

Санкт-Петербургский государственный университет
Прикладная математика и информатика

Отчет по учебной практике 4 (научно-исследовательской работе) (семестр 7)

МЕТОД МОНТЕ-КАРЛО SSA ДЛЯ МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

Выполнил:

Потешкин Егор Павлович

группа 20.Б04-мм

Научный руководитель:

д. 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
Глава 2. Метод Toeplitz MSSA	7
2.1. Описание метода	7
2.2. Выбор длины окна (численное исследование)	8
Глава 3. Метод Monte-Carlo MSSA	10
3.1. Постановка задачи	10
3.2. Одиночный тест	10
3.3. Множественный тест	11
3.4. Численное сравнение методов	13
3.5. Выводы	13
3.6. Оценка параметров красного шума	16
3.7. Случай известного сигнала	17
Заключение	25
Список литературы	26

Введение

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).

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

Глава 2

Метод Toeplitz MSSA

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

Пусть $\mathbf{X} = \{\mathbf{X}^{(d)}\}_{d=1}^D$ — D -канальный временной ряд с длинами $N_d = N$. Зафиксируем $1 < M < N$.

Определение 4. Матрица $\mathbf{T}_{l,k}^{(M)} \in \mathbb{R}^{M \times M}$ с элементами

$$\left(\mathbf{T}_{l,k}^{(M)}\right)_{ij} = \frac{1}{N - |i - j|} \sum_{n=1}^{N-|i-j|} x_n^{(l)} x_{n+|i-j|}^{(k)}, \quad 1 \leq i, j \leq M,$$

является оценкой ковариационной матрицы l и k -го каналов.

2.1. Описание метода

Существует два варианта Toeplitz MSSA: Block \parallel и Sum.

1. Рассмотрим блочную матрицу

$$\mathbf{T}_{\text{Block}} = \begin{pmatrix} \mathbf{T}_{1,1}^{(K)} & \mathbf{T}_{1,2}^{(K)} & \cdots & \mathbf{T}_{1,D}^{(K)} \\ \mathbf{T}_{2,1}^{(K)} & \mathbf{T}_{2,2}^{(K)} & \cdots & \mathbf{T}_{2,D}^{(K)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_{D,1}^{(K)} & \mathbf{T}_{D,2}^{(K)} & \cdots & \mathbf{T}_{D,D}^{(K)} \end{pmatrix} \in \mathbb{R}^{DK \times DK},$$

где $K = N - L + 1$. Найдя ортонормированные собственные векторы Q_1, \dots, Q_{DK} матрицы \mathbf{T} , получаем разложение траекторной матрицы \mathbf{X} :

$$\mathbf{X} = \sum_{i=1}^{DK} \sigma_i P_i Q_i^T = \mathbf{X}_1 + \dots + \mathbf{X}_{DK}, \quad (2.1)$$

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

2. Рассмотрим матрицу $\mathbf{T}_{\text{Sum}} = \sum_{d=1}^D \mathbf{T}_{d,d}^{(L)} \in \mathbb{R}^{L \times L}$. Найдём ортонормированные собственные векторы P_1, \dots, P_L матрицы \mathbf{T}_{Sum} и разложим траекторную матрицу \mathbf{X} следующим образом:

$$\mathbf{X} = \sum_{i=1}^L \sigma_i P_i Q_i^T = \mathbf{X}_1 + \dots + \mathbf{X}_L, \quad (2.2)$$

где $S_i = \mathbf{X}^T P_i$, $Q_i = S_i / \|S_i\|$, $\sigma_i = \|S_i\|$.

Замечание 2. 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.

Замечание 3. Toeplitz Sum MSSA можно использовать для временных рядов с разными длинами каналов, в отличие от Toeplitz Block MSSA.

