

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	5
1.1. Периодограмма временного ряда	7
1.2. Тренд: аналитический и стохастический, периодограмма тренда	8
1.3. Периодическая компонента и ее периодограмма	9
1.4. Шум и его периодограмма	10
1.5. Периодограмма как оценка спектральной плотности, распределение значений, сглаживание периодограммы	11
1.6. Различия между сглаживанием и выделением тренда	12
1.7. Линейный фильтр, импульсная характеристика, причинный фильтр, FIR	13
1.8. Характеристики фильтра через его воздействие на $\cos(2\pi\omega n)$ (или на комплексную экспоненту) – АЧХ, ФЧХ	14
1.9. АЧХ фильтра скользящего среднего, зависимость от длины окна	15
1.10. АЧХ фильтра перехода к разностям (дифференцирования)	17
1.11. Что такое запаздывание и отчего оно может возникать (на примере скользящего среднего)?	18
1.12. Смещение при сглаживании фильтром скользящего среднего, роль второй производной	19
1.13. Фильтр для подавления шума	20
1.14. Связь между периодограммами ряда до и после применения фильтра	21
1.15. Модели данных — аддитивная и мультипликативная	23
1.16. Методы стабилизации дисперсии в разных моделях	24
1.17. Выделение тренда у ряда с сезонностью (выбор длины окна в скользящем среднем)	25
1.18. Переход к разностям – плюсы и минусы (устранение тренда, превращение ряда в стационарный, усиление вклада высоких частот)	26
1.19. Скользящее среднее и скользящая медиана	27
1.20. Растекание частоты в периодограмме и методы его устранения	28
1.21. Выделение тренда с помощью параметрической регрессии	30
1.22. Выделение тренда с помощью метода LOESS	32
1.23. Нахождение огибающей периодического ряда с помощью выделения тренда	35

1.24. Оценивание поведения дисперсии шума с помощью выделения тренда	36
1.25. Метод разложения Classical seasonal decomposition	38
1.26. Метод разложения STL	39
Глава 2. Метод SSA	41
2.1. Метод SSA. Выбор параметра L	43
2.1.1. Первый этап: разложение	43
2.1.2. Второй этап: восстановление	43
2.2. Последовательный SSA	45
2.3. Слабая и сильная разделимость	46
2.3.1. Слабая разделимость	46
2.3.2. Сильная разделимость	46
2.4. Определение сильной и слабой разделимости при смешивании компонент	47
2.5. Идентификация тренда	48
2.6. Идентификация периодичности (сезонности)	49
2.7. Использование матрицы взвешенных корреляций	50
2.8. Элементарные восстановленные компоненты	51
2.9. Корни характеристического полинома: сигнальные и лишние	52
2.9.1. Как отличить сигнальные корни	53
2.10. Оценка параметров в SSA	54
2.11. Прогноз	55
2.12. Доверительные интервалы	56
2.13. Автоматическая идентификация	57
2.14. Заполнение пропусков	58
2.15. Теплицев SSA для стационарных рядов	59
2.16. Projection SSA, выделение линейного тренда	60
2.17. Iterative O-SSA и DerivSSA	61
2.17.1. Oblique SSA	61
2.17.2. Deriv-SSA	61
2.18. Аппроксимация Cadzow SSA	63
2.19. MSSA для анализа многомерных временных рядов	64
2.20. 2D-SSA для разложения изображения	65

Глава 3. ARIMA	66
3.1. AR(p) – модель, запись в виде с оператором сдвига	67
3.2. AR(p) и модель сигнала в SSA	68
3.3. Вид автоковариационной функции acf для AR(p)	69
3.4. Вид pacf для AR(p)	70
3.5. Модель MA(q), вид acf и pacf	71
3.6. ARMA(p,q)	72
3.7. Дифференцирование, ARIMA	73
3.8. Seasonal ARIMA(p,d,q)(P,D,Q)	74
3.9. Exponential smoothing, модели тренда, ES и ARIMA	75
3.9.1. Простое экспоненциальное сглаживание	75
3.9.2. Выделение линейного тренда	76
3.9.3. Замедление скорости роста	76
3.9.4. Holt-Winters сезонное сглаживание	77
3.9.5. State Space Model	77
3.9.6. Выбор модели	78

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

Пусть X — ряд длины N . Рассмотрим ортогональный тригонометрический базис:

$$\left\{ 1, \sin\left(2\pi \frac{x}{N} \cdot 1\right), \cos\left(2\pi \frac{x}{N} \cdot 1\right), \dots, \sin\left(2\pi \frac{x}{N} \cdot \left\lfloor \frac{N}{2} \right\rfloor\right), \cos\left(2\pi \frac{x}{N} \cdot \left\lfloor \frac{N}{2} \right\rfloor\right) \right\}, \quad x \in \mathbb{Z}$$

Размер базиса равен N . Для $N = 2k + 1$ получим размер $1 + 2\lfloor \frac{2k+1}{2} \rfloor = 1 + 2k = N$. Для $N = 2k$ получим $1 + 2\lfloor \frac{2k}{2} \rfloor = 1 + 2k = N + 1$, но последний синус будет иметь аргумент $2\pi x$ при $x \in \mathbb{Z}$, поэтому равен нулю, а значит размер базиса равен N .

Нам интересно разложить функцию ряда по этому базису, чтобы определить влияние конкретных частот на значения ряда. Пусть

$$f(x) : \{1, \dots, N\} \rightarrow \mathbb{R}$$

функция, которая задаёт наш ряд. Нам известны значения этой функции, и мы хотим получить представление:

$$f(x) = c_0 + \sum_{i=1}^{\lfloor \frac{N}{2} \rfloor} c_i \cdot \cos\left(2\pi \frac{x}{N} i\right) + s_i \cdot \sin\left(2\pi \frac{x}{N} i\right),$$

где $x \in \{1, \dots, N\}$. Нахождение коэффициентов $\{c_i\}, \{s_i\}$ можно рассмотреть как задачу регрессии. В результате мы получим функцию:

$$\Pi\left(\frac{k}{N}\right) = \frac{N}{2} \begin{cases} c_0^2 & \text{if } k = 0 \\ c_{\frac{N}{2}}^2 & \text{if } N \text{ is even and } k = \frac{N}{2} \\ c_k^2 + s_k^2 & \text{otherwise} \end{cases}$$

Эта функция и есть *периодограмма ряда*.

В `R` периодограмму можно построить с помощью функции `spec.pgram`

1.2. Тренд: аналитический и стохастический, периодограмма тренда

С аналитической точки зрения, тренд — это аппроксимация полиномом небольшого порядка (то есть медленно меняющаяся составляющая ряда). Если рассматривать ряд, как случайный процесс, то тренд можно определить как математическое ожидание этого процесса.

В периодограмме тренду соответствуют низкие частоты. Проблема может быть в том, что значения тренда могут быть намного больше сезонных отклонений, поэтому мощности низких частот подавят на графике частоты сезонностей:

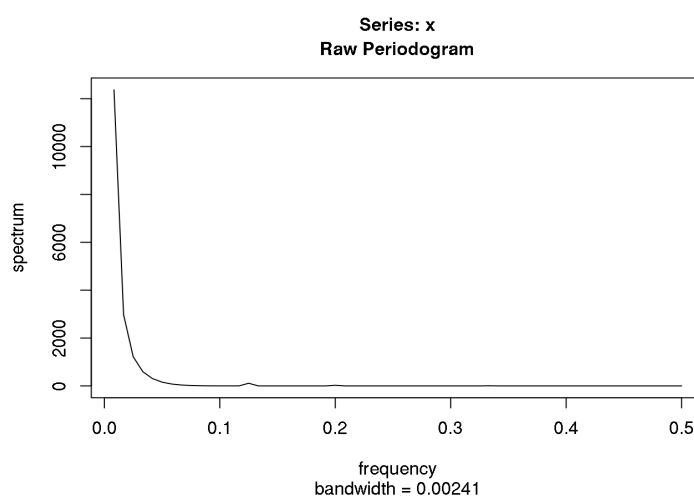


Рис. 1.1. Периодограмма ряда без вычитания тренда (опция `detrend=F`)

Увидеть частоты гармоник на графике довольно сложно.

1.3. Периодическая компонента и ее периодограмма

Периодическая компонента — сумма гармоник ряда, периодически повторяющаяся его компонента. Чем больше амплитуда у гармоники, тем больше значение мощности частоты на периодограмме.

Как мы говорили ранее, мощности низких частот (тренд) обычно подавляют на графике частоты гармоник. Чтобы исправить ситуацию, при первичном анализе ряда в `spec.pgram` можно использовать параметр `detrend`, чтобы избавиться от тренда в первом приближении (вычитается линейный тренд).

