Санкт-Петербургский государственный университет
Прикладная математика и информатика
Статистическое моделирование

Федоров Никита, Понизова Вероника

Статистический анализ временных рядов

Конспект

Санкт-Петербург
2020

Оглавление

Глава 1

Общая часть

- Что такое периодограмма?
- Что такое тренд (разные варианты определения)? Как выглядит периодограмма тренда?
- Что такое периодическая компонента? Как выглядит ее периодограмма?
- Что такое шум? Как выглядит его периодограмма?
- Периодограмма как оценка спектральной плотности. Распределение значений. Сглаживание периодограммы.
- Чем отличается сглаживание от выделения тренда?
- Что такое линейный фильтр, импульсная характеристика? Причинный фильтр, FIR.
- Характеристики фильтра через его воздействие на $\cos(2\pi\omega n)$ (или на комплексную экспоненту) – АЧХ, ФЧХ.
- АЧХ фильтра скользящего среднего, зависимость от длины окна.
- АЧХ фильтра перехода к разностям (дифференцирования).
- Что такое запаздывание и отчего оно может возникать (на примере скользящего среднего)?
- Смещение при сглаживании фильтром скользящего среднего. Роль второй производной.
- Фильтр для подавления шума. Роль нормы коэффициентов фильтра.
- Как связаны периодограммы ряда до применения фильтра и после применения фильтра?
- Модели данных – аддитивная и мультипликативная.

- Методы стабилизации дисперсии в разных моделях (логарифмирование, извлечение квадратного корня, …).
- Выделение тренда у ряда с сезонностью (выбор длины окна в скользящем среднем).
- Переход к разностям – плюсы и минусы (устранение тренда, превращение ряда в стационарный, усиление вклада высоких частот).
- Скользящее среднее и скользящая медиана.
- Растекание частоты в периодограмме. Подправка длины ряда для ее устранения.
- Выделение тренда с помощью параметрической регрессии.
- Выделение тренда с помощью метода LOESS
- Нахождение огибающей периодического ряда с помощью выделения тренда.
- Оценивание поведения дисперсии шума с помощью выделения тренда.
- Метод разложения Classical seasonal decomposition.
- Метод разложения STL.

1.1. Периодограмма временного ряда

1.2. Тренд: аналитический и стохастический, периодограмма тренда

1.3. Периодическая компонента и ее периодограмма

1.4. Шум и его периодограмма

1.5. Периодограмма как оценка спектральной плотности, распределение значений, сглаживание периодограммы

1.6. Различия между сглаживанием и выделением тренда

1.7. Линейный фильтр, импульсная характеристика, причинный фильтр, FIR

1.8. Характеристики фильтра через его воздействие на $\cos(2\pi\omega n)$ (или на комплексную экспоненту) – АЧХ, ФЧХ

1.9. АЧХ фильтра скользящего среднего, зависимость от длины окна

1.10. АЧХ фильтра перехода к разностям (дифференцирования)

1.11. Что такое запаздывание и отчего оно может возникать (на примере скользящего среднего)?

1.12. Смещение при сглаживании фильтром скользящего среднего

1.13. Фильтр для подавления шума

1.14. Связь между периодограммами ряда до и после применения фильтра

1.15. Модели данных – аддитивная и мультипликативная

1.16. Методы стабилизации дисперсии в разных моделях

1.17. Выделение тренда у ряда с сезонностью (выбор длины окна в скользящем среднем)

1.18. Переход к разностям – плюсы и минусы (устранение тренда, превращение ряда в стационарный, усиление вклада высоких частот)

1.19. Скользящее среднее и скользящая медиана

1.20. Растекание частоты в периодограмме и методы его устранения

1.21. Выделение тренда с помощью параметрической регрессии

1.22. Выделение тренда с помощью метода LOESS

1.23. Нахождение огибающей периодического ряда с помощью выделения тренда

1.24. Оценивание поведения дисперсии шума с помощью выделения тренда

1.25. Метод разложения Classical seasonal decomposition

1.26. Метод разложения STL

Глава 2

Метод SSA

- Как выбирать L.
- Последовательный SSA.
- Слабая и сильная разделимость.
- Компоненты смешались. Как понять, это слабая или сильная разделимость?
- Как идентифицировать тренд?
- Как идентифицировать периодичность?
- Использование матрицы взвешенных корреляций.
- Элементарные восстановленные компоненты.
- Корни хар.полинома, сигнальные и лишние.
- Оценка параметров в SSA.
- Прогноз.
- Доверительные интервалы.
- Автоматическая идентификация.
- Заполнение пропусков.
- Теплицев SSA для стац. рядов.
- Projection SSA, выделение лин.тренда.
- Улучшение разделимости с помощью вращения в выбранном подпространстве: Iterative O-SSA (слабая и сильная разделимость) и DerivSSA (сильная разделимость).
- Аппроксимация Cadzow SSA.

- MSSA для анализа многомерных временных рядов. Когда лучше анализировать ряды вместе, а когда отдельно?

- 2D-SSA для разложения изображения.

2.1. Метод SSA. Выбор параметра L

Пусть есть временной ряд X длины $N > 2$, который является суммой нескольких временных рядов ($\mathsf{X} = \sum \mathsf{X}_k$). Рассмотрим алгоритм SSA разложения временного ряда.