2.2. Выбор длины окна (численное исследование)

Посмотрим на точность базового и модифицированных методов для разных значений параметра L . Рассмотрим следующий двухканальный временной ряд длины $N = 71$:

$$\{\mathbf{F}^{(1)}, \mathbf{F}^{(2)}\} = \{\mathbf{S}^{(1)}, \mathbf{S}^{(2)}\} + \{\mathbf{N}^{(1)}, \mathbf{N}^{(2)}\},$$

где $\mathbf{S}^{(1)}, \mathbf{S}^{(2)}$ — некоторые сигналы, а $\mathbf{N}^{(1)}, \mathbf{N}^{(2)}$ — независимые реализации гауссовского белого шума с $\sigma = 5$. Рассмотрим 3 случая, первые два из которых рассматривались ранее в [5]:

1. Косинусы с одинаковыми частотами:

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

2. Косинусы с разными частотами:

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

3. Полиномы первой степени (нестационарные ряды):

$$s_n^{(1)} = 1.2n, \quad s_n^{(2)} = 0.8n, \quad n = 1, \dots, N.$$

В качестве оценки точности восстановления сигнала было взято среднеквадратичное отклонение от истинного значения. В таблице 2.1 представлены результаты на основе 10000 реализаций шума. Наиболее точные результаты для каждого метода были выделены жирным шрифтом. Лучший результат для каждого случая выделен отдельно синим.

Как видно из таблицы 2.1, в первом случае метод Block лучше всего выделял сигнал. В случае разных частот каналы имеют разную структуру, поэтому наиболее оптимальным является использовать Toeplitz SSA для каждого канала по отдельности. В третьем случае мы имеем дело с нестационарными рядами одинаковой структуры, поэтому стандартный MSSA справляется лучше всего.

Таблица 2.1. MSE восстановления сигнала.

Случай 1 ($\omega_1 = \omega_2$)	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
SSA	3.25	2.01	2.00	2.01	3.25
Toeplitz SSA	3.2	1.87	1.63	1.59	1.67
MSSA	3.18	1.83	1.59	1.47	2.00
Toeplitz Sum MSSA	3.17	1.75	1.44	1.32	1.33
Toeplitz Block MSSA	1.39	1.26	1.25	1.33	1.97
Случай 2 ($\omega_1 \neq \omega_2$)	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
SSA	3.25	2.01	2.00	2.01	3.25
Toeplitz SSA	3.2	1.87	1.63	1.59	1.67
MSSA	6.91	3.77	3.07	2.88	3.84
Toeplitz Sum MSSA	6.88	3.65	2.64	2.37	2.27
Toeplitz Block MSSA	4.47	3.67	3.22	3.23	3.8
Случай 3 (тренд)	$L = 12$	$L = 24$	$L = 36$	$L = 48$	$L = 60$
SSA	3.65	2.08	1.96	2.08	3.65
Toeplitz SSA	3.33	2.43	3.74	7.84	16.29
MSSA	3.42	1.94	1.63	1.57	2.27
Toeplitz Sum MSSA	3.32	2.24	3.04	5.91	11.95
Toeplitz Block MSSA	12.55	6.18	2.97	1.78	1.97

Глава 3

Метод Monte-Carlo MSSA

3.1. 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.

3.2. 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 $\boldsymbol{\Xi}$. Рассмотрим вектор $W \in \mathbb{R}^L$ такой, что $\|W\| = 1$. Введем величину

$$p = \|\boldsymbol{\Xi}^T W\|^2.$$

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 $1 - \alpha$.

Замечание 4. 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.

3.3. 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

$$p_{ki} = \|\Xi_i^T W_k\|^2, \quad i = 1, \dots, G; \quad k = 1, \dots, H, \quad (3.1)$$

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.

В данной работе в качестве W_1, \dots, W_H , если не сказано иначе, берутся собственные векторы матрицы \mathbf{X} . Такой способ выбора векторов для проекции самый распространенный, поскольку, если есть значимые векторы, можно восстановить сигнал с помощью SSA на их основе. Но этот вариант, вообще говоря, дает радикальный критерий. Однако, используя поправку, описанную в работе [2], можно сделать радикальный критерий точным.

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 $\alpha_I(\alpha) = \alpha$.

3.4. Численное сравнение методов

В одномерном случае было установлено [2], 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 главе 2.

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 $\omega = 0.075$, $N = 100$.