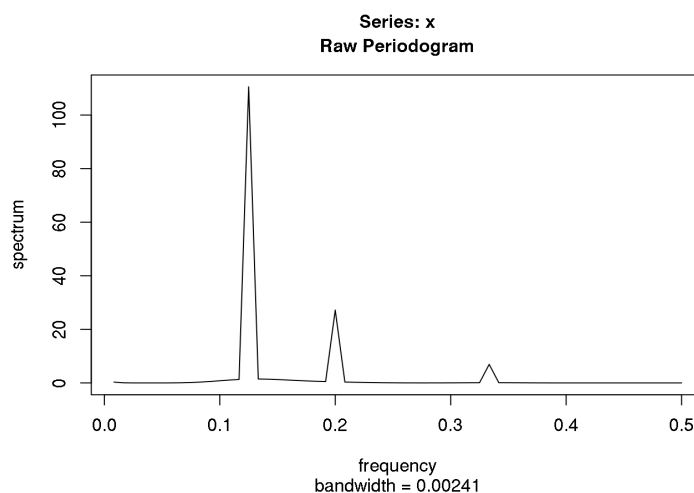


Рис. 1.2. Периодограмма ряда с вычитанием тренда (опция `detrend=T`)

1.4. Шум и его периодограмма

Шум — случайная составляющая временного ряда. Шум в данных может возникнуть из-за погрешности измерений или влияния посторонних факторов.

Периодограмма белого шума (случайный процесс с нулевым математическим ожиданием, постоянной дисперсией и некоррелированный во времени) выглядит следующим образом:

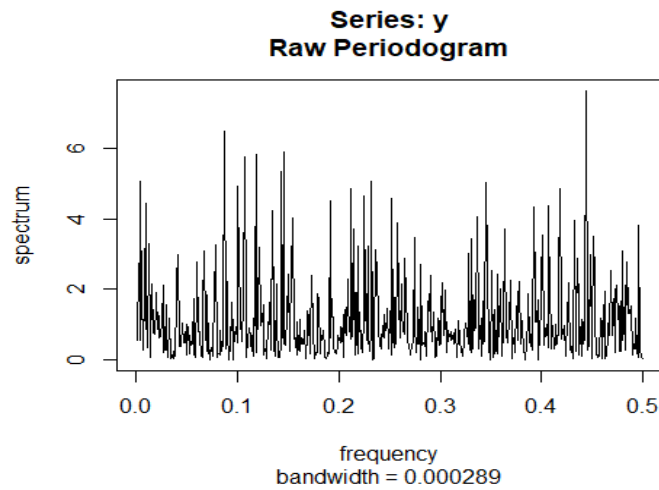


Рис. 1.3. Периодограмма белого шума

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

Если процесс стационарный, периодограмма является оценкой спектральной плотности ряда (впрочем, эта оценка не является состоятельной). Для анализа рядов вроде красного/белого шума имеет смысл рассматривать сглаженную периодограмму. Иначе, если в ряде есть сезонность, то при сглаживании можно случайно убрать пики. В R в функции `spec.pgram` сглаживание осуществляется с помощью скользящего среднего.

1.6. Различия между сглаживанием и выделением тренда

В результате сглаживания действительно получится медленно меняющаяся компонента. Если понимать тренд как некоторую неслучайную функцию, то можем получить тренд. Если под трендом подразумевается среднее, то тренд не получим.

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

Рассмотрим бесконечный временной ряд $\mathbf{X}^\infty = (\dots, x_{-1}, x_0, x_1, \dots)$. Тогда:

- $y_j = (\Phi(\mathbf{x}^N))_j = \sum_{i=-\infty}^{\infty} h_i \mathbf{x}_{j-i}^N$ — линейный фильтр;
- набор коэффициентов $\{h_i\}$ — импульсная характеристика фильтра (impulse response)
- набор коэффициентов $\{h_i : h_i \neq 0\}$, то есть ненулевые коэффициенты импульсной характеристики — finite impulse response (FIR). Тогда $y_j = (\Phi(X))_j = \sum_{i=-r_1}^{r_2} h_i x_{j-i}$
- фильтр $y_j = (\Phi(\mathbf{x}^N))_j = \sum_{i=0}^{r_2} h_i \mathbf{x}_{j-i}^N$ — причинный фильтр (casual), смотрим только в прошлое.

На практике мы никогда не располагаем бесконечным рядом. Поэтому, при рассмотрении ряда конечной длины $\mathbf{X}_N = (x_1, \dots, x_N)$ делается следующее: вкладываем его в бесконечную последовательность $\mathbf{x}^\infty = (\dots, 0, \dots, 0, x_1, \dots, x_N, 0, \dots)$ и рассматриваем её как бесконечный временной ряд.

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

Напоминание: $y_j = (\Phi(\mathbf{x}^N))_j = \sum_{i=-\infty}^{\infty} h_i \mathbf{x}_{j-i}^N$ — линейный фильтр. Тогда:

- $H_{\Phi}(z) = \sum_i h_i z^{-i}$ — передаточная функция (transfer function);
- $|H_{\Phi}(e^{i2\pi\omega})| = A_{\Phi}(\omega)$ — амплитудно-частотная характеристика фильтра (АЧХ);
- $\phi_{\Phi}(\omega) = \text{Arg}(H_{\Phi}(e^{i2\pi\omega}))$ — фазово-частотная характеристика (ФЧХ).

Пусть $\mathbf{x}^N = \cos(2\pi\omega j)$. Тогда:

$$y_j = (\Phi(\mathbf{x}^N))_j = A_{\Phi}(\omega) \cos(2\pi\omega j + \phi_{\Psi}(\omega)).$$

АЧХ показывает как меняется амплитуда в зависимости от частоты.

1.9. АЧХ фильтра скользящего среднего, зависимость от длины окна

Скользящее среднее — частный случай линейного фильтра. Оно имеет один параметр M — ширина окна (порядок скользящего среднего). Для $M = 2k + 1$ определим результат применения фильтра к ряду X_N следующим образом:

$$y_n = \frac{1}{M} \sum_{i=-k}^k x_{n+i}.$$

Таким образом, если ширина окна является нечетным числом, проблем не возникает. Если длина окна является четным числом, то процедура применения фильтра немного усложняется:

$$y_n = \frac{1}{M} \sum_{i=-k}^{k-1} x_{n+i}.$$

Теперь применим фильтр порядка 2 к полученному ряду:

$$z_n = \frac{y_{n-1} + y_n}{2}.$$

Перепишем все через значения исходного ряда:

$$z_n = \frac{x_{n-k} + x_{n+k}}{2M} + \frac{1}{M} \sum_{i=-k}^{k-1} x_{n+i}.$$

Теперь окно стало симметричным.

Характеристики фильтра:

- FIR: $h_i = 1/M, i = 0, \dots, M - 1$;
- $A_\Phi(\omega) = \frac{\sin(M\pi\omega)}{M \sin(\pi\omega)}$

Таким образом, скользящее среднее подавляет частоты периодик, которые имеют период, делящий длину окна нацело. Вообще, чем больше окно, тем сильнее скользящее среднее подавляет высокие частоты.

АЧХ скользящего среднего с длиной окна $M = 12$ для ряда, имеющего сезонность с периодом 12:

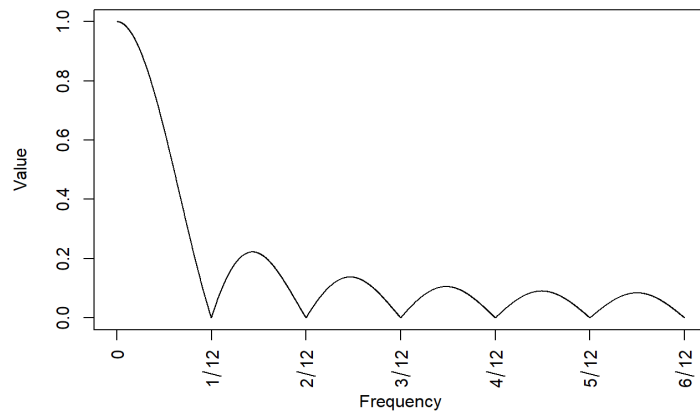


Рис. 1.4. АЧХ, $M = 12, T = 12$

1.10. АЧХ фильтра перехода к разностям (дифференцирования)

Фильтр выглядит следующим образом:

$$y_n = \frac{x_n - x_{n-1}}{2}$$

АЧХ такого фильтра:

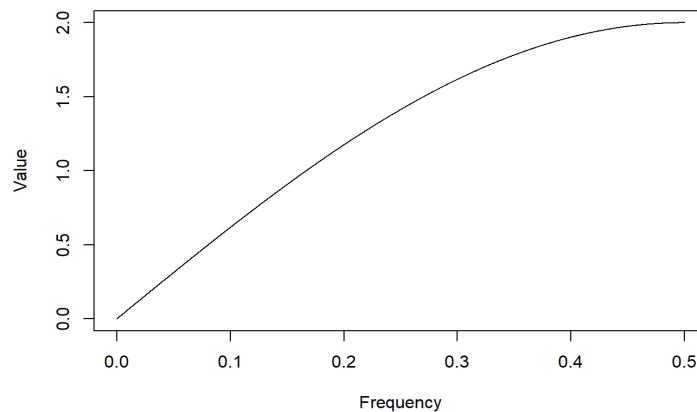


Рис. 1.5. АЧХ фильтра перехода к разностям

Действие этого фильтра обратное скользящему среднему: будут подавлены низкие частоты, полностью должна уйти константа (что можно было предположить из вида фильтра — он напоминает формулу численного дифференцирования).

