

Санкт-Петербургский государственный университет
Прикладная математика и информатика

Отчет по учебной практике 3 (научно-исследовательской работе) (семестр 6)

МЕТОД МОНТЕ-КАРЛО SSA ДЛЯ МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

Выполнил:

Потешкин Егор Павлович

группа 20.Б04-мм

Научный руководитель:

к. ф.-м. н., доцент

Голяндина Нина Эдуардовна

Кафедра Статистического Моделирования

Санкт-Петербург

2023

Оглавление

Введение	3
Глава 1. Метод MSSA	4
1.1. Описание метода	4
1.2. Модификации метода	6
1.3. Выбор длины окна (численное исследование)	7
Глава 2. Метод Monte-Carlo MSSA	9
2.1. Постановка задачи	9
2.2. Одиночный тест	9
2.3. Множественный тест	10
2.4. Выбор векторов для проекции	11
2.5. Численное сравнение методов	12
2.6. Выводы	15
Заключение	16
Список литературы	17

Введение

Метод Singular Spectrum Analysis (SSA) является мощным инструментом для анализа временных рядов. Он позволяет разложить ряд на интерпретируемые компоненты, такие как тренд, периодические колебания и шум, что значительно упрощает процесс анализа. Метод Monte-Carlo SSA, в свою очередь, решает задачу обнаружения сигнала в шуме [1].

Однако, вариант Monte-Carlo SSA для анализа многомерных временных рядов мало исследован. В работе [2] рассматривается применение метода Monte Carlo SSA для анализа многомерных временных рядов, и авторы сталкиваются с проблемой отсутствия реализации Тёплицева варианта MSSA в пакете Rssa [3].

В этой работе была поставлена задача реализовать двумя способами метод Toeplitz MSSA, сравнить их между собой и с обычным MSSA как для методов оценки сигнала, так и для использования в Monte-Carlo MSSA.

Глава 1

Метод MSSA

Метод Multivariate Singular Spectrum Analysis (сокращенно MSSA) состоит из четырех этапов: *вложения, разложения, группировки и диагонального усреднения*. Давайте начнем с общих частей всех версий алгоритмов SSA. Этими общими частями являются процедура вложения и диагонального усреднения (ганкелизации).

Определение 1. Пусть \mathbf{X} — одномерный временной ряд длины N . Выберем параметр L , называемый *длиной окна*, $1 < L < N$. Рассмотрим $K = N - L + 1$ векторов вложения $X_i = (x_i, \dots, x_{i+L-1})^T$, $1 \leq i \leq K$. Определим оператор вложения \mathcal{T} следующим образом:

$$\mathcal{T}(\mathbf{X}) = \mathbf{X} = [X_1 : \dots : X_K] = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix}. \quad (1.1)$$

Определение 2. Матрицу \mathbf{X} из (1.1) называют траекторной матрицей.

Заметим, что матрица \mathbf{X} является *ганкелевой*, т.е. на всех ее побочных диагоналях стоят одинаковые элементы, а оператор \mathcal{T} задает взаимно-однозначное соответствие между множеством временных рядов длины N и множеством ганкелевых матриц $L \times K$.

Определение 3. Пусть $\mathbf{Y} = \{y_{ij}\}_{i,j=1}^{L,K}$ — некоторая матрица. Определим оператор ганкелизации \mathcal{H} :

$$(\mathcal{H}(\mathbf{Y}))_{ij} = \sum_{(l,k) \in A_s} y_{lk} / w_s, \quad (1.2)$$

где $s = i+j-1$, $A_s = \{(l, k) : l+k = s+1, 1 \leq l \leq L, 1 \leq k \leq K\}$ и $w_s = |A_s|$ — количество элементов в множестве A_s . Это соответствует усреднению элементов матрицы \mathbf{Y} по побочным диагоналям.

1.1. Описание метода

Рассмотрим вещественнозначные одномерные временные ряды $\mathbf{X}^{(d)} = (x_1^{(d)}, x_2^{(d)}, \dots, x_{N_d}^{(d)})$ длины $N_d > 2$, $d = 1, \dots, D$. Составим из этих рядов $\mathbf{X} = \{\mathbf{X}^{(d)}\}_{d=1}^D$ — D -канальный временной ряд с длинами N_d .