2.1.1. Первый этап: разложение

Шаг 1. Вложение

На шаге вложения выбирается параметр $1 \leq L \leq N$, называемый *длиной окна.* Затем к ряду X применяется *оператор SSA-вложения*

$$\mathcal{T}^{(L)}_{\mathrm{SSA}}(\mathsf{X}) = \mathbf{X} = [X_1 : \ldots : X_K],$$

сопоставляющий временному ряду X его *L-траекторную матрицу* $\mathbf{X}$, состоящую из $K = N - L + 1$ *векторов вложения* $X_i = (x_{i-1}, \ldots, x_{i+L-2})^{\mathsf{T}}, \quad 1 \leq i \leq K$, имеющих размерность L. Обозначим $d = \mathrm{rank}\mathbf{X}$.

Шаг 2. Сингулярное разложение

На данном шаге происходит SVD (Singular Value Decomposition) — сингулярное разложение траекторной матрицы ряда. Обозначим $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathsf{T}}$, $\lambda_1 \geq \ldots \geq \lambda_L \geq 0$ — *собственные числа* матрицы $\mathbf{S}$, $U_1, \ldots, U_L$ — ортонормированная система *собственных векторов* матрицы $\mathbf{S}$, соответствующих $\lambda_1, \ldots, \lambda_L$, $V_i = \mathbf{X}^{\mathsf{T}} U_i / \sqrt{\lambda_i}, \quad i \in \{1, \ldots, L\}, \lambda_i \neq 0$.

Тогда результат сингулярного разложения — набор *собственных троек*

$$\mathrm{SVD}(\mathbf{X}) = \{(\sqrt{\lambda_i}, U_i, V_i),\ i \in \{1, \ldots, L\}\},\ \mathbf{X} = \sum_i \mathbf{X}_i,\ \mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathsf{T}}.$$

2.1.2. Второй этап: восстановление

Шаг 3. Группировка

Введём $I = \{1, \ldots, d\}$ — множество индексов ненулевых собственных чисел матрицы $\mathbf{S}$. Тогда на шаге группировки I разбивается на непересекающиеся подгруппы

$$I = I_1 \cup \ldots \cup I_m,\ I_j \cap I_r = \varnothing\ \forall j, r \in \{1, \ldots, m\},\ j \neq r,$$

а траекторная матрица $\mathbf{X}$ представляется в виде

$$\mathbf{X} = \mathbf{X}_{I_1} + \cdots + \mathbf{X}_{I_m},\ \mathbf{X}_{I_j} = \sum_{I_j} \mathbf{X}_i.$$

Шаг 4. Диагональное усреднение

На данном шаге происходит две процедуры:

1. К каждой матрице сгруппированного разложения $\mathbf{X}_{I_j}$ применяется *оператор ганкелизации* $\mathcal{H}$, который усредняет матрицы $\mathbf{X}_{I_j}$ вдоль «диагоналей» $i + j = k + 2$, $0 \le k < N$
$$\mathcal{H}(\mathbf{X}_{I_j}) = \widehat{\mathbf{X}}_j,\ j \in \{1, \ldots, m\}.$$
2. К матрицам $\widehat{\mathbf{X}}_j$ применяется оператор $(\mathcal{T}_{\text{SSA}}^{(L)})^{-1}$, проводящий процедуру, обратную процедуре вложения
$$(\mathcal{T}_{\text{SSA}}^{(L)})^{-1}(\widehat{\mathbf{X}}_j) = \widetilde{\mathsf{X}}_j = (\tilde{x}_0^{(j)}, \ldots, \tilde{x}_{N-1}^{(j)}),\ j \in \{1, \ldots, m\}.$$

Таким образом получается разбиение исходного ряда в сумму m рядов длины N
$$\mathsf{X} = \sum_{k=1}^{m} \widetilde{\mathsf{X}}_n^{(k)}.$$

Для выбора параметра L существуют следующие рекомендации:

- В случае, если временной ряд X имеет период m, параметр L стоит брать кратным m.
- Параметр L стоит брать близким к половине длины ряда.

2.2. Последовательный SSA

Пусть $\mathrm{SSA}(\mathsf{X}, L, I = \{I_1, \ldots, I_r\})$ — разложение временного ряда X методом SSA с длиной окна L на r временных рядов с группировкой компонент I. Заметим, что метод SSA можно применять к ряду последовательно, например, применив метод SSA с одной длиной окна выделить конкретную компоненту ряда, а затем применить метод SSA с другой длиной окна к остатку:

$$\mathrm{SSA}(\mathsf{X}, L_1, \{I_1, I_2\}) = \{\mathsf{X}_1, \mathsf{X}_2\}, \ \mathrm{SSA}(\mathsf{X}_2, L_2, I = \{I_{21}, \ldots, I_{2r}\}) = \{\mathsf{X}_{21}, \ldots, \mathsf{X}_{2r}\}.$$

Данная техника называется Sequential SSA. Основное применение этого метода следующее: выделить тренд, применив SSA с маленькой (но не слишком) длиной окна, кратной периоду ряда и последующим применением SSA с длиной окна, близкой к половине длины ряда, для выделения сезонности и разложения её на отдельные периоды.