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

Рассмотрим какой-нибудь причинный фильтр, например скользящее среднее, которое считается только по «прошлому». Сравним его с обычным скользящим средним:

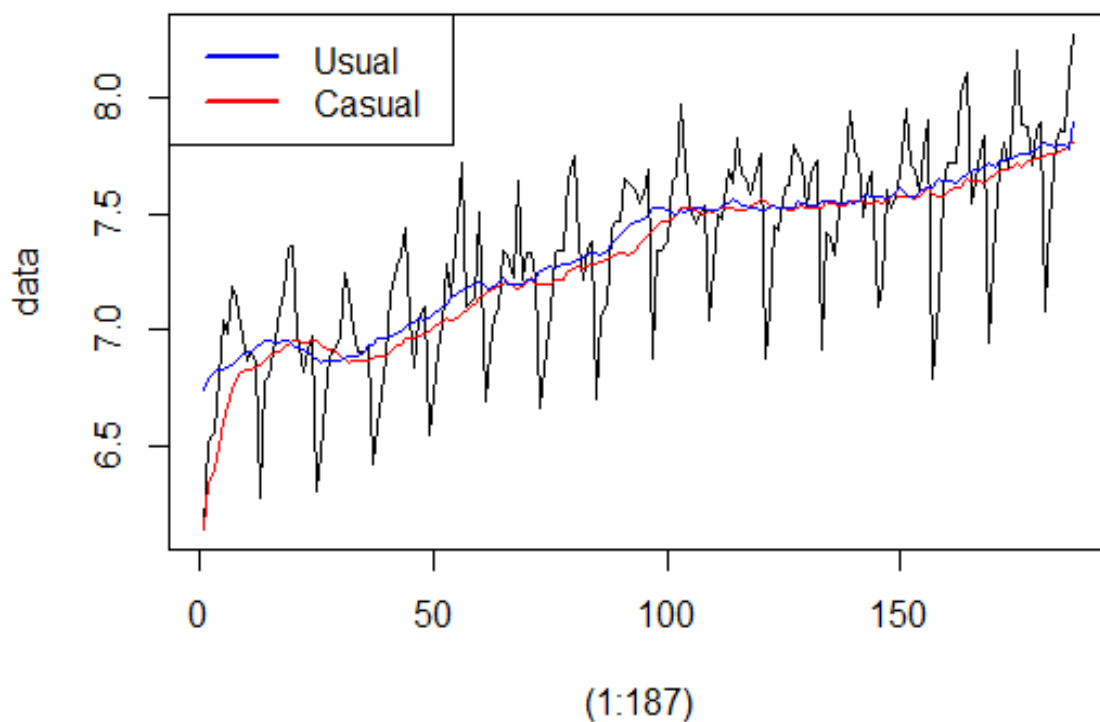


Рис. 1.6. МА, $M = 12$, $T = 12$

Видно, что причинный фильтр повторяет обычное скользящее среднее с запаздыванием (запаздывание равно $M/2$).

1.12. Смещение при сглаживании фильтром скользящего среднего, роль второй производной

Рассмотрим непрерывный аналог фильтрации:

$$y(a) = \int_{-\delta}^{\delta} f(a+x)w(x)dx,$$

где $w(x)$ — весовая функция, причем $\int_{-\delta}^{\delta} w(x)dx = 1$ и $\int_{-\delta}^{\delta} xw(x)dx = 0$. Под интегралом разложим функцию f в ряд Тейлора (предполагаем, что f — гладкая):

$$y(a) \approx f(a) \times 1 + f'(a) \times 0 + \frac{f''(a)}{2} \int_{-\delta}^{\delta} x^2 w(x)dx.$$

Последнее слагаемое в формуле является смещением при применении фильтра. Если скользящее среднее с равными весами, то $w(x) = \frac{1}{2\delta}$ и

$$y(a) \approx f(a) + \frac{f''(a)}{4\delta} \int_{-\delta}^{\delta} x^2 dx = f(a) + \frac{f''(a)\delta^2}{6},$$

то есть смещение зависит от второй производной. Знак смещения зависит, соответственно, от знака второй производной, то есть от характера перегиба функции f . Понятно, что такой фильтр не имеет смысла, если вторая производная большая.

1.13. Фильтр для подавления шума

Рассмотрим последовательность следующего вида: $x_n = s_n + \epsilon_n$, причем $\mathbb{E}\epsilon_n = 0$, $\mathbb{D}\epsilon_n = \sigma^2$, ϵ_i — независимы. Рассмотрим следующий линейный фильтр:

$$y_n = \sum_{i=-r_1}^{r_2} h_i x_i.$$

Предположим, что для сигнала выполняется следующее: $s_n = \sum_{i=-r_1}^{r_2} h_i s_i$. Тогда $\hat{s}_n = \sum_{i=-r_1}^{r_2} h_i x_i$. Посчитаем дисперсию этой оценки:

$$\mathbb{D}\hat{s}_n = \mathbb{D} \sum_{i=-r_1}^{r_2} h_i (s_i + \epsilon_i) = \mathbb{D} \sum_i h_i \epsilon_i = \sum_i h_i^2 \sigma^2 = \sigma^2 \sum_i h_i^2.$$

Если хотим подавлять шум, то логично требовать $\sum_i h_i^2 = \|\mathbf{h}\|^2 \rightarrow \min$. Для скользящего среднего с шириной окна M $\mathbb{D}\hat{s}_n = \frac{\sigma^2}{M}$, т.е. чем больше окно, тем больше подавляем шум.

Здесь, конечно, проблема в начальном предположении о том, что сигнал представим как линейный фильтр. На практике это верно лишь для константного и линейного сигналов.

1.14. Связь между периодограммами ряда до и после применения фильтра

Рассмотрим периодограмму ряда без применения какого-либо фильтра (в т.ч. опция `detrend=FALSE`):

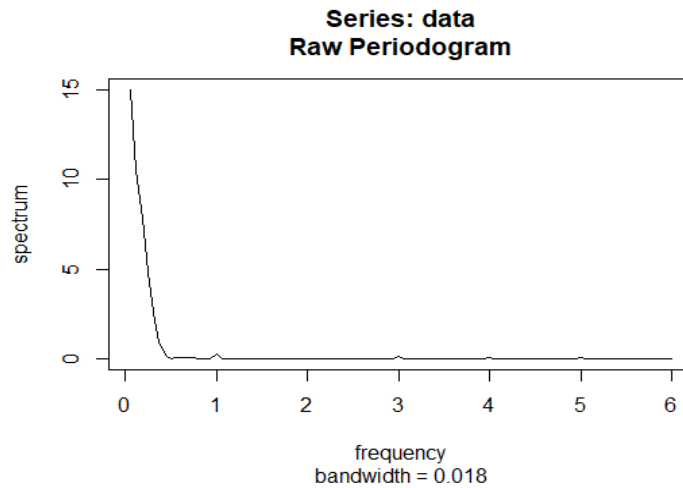


Рис. 1.7. Периодограмма ряда, `detrend = FALSE`

Видим сильное воздействие низких частот, то есть тренда. Влияние гармоник проявляется очень слабо. Применим к ряду фильтр скользящего среднего с $M = 12$, вычтем полученное из исходного ряда и для получившегося остатка построим периодограмму:

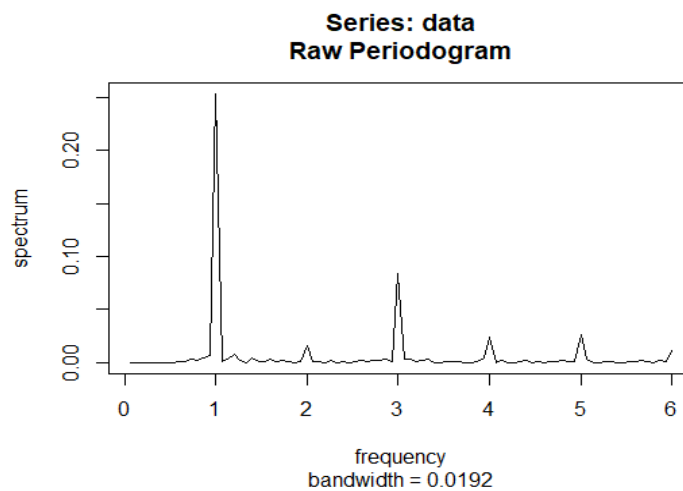


Рис. 1.8. Периодограмма остатка после вычитания результата применения фильтра, `detrend = FALSE`

Воздействие тренда полностью убрано, частоты периодик видны куда явнее.

