist}(x, x')$, т.е. траектории динамической системы $\{T\}$ не могут разбегаться быстрее, чем с экспоненциальной скоростью.

Пространство всех временных рядов, определенное как множество всевозможных последовательностей \hat{y} , — это *банахово пространство*, задавая отображение сдвига в котором, получим универсальную динамическую систему, порождающую любой ограниченный ряд. Теперь, вводя предельное множество, можно определить *емкость*.

В ряде работ был проведен анализ некоторых временных рядов и показано, что многие из них имеют *конечную емкость*. Таким образом, эти ряды могут быть описаны обыкновенным дифференциальным уравнением. В общем случае такое уравнение представляется в виде

$$\frac{d^n p}{dt^n} = F \left(\frac{d^{n-1} p}{dt^{n-1}}, \frac{d^{n-2} p}{dt^{n-2}}, \dots, p, a_i \right), \quad (3)$$

где F — некоторая функция, t — время и a_i — совокупность параметров.

Для описания динамики временного ряда необходимо восстановить правую часть в уравнении (3). В некоторых работах проводились оценки длины временного ряда, необходимой для построения функции F . Эти оценки показывают, что в большинстве случаев имеющихся данных *недостаточно* для определения правой части уравнения (3).

Для решения этой проблемы можно предложить подход, который сводится к следующему. Исследования многих рядов естественного происхождения позволяют подметить две **характерные особенности**:

— зачастую ряд колеблется в ограниченном интервале, границы которого называются уровнями поддержки и сопротивления. Такое поведение называется **гомодинамическим**, т.е. соответствующим практически неизменному закону движения;

— время от значения “пробивают” эти уровни и после переходного процесса ряд выходит на другой гомодинамический участок.

Интерпретация этих особенностей с точки зрения теории динамических систем составляет суть данной лекции.

19.1 Линейные гомодинамические модели

Обыкновенное дифференциальное уравнение может быть представлено в виде системы из n дифференциальных уравнений первого порядка. Фазовым пространством такой системы является n -мерное евклидово пространство. Каждому мгновенному состоянию системы отвечает фазовая точка этого пространства, и каждой точке пространства соответствует определенное и единственное состояние системы.

Динамику можно представить как последовательное изменение положения фазовых точек, т.е. траектории движения этих точек в фазовом пространстве. Таким образом, поведению исходной системы (т.е. исходного процесса) будет отвечать решение n обыкновенных дифференциальных уравнений и соответствующая фазовая траектория.

Вследствие существования уровней поддержки и сопротивления, в определенные интервалы времени временные ряды имеют достаточно большие участки, соответствующие мало изменяющемуся закону движения. Такие участки называются *гомодинамическими*. Как правило, гомодинамика (при некоторых дополнительных допущениях, см. ниже) может наблюдаться в определенные иногда достаточно длительные периоды. Рассмотрим довольно эффективный подход к описанию гомодинамических участков временных рядов.

19.2 Исходные гипотезы

Для построения системы обыкновенных дифференциальных уравнений, описывающих динамику гомодинамического временного ряда, примем следующие исходные гипотезы.

1) Гомодинамические участки соответствуют движению системы в малой области фазового пространства.

2) Достаточно длительное движение системы в малой области фазового пространства возможно по следующим причинам:

- траектория находится в области притяжения неподвижной точки. Такое движение называют гомодинамическим движением *первого* типа.
- траектория находится в области неподвижной точки, притягивающей по одним направлениям и отталкивающей по другим. Такое движение обычно называют гомодинамическим движением *второго* типа.

3) Систему дифференциальных уравнений вблизи неподвижных точек можно линеаризовать. В этом случае, переходя обратно от системы к дифференциальному уравнению n -го порядка, получим линейное уравнение. Следовательно, гомодинамическое движение может быть описано *линейным* дифференциальным уравнением.

4) Прекращение гомодинамического движения может происходить

- в результате внешнего шока, т.е. воздействия факторов, неучтенных в системе; очевидно, такая ситуация допустима для гомодинамического движения как первого так и второго типов;
- в результате выхода из области, в которой справедлива линеаризованная система (см. пункт 3)); этот случай может иметь место лишь для гомодинамического движения второго типа.