2.3. Слабая и сильная разделимость

2.4. Определение сильной и слабой разделимости при смешивании компонент

2.5. Идентификация тренда

Для определения тренда можно воспользоваться:

1. Графиком собственных чисел (первые одиночные компоненты)
2. Графиком собственных векторов (медленно меняющиеся компоненты)
3. Графиком взвешенных корреляций (одиночные компоненты, коррелирующие сами с собой)

2.6. Идентификация периодичности (сезонности)

Для определения сезонности можно воспользоваться:

1. Графиком собственных чисел (ступеньки из двух точек на одинаковом уровне)

2. Графиком собственных векторов (гармонические компоненты)

3. Графиком пар собственных векторов (образуют правильные многоугольники с количеством вершин, соответствующим периоду)

4. Графиком взвешенных корреляций (пары последовательных компонент, коррелирующие друг с другом)

Отдельный случай: частота $\frac{1}{2}$.

2.7. Использование матрицы взвешенных корреляций

Матрица взвешенных корреляций вычисляет корреляции между отрезками ряда, используя веса, пропорциональные количеству появлений элемента в траекторной матрице. Если матрица взвешенных корреляций состоит из отдельных сильно коррелирующих блоков, которые не коррелируют между собой, и каждый из блоков соответствует одной из компонент, это показывает на то, что компоненты хорошо разделились.

По матрице взвешенных корреляций можно определить:

- Компоненты тренда: одиночные компоненты, не коррелирующие с остальными
- Гармонические компоненты: пары последовательных компонент, коррелирующие друг с другом
- Компоненты шума: область, начинающаяся с определённой компоненты, в которой компоненты коррелируют друг с другом, образуя «облако».
- Смешение компонент: мутные блоки, где есть корреляции между компонентами, которые принадлежат разным слагаемым.

2.8. Элементарные восстановленные компоненты

По графикам элементарных восстановленных компонент можно также пытаться определять:

- Компоненты тренда (медленно меняющиеся компоненты)
- Пилообразную компоненту, соответствующую частоте $\frac{1}{2}$.
- Гармонические компоненты (пары похожих по поведению гармонических элементарных компонент) и то, не смешались ли они (при смешении двух компонент с близкими частотами образуются «виолончели» — компоненты, огибающие которых тоже являются гармоническими компонентами)

2.9. Корни характеристического полинома: сигнальные и лишние

Говорят, что ряд порожден линейным рекуррентным соотношением, если

$$\exists(a_1, a_2, \dots, a_d) : \ \forall n \geq d+1 : \ x_n = \sum_{i=1}^{d} a_i x_{n-i}$$

Мы будем говорить, что размерность ряда не превосходит d, $\mathrm{fdim}(\mathsf{X}_N) \leq d$, если $a_d \neq 0$. Если $(a_1, \dots, a_d)$ – минимальный набор коэффициентов, то $\mathrm{fdim}(\mathsf{X}_N) = d$. Для любого ряда, $\mathrm{fdim}(\mathsf{X}_N) \geq 1$.

Мы можем решить линейное рекуррентное соотношение, если найдем явную формулу для x_n, зависящую только от n. Для этого рассмотрим полином вида:

$$\lambda^n = \sum_{i=1}^{d} a_i \lambda^i,$$

коэффициенты этого многочлена задают LRR из определения. Такой полином называется характеристическим. Заметим, что если число λ_i – корень такого полинома, то ряд $x_n = \lambda_i^n$ удовлетворяет LRR.

Вообще говоря, линейная комбинация степенных функций корней многочлена будет удовлетворять LRR. Решение LRR сводится к нахождению или оценке корней, а также коэффициентов линейной комбинации, задающей ряд. В общем виде член ряда будет задаваться следующей формулой:

$$x_n = \sum_{i=1}^{p} \left(\sum_{j=0}^{k_{i-1}} C_j n^j \right) \lambda_i^n$$

где k_i – кратность корня λ_i, а p – количество различных корней характеристического полинома, C_j — неизвестные коэффициенты, которые определяются первыми d членами ряда, или оцениваются по МНК, если присутствует шум.

Заметим, что только ряд, представимый в виде сумм и произведений экспонент, полиномов и гармоник может быть точно задан с помощью LRR.

Если LRR имеет порядок $d = \mathrm{fdim}(\mathsf{X}_N)$, то все корни её характеристического многочлена — сигнальные. В случае, когда ряд представлен LRR большего порядка, у характеристического многочлена данной LRR появляются дополнительные (лишние) корни.

2.9.1. Как отличить сигнальные корни

Поскольку LRR, получающаяся при применении SSA получается min-norm LRR, лишние корни по модулю будут меньше 1. Если изобразить все корни характеристического полинома на комплексной плоскости, лишние корни будут внутри единичной окружности. Однако, сигнальные корни тоже могут быть по модулю меньше 1. В таком случае можно рассмотреть корни характеристическо многочлена ряда Y_N, где $y_n = x_{N+1-n}$ (ряд со временем, идущим в обратную сторону). У характеристического многочлена такого ряда сигнальные корни будут иметь модуль $\dfrac{1}{\lambda}$, где λ — модуль корня в изначальном многочлене. Таким образом можно отличить сигнальные корни от лишних, поскольку и для такого ряда мы строим min-norm LRR, и лишние корни будут по прежнему меньше 1 по модулю.

2.10. Оценка параметров в SSA

Для оценки параметров сигнала можно воспользоваться методом ESPRIT, идея которого заключается в следующем: пусть имеется временной ряд $\mathsf{X} = \mathsf{S} + \mathsf{R}$, где S описывается LRR порядка r, а R — шум. Тогда можно использовать $\mathrm{span}\{U_1, \ldots, U_r\}$, получающиеся в процессе применения SSA к X, как оценку сигнального подпространства. Таким образом, данный метод позволяет получить оценки корней LRR сигнала, если нам известен порядок сигнала.