Построим периодограмму компоненты, непосредственно получившейся в результате сглаживания:

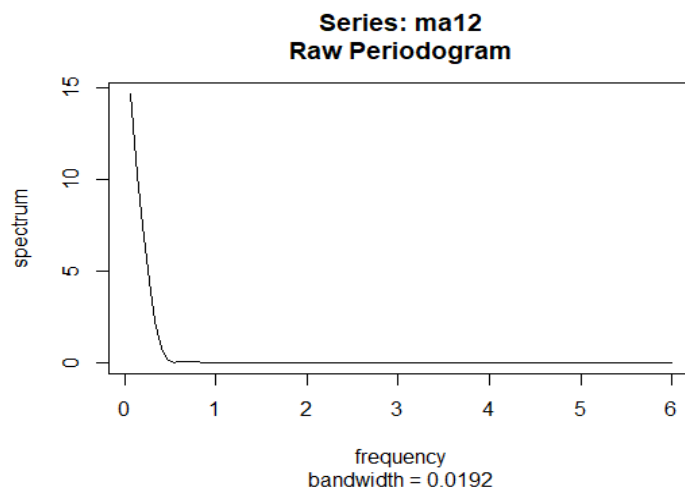


Рис. 1.9. Периодограмма фильтра, `detrend = FALSE`

Здесь наоборот, остались только низкие частоты (тренд), все остальные больше не видны.

1.15. Модели данных — аддитивная и мультипликативная

Аддитивная модель:

$$X_n = T_n + S_n + R_n,$$

где T_n — тренд, S_n — сезонность, R_n — шум. У аддитивной модели постоянная амплитуда сезонной компоненты и постоянная дисперсия шума.

Мультипликативная:

$$X_n = T_n (1 + S_n) (1 + R_n).$$

Здесь размеры сезонных отклонений зависит от величины тренды (амплитуда растет пропорционально). После логарифмирования из мультипликативной модели можем получить аддитивную.

1.16. Методы стабилизации дисперсии в разных моделях

Предположим, что X_n — временной ряд с $\mathbb{E}X_n = \mu$ and $\mathbb{D}X_n = h(\mu)$. Наша цель — получить временной ряд $Y_n = g(x_n)$, где $\mathbb{D}Y_n$ будет (приблизительно) постоянен во времени.

Мы можем аппроксимировать Y_n через приближение Тейлора первого порядка:

$$Y_n = g(x_n) \sim g(\mu) + g'(\mu)(x_n - \mu).$$

Посмотрим на среднее и дисперсию Y_n :

$$\mathbb{E}Y_n = g(\mu)\mathbb{D}Y_n = h(\mu)g'(\mu)^2$$

Из этого можно сделать вывод, что для стабилизации дисперсии необходимо найти преобразование g такое, что $h(\mu)g'(\mu)^2 = \text{const}$ или хотя бы приблизительно константа.

Если $h(\mu)g'(\mu)^2 = c$, и $g(\mu) = \log \mu$, то $h(\mu) = c\mu^2$. Это выполняется, если модель мультипликативна (*log-преобразование*), в рядах с такой моделью $Y = \log X$.

Если $h(\mu)g'(\mu)^2 = c$, и $g(\mu) = \sqrt{\mu}$, то $h(\mu) = 4c\mu$. Это *sqrt-преобразование*, имеет смысл, когда шум пуассоновский, в таком случае: $Y = \sqrt{X}$.

Если данные от 0 до 1 (вероятности, например): $h(\mu) = \sqrt{\mu(1-\mu)}$, $g(\mu) = \arcsin(\mu)$. Тогда $Y = \arcsin X$.

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

Для выделения тренда у ряда с сезонностью с помощью скользящего среднего нужно взять окно кратное периоду.

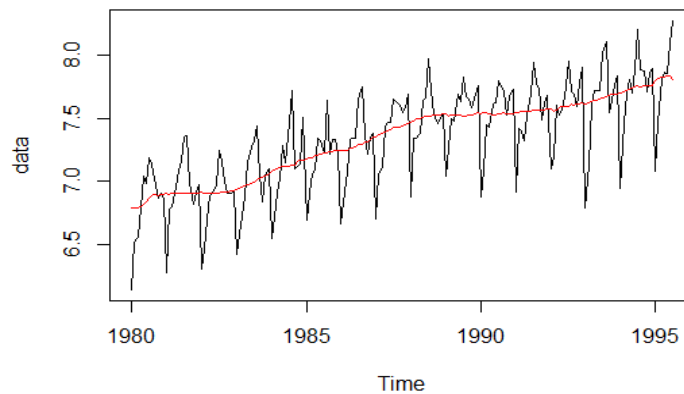


Рис. 1.10. Исходный ряд и выделенный с помощью МА ($M = 24$) тренд

Чтобы судить о том, насколько хорошо мы выделили тренд, стоит посмотреть на периодограмму остатка.

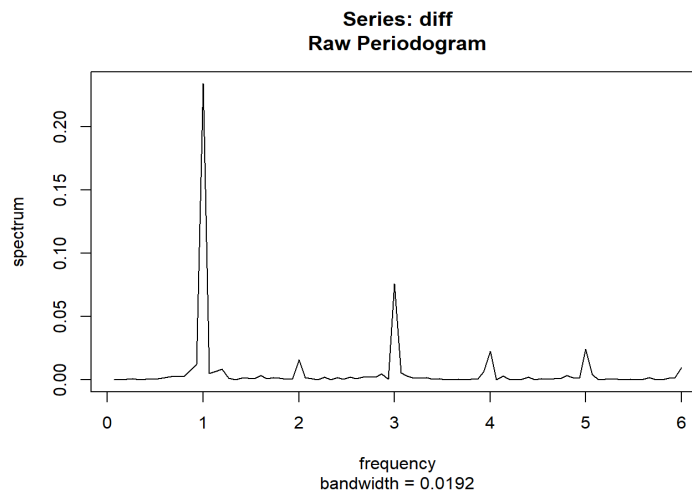


Рис. 1.11. Периодограмма остатка после вычитания тренда, полученного с пом. МА

Видим отсутствие влияния тренда (нет низких частот), можно сравнить с рис. 1.14.

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

Плюсы:

- Если модель нашего ряда имеет вид $x_n = s_n + y_n$, где s_n — тренд, y_n — стационарный процесс, то получим стационарный ряд. Если переход к разностям первого порядка, то уберем линейный тренд, если второго порядка - квадратичный и т.д.;
- Увеличим вклад высоких частот (возможно, легче будет заметить сезонность).

Минусы:

- Увеличим вклад шума.

Рассмотрим белый шум x_n — н.о.р.с.в. с $\mathbb{E}x_i = 0$ и $\mathbb{D}x_i = \sigma^2$. Обозначим $y_n = x_n - x_{n+1}$ — ряд, где перешли к разностям. Тогда:

$$\mathbb{E}(x_n - x_{n+1})(x_{n+1} - x_{n+2}) = -\mathbb{E}x_{n+1}^2,$$

поскольку в силу независимости $\mathbb{E}x_n x_{n+2} = \mathbb{E}x_n x_{n+1} = \mathbb{E}x_{n+1} x_{n+2} = 0$. Кроме того, $\mathbb{E}x_{n+1}^2 = \mathbb{D}x_{n+1} = \sigma^2$. Тогда $\text{cov}(y_n, y_{n+1}) = -\sigma^2$. Дисперсия из тех же соображений:

$$\mathbb{D}y_n = \mathbb{E}(x_n - x_{n+1})^2 = \mathbb{E}x_n^2 + \mathbb{E}x_{n+1}^2 - \mathbb{E}x_n x_{n+1} - \mathbb{E}x_{n+1} x_n = 2\sigma^2.$$

Тогда $\rho(y_n, y_{n+1}) = -1/2$.

1.19. Скользящее среднее и скользящая медиана

Как известно, среднее, как оценка математического ожидания, не устойчива к выбросам. Аналогичное утверждение справедливо и в отношении скользящего среднего.

Если в данных есть выбросы, то можно использовать медиану вместо среднего — эта оценка более устойчива к сильно отклоняющимся значениям. Однако она не является оценкой математического ожидания в случае, если распределение случайного шума несимметричное.

Можно применять эти два метода в паре: вначале, например, убрать выбросы с помощью скользящей медианы, затем сгладить ряд с помощью скользящего среднего.