1.1.1. Вложение

Зафиксируем L , $1 < L < \min(N_1, \dots, N_D)$. Для каждого ряда $\mathbf{X}^{(d)}$ составим траекторную матрицу $\mathbf{X}^{(d)}$. Обозначим $K = \sum_{d=1}^D K_d$. Результатом этапа вложения является траекторная матрица многоканального временного ряда

$$\mathbf{X} = [\mathcal{T}(\mathbf{X}^{(1)}) : \dots : \mathcal{T}(\mathbf{X}^{(D)})] = [\mathbf{X}^{(1)} : \dots : \mathbf{X}^{(D)}]. \quad (1.3)$$

1.1.2. Разложение

Задача этапа разложения — разбить траекторную матрицу \mathbf{X} в сумму матриц ранга 1. В базовой версии MSSA используется сингулярное разложение (SVD).

Положим $\mathbf{S} = \mathbf{X}\mathbf{X}^T$. Пусть λ_i — собственные числа, а U_i — ортонормированная система векторов матрицы \mathbf{S} . Упорядочим λ_i по убыванию и найдем p такое, что $\lambda_p > 0$, а $\lambda_{p+1} = 0$. Тогда

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

где $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$. Тройку $(\sqrt{\lambda_i}, U_i, V_i)$ принято называть i -й собственной тройкой сингулярного разложения, $\sqrt{\lambda_i}$ — сингулярным числом, U_i — левым сингулярным вектором, а V_i — правым сингулярным вектором. Отметим, что левые сингулярные векторы имеют размерность L , а правые сингулярные вектора — размерность K .

1.1.3. Группировка

На этом шаге множество индексов $I = \{1, \dots, p\}$ разбивается на m непересекающихся множеств I_1, \dots, I_m и матрица \mathbf{X} представляется в виде суммы

$$\mathbf{X} = \sum_{k=1}^m \mathbf{X}_{I_k},$$

где $\mathbf{X}_{I_k} = \sum_{i \in I_k} \mathbf{X}_i$.

1.1.4. Диагональное усреднение

Финальным шагом MSSA является преобразование каждой матрицы \mathbf{X}_{I_k} , составленной в разделе 1.1.3, в D -канальный временной ряд следующим образом:

$$\tilde{\mathbf{X}}_{I_k} = \left\{ \mathcal{T}^{-1} \circ \mathcal{H} \left(\mathbf{X}_{I_k}^{(d)} \right) \right\}_{d=1}^D, \quad (1.4)$$

где \mathcal{T} — оператор вложения (1.1), \mathcal{H} — оператор ганкелизации (1.2).

1.1.5. Частный случай

При $D = 1$ \mathbf{X} — одномерный временной ряд, и приведенный выше алгоритм совпадает с алгоритмом Basic SSA, описанный в [4].

1.2. Модификации метода

1.2.1. Тёплицев MSSA

В случае анализа стационарных рядов можно улучшить базовый метод, используя другое разложение матрицы \mathbf{X} . Для начала введем следующее понятие.

Определение 4. Пусть $\mathbf{X} = (x_1, \dots, x_N)$ — одномерный временной ряд и L — фиксированное. Тёплицевой L -ковариационной матрицей называют матрицу $\tilde{\mathbf{C}}$ с элементами

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

Пусть теперь $\mathbf{X} = \{\mathbf{X}^{(d)}\}_{d=1}^D$ — D -канальный временной ряд с длинами N_d . Тогда можно получить разложение \mathbf{X} двумя способами:

1. Пусть $\tilde{\mathbf{C}}_1, \dots, \tilde{\mathbf{C}}_D$ — тёплицевы матрицы для каждого канала. Рассмотрим $\tilde{\mathbf{C}} = \sum_{d=1}^D \tilde{\mathbf{C}}_d$. Найдем ортонормированные собственные векторы H_1, \dots, H_L матрицы $\tilde{\mathbf{C}}$ и разложим траекторную матрицу \mathbf{X} следующим образом:

$$\mathbf{X} = \sum_{i=1}^L \sigma_i H_i Q_i^T, \quad (1.5)$$

где $Z_i = \mathbf{X}^T U_i$, $Q_i = Z_i / \|Z_i\|$ и $\sigma_i = \|Z_i\|$.