2.11. Прогноз

В SSA есть два вида прогноза — рекуррентный и векторный. Также можно рассматривать разные сигнальные подпространства — строковое (row forecast) и столбцовое (column forecast).

В рекуррентном прогнозе строится проекция вектора $e_L = (\mathbf{0}_{L-1}, 1)^\mathsf{T}$ на ортогональное дополнение сигнального подпространства. Полученный вектор состоит из коэффициентов min-norm LRR, описывающей сигнал. Предсказания в следующих точках образуются применением LRR.

В векторном прогнозе рассматривается пространство, натянутое на вектора, полученные удалением последней координаты из SSA-оценок сингулярных векторов исходного сигнального подпространства. Затем последующие вектора, содержащие прогнозируемые точки, образуются путём применения оператора, построенного с помощью проектора это подпространство. После того, как получена матрица, содержащая новые вектора, производится процедура, аналогичная обратному преобразованию траекторной матрицы во временной ряд при применении SSA.

В пакете Rssa данные алгоритмы реализованы функциями rforecast и vforecast.

2.12. Доверительные интервалы

Доверительные интервалы для прогнозов в SSA строятся посредством вычисления доверительных интервалов для средних bootstrap-оценок, используя распределение предсказанных значений в предположении стационарности шума. В Rssa данная процедура реализована функцией bforecast.

2.13. Автоматическая идентификация

Автоматическую группировку в SSA можно осуществлять либо на основе матрицы взвешенных корреляций, либо на основе периодограммы.

Группировка по матрице взвешенных корреляций основана на применении процедуры иерархической кластеризации, где расстояние между кластерами зависит от взвешенной корреляции между компонентами ($\mathrm{dist}(\mathsf{X}_i, \mathsf{X}_j) = 0.5(1 - \rho_{ij}^{(w)})$). Для данной группировки нужно задать необходимое количество кластеров.

Группировка по периодограмме основана на разбиении периодограммы на интервалы и вычисление вклада элементарных компонент разложения в эти интервалы. Также можно задать минимальный порог вклада компоненты.

В Rssa данные методы реализованы функцией grouping.auto.

2.14. Заполнение пропусков

Заполнение пропусков в SSA осуществляется двумя методами — итеративно и с помощью прогнозов.

В итеративном методе заполнения пропусков (igapfill в Rssa) пропущенные значения инициализируются, затем производится восстановление сигнала. Данная процедура продолжается, пока значения на месте пропусков не сойдутся к определённому значению (изменения при посоледующих итерациях незначительны). Данная процедура сильно зависит от способа инициализации значений.

В методе заполнения пропусков с помощью прогнозов происходит процедура, аналогичная рекуррентному прогнозу: пропуски заполняются с помощью LRR. Прогнозы пропусков происходят с обеих сторон со взятием (взвешенного) среднего значения.

2.15. Теплицев SSA для стационарных рядов

В случае, когда ряд стационарен, рекомендуется использовать модификацию обычного SSA, которая назвается Теплицев SSA (Toeplitz SSA). Идея данной модификации заключается в том, чтобы вместо матрицы $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathsf{T}}$ подставлять теплицеву матрицу $\widetilde{\mathbf{S}}$, образованную автоковариациями $\tilde{c}_{ij} = \dfrac{1}{N-|i-j|} \sum\limits_{0}^{N-|i-j|-1} x_m x_{m+|i-j|}$ (Сразу приводим траекторную матрицу к теплицеву виду). Также рекомендуется произвести центрирование ряда перед применением данной процедуры (чтобы «корреляции» действительно были корреляциями).

2.16. Projection SSA, выделение линейного тренда

Идея SSA с проекциями заключается в следующем: рассмотрим временной ряд X длины N и его траекторную матрицу $\mathbf{X}$. Тогда схема метода SSA with projection может быть описана следующим образом:

- Построим матрицу $\mathbf{C} = \mathbf{C}_{\mathbf{X}}$, основываясь на некоторой априорной информации о временном ряде X
- Вычислим $\mathbf{X}' = \mathbf{X} - \mathbf{C}$
- Проведём SVD-разложение матрицы $\mathbf{X}' = \sum\limits_i \sqrt{\lambda_i'} U_i' V_i'^{\mathsf{T}}$

В результате получим разложение матрицы $\mathbf{X}$: $\mathbf{X} = \mathbf{C} + \mathbf{X}' = \mathbf{C} + \sum\limits_i \sqrt{\lambda_i'} U_i' V_i'^{\mathsf{T}}$

Особенность метода заключается в способе построения матрицы $\mathbf{C}$. Если в качестве матрицы $\mathbf{C}$ использовать проектор на пространство, порождённое вектором из единиц, получим SSA с однократным центрированием (по строкам или по столбцам). Таким же образом можно получить SSA с двукратным центрированием (как по строкам, так и по столбцам). Данные применения хороши для выделения константного и линейного трендов. В общем случае можно строить проектор исходя из дополнительной информации о структуре сигнала, например, если имеется ряд с сигналом похожей структуры.

2.17. Iterative O-SSA и DerivSSA

2.17.1. Oblique SSA

Ортогональность отрезков некоторой длины L и $K = N - L + 1$ рядов $\mathsf{X}_N^{(1)}$ и $\mathsf{X}_N^{(2)}$ является определением слабой разделимости этих рядов. Эта ортогональность нужна нам для построения SVD разложения и последующей группировки слагаемых.