1.20. Растекание частоты в периодограмме и методы его устранения

Если длина ряда не будет кратна частоте периодики, то мы не сможем её выделить на графике, и получим растекание частоты (frequency leakage).

Для примера рассмотрим ряд длиной 50 точек: $y_i = \cos(\frac{2\pi x_i}{11})$. Построим его периодограмму:

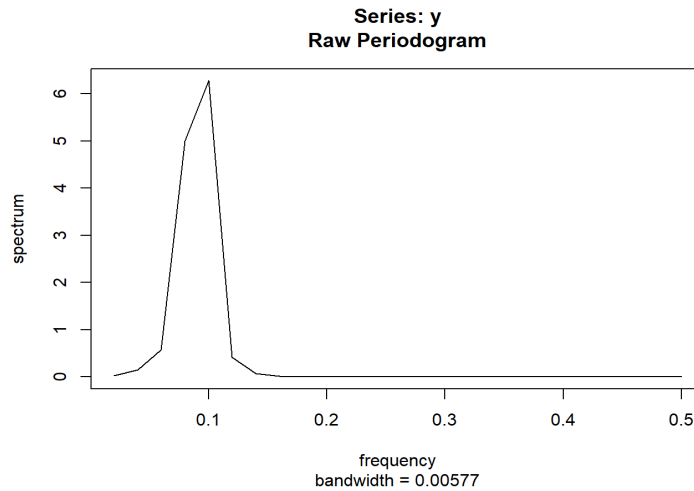


Рис. 1.12. Периодограмма исходного ряда

Первый вариант: отрезать начальных точек так, чтобы длина ряда была кратна 11:

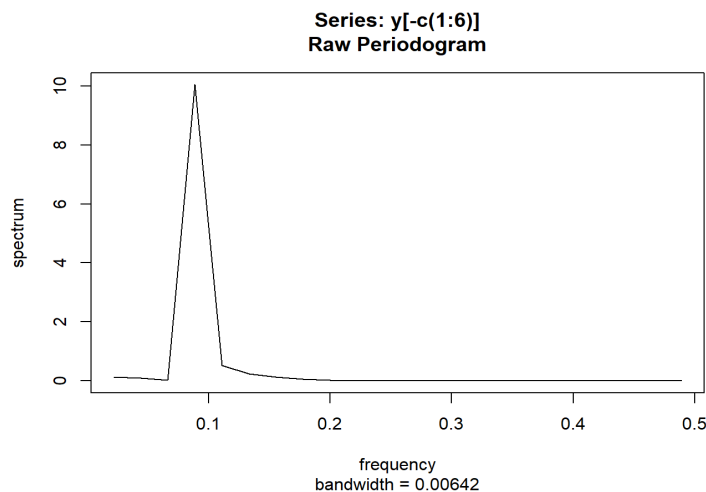


Рис. 1.13. Коррекция периодограммы

Второй вариант: добавить нулей так, чтобы длина ряда была кратна 11:

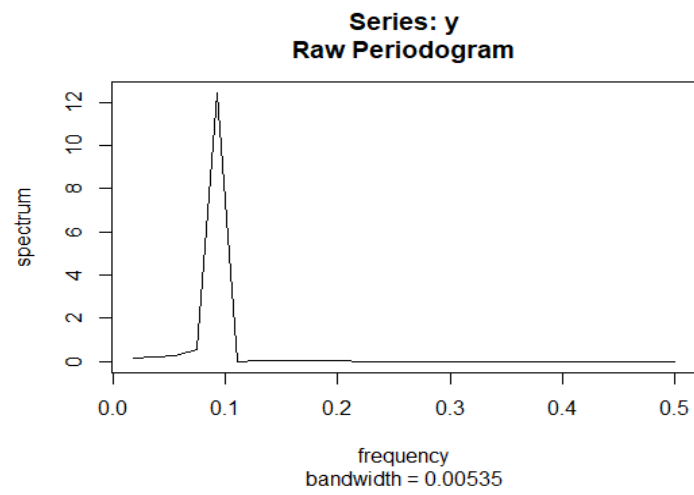


Рис. 1.14. Коррекция периодограммы

1.21. Выделение тренда с помощью параметрической регрессии

Выделить тренд можно с помощью регрессии. В простейшем случае предиктором будет номер наблюдения. Модель можно усложнить, добавив степени номеров наблюдений, тогда получится полиномиальная регрессия.

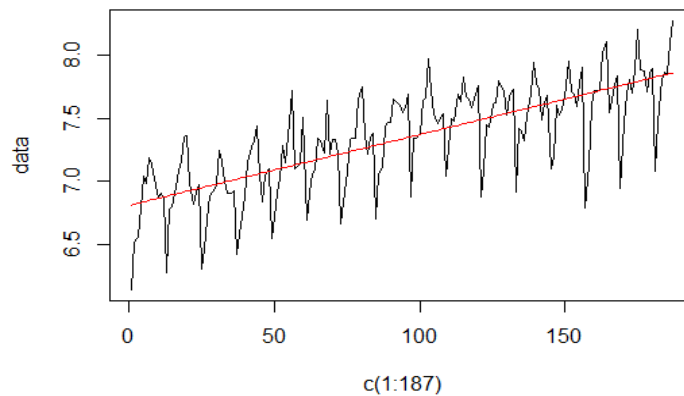


Рис. 1.15. Выделение тренда с помощью линейной регрессии

Видно, что тренд здесь все же не является линейным (хоть и близко). Чтобы убедиться в том, что можно лучше, взглянем на периодограмму остатка:

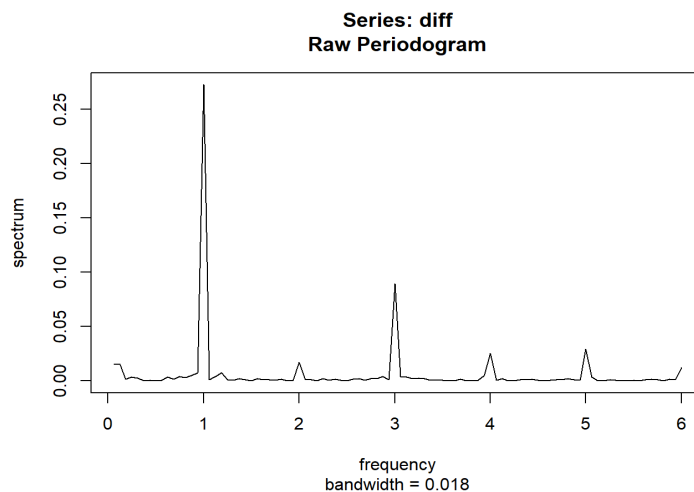


Рис. 1.16. Периодограмма остатка после вычитания тренда, полученного с пом. линейной регрессии

Видим, что низкие частоты остались. Более того, с точки зрения периодограммы, это почти как задать опцию `detrend = TRUE`, т.е. просто вычитать среднее. Попробуем применить полиномиальную регрессию:

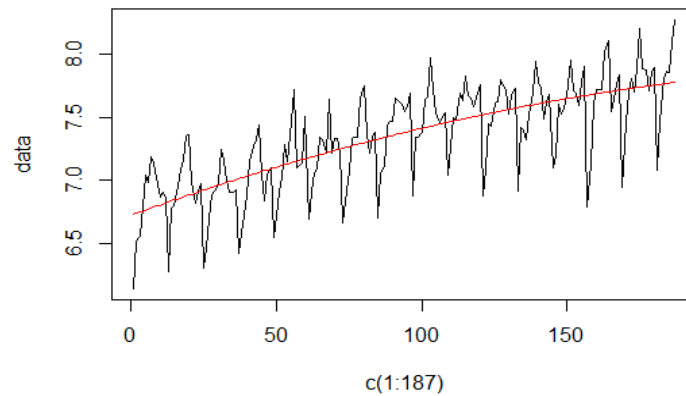


Рис. 1.17. Выделение тренда с помощью полиномиальной регрессии ($\text{deg} = 2$)

Визуально тренд выглядит значительно лучше. Периодограмма остатка:

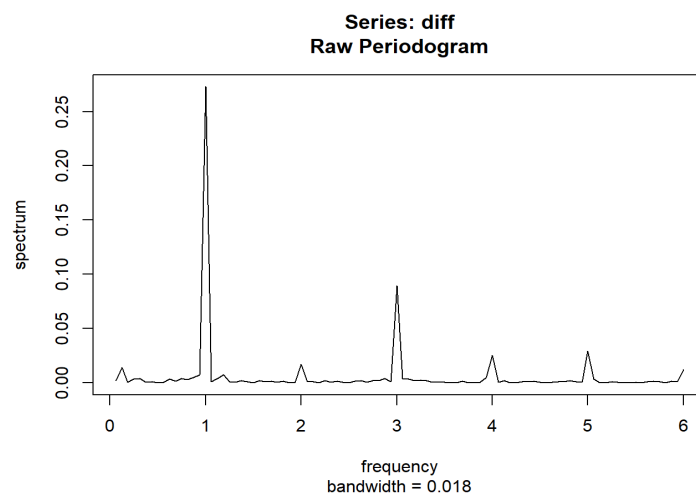


Рис. 1.18. Периодограмма остатка после вычитания тренда, полученного с пом. полиномиальной регрессии

Лучше, но все еще не идеально. Как мы видели ранее, даже скользящее среднее, например, дает более качественный результат.