2. Если все каналы одинаковой длины N , можно рассмотреть блочную матрицу размера $DK \times DK$:

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{1,1} & \mathbf{T}_{1,2} & \cdots & \mathbf{T}_{1,D} \\ \mathbf{T}_{2,1} & \mathbf{T}_{2,2} & \cdots & \mathbf{T}_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_{D,1} & \mathbf{T}_{D,2} & \cdots & \mathbf{T}_{D,D} \end{pmatrix},$$

где $K = N - L + 1$. Элементы каждого блока \mathbf{T}_{lk} имеют вид

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

где $\tilde{N} = \min(N, N + i - j) - \max(1, 1 + i - j) + 1$. Найдя ортонормированные собственные векторы Q_1, \dots, Q_{DK} матрицы \mathbf{T} , получаем разложение

$$\mathbf{X} = \sum_{i=1}^{DK} \sigma_i H_i Q_i^T, \quad (1.6)$$

где $Z_i = \mathbf{X}Q_i$, $H_i = Z_i / \|Z_i\|$ и $\sigma_i = \|Z_i\|$.

Шаги группировки и диагонального усреднения можно оставить в том виде, в котором они представлены в разделе 1.1.3 и в разделе 1.1.4.

Для конкретности, будем называть первый метод Sum, а второй — Block. Стоит отметить, что в Sum собственные векторы матрицы $\tilde{\mathbf{C}}$ — аналоги левых сингулярных векторов матрицы \mathbf{X} , в то время как в Block собственные векторы матрицы \mathbf{T} — аналоги правых сингулярных векторов. В дальнейшем будем просто называть их *левыми* и *правыми векторами*.

1.3. Выбор длины окна (численное исследование)

Посмотрим на точность базового и модифицированных методов, для разных значений параметра L , наподобие работы [5]. Рассмотрим следующий двухканальный временной ряд: $(F^{(1)}, F^{(2)}) = (H^{(1)}, H^{(2)}) + (N^{(1)}, N^{(2)})$, где $H^{(1)}, H^{(2)}$ — гармоники, а $N^{(1)}, N^{(2)}$ — независимые реализации гауссовского белого шума. Гауссовский белый шум — стационарный случайный процесс, имеющий нормальное распределение. Как и в [5], пусть $N = 71$, дисперсия шумовых компонент $\sigma^2 = 25$, число повторений равно 10000. Рассмотрим 2 случая:

1. Одинаковые периоды:

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

2. Разные периоды:

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

В таблице 1.1 представлены результаты восстановления сигнала для разных L . Данные для методов SSA и MSSA были взяты из работы [5]. Наиболее точные результаты для каждого метода были выделены жирным шрифтом. Как видим из таблицы 1.1, в обоих

Таблица 1.1. 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

случаях метод Sum показывал наилучший результат для $L > (N + 1)/2$, в то время как метод Block наиболее точен при длине окна, близкой к половине длины ряда, причем оба метода в случае одинаковых периодов показывают лучше результат, чем MSSA.

Глава 2

Метод Monte-Carlo MSSA

2.1. Постановка задачи

Рассмотрим задачу поиска сигнала (не случайной составляющей) в многоканальном временном ряде. Нулевая гипотеза H_0 — отсутствие сигнала (ряд состоит из чистого шума). Тогда альтернатива H_1 — ряд содержит сигнал, например, периодическую составляющую.

Определение 5. Случайный вектор $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)$ называют красным шумом с параметрами φ и δ , если $\xi_n = \varphi \xi_{n-1} + \delta \varepsilon_n$, где $0 < \varphi < 1$, ε_n — белый гауссовский шум со средним значением 0 и дисперсией 1 и ξ_1 имеет нормальное распределение с нулевым средним и дисперсией $\delta^2/(1 - \varphi^2)$.

В данной главе под шумом будем подразумевать именно красный, причем с известными параметрами. Также будем рассматривать только односторонние критерии.