Таким образом, мы допускаем, что большую часть времени траектория системы проводит вблизи неподвижных точек. Кроме того, по тем или иным причинам траектория может переходить из окрестности одной неподвижной точки к другой.

Следуя принятым исходным гипотезам, представим изменение на некотором гомодинамическом участке в следующем образом:

$$P(t) = p(t) + \varepsilon(t) ,$$

$$\frac{d^n p(t)}{dt^n} + \sum_{m=0}^{n-1} a_m \frac{d^m p(t)}{dt^m} = 0 , \quad (4)$$

где $P(t)$ — наблюдаемое значение цены в момент времени t , $\varepsilon(t)$ — белый шум с нулевым средним и дисперсией ν , $p(t)$ — математическое ожидание исследуемой величины в момент времени t , a_m — постоянные коэффициенты. Присутствие в наблюдаемых случайной составляющей в реальности неизбежно.

Решение дифференциального уравнения (4) можно записать в виде:

$$p(t) = \sum_{m=1}^n z_m e^{w_m t} .$$

Таким образом, для прогнозирования соответствующего временного ряда необходимо уметь оценивать неизвестные параметры z_j и w_j .

19.3 Техника оценки параметров

Поскольку случайный аддитивный член $\varepsilon(t)$ является *белым шумом*, оценку параметров можно производить *методом наименьших квадратов*. Следуя этому методу необходимо произвести минимизацию по неизвестным параметрам функции следующего вида:

$$\sum_{s=1}^T \left(P(t_s) - \sum_{m=1}^n z_m e^{w_m t_s} \right)^2 \longrightarrow \min . \quad (5)$$

Минимизация функции (5) затрудняется тем, что параметры z_j и w_j являются, вообще говоря, комплексными числами. Кроме того, они не являются независимыми величинами, т.е. если $\text{Im } w_j \neq 0$, то существует $k \neq j$ такое, что $z_j = z_k^*$, $w_j = w_k^*$. Здесь и далее знак * обозначает комплексное сопряжение. В связи с этим минимизация функции (5) должна проводиться поэтапно.

На первом этапе предположим, что комплексные параметры отсутствуют. Тогда необходимо найти минимальное значение функции (5) и набор параметров z_j , w_j , соответствующий этому значению.

На втором этапе допустим, что существует пара сопряженных друг другу комплексных параметров w_1 и w_2 , и им соответствует пара сопряженных друг другу комплексных параметров z_1 и z_2 . Остальные парамет-

ры действительные. В этом случае функция (5) запишется как

$$\sum_{s=1}^T \left(P(t_s) - 2R_1 e^{\nu_1 t_s} \cos(u_1 t_s + \varphi_1) - \sum_{m=3}^n z_m e^{w_m t_s} \right)^2 \longrightarrow \min . \quad (6)$$

Здесь $z_1 = R_1 \exp(i\varphi_1)$, $z_2 = R_1 \exp(-i\varphi_1)$, $w_1 = \nu_1 + iu_1$, $w_2 = \nu_1 - iu_1$. Минимизируем функцию (6) по действительным параметрам R_1 , φ_1 , ν_1 , u_1 , z_3, \dots, z_n , w_3, \dots, w_n .

На k -м этапе ($k \leq [n/2] + 1$, где операция $[\cdot]$ означает взятие целой части) рассмотрим $k - 1$ пар сопряженных друг другу комплексных параметров, w_1 и w_2, \dots, w_{2k-3} и w_{2k-2} . Им соответствуют $k - 1$ пар сопряженных друг другу комплексных параметров z_1 и z_2, \dots, z_{2k-3} и z_{2k-2} . Остальные параметры будут действительными. Тогда функция (5) примет вид:

$$\sum_{s=1}^T \left(P(t_s) - 2 \sum_{l=1}^{k-1} R_l e^{\nu_l t_s} \cos(u_l t_s + \varphi_l) - \sum_{m=2k-1}^n z_m e^{w_m t_s} \right)^2 \longrightarrow \min . \quad (7)$$

Здесь $z_l = R_l \exp(i\varphi_l)$, $z_{l+1} = R_l \exp(-i\varphi_l)$, $w_l = \nu_l + iu_l$, $w_{l+1} = \nu_l - iu_l$. Теперь необходимо минимизировать функцию (7) по действительным параметрам R_1, \dots, R_{k-1} , $\varphi_1, \dots, \varphi_{k-1}$, ν_1, \dots, ν_{k-1} , u_1, \dots, u_{k-1} , z_{2k-1}, \dots, z_n , w_{2k-1}, \dots, w_n .