1.22. Выделение тренда с помощью метода LOESS

Для выделения тренда можно рассматривать не только параметрическую регрессию, но и непараметрическую (LOESS, LOWESS)

LOESS метод основан на сглаживании с помощью построения локальных (взвешенных) регрессий. Алгоритм получения значения в точке n выглядит следующим образом:

1. выделяем окно длины h (параметр алгоритма);
2. по точкам в радиусе $n \pm h/2$ строим регрессию (линейную/полиномиальную, это LOESS, или линейную/полиномиальную взвешенную, это LOWESS). Значение в точке n предсказываем с помощью этой модели.

Для LOWESS стандартный пример весовой функции — трикубическая:

$$W(z) = \begin{cases} (1 - |z|^3)^3 & \text{for } |z| < 1 \\ 0 & \text{for } |z| \geq 1 \end{cases}$$

Выделение тренда с помощью LOESS (ширина окна — половина длины ряда, степень локальных регрессий — 2):

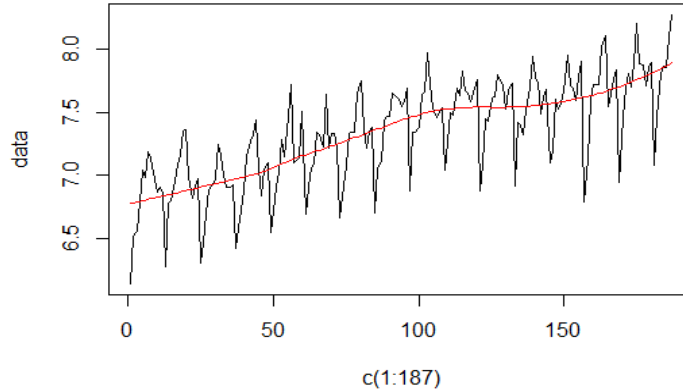


Рис. 1.19. Выделение тренда с помощью локальной регрессии

Как обычно, анализируем периодограмму остатка:

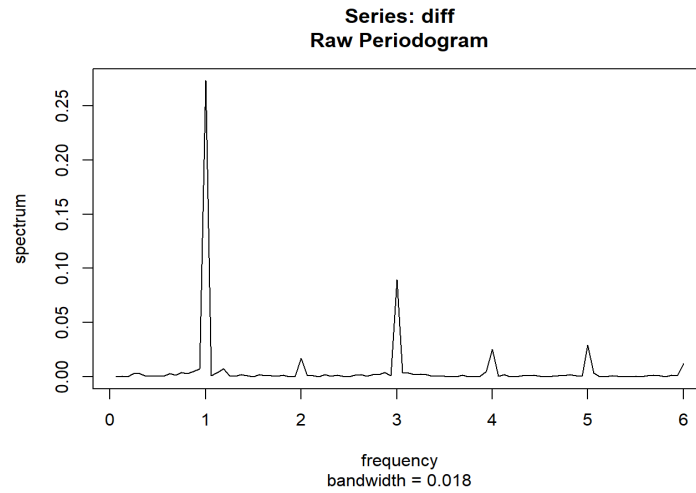


Рис. 1.20. Периодограмма остатка ряда

Видим, что как и визуально на графике, так и с точки зрения периодограммы, тренд выделился хорошо.

Рассмотрим также выделение тренда с помощью фильтра Холдрика–Прескотта (Hodrick–Prescott filter). Решается следующая оптимизационная задача:

$$\sum_{i=1}^N (X_i - s_i)^2 - \lambda \sum_{i=2}^{N-2} ((s_{i+1} - s_i) - (s_i - s_{i-1}))^2 \rightarrow \min_{\{s_i\}}$$

Здесь под s_i подразумевается тренд. Эмпирическая рекомендация: если период составляет $p = 12$, то $\lambda = 129606$.

Результат применения:

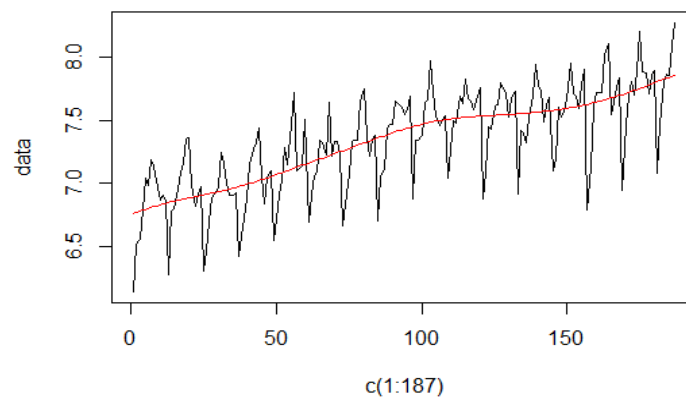


Рис. 1.21. Выделение тренда с помощью HP-filter

Смотрим на периодограмму остатка:

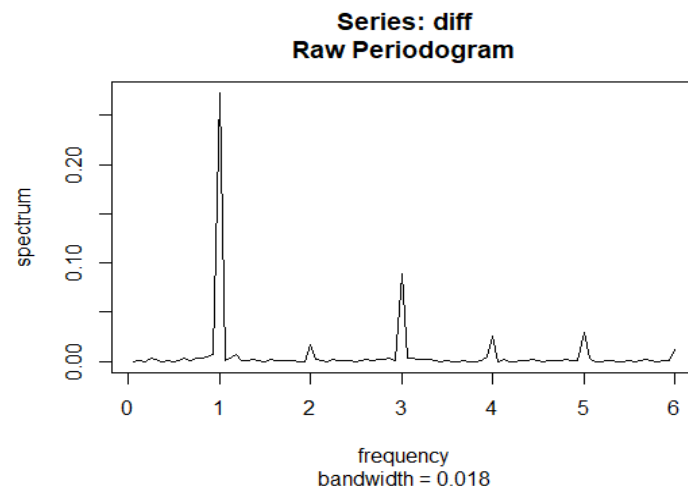


Рис. 1.22. Выделение тренда с помощью HP-filter

1.23. Нахождение огибающей периодического ряда с помощью выделения тренда

Рассмотрим ряд в форме $x_n = A(n) \cos(2\pi\omega n + \phi)$, где $A(n)$ ведет себя как низкочастотный косинус (для примера и рассмотрим такую амплитуду).

Возведя в квадрат, получим $x_n^2 = A^2(n) + A^2(n) \cos(2(2\pi\omega n + \phi))$ с точностью до умножения на 2. То есть, амплитуда в квадрате стала трендом у нового ряда.

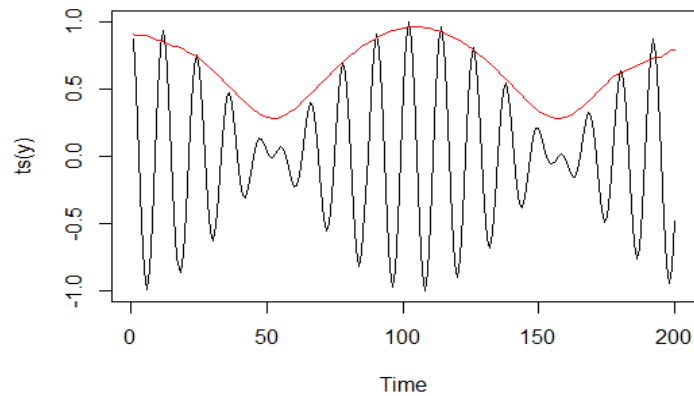


Рис. 1.23. Огибающая периодического ряда

1.24. Оценивание поведения дисперсии шума с помощью выделения тренда

предположим, что ряд задаётся формулой $x_n = f(n)\epsilon_n$, где ϵ_n — белый шум с $\mathbb{D}\epsilon_n = \sigma^2$ и $\mathbb{E}\epsilon_n = 0$. Возведем в квадрат каждый член ряда: $x_n^2 = f^2(n)\epsilon_n^2$. Заметим, что $E\epsilon_n^2 = D\epsilon_n = \sigma^2$, причем $\epsilon_n^2 \geq 0$.

Мы можем переписать формулу для члена ряда в виде:

$$x_n^2 = f^2(n) (\sigma^2 + \bar{\epsilon}_n^2) = \sigma^2 f^2(n) + f^2(n) \bar{\epsilon}_n^2.$$

При этом $E\bar{\epsilon}_n = 0$. Теперь, если мы выделим тренд из исходного ряда, мы получим оценку $\sigma^2 f^2(n)$. С помощью простых преобразований мы найдем оценку огибающей для шума.

