

Метод Монте-Карло SSA для многомерных временных рядов

Потешкин Егор Павлович, гр.20.Б04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Отчет по производственной практике (научно-исследовательская
работа) (6 семестр)

Санкт-Петербург, 2023

Временной ряд $X = (x_1, \dots, x_N)$.

Дано: $X = S + R$, где S — сигнал, R — шум.

Проблема: проверить, есть ли в сигнал в ряде, и, если есть, выделить его.

H_0 : ряд состоит из чистого шума ($S = 0$).

Критерий проверки H_0 : метод Monte-Carlo SSA (MC-SSA).

Метод SSA позволяет выделить обнаруженный сигнал.

Будем рассматривать часто используемую на практике версию MC-SSA, которая использует метод SSA. Назовем ее me1 [Golyandina, 2023].

Такая версия критерия радикальная, это можно исправить поправкой до точного критерия, но очень сильная радикальность мешает это сделать.

Решение: вместо SSA рассмотреть модификацию Toeplitz SSA. В [Ларин, 2022] показано, что Toeplitz MC-SSA менее радикальный.

Введение. Обобщение на многомерный случай

$X = \{X^{(d)}\}_{d=1}^D$ — D -канальный временной ряд.

MSSA и Monte-Carlo MSSA — обобщение SSA и MC-SSA на многомерный случай.

Проблема: me1, обобщенный на многомерный случай, очень радикальный [Ларин, 2022].

Решение: использовать Toeplitz MSSA.

Но: отсутствует реализация этой модификации.

Задача:

- 1 Реализовать Toeplitz MSSA.
- 2 Сравнить с обычным MSSA.

Обозначения и известные результаты: оператор вложения и ганкелизации

$X = (x_1, \dots, x_N)$. Зафиксируем *длину окна* L , $1 < L < N$.

Оператор вложения \mathcal{T} :

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

Результат \mathbf{X} — *траекторная матрица*.

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

Обозначения и известные результаты: метод MSSA

Дано: D -канальный временной ряд $\mathbf{X} = \{\mathbf{X}^{(d)}\}_{d=1}^D$.

Параметр: длина окна L .

Результат: m восстановленных D -канальных временных рядов.

❶ Вложение: $\mathbf{X} = [\mathcal{T}(\mathbf{X}^{(1)}) : \dots : \mathcal{T}(\mathbf{X}^{(D)})] = [\mathbf{X}^{(1)} : \dots : \mathbf{X}^{(D)}]$

❷ Сингулярное разложение (SVD):

$$\mathbf{X} = \sum_{i=1}^p \sqrt{\lambda_i} U_i V_i^T = \sum_i^p \mathbf{X}_i.$$

❸ Группировка: $\mathbf{X} = \sum_{k=1}^m \mathbf{X}_{I_k}$

❹ Диагональное усреднение: $\{\mathcal{T}^{-1} \circ \mathcal{H}(\mathbf{X}_{I_k}^{(d)})\}_{d=1}^D$

Обозначения и известные результаты: Toeplitz SSA

Стационарный ряд $X = (x_1, \dots, x_N)$.

Тёплицева L -ковариационная матрица \tilde{C} с элементами

$$\tilde{c}_{ij} = \frac{1}{N - |i - j|} \sum_{n=1}^{N-|i-j|} x_n x_{n+|i-j|}, \quad 1 \leq i, j \leq L, \quad (2)$$

является оценкой ковариационной матрицы.

Toeplitz SSA отличается от базового SSA на шаге разложения тем, что U_i — собственные векторы \tilde{C} . Причем важно, что X — стационарный ряд.