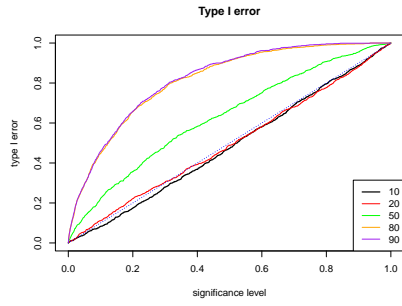
Построим графики ошибки первого рода и ROC-кривые для каждой длины окна $L = 10, 20, 50, 80, 90$. Будем воспринимать ROC-кривую как график мощности критерия, к которому была применена поправка, описанная в работе [2].

На рис. 3.1 и 3.2 векторы для проекции были взяты из разложения (2.2). На рис. 3.1, *a* видно, что при $L > 20$ метод радикальный, а наибольшая мощность достигается при $L = 90$. На рис. 3.2, *a* 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.

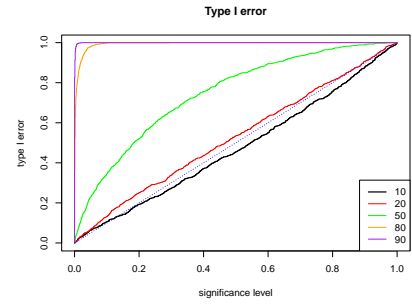
На рис. 3.3 и 3.4 векторы для проекции были взяты из разложения (2.1). Если рассматривать проекцию на левые векторы, то на рис. 3.3, *a* видно, что метод радикальный, а наибольшая мощность достигается при $L = 20$. Проекция на правые векторы также дает радикальный критерий, как видно на рис. 3.4, *a*. 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 бóльшую мощность, удалось построить не на всем промежутке.

3.5. Выводы

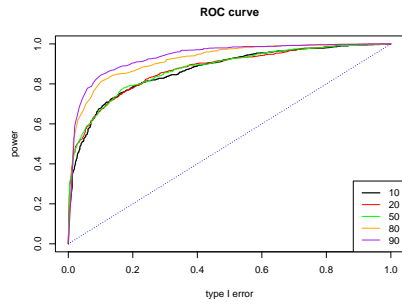
Подведем итоги. На данный момент для метода Sum оптимальной длиной окна является $L = 90$, если рассматривать проекцию как на левые, так и на правые векторы.



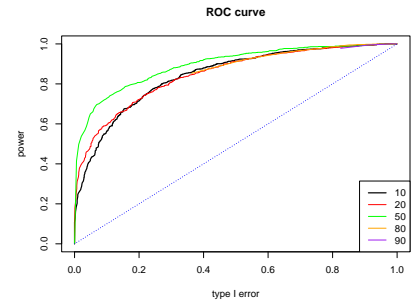
(a) Ошибка первого рода (Sum).



(б) Ошибка первого рода (базовый MSA).

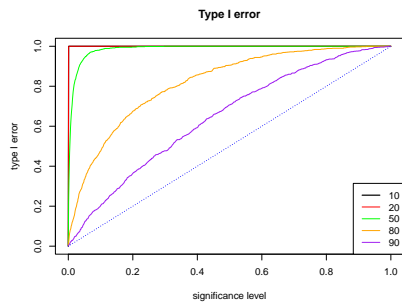


(в) ROC-кривая (Sum).

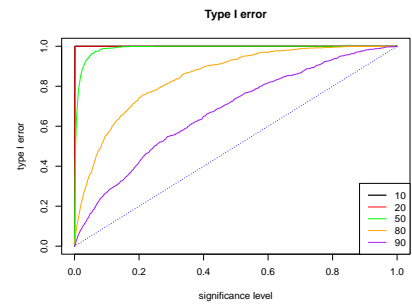


(г) ROC-кривая (базовый MSA).

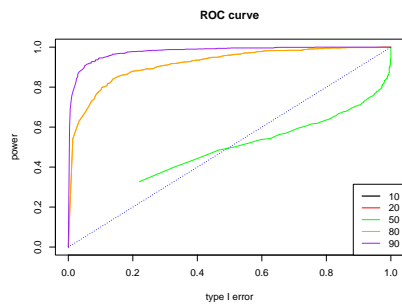
Рис. 3.1. Сравнение методов Sum и базового MSA (проекция на левые векторы).



