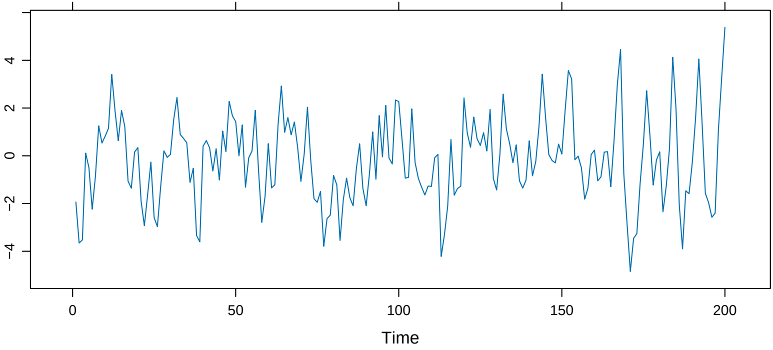
Выделение сигнала на основе критерия Monte Carlo SSA

Потешкин Егор Павлович, Голяндина Нина Эдуардовна

Санкт-Петербургский государственный университет
Кафедра статистического моделирования

Мат-Мех. Наука 2025
30 апреля 2025, Санкт-Петербург

Есть ли сигнал?



Вопрос: это чистый шум или там есть сигнал?

Постановка задачи

$\mathsf{X} = (x_1, \ldots, x_N)$, $x_i \in \mathbb{R}$ — временной ряд.

Дано: $\mathsf{X} = \mathsf{T} + \mathsf{H} + \mathsf{R}$, где T — тренд, H — периодическая компонента и R — шум.

Проблемы:

1. Как проверить наличие сигнала $\mathsf{S} = \mathsf{T} + \mathsf{H}$?
2. Как выделить сигнал S, если он есть?

Методы:

1. Monte-Carlo SSA (MC-SSA) [Allen and Smith, 1996] проверяет $H_0 : \mathsf{S} = 0$.
2. Singular spectrum analysis (SSA) [Golyandina, Nekrutkin and Zhigljavsky, 2001].

Задача: реализовать алгоритм автоматического выделения сигнала на основе MC-SSA.

Обозначения и известные результаты: оператор вложения и ганкелизации

$\mathsf{X} = (x_1, \ldots, x_N)$. Зафиксируем L $(1 < L < N)$.

Оператор вложения $\mathcal{T}_{\mathsf{SSA}}$:

$$\mathcal{T}_{\mathsf{SSA}}(\mathsf{X}) = \mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix},$$

где $K = N - L + 1$.

Оператор ганкелизации $\mathcal{H}$ — усреднение матрицы по побочным диагоналям.

Обозначения и известные результаты: алгоритм SSA

Входные данные: временной ряд $\mathsf{X} = (x_1, \ldots, x_N)$.
Параметры: длина окна L, набор индексов $I \subset \{1, \ldots, d\}$.
Выходные данные: оценка сигнала.

Входные данные: X – временной ряд

$\downarrow$ 1. Вложение

Траекторная матрица
$\mathbf{X} = \mathcal{T}_{\mathsf{SSA}}(\mathsf{X})$

$\rightarrow$ 2. Разложение

Сумма матриц единичного ранга
$$\mathbf{X} = \sum_{j=1}^{d} \mathbf{X}_j$$

$\downarrow$ 3. Группировка

Группировка матриц, соответствующих сигналу
$$\mathbf{X} = \mathbf{X}_I + (\mathbf{X} - \mathbf{X}_I)$$
$$\mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i$$

$\leftarrow$ 4. Восстановление

Результат: SSA разложение
$$\mathsf{X} = \widetilde{\mathsf{S}} + \widetilde{\mathsf{R}}$$
$$\widetilde{\mathsf{S}} = \mathcal{T}_{\mathsf{SSA}}^{-1} \circ \mathcal{H}(\mathbf{X}_I)$$