2.2. Одиночный тест

Пусть $\boldsymbol{\xi} = \{\boldsymbol{\xi}^{(d)}\}_{d=1}^D$ — D -канальный красный шум. Зафиксируем длину окна L и обозначим траекторную матрицу ряда $\boldsymbol{\xi}$ как $\boldsymbol{\Theta}$. Рассмотрим вектор $W \in \mathbb{R}^L$ такой, что $\|W\| = 1$. Введем величину

$$p = \|\boldsymbol{\Theta}^T W_k\|^2.$$

Статистикой критерия является величина

$$\hat{p} = \|\mathbf{X}^T W\|^2.$$

Если вектор W — синусоида с частотой ω , то \hat{p} отражает вклад частоты w в исходный ряд.

Рассмотрим алгоритм статистического критерия проверки наличия сигнала в ряде с проекцией на один вектор W , описанный в работе [1].

Алгоритм 1. Одиночный тест [1]

1. Построить статистику критерия \hat{p} .

2. Построить доверительную область случайной величины p : интервал от нуля до $(1 - \alpha)$ -квантиля, где α — уровень значимости.
3. Если \hat{p} не попадает в построенный интервал — H_0 отвергается.

Построенная доверительная область называется *прогнозируемым интервалом* с уровнем доверия $1 - \alpha$.

В большинстве случаев, распределение p неизвестно. Поэтому оно оценивается методом Монте-Карло: берется G реализаций случайной величины ξ , для каждой вычисляется p и строится эмпирическое распределение. В связи с этим описанный выше алгоритм называют методом Monte-Carlo SSA.

2.3. Множественный тест

Пусть теперь частоты периодических компонент неизвестны (что не редкость на практике), но известен диапазон частот и нужно проверить, что в ряде присутствует сигнал с хотя бы одной частотой из заданного диапазона. Тогда нулевая гипотеза H_0 о том, что ряд не содержит сигнала ни на одной из частот из рассматриваемого диапазона, а альтернатива H_1 — ряд содержит сигнал с хотя бы одной частотой, принадлежащей рассматриваемому диапазону.

Пусть W_1, \dots, W_H — вектора для проекции. В таком случае нужно построить H предсказательных интервалов по выборкам $P_k = \{p_{ki}\}_{i=1}^G$ с элементами

$$p_{ki} = \|\Xi_i^T W_k\|^2, \quad i = 1, \dots, G; \quad k = 1, \dots, H, \quad (2.1)$$

где G — количество суррогатных реализаций ξ , Ξ_i — траекторная матрица i -й реализации ξ .

В работе [1] подробно описана проблема множественного тестирования, когда вероятность ложного обнаружения периодической составляющей для одной из рассматриваемых частот (групповая ошибка I рода) неизвестна и значительно превышает заданный уровень значимости (частота ошибок одиночного теста), и ее решение. Приведем модифицированный алгоритм построения критерия в случае множественного тестирования, который будем использовать в дальнейшем.

Алгоритм 2. Multiple MC-SSA [1]

1. Для $k = 1, \dots, H$ вычисляется статистика \hat{p}_k , выборка $P_k = \{p_{ki}\}_{i=1}^G$, ее среднее μ_k и стандартное отклонение σ_k .
2. Вычисляется $\eta = (\eta_1, \dots, \eta_G)$, где

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

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

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

5. Если H_0 отвергнута, вклад W_k (и соответствующей частоты) существенен, если \hat{p}_k превосходит $\mu_k + q\sigma_k$. Таким образом, $[0, \mu_k + q\sigma_k]$ считаются скорректированными интервалами прогнозирования.

2.4. Выбор векторов для проекции

Для начала отметим, что в одномерном случае можно рассматривать как проекции на собственные вектора, так и на факторные — не имеет значения, поскольку собственные векторы становятся факторными при изменении длины окна с L на $N - L + 1$. А в многомерном случае это не так по построению матрицы (1.3), поэтому их нужно рассматривать по-отдельности.

Перечислим основные способы выбора векторов для проекции. Первый вариант — рассматривать собственные векторы теоретической ковариационной матрицы красного шума. Пусть $\boldsymbol{\xi} = \{\boldsymbol{\xi}^{(d)}\}_{d=1}^D$ — D -канальный красный шум с параметрами φ_d и δ_d . При рассмотрении собственных векторов матрица, разложение которой дает эти собственные векторы, имеет вид $\sum_{d=1}^D \sigma_d \{\varphi_d^{|i-j|}\}_{i,j=1}^L$, а при рассмотрении факторных векторов матрица имеет блочно-диагональный вид с блоками $\sigma_d \{\varphi_d^{|i-j|}\}_{i,j=1}^{K_d}$, где $\sigma_d = \delta_d^2 / (1 - \varphi_d^2)$. Такой вариант в обоих случаях дает точный критерий при любой длине окна.

