

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	34

1.24. Оценивание поведения дисперсии шума с помощью выделения тренда	35
1.25. Метод разложения Classical seasonal decomposition	37
1.26. Метод разложения STL	38
Глава 2. Метод SSA	40
2.1. Последовательный SSA	42
2.2. Слабая и сильная разделимость	43
2.3. Определение сильной и слабой разделимости при смешивании компонент	44
2.4. Идентификация тренда	45
2.5. Идентификация периодичности (сезонности)	46
2.6. Использование матрицы взвешенных корреляций	47
2.7. Элементарные восстановленные компоненты	48
2.8. Корни характеристического полинома: сигнальные и лишние	49
2.9. Оценка параметров в SSA	50
2.10. Прогноз	51
2.11. Доверительные интервалы	52
2.12. Автоматическая идентификация	53
2.13. Заполнение пропусков	54
2.14. Теплицев SSA для стационарных рядов	55
2.15. Projection SSA, выделение линейного тренда	56
2.16. Iterative O-SSA и DerivSSA	57
2.17. Аппроксимация Cadzow SSA	58
2.18. MSSA для анализа многомерных временных рядов	59
2.19. 2D-SSA для разложения изображения	60
Глава 3. ARIMA	61
3.1. AR(p) – модель, запись в виде с оператором сдвига	62
3.2. AR(p) и модель сигнала в SSA	63
3.3. Вид автоковариационной функции ascf для AR(p)	64
3.4. Вид расcf для AR(p)	65
3.5. Модель MA(q), вид ascf и расcf	66
3.6. ARMA(p,q)	67
3.7. Дифференцирование, ARIMA	68

3.8.	Seasonal ARIMA(p,d,q)(P,D,Q)	69
3.9.	Exponential smoothing, модели тренда, ES и ARIMA	70

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:

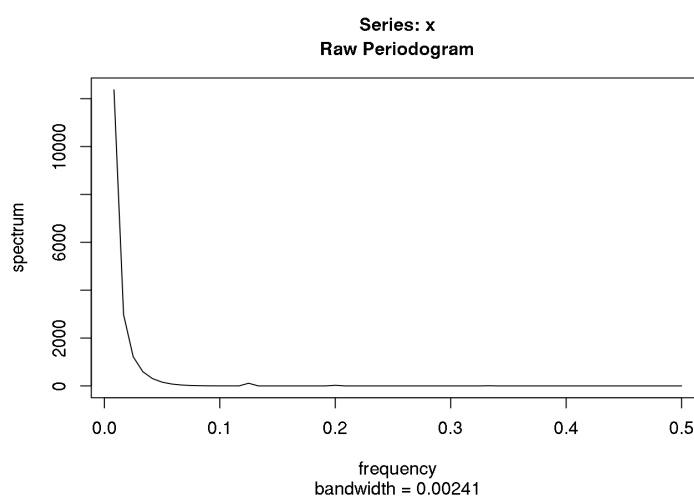


Рис. 1.1. Периодограмма ряда без вычитания тренда (опция `detrend=F`)

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.

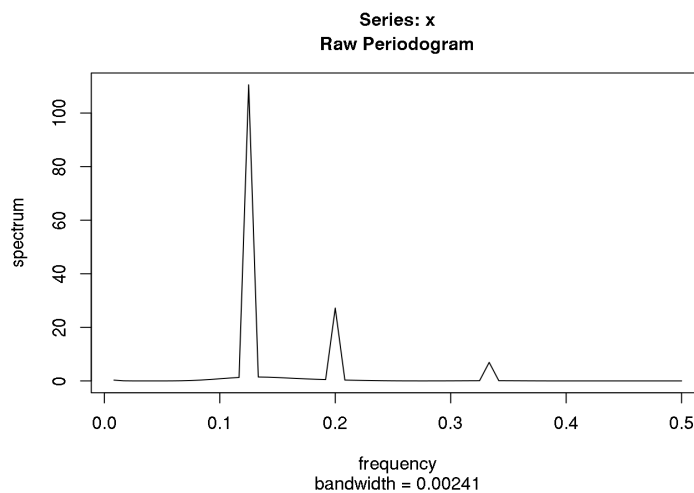


Рис. 1.2. Периодограмма ряда с вычитанием тренда (опция `detrend=T`)

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:

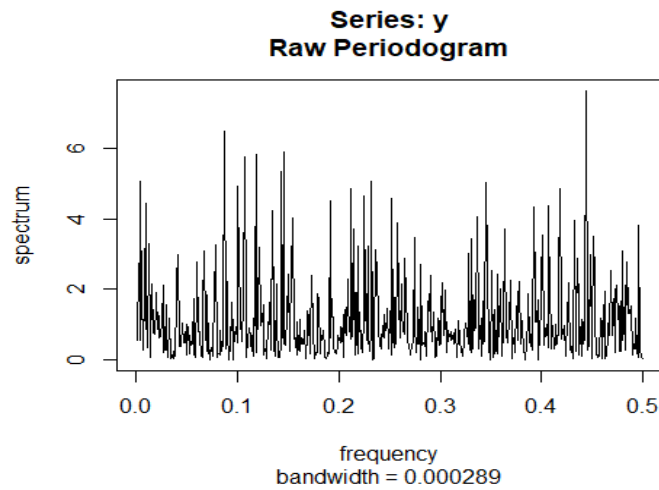


Рис. 1.3. Периодограмма белого шума

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). Поскольку в реальной жизни данные зашумлены, то имеет смысл рассматривать сглаженную периодограмму. Это позволяет выделить частоты с наибольшим вкладом в периодику ряда, при этом уменьшив значение случайных пиков. 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

Пусть $\mathbf{X}_N = (x_1, \dots, x_N)$ — ряд длины N . Вложим его в бесконечную последовательность $\mathbf{x}^N = (\dots, 0, \dots, 0, x_1, \dots, x_N, 0, \dots, 0)$. 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.

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 2π -периодична, описывает действие фильтра на частоту ω . Это информация нам важна при выборе коэффициентов фильтра, когда мы хотим, например, избавиться от сезонности.

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:

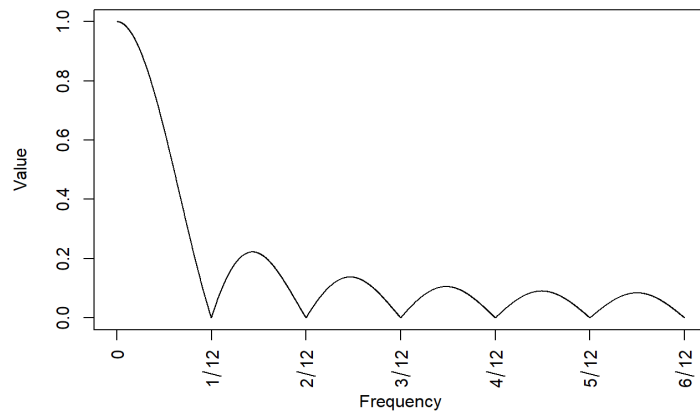


Рис. 1.4. АЧХ, $M = 12, T = 12$

1.10. АЧХ фильтра перехода к разностям (дифференцирования)

Фильтр выглядит следующим образом:

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

АЧХ такого фильтра:

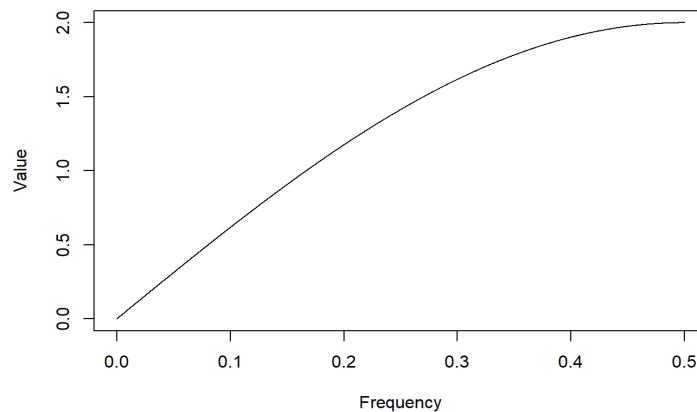


Рис. 1.5. АЧХ фильтра перехода к разностям

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:

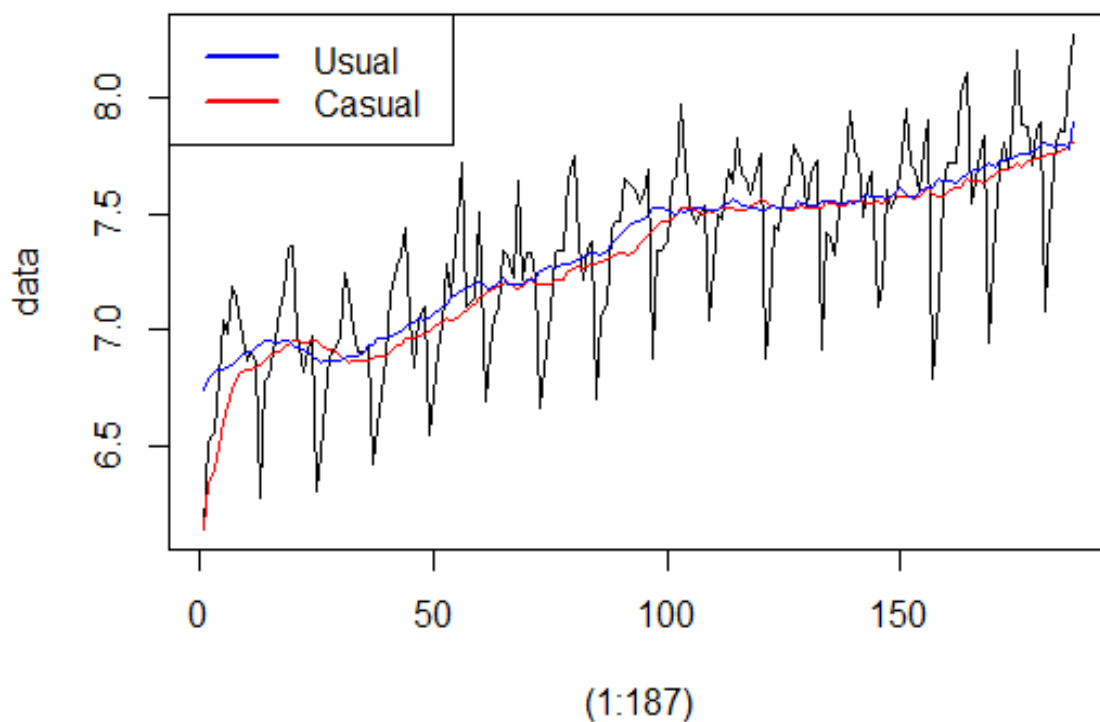


Рис. 1.6. МА, $M = 12$, $T = 12$

Видно, что причинный фильтр повторяет обычное скользящее среднее с запаздыванием.

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. 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`):

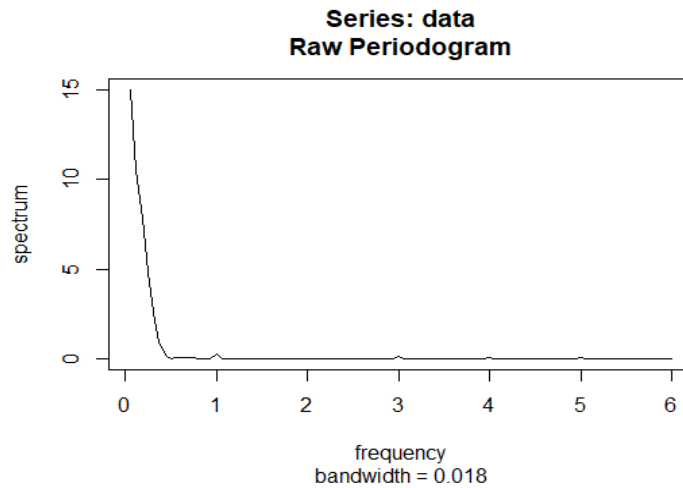


Рис. 1.7. Периодограмма ряда, `detrend = FALSE`

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:

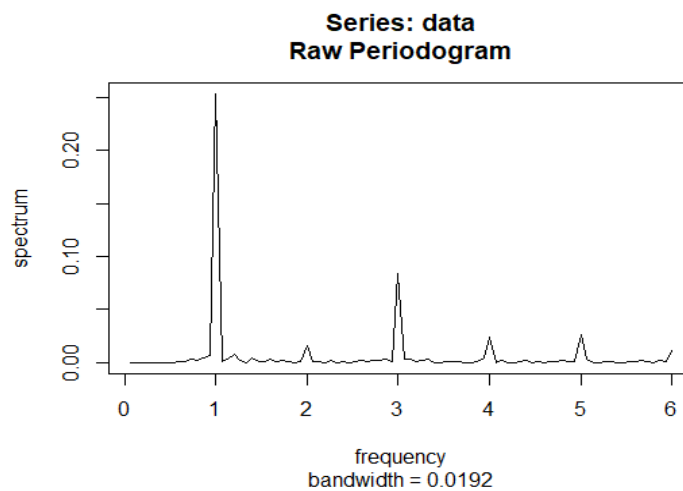


Рис. 1.8. Периодограмма остатка после вычитания результата применения фильтра, `detrend = FALSE`

Воздействие тренда полностью убрано, частоты периодик видны куда явнее.

Построим периодограмму компоненты, непосредственно получившейся в результате сглаживания:

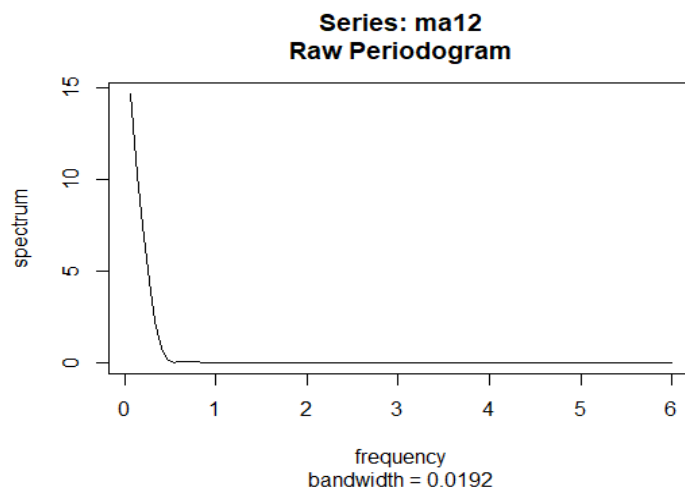


Рис. 1.9. Периодограмма фильтра, `detrend = FALSE`

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

Для мультипликативной модели: $Y = \log X$.

Если шум пуассоновский: $Y = \sqrt{X}$.

Если данные от 0 до 1 (вероятности, например): $Y = \arcsin X$.

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.

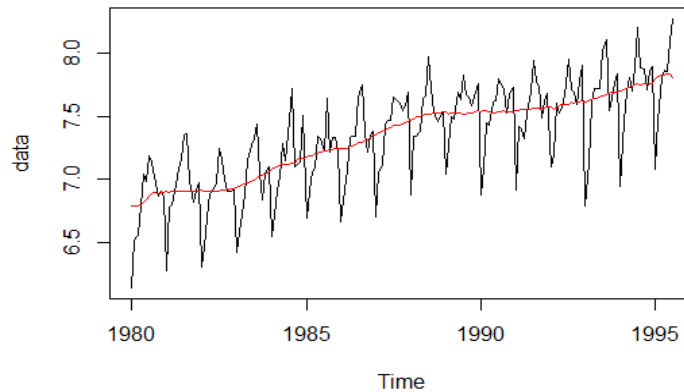


Рис. 1.10. Исходный ряд и выделенный с помощью МА ($M = 24$) тренд

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

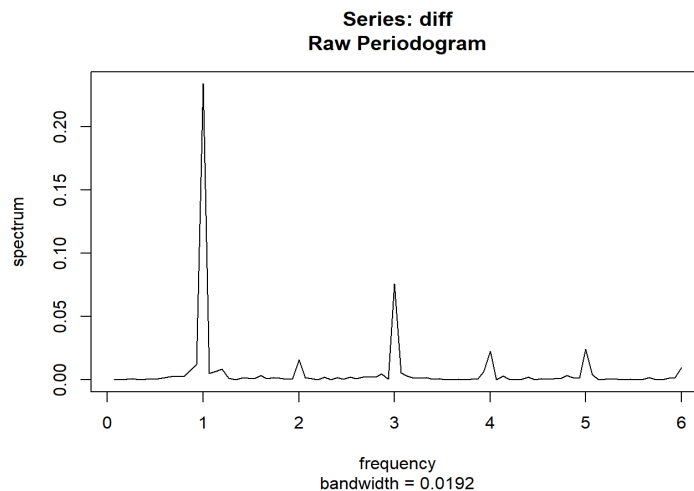


Рис. 1.11. Периодограмма остатка после вычитания тренда, полученного с пом. МА

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.