Таким образом, после $\left[\frac{n}{2} \right] + 1$ этапов будем иметь $\left[\frac{n}{2} \right] + 1$ значений функции (5). Выбирая минимальное из этих значений, тем самым найдем оптимальные величины параметров z_j и w_j .

20 Прогноз временных рядов естественного происхождения

В этой лекции подробно рассмотрены особенности применения одного из описанных выше методов — локальной аппроксимации. В качестве примеров рассмотрены экономические индикаторы — показатели Доу Джонса, S&P и NASDAQ на различных фазах экономических циклов.

20.1 Краткий обзор LA

Основная идея метода локальной аппроксимации состоит в разбиении области определения функции на несколько локальных областей, построении аппроксимирующих моделей и оценке параметров этих моделей отдельно в каждой области. Если функция гладкая, то области могут быть достаточно малыми так, чтобы функция в каждой из них не изменялась слишком резко. Это позволяет применять в каждой области достаточно простые модели, например, линейные.

Главное условие эффективного применения локальной аппроксимации — это удачный выбор размера локальной области, т.е. числа соседей. Это число надо выбрать так, чтобы в каждой области их было бы достаточно для устойчивой оценки параметров и прибавление небольшого числа новых соседей не давало бы существенного изменения оцениваемых параметров.

Проще всего представить суть метода LA на примере прогнозирования биржевых котировок. Для того чтобы сделать прогноз котировок на завтра в соответствии с методом LA, надо просмотреть данные о котировках за весь период наблюдений и найти день наиболее похожий на сегодняшний. Значение котировки в следующий за ним день и будет прогнозом на завтра.

В определенном смысле этот метод напоминает технический анализ, однако он основан на иных представлениях. Использование локальной аппроксимации в окрестности точки в пространстве состояний, задаваемой временным рядом, включает в себя три основных шага:

- **выбор локального представления**, т.е. модели описания эволюции соседей;
- **определение окрестности** (числа соседей);
- **оценка параметров выбранной модели и построение прогноза** в предположении, что изменяется по тому же закону и с теми же параметрами, что и его соседи.

Самый простой способ задания областей состоит в разбиении всей области определения функции на отдельные непересекающиеся подобласти. Например, с помощью прямоугольной сетки. Этот подход имеет *два основных недостатка*.

Во-первых, при таком разграничении отсутствуют перекрытия между отдельными подобластями. Значит, нет непрерывного перехода от модели с параметрами, выбранными для одной области, к модели, полученной для соседней области. Поэтому точки вблизи границ областей будут приближаться плохо. *Второй недостаток* — возможное неравномерное распределение точек по областям, при котором в некоторых из них может оказаться так мало соседей, что будет невозможно оценить все параметры.

Альтернативой разбиению всей области на непересекающиеся прямоугольные подобласти будет **выбор соседей в соответствии с некоторым критерием**. Самый простой из них — **критерий близости**.

Определим $N^{(n)}$ –точечную окрестность (область) \mathbf{x} в пространстве задержек как совокупность точек $\{\mathbf{y}^t\}$, т.е. соседей точки \mathbf{x} , которые находятся "достаточно близко" к \mathbf{x} . Близость некоторой точки \mathbf{y}^t к \mathbf{x} , *хотя мы и рассматриваем динамику изменения наблюдаемой, не подразумевает близость во времени*. Расстояние между точками определяется в соответствии с некоторой мерой в пространстве задержек. Тогда для данной метрики $\|\cdot\|$ при заданном числе соседей $N^{(n)}$ набор $\{\mathbf{y}^t\}$ будет набором ближайших соседей \mathbf{x} , если он обеспечивает минимум $\sum_{t=1}^{N^{(n)}} \|\mathbf{y}^t - \mathbf{x}\|$.

Этот критерий *не оптимален и имеет много недостатков*. Например, уже в двумерном случае понятно, что если треугольник, определенный тремя самыми близкими соседями не включает в себе \mathbf{x} , то прогноз может получиться слишком неточным. Другой недостаток связан с опре-

делением числа соседей: можно указать лишь его нижнюю оценку. Это число должно быть не меньше количества параметров модели, но, сколько точно соседей брать — однозначного рецепта нет. При этом выбор "хороших" соседей — один из важнейших факторов, влияющих на качество предсказаний.