Рассмотрим реальные данные и остаток после вычитания из ряда тренда и сезонности:

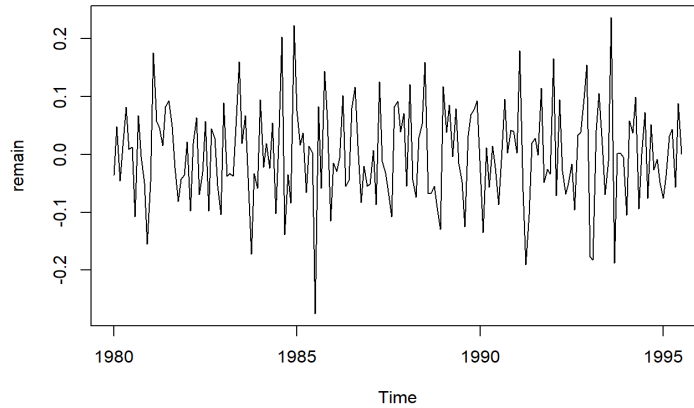


Рис. 1.24. Шум после выделения сигнала

Получили шум с какой-то дисперсией. Чтобы её оценить, возведем шум в квадрат и выделим тренд (любым подходящим способом):

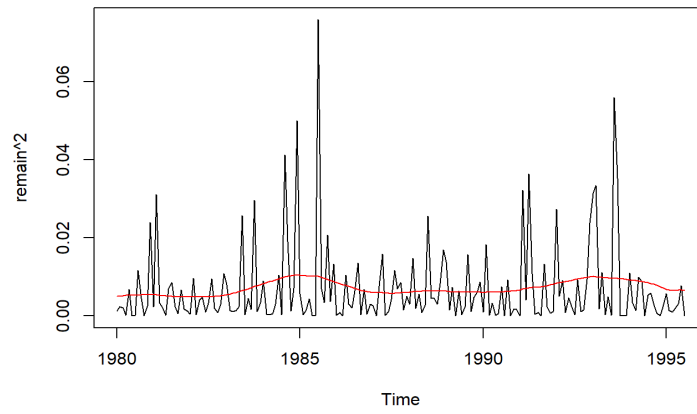


Рис. 1.25. Квадрат шума с выделенным трендом

1.25. Метод разложения Classical seasonal decomposition

Рассматриваем аддитивную модель ряда:

$$X_t = T_t + S_t + N_t.$$

Зафиксируем период p у сезонной компоненты S_t . Шестишаговый алгоритм декомпозиции выглядит следующим образом:

1. применяем скользящее среднее с длиной окна p , получаем оценку тренда \widetilde{T}_t ;
2. рассматриваем $X_t - \widetilde{T}_t$, получаем оценку $\widetilde{S_t + N_t}$;
3. $\widetilde{S_t + N_t}$ усредняется по периоду p : вычисляем средние для каждой p -ой точки. Получим среднее значение за каждый период. Выворачиваем эти значения так, чтобы их среднее равнялось нулю. Это и есть наша аппроксимация сезонности \widetilde{S}_t ;
4. рассматриваем $X_t - \widetilde{S}_t$ и получаем оценку $\widetilde{T_t + N_t}$;
5. применяем фильтр с небольшой длиной окна к $\widetilde{T_t + N_t}$ и получаем $\widetilde{\widetilde{T}}_t$ (еще одну оценку тренда);
6. $\widetilde{N}_t = \widetilde{T_t + N_t} - \widetilde{\widetilde{T}}_t$.

В результате получаем \widetilde{N}_t , $\widetilde{\widetilde{T}}_t$ и \widetilde{S}_t . В R эта процедура реализована в функции `decompose`. Результат её работы выглядит следующим образом:

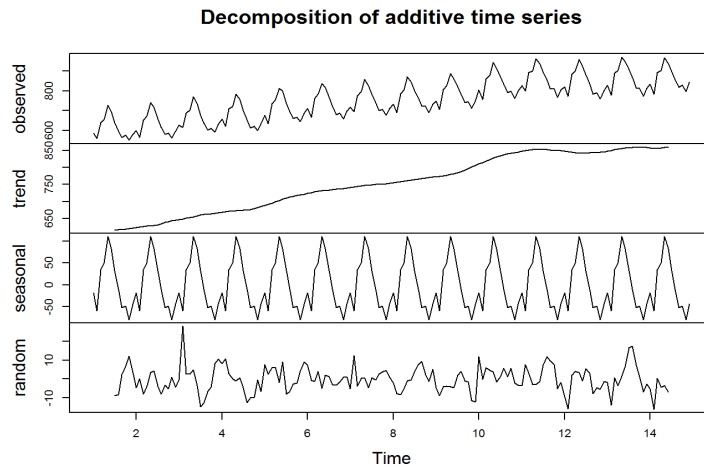


Рис. 1.26. Декомпозиция ряда на аддитивные составляющие

1.26. Метод разложения STL

Усовершенствованный classical seasonal decomposition. Расшифровывается как "Seasonal and Trend decomposition using Loess". В R это функция `stats::stl`.

Рассматриваем аддитивную модель ряда:

$$X_t = T_t + C_t + S_t + N_t.$$

Здесь C_t — циклическая компонента, ранее мы неявно включали её в периодическую.

Метод состоит из двух циклов: внешнего и внутреннего. Внешний цикл нужен для решения проблем с выбросами и пропущенными значениями, он такой же, как в LOESS. На каждой итерации внешнего цикла выполняется внутренний цикл, который и заключается в разложении ряда на тренд, сезонность и остаток. Опишем итерацию внутреннего цикла.

На вход на $k + 1$ внутренний цикл поступает ряд X_N длины N , T_N^k — выделенный на предыдущем шаге тренд. Полагаем $T_N^0 \equiv 0$.

1. Удаляем тренд: $Y_N = X_N - T_N^k$;
2. Сглаживаем ряд Y_N с помощью $LOESS(span = n_s, d = 1)$, мы также предсказываем n_p значений до начала ряда и после. В результате получаем ряд $C_{N+2n_p}^{k+1}$ — сглаженную циклическую составляющую;
3. Применяем низкочастотный фильтр к $C_{N+2n_p}^{k+1}$: дважды скользящее среднее с окном n_p , и ещё раз — с окном 3. К полученному после усреднения ряду применяем $LOESS(span = n_l, d = 1)$. В результате получаем ряд L_N^{k+1} длины N ;
4. Выделяем сезонную составляющую: $S_N^{k+1} = C_N^{k+1} - L_N^{k+1}$. Здесь мы игнорируем первые и последние n_p значений ряда $C_{N+2n_p}^{k+1}$;
5. Удаляем сезонность из исходного ряда: $X_N - S_N^{k+1}$;
6. Сглаживаем получившийся после удаления сезонности ряд с помощью $LOESS(span = n_t, d = 1)$.

Параметры $LOESS$ — это значение $span$ и d , $span$ — количество ближайших наблюдений, которые берутся для предсказания значения в некоторой точке x . На этом множестве

строится полином степени d , который и используется для предсказания x . (На самом деле, наблюдениям ещё задаются веса в зависимости от расстояния от x . С помощью этих весов мы можем, например, уменьшить влияние выбросов).

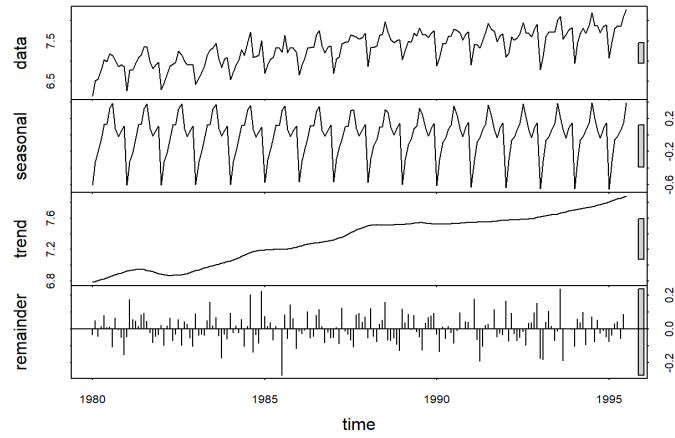


Рис. 1.27. Декомпозиция ряда на аддитивные составляющие

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 \mathbf{X} длины $N > 2$, который является суммой нескольких временных рядов ($\mathbf{X} = \sum \mathbf{X}_k$). 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 \mathbf{X} применяется *оператор SSA-вложения*