Пример: применение SSA

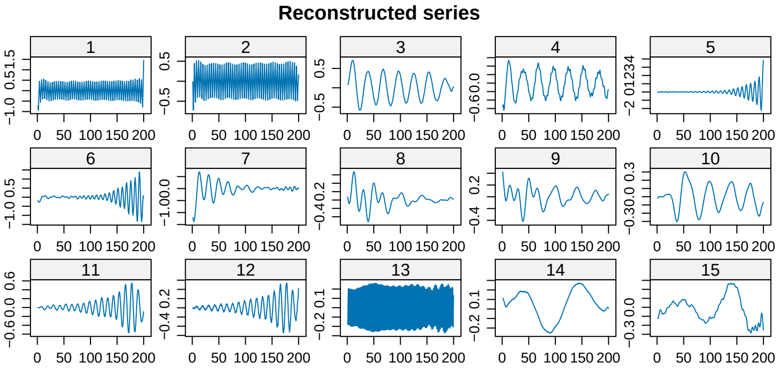


Рис.: Элементарные восстановленные компоненты $(L = 100)$

Компоненты, соответствующие сигналу: 1, 2, 5, 6 и 13.

Обозначения и известные результаты: Monte-Carlo SSA

Входные данные: $\mathsf{X} = \mathsf{S} + \mathsf{R}$, где S — сигнал, R — реализация стационарного процесса $\boldsymbol{\xi}$ с нулевым средним и со спектральной плотностью $f_{\boldsymbol{\theta}}$.

Параметры: длина окна L, $W_1, \ldots, W_M \in \mathbb{R}^L$ — нормированные векторы, соответствующие определенным частотам.

Статистика критерия: величины

$$\widehat{p}_k = \left\| \mathbf{X}^{\mathrm{T}} W_k \right\|^2.$$

Распределение $\widehat{p}_k$ при верной H_0, вообще говоря, неизвестно — оно оценивается с помощью метода Monte Carlo.

Multiple MC-SSA [Golyandina, 2023]: модификация MC-SSA с поправкой на множественные сравнения.

В качестве W_k рассматриваются косинусы с равностоящими частотами $\omega_k = k/(2L)$, $k = 1, \ldots, L$.

Обозначения и известные результаты: оценка параметров шума

Параметры шума $\boldsymbol{\theta}$, вообще говоря, неизвестны, поэтому их нужно оценивать.

Получить оценки параметров можно, максимизируя правдоподобие Whittle [Whittle, 1953]:

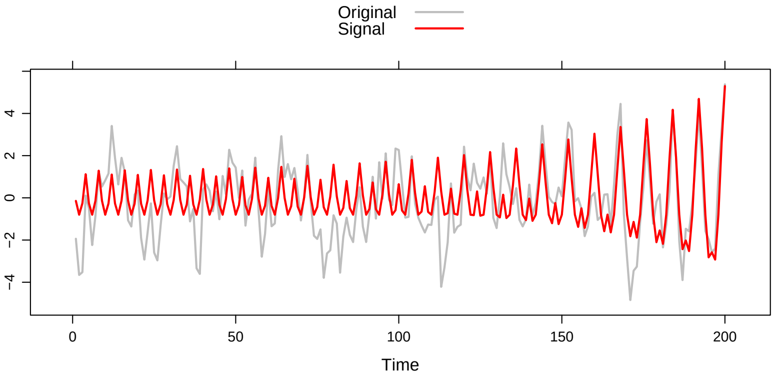
$$\ell_W(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{j=1}^{m}\left(\ln f_{\boldsymbol{\theta}}(\omega_j) + \frac{I_N(\omega_j)}{f_{\boldsymbol{\theta}}(\omega_j)}\right),$$

где $m = \lfloor (N-1)/2 \rfloor$, $f_{\boldsymbol{\theta}}$ — спектральная плотность $\boldsymbol{\xi}$, I_N — периодограмма исходного ряда, $\omega_j = j/N$.

Оценивать параметры можно по части спектра: пусть $J = \{j_1, \ldots, j_p\}$ — индексы частот, которые мы не хотим учитывать при оценке параметров. Тогда при вычислении $\ell_W(\boldsymbol{\theta})$ рассматриваются только индексы $j \notin J$.