Второй вариант — рассматривать левые или правые векторы матрицы \mathbf{X} . Этот вариант, вообще говоря, радикальный, поскольку, если рассматривать реализацию шума $\boldsymbol{\xi}$, левые и правые векторы его траекторной матрицы всегда будут разными (будут подстраиваться под конкретную реализацию), в то время как в первом варианте векторы

зависят только от его параметров. Однако, используя поправку, описанную в работе [2], можно сделать радикальный критерий точным.

Если рассматривать левые векторы, то их длина равна L , а сами векторы отображают общую структуру для всех рядов. Если же рассматривать правые, то их длина равна $D(N - L + 1)$, а сами векторы имеют составную структуру, где каждому ряду соответствует вектор длины $N - L + 1$.

Определение 6. 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. Отметим, что для точного критерия ROC-кривая совпадает с графиком мощности, так как $\alpha_I(\alpha) = \alpha$.

2.5. Численное сравнение методов

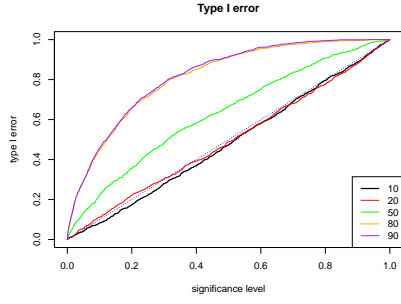
В одномерном случае было установлено, что если вместо SVD разложения матрицы \mathbf{X} использовать тёплицево, то радикальность критерия уменьшается. Установим, что будет в многомерном случае, если использовать модификации, описанные в разделе 1.2.1.

Пусть количество каналов равно двум, количество суррогатных реализаций красного шума $G = 1000$. Для оценки ошибки первого рода, будем рассматривать красный шум $\boldsymbol{\xi}$ с параметрами $\varphi = 0.7$ и $\delta = 1$, а для оценки мощности будем рассматривать временной ряд $\mathbf{X} = \mathbf{S} + \boldsymbol{\xi}$, где \mathbf{S} — сигнал с элементами

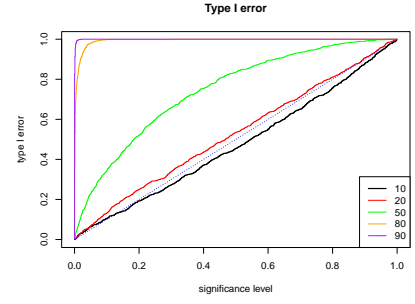
$$s_n^{(1)} = s_n^{(2)} = \cos(2\pi\omega n), \quad n = 1, \dots, N,$$

где $\omega = 0.075$, $N = 100$.

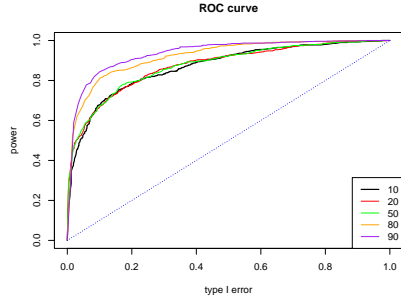
Построим графики ошибки первого рода и ROC-кривые для каждой длины окна $L = 10, 20, 50, 80, 90$. Будем воспринимать ROC-кривую как график мощности критерия, к которому была применена поправка, описанная в работе [2].



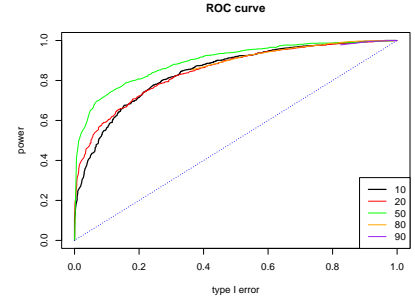
(a) Ошибка первого рода (Sum).



(б) Ошибка первого рода (базовый MSA).



(в) ROC-кривая (Sum).