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

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

Шаг 2. Сингулярное разложение

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

Тогда результат сингулярного разложения — набор *собственных троек*

$$\text{SVD}(\mathbf{X}) = \{(\sqrt{\lambda_i}, U_i, V_i), i \in \{1, \dots, L\}\}, \quad \mathbf{X} = \sum_i \mathbf{X}_i, \quad \mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^\top.$$

2.1.2. Второй этап: восстановление

Шаг 3. Группировка

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

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

а траекторная матрица \mathbf{X} представляется в виде

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

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» $i + j = k + 2$, $0 \leq k < N$

$$\mathcal{H}(\mathbf{X}_{I_j}) = \widehat{\mathbf{X}}_j, \quad j \in \{1, \dots, m\}.$$

2. К матрицам $\widehat{\mathbf{X}}_j$ применяется оператор $(\mathcal{T}_{\text{SSA}}^{(L)})^{-1}$, проводящий процедуру, обратную процедуре вложения

$$(\mathcal{T}_{\text{SSA}}^{(L)})^{-1}(\widehat{\mathbf{X}}_j) = \widetilde{\mathbf{X}}_j = (\tilde{x}_0^{(j)}, \dots, \tilde{x}_{N-1}^{(j)}), \quad j \in \{1, \dots, m\}.$$

Таким образом получается разбиение исходного ряда в сумму m рядов длины N

$$\mathbf{X} = \sum_{k=1}^m \widetilde{\mathbf{X}}_k^{(k)}.$$

Для выбора параметра L существуют следующие рекомендации:

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

2.2. Последовательный SSA

Пусть $\text{SSA}(\mathbf{X}, L, I = \{I_1, \dots, I_r\})$ — разложение временного ряда \mathbf{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:

$$\text{SSA}(\mathbf{X}, L_1, \{I_1, I_2\}) = \{\mathbf{X}_1, \mathbf{X}_2\}, \quad \text{SSA}(\mathbf{X}_2, L_2, I = \{I_{21}, \dots, I_{2r}\}) = \{\mathbf{X}_{21}, \dots, \mathbf{X}_{2r}\}.$$

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

2.3.1. Слабая разделимость

При фиксированном L два ряда X_1 и X_2 являются слабо разделимыми, если их траекторные столбцовые (и строковые) пространства ортогональны.

2.3.2. Сильная разделимость

При фиксированном L два ряда X_1 и X_2 являются сильно разделимыми, если они слабо разделимы, и наборы их сингулярных чисел не пересекаются.

Поскольку при таких определениях данные условия далеко не всегда выполняются в реальных приложениях, имеются асимптотические аналоги разделимости.

2.4. Определение сильной и слабой разделимости при смешивании компонент

Определить, что компоненты смешались, можно с помощью матрицы взвешенных корреляций.

Отсутствие сильной разделимости (при наличии слабой) может быть вызвано следующими факторами:

- Собственные числа компонент слишком близки (для определения можно воспользоваться графиком собственных чисел).

Отсутствие слабой разделимости может быть вызвано следующими факторами:

- Ряд слишком короткий (не хватает длины ряда для сходимости)
- Ряд слишком зашумлён (компоненты шума смешиваются с компонентами сигнала)

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 , $\text{fdim}(\mathbf{X}_N) \leq d$, если $a_d \neq 0$. Если (a_1, \dots, a_d) – минимальный набор коэффициентов, то $\text{fdim}(\mathbf{X}_N) = d$. Для любого ряда, $\text{fdim}(\mathbf{X}_N) \geq 1$.

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 $d = \text{fdim}(\mathbf{X}_N)$, 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 характеристического 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 $\frac{1}{\lambda}$, 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 $\mathbf{X} = \mathbf{S} + \mathbf{R}$, где \mathbf{S} описывается LRR порядка r , а \mathbf{R} — шум. Тогда можно использовать $\text{span}\{U_1, \dots, U_r\}$, получающиеся в процессе применения SSA к \mathbf{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 $e_L = (\mathbf{0}_{L-1}, 1)^T$ 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 ($\text{dist}(\mathbf{X}_i, \mathbf{X}_j) = 0.5(1 - \rho_{ij}^{(w)})$). 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 (`igarfill` в `Rssa`) пропущенные значения инициализируются, затем производится восстановление сигнала. Данная процедура продолжается, пока значения на месте пропусков не сойдутся к определённому значению (изменения при последующих 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 называется Теплицев SSA (Toeplitz SSA). Идея данной модификации заключается в том, чтобы вместо матрицы $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ подставлять теплицеву матрицу $\tilde{\mathbf{S}}$, образованную автоковариациями $\tilde{c}_{ij} = \frac{1}{N - |i - j|} \sum_0^{N - |i - j| - 1} x_m x_{m + |i - j|}$ (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 \mathbf{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 \mathbf{X}
- Вычислим $\mathbf{X}' = \mathbf{X} - \mathbf{C}$
- Проведём SVD-разложение матрицы $\mathbf{X}' = \sum_i \sqrt{\lambda'_i} U'_i V_i'^T$

В результате получим разложение матрицы \mathbf{X} : $\mathbf{X} = \mathbf{C} + \mathbf{X}' = \mathbf{C} + \sum_i \sqrt{\lambda'_i} U'_i V_i'^T$

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 $X_N^{(1)}$ и $X_N^{(2)}$ 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} : $\langle X, Y \rangle_{\mathbf{A}} = X^T \mathbf{A} Y$. 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 строк траекторных матриц относительно соответствующих 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 чисел 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, соответствующие 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 дифференцирование, чтобы получить нужные компоненты. Действительно, $(A \cos(2\pi\omega n))' = A(2\pi\omega) \sin(2\pi\omega n)$, $(Ae^{an})' = Aae^{an}$ — дифференцирование 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 Σ).
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:

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

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

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

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 независимо с x_k при $k < n$.

Выражение для авторегрессионного процесса 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 $s_n = \sum_{i=1}^p \phi_i s_{n-i}$ — сигнал и ε_n — шум. Модель авторегрессии имеет вид $x_n = \sum_{i=1}^p \phi_i x_{n-i} + \varepsilon_n$. 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 asf для AR(p)

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

$$\Phi(L)x_n = \varepsilon_n \therefore \mathbb{E}\Phi(L)x_n x_{n-k} = \mathbb{E}\varepsilon_n x_{n-k} = 0,$$

поэтому $\Phi(L)\rho_k = 0$ (уравнения Юла-Уолкера). Из этого следует, что ρ_k можно представить в виде $\sum_{i=1}^p A_i \lambda_i^k$, где λ_i — корни характеристического полинома $\Phi(L)$. Данная функция экспоненциально убывает. В случае наличия комплексных корней появляется цикличность asf.

3.4. Вид расf для $\text{AR}(p)$

Исходя из вида асf для $\text{AR}(p)$ можно получить, что расf перестанет значимо отличаться от 0 после сдвига p .

3.5. Модель МА(q), вид asf и расf

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 asf и расf для МА(q) по сравнению с AR(q) «меняются местами».

3.6. ARMA(p,q)

Моделью $\text{ARMA}(p, q)$, 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

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

Заметим, что для того, чтобы процесс $\text{ARMA}(p, q)$ был реализуем относительно шума ε_n (то есть чтобы выполнялись условия стационарности) необходимо и достаточно, чтобы все корни характеристического полинома лежали вне единичной окружности.

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

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

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 $(1 - L)$) используется для того, чтобы нестационарный ряд привести к стационарному виду. Модель $ARIMA(p, d, q)$ означает, что разности временного ряда порядка d подчиняются модели $ARMA(p, q)$.

Запись модели через оператор сдвига

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

Исходя из условий реализуемости процесса $ARMA(p, q)$ можно заключить, что если у характеристического полинома есть корни не вне единичной окружности, условия стационарности не выполнены.

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:

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

Мы можем ввести сезонное дифференцирование: $(1 - L^m)x_n$. Если ряд один раз сезонно-дифференцируем, 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 применена 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 сглаживание

Ранее 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:

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

- 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:

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

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 $t \rightarrow \infty$. Мы хотели бы предсказывать не столь большие значения для далеких предсказаний. Для этого мы немного изменим формулу:

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

Полагая $0 < \phi < 1$, получим что сумма $\sum_{i=1}^t \phi^i$ оказывается конечной для $t \rightarrow \infty$ 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 – $m = 4$.

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

Добавилось ещё $m+1$ параметров: s_1, \dots, 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 $t \rightarrow \infty$, заменив $t \cdot b_{N+1}$ на $\sum_{i=1}^t \phi^i b_{N+1}$.

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:

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

Мы можем переписать второе уравнение в следующем виде:

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

Заметим, что второе слагаемое в правой 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 элементами 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:

$$\begin{aligned} \text{AIC} &= -2\log(L) + 2k, \\ \text{AIC}_c &= \text{AIC} + \frac{k(k+1)}{N-k-1}, \\ \text{BIC} &= \text{AIC} + k[\log(N) - 2], \end{aligned}$$

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).