Следующий шаг после выбора соседей — *оценка параметров выбранной для аппроксимации модели*. Выбор модели представляет собой отдельную задачу. Как правило, это делается эмпирически по результатам сравнения точности прогнозов на различных примерах.

В отсутствие априорной информации о системе, породившей прогнозируемый временной ряд, используются **полиномиальные аппроксимации** $F^{t+T}(\mathbf{x}^t) = P_q(x_1^t, x_2^t, \dots, x_M^t)$, где T — время прогноза и q — степень полинома. Поэтому схемы локальной аппроксимации можно классифицировать в соответствии с порядком аппроксимирующего полинома.

Выбрав порядок аппроксимации, можно определить минимально необходимое число соседей, которое равно числу свободных параметров. Приближение, полученное по такому количеству соседей крайне неустойчиво, добавление одного нового соседа радикально меняет результат, и получаемый на основе этого приближения прогноз редко оказывается удачным. Поэтому для повышения устойчивости прогноза количество соседей берут в несколько раз больше, чем необходимое их число.

Казалось бы, переход к более высоким порядкам аппроксимации должен сопровождаться повышением точности прогноза. Однако здесь мы сталкиваемся с той же проблемой, что и при использовании глобальной аппроксимации — *повышение степени аппроксимирующего полинома требует резкого увеличения количества соседей, а большее число соседей — больший размер области — влечет усложнение поведения функции*.

Таким образом, увеличение порядка должно приводить к повышению точности аппроксимации, но его следствие — рост числа соседей к её снижению. Поиск компромисса между этими двумя эффектами — центральная проблема в методе LA.

Существует **три основных способа улучшения прогноза методом**

LA. Рассмотрим их на примере с котировками акций на бирже.

Во-первых, можно выбрать несколько "соседей", т.е. дней с такими же котировками как сегодня, и усреднить их предсказания.

Ещё *большого улучшения* можно достичь, если при выборе "соседей" руководствоваться не только сегодняшним значением, но и динамикой котировок за последние несколько дней, например, за неделю. В этом случае "соседней" будет уже неделя, наиболее близкая по значениям котировок к текущей. Этот переход соответствует увеличению размерности вложения. Теперь мы сравниваем уже не отдельные (скалярные) значения, а векторы, и среди получающихся таким образом "соседей" меньше вероятность появления ложных.

Третий способ улучшения прогноза состоит в коррекции прогноза в зависимости от того, насколько тот или иной сосед отличается от сегодняшнего, стартового вектора (вектор из значений котировок сегодня, вчера, позавчера и т.д.). Метод LA первого порядка (LA1) предполагает линейную зависимость корректирующей добавки от отклонений векторов соседей от стартового вектора. Если учитывать квадратичную зависимость от отклонений, то получим LA2. И так далее. Однако при использовании высоких порядков аппроксимации для оценки необходимых параметров требуется очень большое количество соседей и, следовательно, очень длинные ряды. По этой причине ограничиваются порядком аппроксимации не выше второго.

Оценка вектора параметров, как правило, проводится **методом наименьших квадратов**. Однако его применение в данной задаче имеет ряд особенностей, связанных с необходимостью выбора ближайших соседей стартового вектора, которые в этом случае естественно оказываются близкими между собой. Эта мультиколлинеарность факторов требует применения специальных методов для оценки вектора параметров. Обычно в качестве такого метода используется **SVD-разложение**.

Для прогноза на несколько шагов (дней, если продолжать пример, рассмотренный выше) существует три варианта LA. **Итеративный**, когда параметры (коэффициент влияния отклонений "соседей", т.е. среднее зна-

чение прогнозов от каждого соседа) рассчитываются лишь однажды и на последующих шагах не пересчитываются, а только что прогнозируемый вектор используется в качестве нового стартового вектора.

Модификация этого варианта — **итеративный вариант с пересчётом**. В нём спрогнозированный вектор, также как и в предыдущем случае, становится стартовым, но все соседи выбираются заново. Затем пересчитываются все параметры.