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:

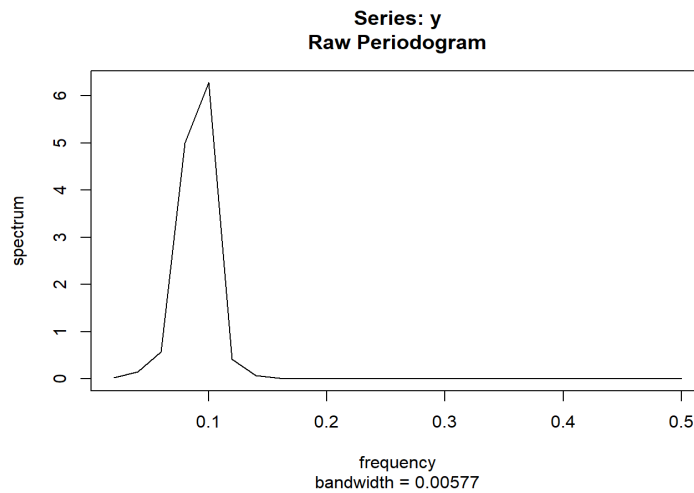


Рис. 1.12. Периодограмма исходного ряда

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

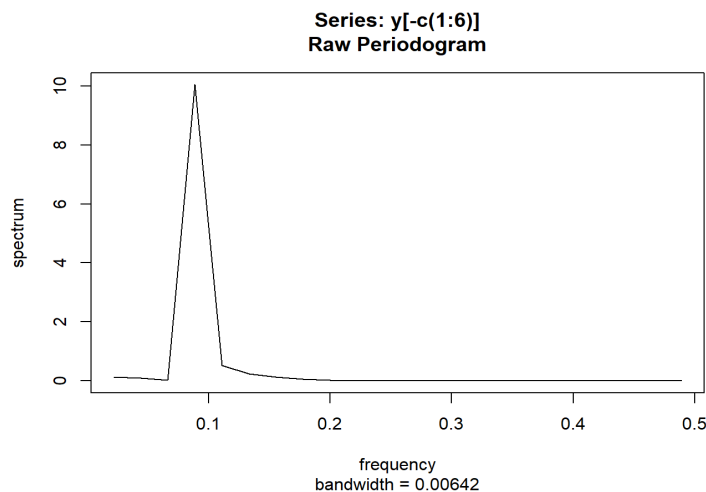


Рис. 1.13. Коррекция периодограммы

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

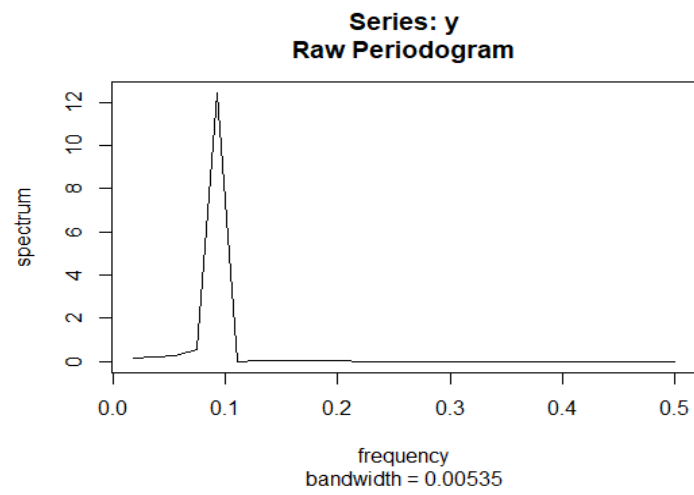


Рис. 1.14. Коррекция периодограммы

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.