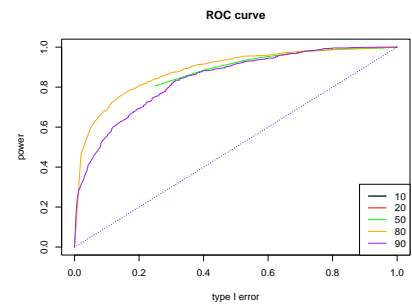
(a) Ошибка первого рода (Sum).



(б) Ошибка первого рода (базовый MSA).

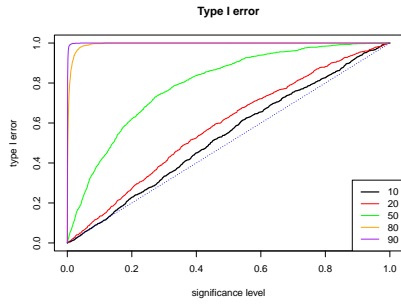


(в) ROC-кривая (Sum).

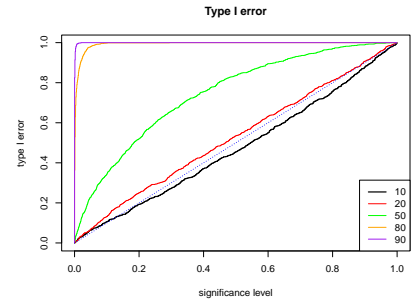


(г) ROC-кривая (базовый MSA).

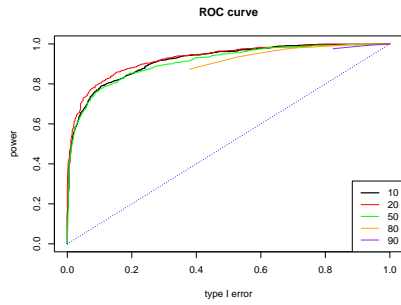
Рис. 3.2. Сравнение методов Sum и базового MSA (проекция на правые векторы).



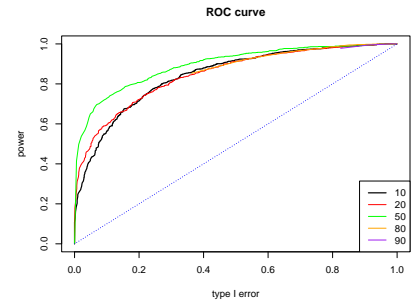
(a) Ошибка первого рода (Block).



(б) Ошибка первого рода (базовый MSSA).

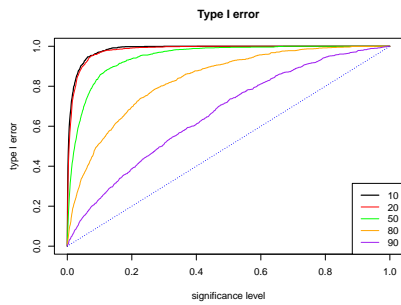


(в) ROC-кривая (Block).

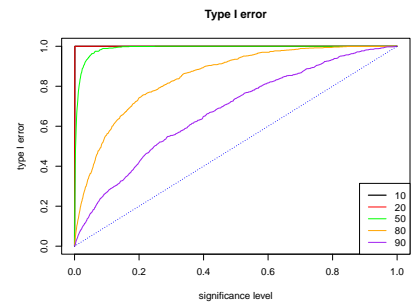


(г) ROC-кривая (базовый MSSA).

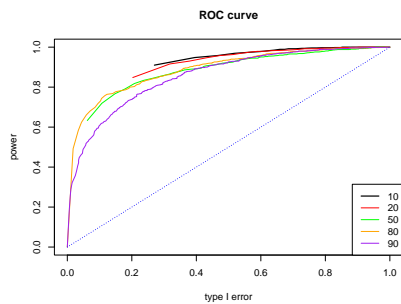
Рис. 3.3. Сравнение методов Block и базового MSSA (проекция на левые векторы).