В третьем, "**прямом**" варианте, прогноз строится независимо для всех интересующих моментов времени, как будто это прогнозы на один шаг, без добавления спрогнозированных векторов к исходным данным. Преимущество этого варианта состоит в том, что не происходит накопления ошибки прогноза, т.к. ошибка в одном из шагов никак не влияет на прогноз значения в следующий момент времени.

С точки зрения анализа системы, отражением динамики которой для нас является исследуемый временной ряд, метод ЛА менее информативен по сравнению с глобальной аппроксимацией, так как оценка параметров в глобальном методе может дать дополнительные сведения о самой системе. Однако если главный критерий выбора модели — точность, то в большинстве случаев предпочтительной оказывается локальная аппроксимация, преимущества которой особенно заметны для сложных непериодических функций, когда невозможно повышение точности за счёт накопления и усреднения данных наблюдений.

20.2 Примеры прогнозов экономических показателей

Настоящая лекция — последняя в курсе анализа временных рядов. Поскольку весь материал, представленный выше, достаточно сложен, а большинство исследователей, занимающихся прикладным анализом и прогнозом, интересуются вопросом, как данную теорию применить на практике, в данной лекции представлена только выдержка, где описаны достоинства и недостатки наиболее перспективного метода прогноза — ЛА, поэтому представленный ниже материал дается достаточно подробно.

При построении прогноза для реальных экономических рядов предпо-

чтение было отдано **итеративному варианту метода локальной аппроксимации первого порядка** (итеративный LA1).

Выбор именно *первого порядка аппроксимации* для большинства представленных прогнозов — следствие уже упоминавшегося компромисса между точностью аппроксимации и близостью соседей между собой. Последнее требование особенно важно при прогнозе экономических рядов. Здесь, как правило, сильно выражен тренд и, как следствие редко удается найти необходимое для методов старших порядков число соседей стартового вектора.

Выбор *итеративного варианта* при прогнозе на несколько шагов вперед обоснован для анализа недетерминированных временных рядов.

Количество соседей выбиралось втрое превышающим размерность вложения. В свою очередь, размерность вложения рассчитывалась отдельно для каждого ряда.

Рассмотрим достоинства и недостатки данного метода на примере анализа реальных экономических рядов.

Для оценки эффективности прогнозов имеющийся ряд разбивался на две части. На основе анализа первой части строился прогноз значений ряда для второго, значительно более короткого отрезка.

На рис.24 представлена динамика индекса Доу-Джонса (левее вертикальной синей черты), его прогноз на 80 дней (зеленая линия правее вертикальной черты) и истинные значения (точки правее вертикальной черты). Для построения прогноза использовался ряд, состоящий из 4100 точек, отражающий динамику индекса с 1901 по 1915 годы (на рисунке серыми точками показаны последние 280 значений ряда). Start Point - точка начала прогноза, Dimension - размерность вложения, Order - порядок LA (в данном случае LA1). Как видно из рисунка, удается достаточно точно построить прогноз в течение длительного отрезка времени. Затем начинается расхождение и дальнейшее совпадение с динамикой исходного ряда больше похоже на случайное, чем на реальный прогноз.

Прогноз индекса Доу-Джонса, представленный на рис.24, — это один из лучших результатов, которого удалось достичь с помощью метода LA.