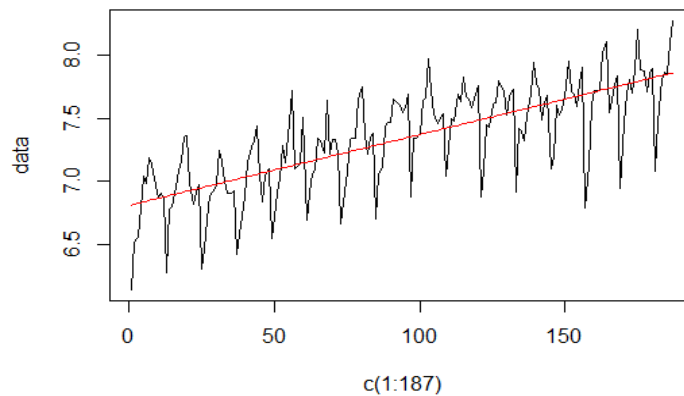


Рис. 1.15. Выделение тренда с помощью линейной регрессии

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

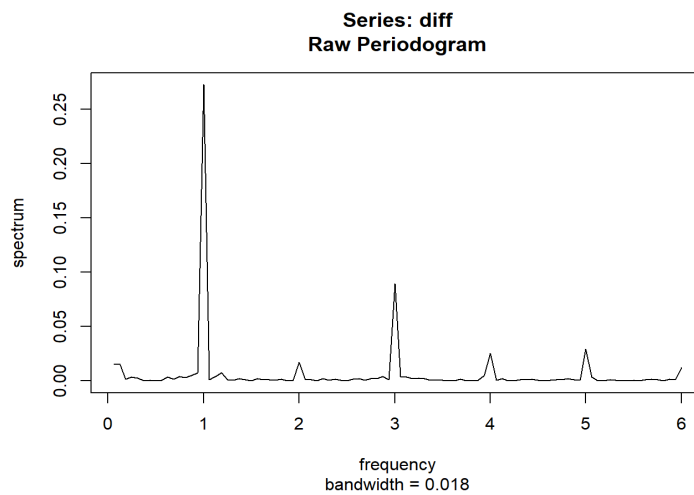


Рис. 1.16. Периодограмма остатка после вычитания тренда, полученного с пом. линейной регрессии

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:

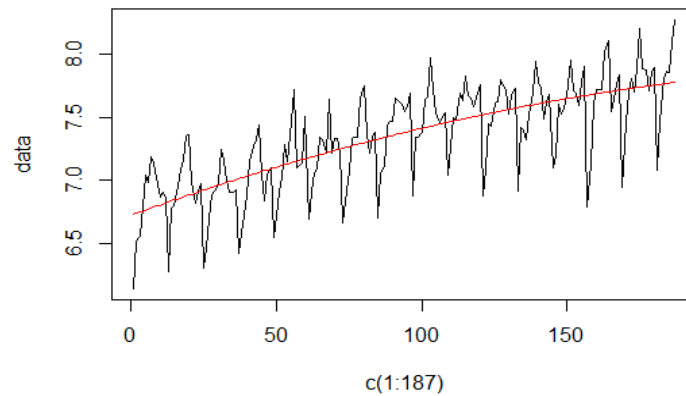


Рис. 1.17. Выделение тренда с помощью полиномиальной регрессии ($\text{deg} = 2$)

Визуально тренд выглядит значительно лучше. Периодограмма остатка:

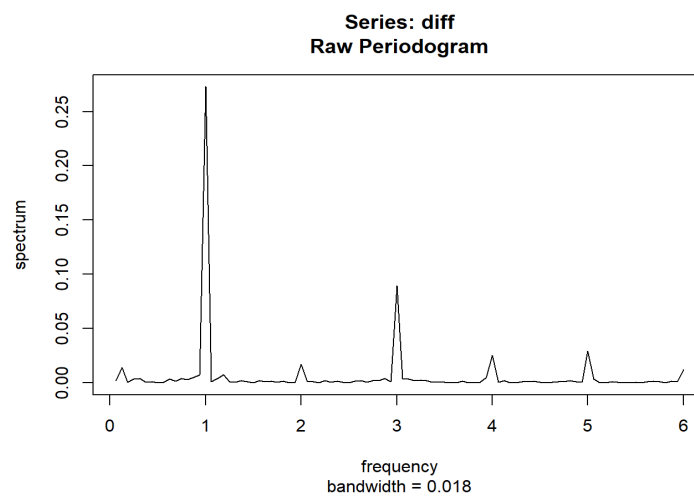


Рис. 1.18. Периодограмма остатка после вычитания тренда, полученного с пом. полиномиальной регрессии

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):