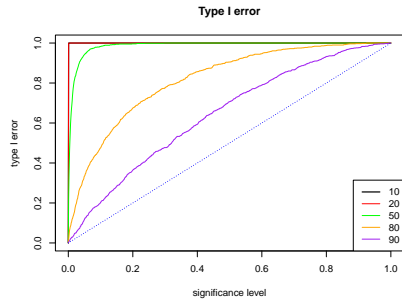


(г) ROC-кривая (базовый MSA).

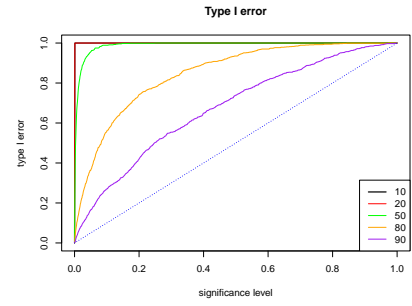
Рис. 2.1. Сравнение методов Sum и базового MSA (проекция на левые векторы).

На рис. 2.1 и 2.2 векторы для проекции были взяты из разложения (1.5). На рис. 2.1, а видно, что при $L > 20$ метод радикальный, а наибольшая мощность достигается при $L = 90$. На рис. 2.2, а отчетливо заметно, что метод радикальный для всех L . Наибольшая мощность наблюдается при $L = 90$, но отметим, что из-за слишком большой ошибки первого рода построить ROC-кривую на промежутке $[0,3)$ для $L = 50$ и на всем промежутке для $L = 10$ и $L = 20$ не получилось.

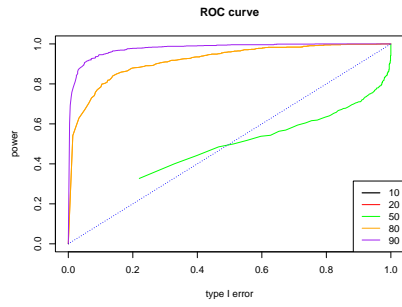
На рис. 2.3 и 2.4 векторы для проекции были взяты из разложения (1.6). Если рассматривать проекцию на левые векторы, то на рис. 2.3, а видно, что метод радикальный, а наибольшая мощность достигается при $L = 20$. Проекция на правые векторы также дает радикальный критерий, как видно на рис. 2.4, а. Наибольшая мощность наблюдается при $L = 80$, но из-за слишком большой ошибки первого рода ROC-кривую для $L = 10$ и $L = 20$, для которых метод, предположительно, имеет большую мощность, удалось построить не на всем промежутке.



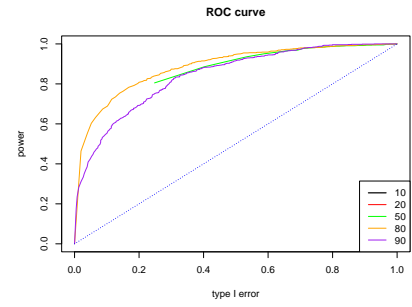
(a) Ошибка первого рода (Sum).



(б) Ошибка первого рода (базовый MSA).

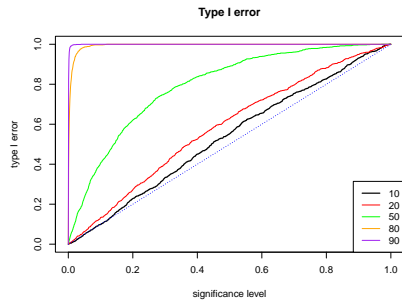


(в) ROC-кривая (Sum).

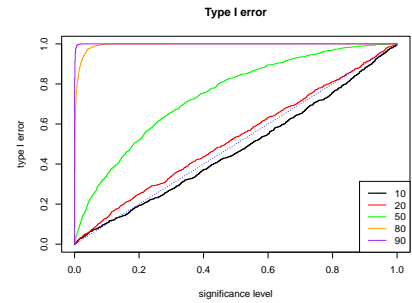


(г) ROC-кривая (базовый MSA).

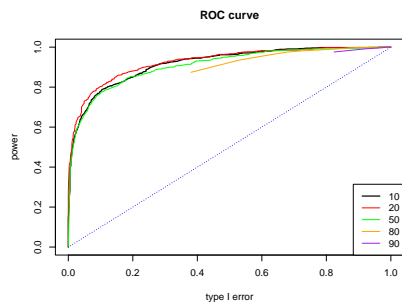
Рис. 2.2. Сравнение методов Sum и базового MSA (проекция на правые векторы).