Полученные результаты: Toeplitz MSSA

SVD: $\mathbf{X} = \sum_{i=1}^p \sqrt{\lambda_i} U_i V_i^T$, где U_i — с.в. $\mathbf{X}\mathbf{X}^T$, V_i — с.в. $\mathbf{X}^T\mathbf{X}$.
 $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}\}$.

$$\begin{aligned}\mathbf{X}\mathbf{X}^T &= \mathbf{X}^{(1)}(\mathbf{X}^{(1)})^T + \mathbf{X}^{(2)}(\mathbf{X}^{(2)})^T \\ \mathbf{X}^T\mathbf{X} &= \begin{pmatrix} (\mathbf{X}^{(1)})^T \mathbf{X}^{(1)} & (\mathbf{X}^{(1)})^T \mathbf{X}^{(2)} \\ (\mathbf{X}^{(2)})^T \mathbf{X}^{(1)} & (\mathbf{X}^{(2)})^T \mathbf{X}^{(2)} \end{pmatrix}.\end{aligned}$$

В связи с этим возникает два метода Toeplitz MSSA:

- 1 Сумма тёплицевых L -ковариационных матриц — метод Sum.
- 2 Блочная матрица — метод Block.

- 1 Построить $\tilde{\mathbf{C}} = \sum_{d=1}^D \tilde{\mathbf{C}}_d$, где $\tilde{\mathbf{C}}_1, \dots, \tilde{\mathbf{C}}_D$ — тёплицевы L -ковариационные матрицы для каждого канала.
- 2 Найти ортонормированные собственные векторы H_1, \dots, H_L матрицы $\tilde{\mathbf{C}}$.
- 3 Получить разложение

$$\mathbf{X} = \sum_{i=1}^L H_i Z_i^T, \quad (3)$$

где $Z_i = \mathbf{X}^T H_i$.

- ❶ $K = N - L + 1$. Рассмотрим блочную матрицу \mathbf{T} , где элементы каждого блока \mathbf{T}_{lk} , $1 \leq l, k \leq D$, имеют вид

$$t_{ij}^{(lk)} = \frac{1}{\tilde{N}} \sum_{n=\max(1, 1+i-j)}^{\min(N, N+i-j)} x_n^{(l)} x_{n+j-i}^{(k)}, \quad 1 \leq i, j \leq K, \quad (4)$$

где $\tilde{N} = \min(N, N + i - j) - \max(1, 1 + i - j) + 1$.

- ❷ Найти ортонормированные собственные векторы Q_1, \dots, Q_{DK} матрицы \mathbf{T} и получить разложение

$$\mathbf{X} = \sum_{i=1}^{DK} (\mathbf{X} Q_i) Q_i^T = \sum_{i=1}^{DK} P_i Q_i^T. \quad (5)$$

Полученные результаты: Toeplitz MSSA. Численное исследование

Дано: $(F^{(1)}, F^{(2)}) = (H^{(1)}, H^{(2)}) + (R^{(1)}, R^{(2)})$, $N = 71$.

Задача: проверить точность базового и модифицированных методов для разных значений параметра L .

Рассмотрим 2 случая:

❶ Одинаковые периоды:

$$h_n^{(1)} = 30 \cos(2\pi n/12), \quad h_n^{(2)} = 20 \cos(2\pi n/12), \quad n = 1, \dots, N.$$

❷ Разные периоды:

$$h_n^{(1)} = 30 \cos(2\pi n/12), \quad h_n^{(2)} = 20 \cos(2\pi n/8), \quad n = 1, \dots, N.$$

Полученные результаты: Toeplitz MSSA. Численное исследование. Результаты

Таблица: MSE восстановления сигнала.

Случай 1	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
MSSA	3.18	1.83	1.59	1.47	2.00
SSA	3.25	2.01	2.00	2.01	3.25
Sum	3.17	1.75	1.44	1.32	1.33
Block	1.39	1.26	1.25	1.33	1.97

Случай 2	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
MSSA	6.91	3.77	3.07	2.88	3.84
SSA	3.23	2.01	2.00	2.01	3.23
Sum	6.88	3.65	2.64	2.37	2.27
Block	4.47	3.67	3.22	3.23	3.8

Дано: $X = S + R$, где S — сигнал, R — красный шум.

Параметр: W — вектор с какой-то частотой.

Задача: проверка гипотезы H_0 , что ряд состоит из чистого шума ($S = 0$).

Решение: Monte-Carlo SSA:

- 1 Построить статистику критерия $\hat{p} = \|X^T W\|^2$.
- 2 Построить доверительную область случайной величины $p = \|\Xi^T W\|^2$: распределение p оценивается методом Монте-Карло и строится интервал от нуля до $(1 - \alpha)$ -квантиля.
- 3 Если \hat{p} не попадает в построенный интервал — H_0 отвергается.