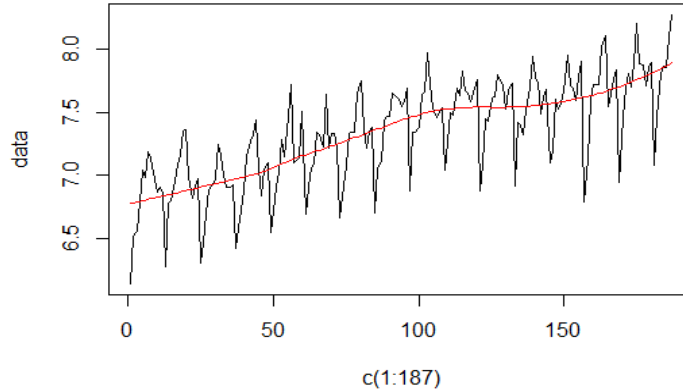


Рис. 1.19. Выделение тренда с помощью локальной регрессии

Как обычно, анализируем периодограмму остатка:

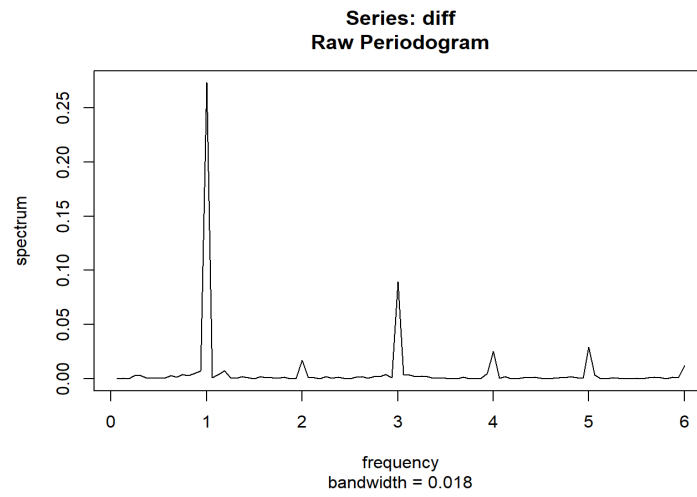


Рис. 1.20. Периодограмма остатка ряда

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

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.

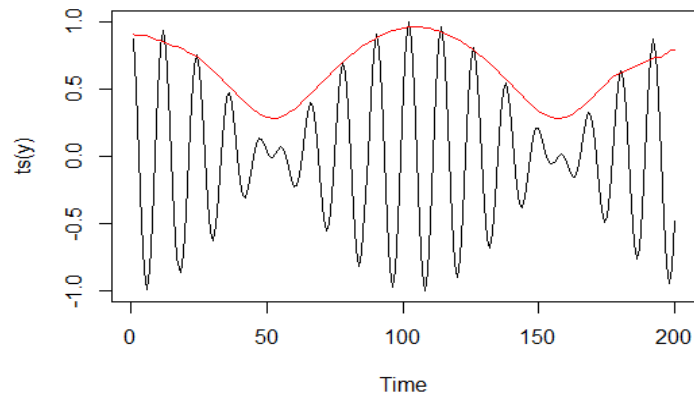


Рис. 1.21. Огибающая периодического ряда

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:

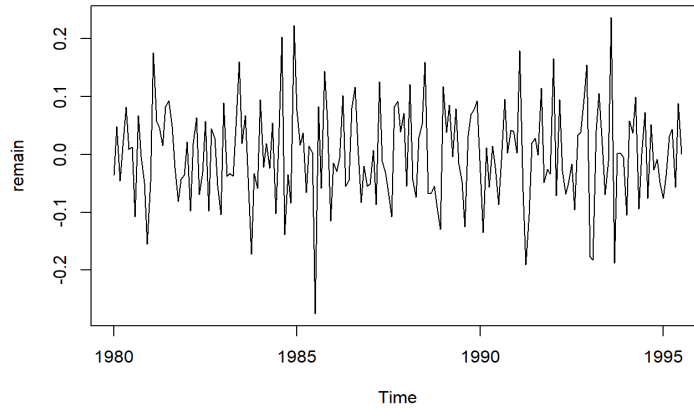


Рис. 1.22. Шум после выделения сигнала

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

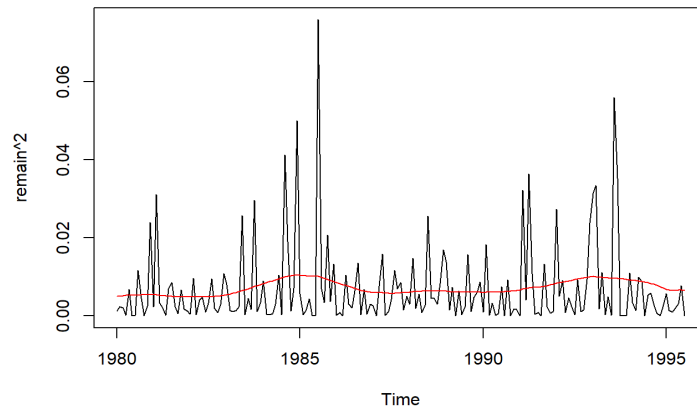


Рис. 1.23. Квадрат шума с выделенным трендом

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:

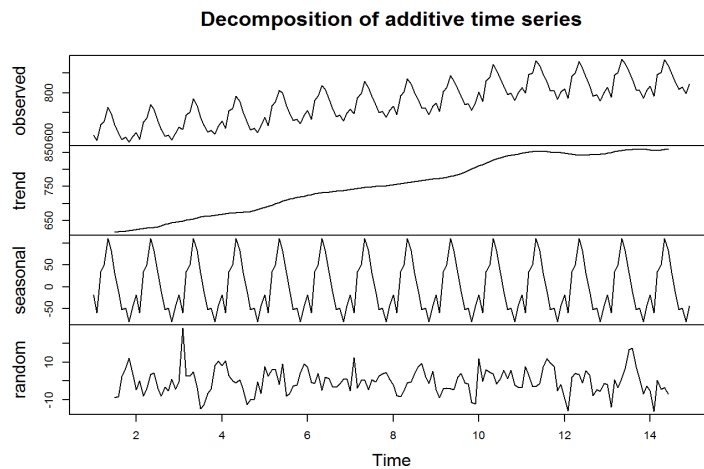


Рис. 1.24. Декомпозиция ряда на аддитивные составляющие

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).

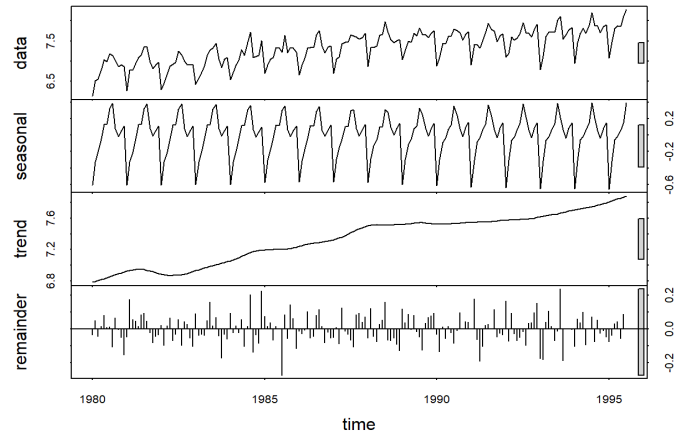


Рис. 1.25. Декомпозиция ряда на аддитивные составляющие

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. Последовательный SSA

2.2. Слабая и сильная разделимость

2.3. Определение сильной и слабой разделимости при смешивании компонент

2.4. Идентификация тренда

2.5. Идентификация периодичности (сезонности)

2.6. Использование матрицы взвешенных корреляций

2.7. Элементарные восстановленные компоненты

2.8. Корни характеристического полинома: сигнальные и лишние

2.9. Оценка параметров в SSA

2.10. Прогноз

2.11. Доверительные интервалы

2.12. Автоматическая идентификация

2.13. Заполнение пропусков

2.14. Теплицев SSA для стационарных рядов

2.15. Projection SSA, выделение линейного тренда

2.16. Iterative O-SSA и DerivSSA

2.17. Аппроксимация Cadzow SSA

2.18. MSSA для анализа многомерных временных рядов

2.19. 2D-SSA для разложения изображения

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

3.2. AR(p) и модель сигнала в SSA

3.3. Вид автоковариационной функции acf для AR(p)

3.4. Вид расф для AR(p)

3.5. Модель $MA(q)$, вид acf и $pacf$

3.6. ARMA(p,q)

3.7. Дифференцирование, ARIMA

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

3.9. Exponential smoothing, модели тренда, ES и ARIMA