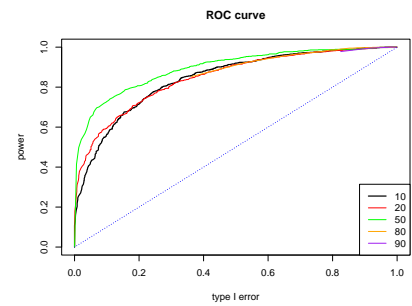
(a) Ошибка первого рода (Block).



(б) Ошибка первого рода (базовый MSA).

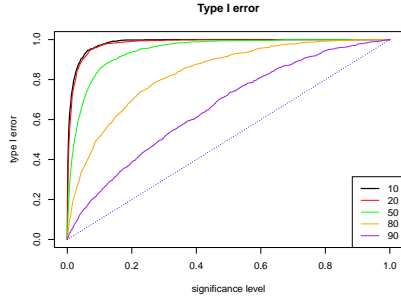


(в) ROC-кривая (Block).

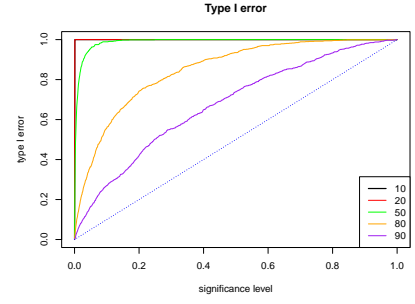


(г) ROC-кривая (базовый MSA).

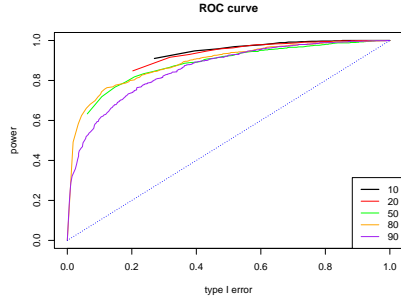
Рис. 2.3. Сравнение методов Block и базового MSA (проекция на левые векторы).



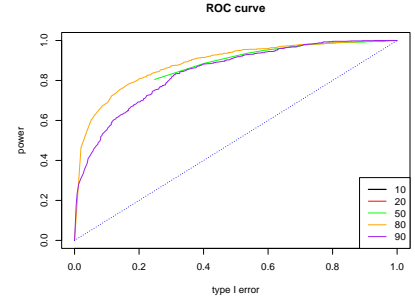
(a) Ошибка первого рода (Block).



(б) Ошибка первого рода (базовый MSSA).



(в) ROC-кривая (Block).



(г) ROC-кривая (Block).

Рис. 2.4. Сравнение методов Block и базового MSSA (проекция на правые векторы).

2.6. Выводы

Подведем итоги. На данный момент для метода Sum оптимальной длиной окна является $L = 90$, если рассматривать проекцию как на левые, так и на правые векторы. Для метода Block оптимальной длиной окна является $L = 20$, если рассматривать проекцию на левые векторы, и $L = 80$, если рассматривать проекцию на правые векторы.

Также все методы, кроме Sum, с проекцией на левые вектора сильно радикальные. Поэтому рекомендуется использовать именно этот вариант модификации.

Заключение

В ходе данной работы для реализации двух методов Тёплицева MSSA был использован язык программирования R. Было получено, что в точности восстановления сигнала оба метода в большинстве случаев показывают лучший результат, чем обычный MSSA. Но в Monte-Carlo SSA метод Sum более предпочтителен, чем метод Block, что важно ввиду его простоты в реализации и структуры, подходящей под пакет Rssa [3].

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

Список литературы

1. 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.
2. Ларин Е. С. Метод SSA для проверки гипотезы о существовании сигнала во временном ряде : квалификационная работа магистра ; СПбГУ. — 2022.
3. Rssa: A Collection of Methods for Singular Spectrum Analysis. — R package version 1.0.5. Access mode: <https://CRAN.R-project.org/package=Rssa>.
4. Голяндина Н. Э. Метод «Гусеница»-SSA: анализ временных рядов. — СПбГУ, 2004. — Учебное пособие.
5. 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.