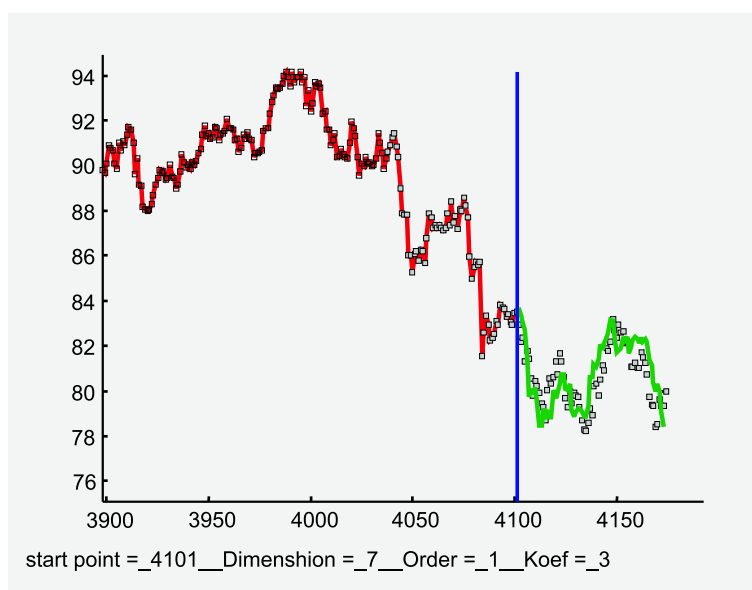


Рис. 24: Прогноз Индекса Доу-Джонса

Этому есть свое объяснение — ряд достаточно длинный (более 4000 точек) и несильно зашумлен. Используемый для прогноза период был выбран не случайно: в это время в динамике ряда можно было уловить элементы квазипериодичности и, что не мало важно, отсутствовал тренд.

При построении прогноза для индекса S&P во время "великой депрессии" (1929 – 1933 гг.) в США (рис.25), было замечено, что лучше всего удается предсказывать дальнейшую динамику ряда, если точку начала прогноза брать в начале подъема (спада) временного ряда. Как видно, можно легко проследить тренд ряда, тогда как локальные спады удается отследить не всегда.

Из рис.25 также следует, что при затяжном подъеме кривой прогноз оказался уже совсем несостоятельным. Сменой порядка метода и подбором числа соседей решить данную проблему не удалось. Подобную картину можно было наблюдать и на большинстве других экономических рядов, имеющих те же особенности.

На следующем рисунке (рис.26) представлены результаты исследования изменения индекса Nasdaq в течение 9 месяцев. Показаны более све-

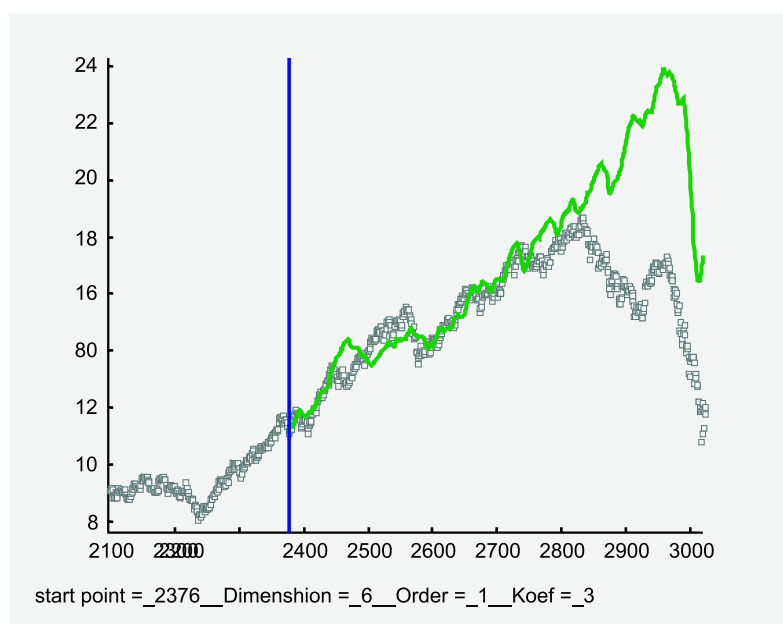


Рис. 25: Прогноз индекса S&P. Период "Великая депрессия".

жие экономические данные (за 2000 год) и даже, несмотря на то, что ряд был коротким, с помощью метода ЛА удалось достичь неплохих результатов. Во-первых, *получилось 23 точки более или менее удачного прогноза до того момента, когда уже началось сильно заметное расхождение*. Во-вторых, *даже после этого момента с помощью ЛА удалось отследить максимальные пики на графике*.

Таким образом, данный метод вполне пригоден для его практического использования в некоторых ситуациях. В каких именно — зависит от природы рассматриваемого ряда. Для точного ответа на этот вопрос необходим глубокий анализ ряда, его предварительная обработка и вычисление некоторых важных инвариантных характеристик, используемых в современной теории динамических систем.

Один из неприятных моментов, с которым пришлось столкнуться при исследовании свойств ЛА помимо того, что для некоторых рядов он оказался просто недееспособен (например, практически для всех курсов акций российских предприятий!), иллюстрируется на рис.27.