$\mathsf{X} = \mathsf{S} + \boldsymbol{\xi}$, где $\boldsymbol{\xi}$ — красный шум с параметрами $\phi = 0.7$ и $\sigma^2 = 1$, $N = 200$,

$$s_n = 0.075\, e^{0.02n} \cos(2\pi n/8) + 2\cos(2\pi n/4) + 0.2 \cdot (-1)^n.$$



Пример: применение Monte Carlo SSA

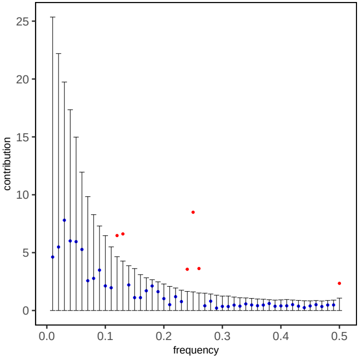


Рис.: Истинная модель шума

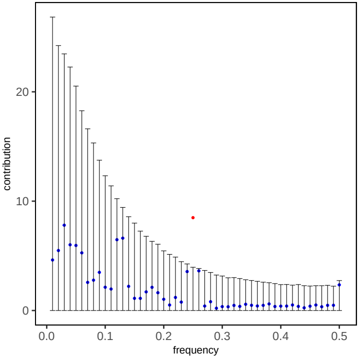


Рис.: Оцененная модель шума

Проблема: при оценивании параметров обнаруживаются не все частоты.

Решение: итеративно применять критерий после выделения одной гармоники, пока гипотеза $H_0 : \mathsf{S} = 0$ отвергается.

Обозначения и известные результаты: автоматическая группировка в SSA

Для ряда X длины N и $0 \leqslant \omega_1 \leqslant \omega_2 \leqslant 0.5$ определим меру

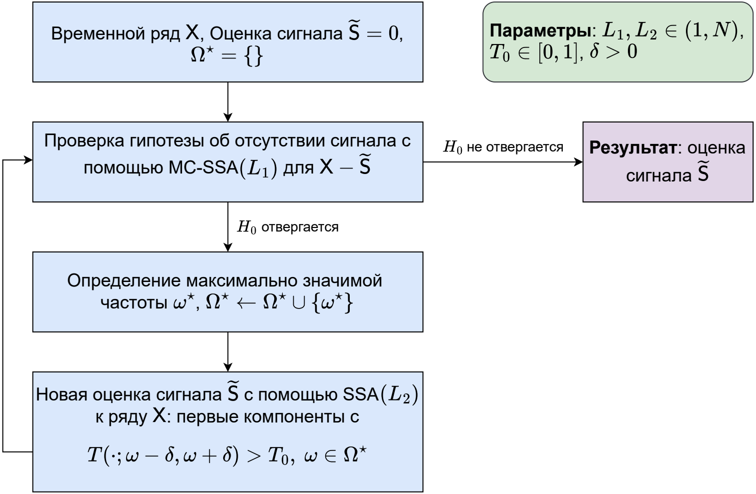
$$T(\mathsf{X}; \omega_1, \omega_2) = \frac{1}{\|\mathsf{X}\|^2} \sum_{k: \omega_1 \leqslant k/N \leqslant \omega_2} I_N(k/N),$$

где I_N — периодограмма X.

Величину $T(\mathsf{X}; \omega_1, \omega_2)$ можно рассматривать как долю вклада частот, содержащегося в интервале $[\omega_1, \omega_2]$.

Пусть $\omega^\star$ — значимая частота. Тогда $[\omega_1, \omega_2] = [\omega^\star - \delta, \omega^\star + \delta]$.