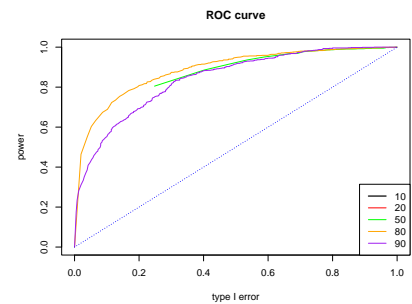
(a) Ошибка первого рода (Block).



(б) Ошибка первого рода (базовый MSSA).



(в) ROC-кривая (Block).



(г) ROC-кривая (Block).

Рис. 3.4. Сравнение методов Block и базового MSSA (проекция 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.

3.6. Оценка параметров красного шума

До сих пор мы предполагали, что параметры красного шума φ и δ известны, но в реальных задачах редко возникает такая ситуация. В этой ситуации можно воспользоваться методом bootstrapping, который позволяет использовать оцененные параметры шума для построения критерия [1]. Оценить параметры шума на основе реальных данных зная, что шум — красный можно, например, с помощью метода ARIMA [?].

В этой работе предлагается строить следующую двухступенчатую оценку параметров: сначала оцениваются параметры исходного ряда, применяется MC-SSA, затем выделяется сигнал, если он обнаружится, и оцениваются параметры «остатка».

Проверим на практике, что такой подход даст результаты лучше, чем обычная оценка параметров без выделения сигнала. За сигнал возьмем

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

с амплитудой $A = 3$ и частотой $\omega \in (0, 0.5)$. Параметры красного шума и длину ряда N возьмем такими же, как в разделе 3.4.

Будем оценивать параметр φ , обозначим оценку за $\hat{\varphi}$. В качестве оценки точности было взято среднеквадратичное отклонение от истинного значения. В таблице 3.1 представлены результаты на основе 100 реализаций шума. Поскольку $\text{MSE}\hat{\varphi} = D\hat{\varphi} + \text{bias}^2\hat{\varphi}$, в таблице также представлены значения дисперсии и смещения оценки. Для MC-SSA были взяты $G = 1000$, $L = 50$, $\alpha = 0.05$.

По таблице 3.1 видно, что основной вклад в ошибку оценки вносит смещение, описанная выше процедура это смещение сильно уменьшает, делая слабо-отрицательным.

ARIMA	$\omega = 0.1$	$\omega = 0.2$	$\omega = 0.3$	$\omega = 0.4$
$\text{MSE}\hat{\varphi}$	0.0057	0.0826	0.5045	1.1622
$D\hat{\varphi}$	0.0004	0.0015	0.0022	0.0085
$\text{bias}\hat{\varphi}$	0.0723	-0.2849	-0.7087	-1.0741
Двухступенчатая оценка	$\omega = 0.1$	$\omega = 0.2$	$\omega = 0.3$	$\omega = 0.4$
$\text{MSE}\hat{\varphi}$	0.0091	0.0068	0.0062	0.0146
$D\hat{\varphi}$	0.0079	0.0066	0.0062	0.0096
$\text{bias}\hat{\varphi}$	-0.0361	-0.0156	-0.0103	-0.0714

Таблица 3.1. Оценка параметров красного шума

3.7. Случай известного сигнала

Пусть известно, что во временном ряде присутствует некоторый сигнал, но, возможно, еще есть какой-то другой. Тогда модель выглядит следующим образом:

$$\mathbf{X} = \mathbf{S} + \mathbf{H} + \boldsymbol{\xi},$$

где \mathbf{S} — известный сигнал, который далее будем называть просто сигналом, \mathbf{H} — неизвестный сигнал и $\boldsymbol{\xi}$ — красный шум. Будем проверять следующую нулевую гипотезу с альтернативой:

$$H_0 : \mathbf{H} = 0,$$

$$H_1 : \mathbf{H} = \{\cos(2\pi\omega n)\}_{n=1}^N,$$

где $\omega = 0.075$. Будем проверять H_0 следующим образом.