Обозначения и известные результаты: MC-SSA: множественные тесты

- В реальных задачах частота сигнала неизвестна и нужно проверить, что в ряде присутствует сигнал.
- Рассматривается набор векторов W_1, \dots, W_H , соответствующих разным частотам.
- Проблема множественного тестирования, нужно ограничить групповую ошибку.
- Решение: Multiple MC-SSA [Golyandina, 2023].

Обозначения и известные результаты: Multiple MC-SSA

Дано: временной ряд X .

Параметр: способ выбора векторов W_1, \dots, W_H , количество суррогатных данных G .

Алгоритм:

- 1 Для $k = 1, \dots, H$ вычисляется статистика \hat{p}_k , выборка $P_k = \{p_{ki}\}_{i=1}^G$, где $p_{ki} = \|\Xi_i^T W_k\|^2$, ее среднее μ_k и стандартное отклонение σ_k .

- 2 Вычисляется $\eta = (\eta_1, \dots, \eta_G)$ с

$$\eta_i = \max_{1 \leq k \leq H} (p_{ki} - \mu_k) / \sigma_k, \quad i = 1, \dots, G.$$

- 3 Находится q_k как выборочный $(1 - \alpha)$ -квантиль η , где α — уровень значимости.
- 4 Нулевая гипотеза не отвергается, если

$$\max_{1 \leq k \leq H} (\hat{p}_k - \mu_k) / \sigma_k < q. \quad (6)$$

Обозначения и известные результаты: Multiple MC-MSSA: рассматриваемые случаи

Basic MC-MSSA: W_i — собственные или факторные векторы матрицы \mathbf{X} .

Toeplitz MC-MSSA: $\mathbf{X} = \sum_i \sigma_i H_i Q_i^T$, $W_i = H_i$ или Q_i .

Пусть $\mathbf{X} = \sum_i \sigma_i H_i Q_i^T$ — любое приведенное ранее разложение. Будем называть H_i *левыми*, а Q_i — *правыми векторами* матрицы \mathbf{X} .

Определение

ROC-кривая — это кривая, задаваемая параметрически

$$\begin{cases} x = \alpha_I(\alpha) \\ y = \beta(\alpha) \end{cases}, \quad \alpha \in [0, 1],$$

где $\alpha_I(\alpha)$ — функция зависимости ошибки первого рода α_I от уровня значимости α , $\beta(\alpha)$ — функция зависимости мощности β от уровня значимости α .

С помощью ROC-кривых можно сравнивать по мощности неточные (в частности, радикальные) критерии, к которым применена поправка. Отметим, что для точного критерия ROC-кривая совпадает с графиком мощности, так как $\alpha_I(\alpha) = \alpha$.

Полученные результаты: MC-MSSA. Численное сравнение методов.

Дано: $X = S + R$, где R — красный шум с параметрами $\varphi = 0.7$ и $\delta = 1$, а S — сигнал с

$$s_n^{(1)} = s_n^{(2)} = \cos(2\pi n\omega), \quad n = 1, \dots, N,$$

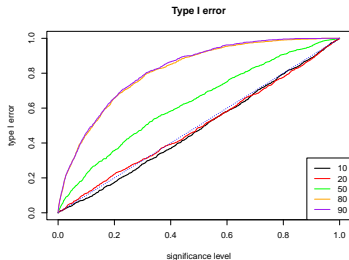
где $w = 0.075$, $N = 100$, $G = 1000$.

H_0 : $S = 0$,

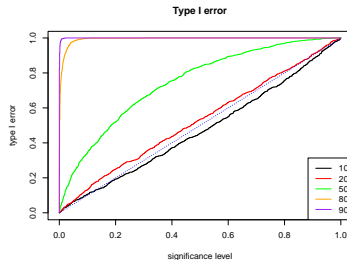
H_1 : $S \neq 0$

Задача: сравнить критерии Basic MC-MSSA и Toeplitz MC-MSSA в двух вариациях с помощью графиков ошибки первого рода и ROC-кривых.

Полученные результаты: MC-MSSA. Численное сравнение методов. Метод Sum.