На практике мы чаще имеем дело не с точной разделимостью, а асимптотической, то есть когда ортогональность достигается при увеличении до бесконечности длины ряда и размера окна. В некоторых случаях, например, для гармоник с близкими частотами, может потребоваться очень большая длина ряда, чтобы получить удовлетворительную разделимость компонент.

До этого мы говорили об ортогональности в евклидовом пространстве. Мы можем рассмотреть общий вид скалярного произведения с симметричной неотрицательно-определенной матрицей $\mathbf{A}$: $\langle X, Y \rangle \mathbf{A} = X^{\mathsf{T}}\mathbf{A}Y$. В евклидовом пространстве $\mathbf{A}$ – единичная матрица.

Для двух произвольных векторов X и Y мы можем подобрать такую $\mathbf{A}$, что они будут ортогональны относительно этой матрицы. Мы можем обобщить и алгоритм SVD, потребовав чтобы вектора $\{U_i\}$ и $\{V_i\}$ образовывали ортонормированный базис относительно некоторых матриц $\mathbf{L}$ и $\mathbf{R}$. Мы назовем такое разложение ($\mathbf{L}$,$\mathbf{R}$)?SVD

Мы также можем ввести понятие слабой и сильной ($\mathbf{L}$,$\mathbf{R}$)-разделимости для рядов, определив ортогональность столбцов и сток траекторных матриц относительно соответсвующих матриц $\mathbf{L}$ и $\mathbf{R}$. Оказывается, требование сильной ($\mathbf{L}$,$\mathbf{R}$)-разделимости слабее исходного определения сильной разделимости.

Идея косоугольного SSA заключается в том, чтобы подобрать такие $\mathbf{L}$ и $\mathbf{R}$, чтобы получить лучшую разделимость компонент.

2.17.2. Deriv-SSA

Для того, чтобы при любом разложении SSA мы могли из троек составить исходные ряды, нам нужна сильная разделимость. Для сильной разделимости требуется, чтобы компонентам соответствовали разные собственные числа.

Мы можем столкнуться с проблемой близких собственных числе даже когда компоненты на самом деле ортогональны. Например, для гармоник с разными частотами, но одинаковыми амплитудами, собственные числа будут близки. И хотя гармоники мы сможем отделить от шума, но их самих мы разделить не сможем.

Мы хотим изменить вклады компонент, и, как следствие, соответсвующие собственные числа, притом не затрагивая подпространства, порожденные этими компонентами. Мы можем это сделать, например добавив точно такую же компоненту в ряд, но мы не можем этого сделать, потому что заранее они нам неизвестны.

Поскольку мы предполагаем, что работаем в основном с экспонентами, гармониками и полиномами, естественное решение — это использовать дифферецирование, чтобы получить нужные компоненты. Действительно, $(A\cos(2\pi\omega n))' = A(2\pi\omega)\sin(2\pi\omega n)$, $(Ae^{\alpha n})' = A\alpha e^{\alpha n}$ — дифферецирование не меняет порожденного подпространства, полином в этом случае — исключение.

Мы работаем с дискретным временем, поэтому вместо дифференцирования используем разности соседних элементов. Из ряда X_N мы получаем новый ряд Y_{N-1}, такой что $y_n = x_n - x_{n-1}$. Этот ряд мы дописываем в конец X_N и применяем SSA уже к этому ряду длины $2N-1$. Можно показать, что это частный случай O-SSA.

Такой фильтр усиливает влияние высоких частот, поэтому применять такой метод лучше всего к уже выделенному сигналу без шума, чтобы улучшить разбиение на отдельные компоненты.

2.18. Аппроксимация Cadzow SSA

Алгоритм аппроксимации сигнала Cadzow SSA:

1. Провести шаги алгоритма SSA вплоть до SVD-разложения.
2. Применить некоторую функцию f к сингулярным значениям полученного разложения (к матрице $\mathbf{\Sigma}$).
3. (Опционально) Отбросить компоненты, соответствующие слишком маленьким собственным числам (в предположении, что они относятся к шуму).
4. Оставить первые r компонент (предполагаем, что ранг сигнала равен r).
5. Повторять процедуру, пока алгоритм не сойдётся по какой-нибудь заранее заданной метрике.

В базовом алгоритме Cadzow функция f соответствует функции id.

2.19. MSSA для анализа многомерных временных рядов

Метод MSSA является обобщением метода SSA на случай многомерных временных рядов.

Основные отличия метода MSSA от базового метода SSA:

1. На шаге вложения траекторная матрица ряда составляется из траекторных матриц его компонент:

$$\mathbf{X} = \mathcal{T}^{(L)}_{\mathrm{MSSA}}(\mathsf{X}) = \left[\mathcal{T}^{(L)}_{\mathrm{SSA}}(\mathsf{X}^{(1)}) : \ldots : \mathcal{T}^{(L)}_{\mathrm{SSA}}(\mathsf{X}^{(s)})\right] = \left[\mathbf{X}^{1} : \ldots : \mathbf{X}^{(s)}\right].$$

2. На шаге диагонального усреднения матрицы $\mathbf{X}_{I_j}$ приводятся к составному ганкелеву виду с помощью покомпонентного применения оператора ганкелизации:

$$\overline{\mathcal{H}}(\mathbf{X}_{I_j}) = \left[\mathcal{H}(\mathbf{X}^{(1)}_{I_j}) : \ldots : \mathcal{H}(\mathbf{X}^{(s)}_{I_j})\right].$$

Метод MSSA может дать результаты лучше, чем применение базового SSA к обоим рядам, если у рядов есть компоненты с общей структурой (т.е. лежащие в одном подпространстве пространства, образованного траекторной матрицей).