На этом рисунке представлены результаты прогноза динамики индек-

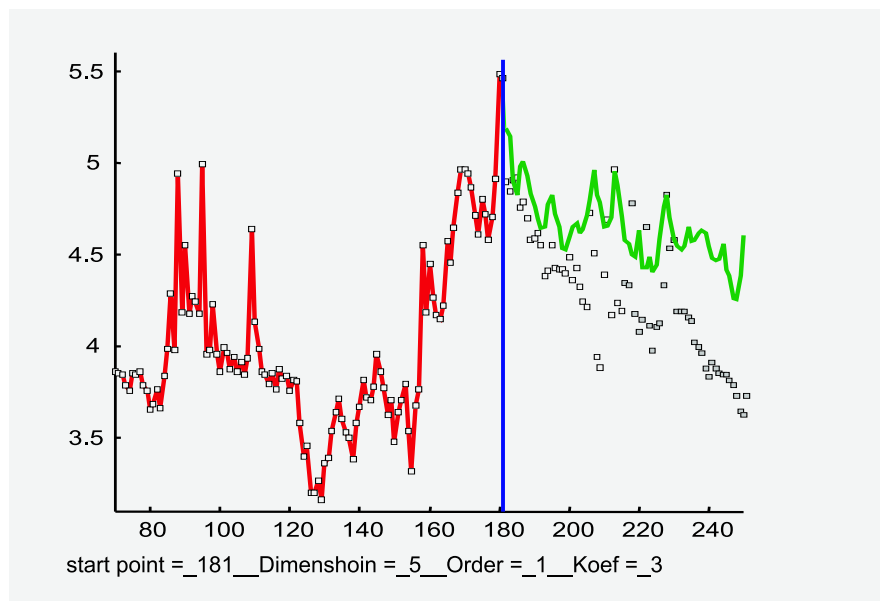


Рис. 26: Прогноз индекса Nasdaq.

са S&P. Видно, что *кривая прогноза по своему характеру напоминает тренд временного ряда, но оказывается смещенной вниз относительно начальной точки прогноза*. Возможно, этот негативный момент связан со скрытыми закономерностями, так как подобный эффект наблюдался еще у некоторых экономических рядов, которые мы исследовали. Кроме того, выбранный период соответствует моменту приближения к "великой депрессии". Вследствие этого большинство процессов, которые "спрятаны" в этом ряде, сильно нестационарны и неустойчивы. Все эти особенности можно попытаться выявить, но это предмет совершенно иного анализа.

Хотелось бы предупредить, что поспешный вывод о нестационарности и неустойчивости процесса на основании результата, представленного на рис.27, может оказаться некорректным. Иными словами, если ряд сильно неустойчив, метод может не работать. Обратное утверждение в общем случае неверно.

На сегодняшний день большинство способов, которые используются для прогнозирования временных рядов в экономике, так или иначе, связаны с построением *авторегрессионных* моделей. Методы выбора пара-



Рис. 27: Прогноз индекса S&P. Приближение к "великой депрессии".

метров этих моделей и их идентификации хорошо проработаны, а сами алгоритмы этих методов входят в большинство статистических и эконометрических программных пакетов.

Рассматриваемый в настоящей работе метод ЛА имеет много общего с авторегрессионными методами. В то же время алгоритм ЛА обладает одним явным *преимуществом* перед обычной авторегрессией. Оно заключается в использовании *не глобально-линейной* (единой для всего ряда), а **кусочно-линейной аппроксимации**. Это позволяет с помощью ЛА успешно прогнозировать *нерегулярные (квазипериодические, хаотические) временные ряды*, для которых линейное авторегрессионное представление неприемлемо.

Основная причина, которая пока ограничивает использование метода ЛА, состоит в том, что его *эффективное применение возможно только для рядов тех показателей, которые фиксировались достаточно продолжительное время*. Этому требованию удовлетворяют лишь немногие ряды финансовых и макроэкономических показателей, например, индекс Доу-Джонса, по которому имеются данные больше чем за 100 последних лет. При анализе столь длинных рядов, конечно, трудно ожидать неизменно-

сти системы в течение всего времени наблюдений. Но для LA, в отличие от авторегрессии, этого и не требуется!

Представленные в работе примеры использования LA для прогнозирования биржевых индексов показывают, что во многих случаях получается действительно неплохой прогноз. Это дает основания считать, что метод LA может рассматриваться как достойная альтернатива авторегрессионным моделям.