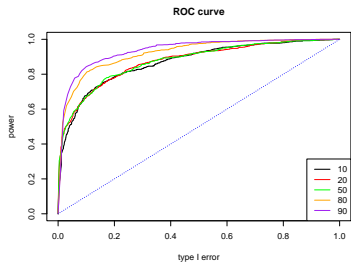
(a) Sum



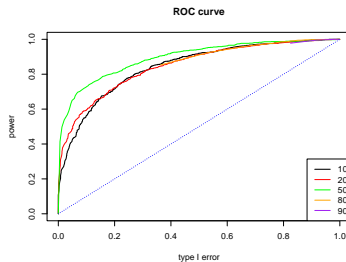
(b) базовый MSSA

Рис.: Ошибка первого рода критерев с проекцией на левые векторы методов Sum и базового MSSA.

Полученные результаты: MC-MSSA. Численное сравнение методов. Метод Sum. Продолжение.



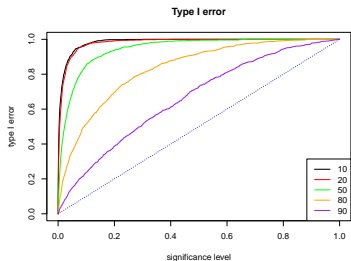
(a) Sum



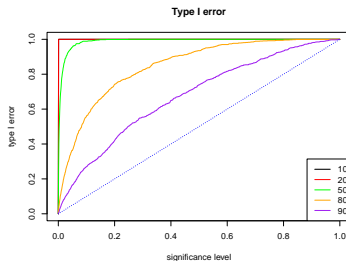
(b) базовый MSSA

Рис.: ROC-кривая критериев с проекцией на левые векторы методов Sum и базового MSSA.

Полученные результаты: MC-MSSA. Численное сравнение методов. Метод Block.



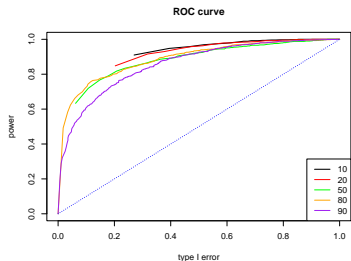
(a) Block



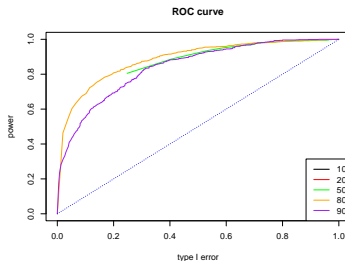
(b) базовый MSSA

Рис.: Ошибка первого рода критериев с проекцией на правые векторы методов Block и базового MSSA.

Полученные результаты: MC-MSSA. Численное сравнение методов. Метод Block. Продолжение.



(a) Block






(b) базовый MSSA

Рис.: ROC-кривая критериев с проекцией на правые векторы методов Block и базового MSSA.

Мои результаты:

- ❶ Был реализован метод Toeplitz MSSA в вариантах block и sum (на языке R), на его основе была построена реализация Toeplitz Multiple MC-MSSA.
- ❷ На основе численных экспериментов были сделаны следующие выводы:
 - Sum и Block версии Toeplitz MSSA для стационарного ряда точнее выделяют сигнал, чем Basic MSSA.
 - Toeplitz MC-MSSA менее радикален, чем Basic MC-MSSA. Версия Sum с проекцией на собственные вектора оказывается наименее радикальной без потери в мощности скорректированного критерия.

В дальнейшем предполагается использовать метод Sum с проекцией на левые векторы, оптимизировать его реализацию и продолжить исследование метода Monte-Carlo MSSA.

-  Golyandina N. Detection of signals by Monte Carlo singular spectrum analysis: multiple testing // Statistics and Its Interface. — 2023. — Vol. 16, no. 1. — P. 147–157.
-  Ларин Е. С. Метод SSA для проверки гипотезы о существовании сигнала во временном ряде : квалификационная работа магистра ; СПбГУ. — 2022.
-  Multivariate and 2D Extensions of Singular Spectrum Analysis with the Rssa Package / Golyandina Nina, Korobeynikov Anton, Shlemov Alex, and Usevich Konstantin // Journal of Statistical Software. — 2015. — Vol. 67, no. 2.