2.20. 2D-SSA для разложения изображения

Отличие 2D-SSA от SSA в том, что на вход подаётся двумерный массив, вместо вектора, являющегося реализацией временного ряда, и окно, с помощью которого образуется траекторная матрица так же является прямоугольником (двумерное) и параметр длины окна $L = (L_x, L_y)$. В качестве столбцов траекторной матрицы выступают векторизации прямоугольных окон.

Данный метод применяется для разложения изображений. Дополнительным расширением данного метода является Shaped 2D-SSA, где окна могут быть не прямоугольными.

Глава 3

ARIMA

- AR(p) – модель, запись в виде с оператором сдвига
- AR(p) и модель сигнала в SSA
- Вид автоковариационной функции acf для AR(p)
- Вид pacf для AR(p)
- Модель MA(q), вид acf и pacf
- ARMA(p,q)
- Дифференцирование, ARIMA
- Seasonal ARIMA(p,d,q)(P,D,Q)
- Exponential smoothing, модели тренда, ES и ARIMA

3.1. AR(p) – модель, запись в виде с оператором сдвига

Авторегрессионный процесс порядка p определяется следующим образом

$$x_n = c + \sum_{i=1}^{p} \phi_i x_{n-i} + \varepsilon_n,$$

где ε_n — i.i.d. с $\mathbb{E}\varepsilon_n = 0$ и $\mathbb{D}\varepsilon_n = \sigma^2$ (как правило $\varepsilon_n \sim N(0, \sigma^2)$), и ε_n назависимо с x_k при $k < n$. Пусть существуют такие параметры, что процесс $\mathsf{X} = \{x_n\}$ стационарный.

Это выражение можно переписать через оператор сдвига. Пусть $Lx_n = x_{n-1}$ — оператор сдвига. Тогда,

$$\varepsilon_n = (1? \sum_{i=1}^{p} \phi_i L^i) = \Phi(L) x_n,$$

где Φ - характеристический полином.

3.2. AR(p) и модель сигнала в SSA

В SSA рассматривается модель ряда $x_n = s_n + \varepsilon_n$, где $s_n = \sum_{i=1}^{p} \phi_i s_{t?i}$ — сигнал и ε_n — шум. Модель авторегрессии имеет вид $x_n = \sum_{i=1}^{p} \phi + ix_{n-i} + \varepsilon_n$. В случае SSA шум добавляется ко всему сигналу, а в случае AR - на каждом шаге. Несмотря на то, что модели выглядят похожими, SSA рассматривает авторегрессию только как модель для шума.

3.3. Вид автоковариационной функции acf для AR(p)

Для модели AR(p) автоковариационная функция экспоненциально убывает и синусоидальна.

3.4. Вид pacf для AR(p)

Для модели AR(p) pacf будет иметь значимый скачок на сдвиге p, но больше значимых скачков после p не будет.

3.5. Модель MA(q), вид acf и pacf

Модель скользящего среднего порядка q определяется следующим образом $x_n = \varepsilon_n + \sum_{j=0}^{q} \theta_j \varepsilon_{n?j}$, где ε_n - белый шум, θ_j - параметры модели.

Представление через оператор сдвига

$$x_n = (1 + \sum_{j=1}^{q} \theta_j L^j)\varepsilon_n = \Theta(L)\varepsilon_n.$$

Виды acf и pacf для MA(q) по сравнению с AR(q) «меняются местами».

3.6. ARMA(p,q)

Моделью ARMA(p, q), где p и q — целые числа, задающие порядок модели, называется следующий процесс

$$x_n = c + \varepsilon_n + \sum_{i=1}^{p} \phi_i x_{n?i} + \sum_{i=1}^{q} \theta_i \varepsilon_{n?i}.$$

Запись через оператор сдвига

$$\Phi_p(L) x_n = \Theta_q(L) \varepsilon_n.$$

Если мы имеем дело со смешанной моделью ARMA(p, q), то графики уже не помогут. Для того, чтобы определить p и q, мы можем воспользоваться информационными критериями. Обозначим $\theta = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q)$ — параметры модели, X_N — ряд, $L = L(\theta|\mathsf{X}_N)$ —- максимум функции правдоподобия для модели со степенями p и q. Мы рассмотрим несколько информационных критериев:

$$\begin{aligned} \mathrm{AIC} &= -2\log L + 2(p+q+1), \\ \mathrm{AIC}_c &= \mathrm{AIC} + \frac{2(p+q+k+1)(p+q+k+2)}{N-p-q-k-2}, \\ \mathrm{BIC} &= \mathrm{AIC} + \lfloor \log N - 2 \rfloor (p+q+k+1). \end{aligned}$$

Мы выбираем какой-нибудь из критериев, перебираем несколько пар значений (p, q) и останавливаемся на той модели, для которой значение выбранного критерия наименьшее.