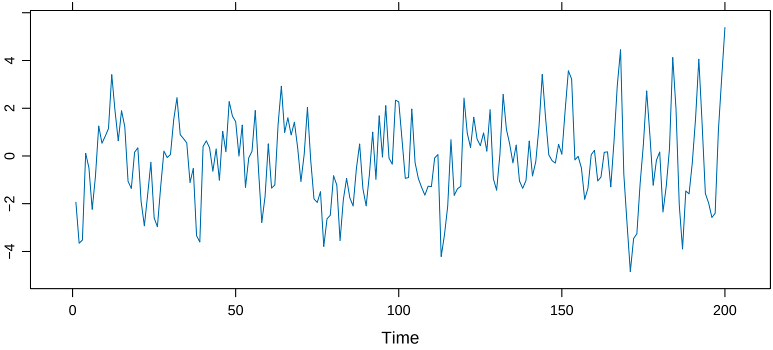
Алгоритм autoMCSSA



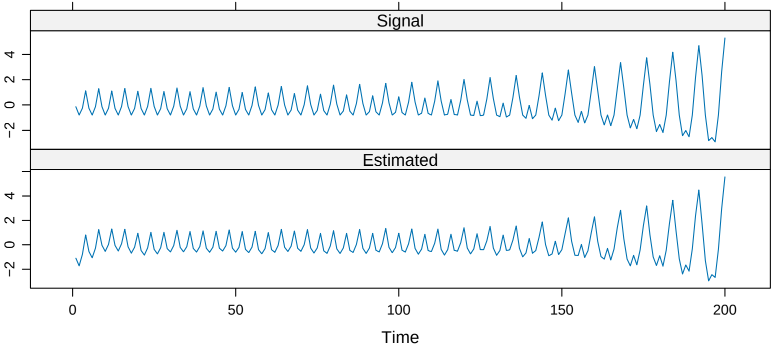
Пример: применение autoMCSSA

$\mathsf{X} = \mathsf{S} + \boldsymbol{\xi}$, где $\boldsymbol{\xi}$ — красный шум с параметрами $\phi = 0.7$ и $\sigma^2 = 1$, $N = 200$,

$$s_n = 0.075\, e^{0.02n} \cos(2\pi n/8) + 2\cos(2\pi n/4) + 0.2 \cdot (-1)^n.$$



Пример: применение autoMCSSA



Параметры: $L_1 = 50$, $L_2 = 100$, $\delta = 1/80$, $T_0 = 0.5$.

Метод autoMCSSA правильно идентифицировал значимые компоненты (1, 2, 5, 6 и 13).

Численное сравнение с autoSSA

Сравним метод autoMCSSA с методом autoSSA [Дудник, 2025].

Рассмотрим временной ряд $\mathsf{X} = \mathsf{S} + \boldsymbol{\xi}$ длины $N = 100$, где

$$s_n = 0.2e^{0.05n}\cos(2\pi n/4) + 2\cos(2\pi n/3) + (-1)^n,$$

$\boldsymbol{\xi}$ — красный шум с параметрами $\phi \in \{0, 0.5\}$, $\sigma^2 = 1$. Для autoMCSSA были выбраны следующие параметры:

- Длины окна $L_1 = 20$ и $L_2 = 50$;
- Радиус промежутка для вычисления меры T $\delta = 1/80$;
- Порог для меры T $T_0 = 0.5$;
- Максимальное количество итераций: 10.

Численное сравнение с autoSSA

Таблица: MSE выделения сигнала ($\phi = 0$)

	Mean MSE	Median MSE
autoMCSSA	0.15196	0.13921
autoSSA	0.14872	0.14003

Таблица: MSE выделения сигнала ($\phi = 0.5$)

	Mean MSE	Median MSE
autoMCSSA	0.11563	0.09038
autoSSA	0.09341	0.08795

Итоги

1. Был реализован метод autoMCSSA, позволяющий автоматически выделить значимый сигнал, а также модификация метода Whittle по части спектра.
2. Получено, что autoMCSSA позволяет выделять сигнал, компоненты которого в SSA необязательно доминируют.
3. При сравнении с autoSSA метод autoMCSSA показал сравнимый результат.
4. Необходимо сформулировать подход к выбору параметров autoMCSSA.