3.7. Дифференцирование, ARIMA

ARIMA — интегрированная модель авторегрессии — скользящего среднего. Дифференцирование (создание нового ряда посредством применения разностного оператора $(1-L)$) используются для того, чтобы нестационарный ряд привести с стационарному виду. Модель ARIMA(p, d, q) означает, что разности временного ряда порядка d подчиняются модели ARMA(p, q).

Запись модели через оператор сдвига

$$\Phi_p(L)(1-L)^d x_n = \Theta_q(L)\varepsilon_n.$$

3.8. Seasonal ARIMA(p,d,q)(P,D,Q)

До сих пор мы предполагали, что у ряда нет сезонной составляющей. Если у ряда есть сезонные колебания, то имеет смысл рассматривать наблюдения, которые относятся к одной и той же части цикла для построения модели. Например, если мы имеем дело с ежемесячными наблюдениями, то имеет смысл рассматривать наблюдения с шагом 12.

Пусть частота сезонности равна m. Мы можем рассматривать m, например, как количество наблюдений за год. Тогда модель ARIMA с учетом сезонности будет выглядеть так:

$$(1-\phi_1 L^{-}\phi_2 L^2 - \ldots - \phi_p L^p)(1-\Phi_1 L^m - \Phi_2 L^{2m} - \ldots - \Phi_P L^{mP})(1-L)^d(1-L^m)^D x_n =$$
$$= c + (1+\theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)(1+\Theta_1 L^m + \Theta_2 L^{2m} + \ldots + \Theta_Q L^{mQ})\varepsilon_n.$$

Мы можем ввести сезонное дифференцирование: $(1-L^m)x_n$. Если ряд один раз сезонно-дифферецируем, то такая операция сведет его к стационарному процессу. Если ряд имеет и сезонность, и тренд, то двойное дифференцирование $(1-L)(1-L^m)x_n$ приведет его к стационарному виду.

3.9. Exponential smoothing, модели тренда, ES и ARIMA

Аббревиатура ETS(M,A,N) расшифровывается как Error, Trend, Seasonal. Далее мы расскажем, что модель может быть примененая для ряда случаев: для аддитивных и мультипликативных ошибок, постоянного или монотонного тренда, мультипликативной или аддитивной сезонности. Каждый из случаев носит название, состоящее из трех букв. Например ETS(M, A, N) будет означать, что у ряда мультипликативные ошибки и монотонный тренд без сезонных колебаний.

3.9.1. Простое экспоненциальное сглаживане

В первом отчете мы рассматривали скользящее среднее. Мы можем использовать его для построения простейшего предсказания: $\hat{x}_{N+1} = \frac{1}{N}\sum_{i=1}^{N} x_i$, где N — длина исходного ряда. Заметим, что в таком определении все точки получают одинаковый вес, одинаково влияют на предсказание. Нам бы хотелось, чтобы влияние старых наблюдений было меньше.

Давайте сделаем такой фильтр. Выберем некоторое $0 < \alpha < 1$ и положим

$$\hat{x}_{N+1} = \alpha x_N + \alpha(1-\alpha)x_{N-1} + \alpha(1-\alpha)^2 x_{N-2} + \dots$$

При таком определении у старых наблюдений вес убывает экспоненциально. Перепишем нашу формулу:

$$\hat{x}_{N+1} = \alpha x_N + (1-\alpha)l_n.$$

Аналогично мы определим и предыдущие элементы ряда:

$$2 \leq n \leq N, l_n = \alpha x_{n-1} + (1-\alpha)l_{n-1}.$$

На самом деле мы определили бесконечный линейный фильтр (Infinite Response Filter), но мы имеем дело с конечными последовательностями, поэтому нам необходимо найти начальное значение, дальше которого формулу мы не разворачиваем: $l_2 = \alpha x_1 + (1-\alpha)l_1$. l_1 – это начальное значение для нашего фильтра, и мы находим его с помощью метода наименьших квадратов вместе с параметром α. Теперь мы можем явно выразить предсказание:

$$\hat{x}_{N+1} = \sum +i = 1^N \alpha(1-\alpha)^{N-i} x_i + (1-\alpha)l_1.$$

Для дальнейших рассуждений нам удобно представить формулу в составном виде:

$$\hat{x}_{N+t}l_{n+1} = l_n + 1, t \geq 1,$$
$$l_{n+1} = \alpha x_n + (1-\alpha)l_n, 1 \leq n \leq N.$$

- x_n – данные, с которыми мы работаем, исходный временной ряд.
- l_n – предсказанное для x_n значение с помощью сглаживания.

С помощью такой формулы мы сможем предсказывать значение только для рядов без тренда и сезонности: мы предсказываем одно и то же будущее значение и никак не учитываем изменения ряда.

3.9.2. Выделение линейного тренда

Хольтом (Holt) в 1957 было предложено расширение метода для рядов с линейным трендом:

$$\hat{x}_{N+t} = l_{N+1} + t \cdot b_{N+1},\ t \geq 1$$
$$l_{n+1} = \alpha x_n + (1-\alpha)(l_n - b_n), 1 \leq n \leq N$$
$$b_{n+1} = \beta^*(l_{n+1} - l_n) + (1-\beta^*)b_n, 1 \leq n \leq N$$

b_{N+1} определяет скорость роста ряда, также как и для l_n ,нам понадобится найти начальное значение параметра b_1 и β^* с помощью метода наименьших квадратов.

3.9.3. Замедление скорости роста

Описанная ранее формула предполагает неограниченный рост значений ряда для t??. Мы хотели бы предсказывать не столь большие значения для далеких предсказаний. Для этого мы немного изменим формулу:

$$\hat{x}_{N+t} = l_{N+t} + (\sum_{i=1}^{t} \phi^t) b_{N+1}$$
$$l_{n+1} = \alpha x_n + (1-\alpha)(l_n - \phi b_n), 1 \leq n \leq N$$
$$b_{n+1} = \beta^*(l_{n+1} - l_n) + (1-\beta^*)\phi b_n, 1 \leq n \leq N$$

Полагая $0 < \phi < 1$, получим что сумма $\sum_{i=1}^{t} \phi^t$ оказывается конечной для $t \to \infty$ и равна $\frac{1}{1-\phi}$. Таким образом мы можем ограничить значения для далеких предсказаний.

3.9.4. Holt-Winters сезонное сглаживание

Винтерс (Winters) в 1960 расширил метод, предложенный Хольтом, добавив сезонную составляющую. Обозначим m – частоту сезонности, $m \geq 1$. Например, для ежемесячных наблюдений $m = 12$, для квартальных — $m = 4$.

$$\hat{x}_{N+t} = l_{N+1} + t \cdot b_{N+1} + s_{N+t-m(\lfloor \frac{t-1}{m} \rfloor + 1)},\ t \geq 1$$
$$l_{n+1} = \alpha(x_n - s_{n-m+1}) + (1-\alpha)(l_n - b_n), 1 \leq n \leq N$$
$$b_{n+1} = \beta^*(l_{n+1} - l_n) + (1-\beta^*)b_n, 1 \leq n \leq N$$
$$s_{n+1} = \gamma(x_{n+1} - l_n - b_n) + (1-\gamma)s_{n-m+1}$$

Добавилось ещё $m{+}1$ параметров: $s_1, \ldots, s_m$ – начальные значения для сезонностей и параметр γ, которые нам нужно определить как и раньше с помощью МНК. Мы также можем замедлить скорость роста тренда при $t \to \infty$, заменив $t \cdot b_{N+1}$ на $\sum_{i=1}^{t} \phi^t b_{N+1}$.

Также метод может быть обобщен для мультипликативных рядов.

3.9.5. State Space Model

До сих пор мы рассматривали методы для точечных предсказаний, но мы хотим также строить доверительные интервалы для них, а также уметь сравнивать разные методы. Нам понадобится ввести статистическую модель для этого.

Для начала рассмотрим простое экспоненциальное сглаживание:

$$\hat{x}_{N+1} = l_{N+1}$$
$$l_{n+1} = \alpha x_n + (1-\alpha) l_n.$$

Мы можем переписать второе уравнение в следующем виде:

$$l_{n+1} = l_n + \alpha(x_n - l_n).$$

Заметим, что второе слагаемое в равной части соответствует ошибке предсказания, давайте обозначим её за ε_n. Тогда $l_{n+1} = l_n + \alpha\varepsilon_n$. Предположим, что $\varepsilon_n \sim N(0, \sigma^2)$, независимы и одинаково распределенные.

Мы получим следующие уравнения:

$$x_n = l_n + \varepsilon_n,$$
$$l_{n+1} = l_n + \alpha \varepsilon_n.$$

Первое позволяет установить связь с элементам исходного ряда, а последнее задаёт правило перехода к предсказанию следующего наблюдения. Заметим, что если $\alpha = 1$, то мы получим простое случайное блуждание.

Для модели с линейным трендом мы обозначим $\varepsilon_n = x_n - l_n - b_n$, тогда мы можем переписать наши формулы в следующем виде:

$$x_n = l_n + b_n + \varepsilon_n,$$
$$l_{n+1} = l_n + b_n + \alpha \varepsilon_n,$$
$$b_{n+1} = b_n + \beta \varepsilon_n,$$

где $\beta = \alpha \beta^*$. Для ряда с сезонностью с периодом m и линейным трендом, предполагая $\varepsilon_n = x_n - l_n - b_n - s_{n-m}$, мы получим следующие формулы:

$$x_n = l_n + b_n + s_{n-m+1} + \varepsilon_n,$$
$$l_{n+1} = l_n + b_n + \alpha \varepsilon_n,$$
$$b_{n+1} = b_n + \beta \varepsilon_n,$$
$$s_{n+1} = s_{n-m+1} + \gamma \varepsilon_n.$$

Мы также можем определить модели для случая мультипликативной сезонности, мультипликативных ошибок, а также модель с подавлением роста тренда.

3.9.6. Выбор модели

Теперь, когда мы определили модель данных, мы можем находить параметры с помощью метода максимального правдоподобия. Кроме того, с помощью метода максимального правдоподобия мы можем сравнивать разные модели.

Определим информационные критерии на основе значения функции правдоподобия:

$$\mathrm{AIC} = -2\log(L) + 2k,$$
$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{k(k+1)}{N-k-1},$$
$$\mathrm{BIC} = \mathrm{AIC} + k[\log(N) - 2],$$

где L — значение функции правдоподобия, k — количество параметров модели. Такие критерии позволяют учитывать сложность модели и количество наблюдений.

Существует эквивалентность между ARIMA и некоторыми моделями экспоненциального сглаживания. Например, модель с постоянным трендом и без сезонности (Simple Exponential smoothing) эквивалентна модели ARIMA(0,1,1). Модель с линейным трендом и без сезонности эквивалентна ARIMA(0,2,2).