Алгоритм 3. MC-SSA с известным сигналом

1. Находится приближенное значение сигнала $\hat{\mathbf{S}}$ и оцениваются параметры $\boldsymbol{\xi}$ на основе остатка $\tilde{\mathbf{X}} = \mathbf{X} - \hat{\mathbf{S}}$.
2. Находятся левые векторы P_1, \dots, P_L траекторной матрицы временного ряда $\tilde{\mathbf{X}}$, полученные из разложения 2.2.
3. Применяется MC-SSA к исходному ряду \mathbf{X} с проекцией на векторы P_1, \dots, P_L , при этом суррогатными рядами являются реализации случайной величины $\boldsymbol{\eta}$:

$$\boldsymbol{\eta} = \boldsymbol{\xi} + \hat{\mathbf{S}}.$$

3.7.1. Периодическая компонента

Рассмотрим в качестве сигнала модулированную синусоиду:

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

где $A = 3$, $\omega = 0.25$.

Будем смотреть на графики ошибки первого рода, мощности и ROC-кривые этого критерия для разных длин окна, как в разделе 3.4. Начнем с ситуации, когда сигнал и параметры шума известны точно, постепенно приходя к случаю, когда и сигнал, и параметры шума нужно оценивать.

На рис. 3.5 представлены графики ошибки первого рода, мощности и ROC-кривые критерия, когда сигнал и параметры шума известны точно. График ошибки первого рода на рис. 3.5, *а* показывает, что при длине окна $L = 10$ критерий консервативный, при $L = 20$ — приблизительно точный, при остальных L — радикальный. Таким образом, имеет смысл рассматривать мощность на рис. 3.5, *б* только для $L = 10, 20$. ROC-кривая на рис. 3.5, *в* похожа на график мощности при $L = 20$, что соответствует теории. После применения поправки критерий становится точным для любой длины окна, при этом ROC-кривая будет графиком мощности поправленного критерия. Наибольшая мощность достигается при длине окна $L = 90$.

На рис. 3.6 представлены графики ошибки первого рода, мощности и ROC-кривые критерия, когда сигнал известен точно, но параметры шума оцениваются. График ошибки первого рода на рис. 3.6, *а* показывает, что при длине окна $L = 10$ при $\alpha < 0.82$ и $L = 20$ при $\alpha < 0.58$ критерий консервативный, в иных случаях — радикальный. Таким образом, имеет смысл рассматривать мощность на рис. 3.6, *б* только для $L = 10$ при $\alpha < 0.82$ и $L = 20$ при $\alpha < 0.58$. Наибольшая мощность поправленного критерия достигается при длине окна $L = 90$. Отклонение от случая с известным сигналом и параметрами шума можно объяснить погрешностями при оценке параметров.

На рис. 3.7 представлены графики ошибки первого рода, мощности и ROC-кривые критерия, когда оценивается и сигнал, и параметры шума. Поскольку неопределенности стало еще больше, отклонение от случая с известным сигналом и параметрами шума стало еще сильнее. График ошибки первого рода на рис. 3.7, *а* показывает, что при длине окна $L = 10$ при $\alpha < 0.67$ и $L = 20$ при $\alpha < 0.2$ критерий консервативный, в иных случаях — радикальный. Наибольшая мощность поправленного критерия достигается

при длине окна $L = 90$, как и в предыдущих случаях.

3.7.2. Тренд

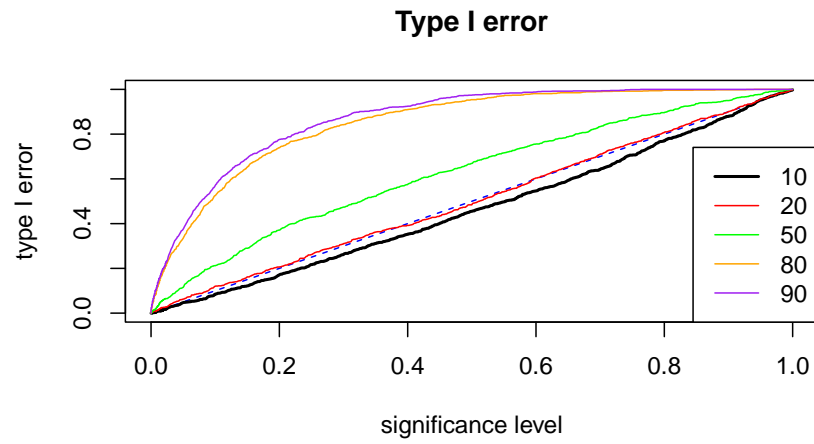
Отдельно рассмотрим вариант, когда сигнал — тренд, т.е. медленно меняющаяся компонента. Рассмотрим следующий экспоненциальный ряд:

$$s_n = Ae^{\alpha n}, \quad n = 1, \dots, N,$$

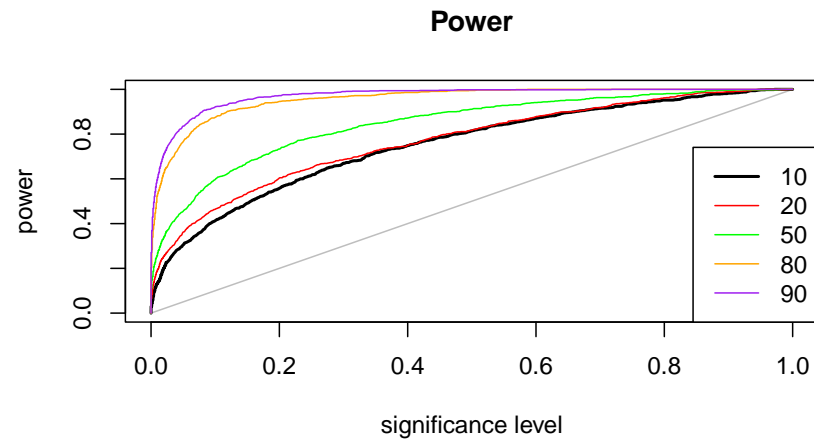
где $A = 0.2$, $\alpha = 0.05$.

На рис. 3.8 представлены графики ошибки первого рода, мощности и ROC-кривые критерия, когда сигнал и параметры шума известны точно. График ошибки первого рода на рис. 3.8, *а* показывает, что при $L = 10$ критерий консервативен, при $L = 20$ — приблизительно точный (слабо радикальный), для остальных L — радикальный. Поэтому график мощности на рис. 3.8, *б* похож на ROC-кривую на рис. 3.8, *в* при $L = 20$. Наибольшая мощность достигается при длине окна $L = 90$.

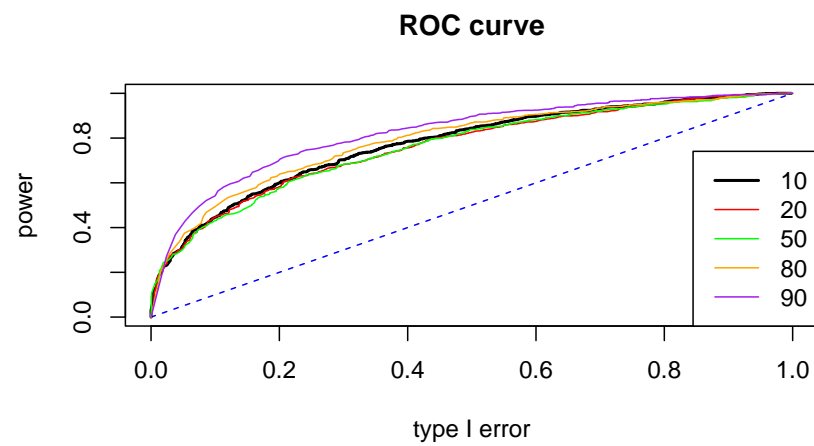
На рис. 3.9 представлены графики ошибки первого рода, мощности и ROC-кривые критерия, когда сигнал известен точно, но параметры шума оцениваются. По графику ошибки первого рода на рис. 3.9, *а* видно, что при длине окна $L = 10$ при $\alpha < 0.7$ и $L = 20$ при $\alpha < 0.43$ критерий консервативный, в остальных случаях — радикальный. Как и в случае периодического сигнала, отклонение от предыдущего случая связано с погрешностями оценок параметров шума. По графику ROC-кривой на рис. 3.9, *в* наибольшая мощность поправленного критерия наблюдается при длине окна $L = 90$.



(a) Ошибка первого рода

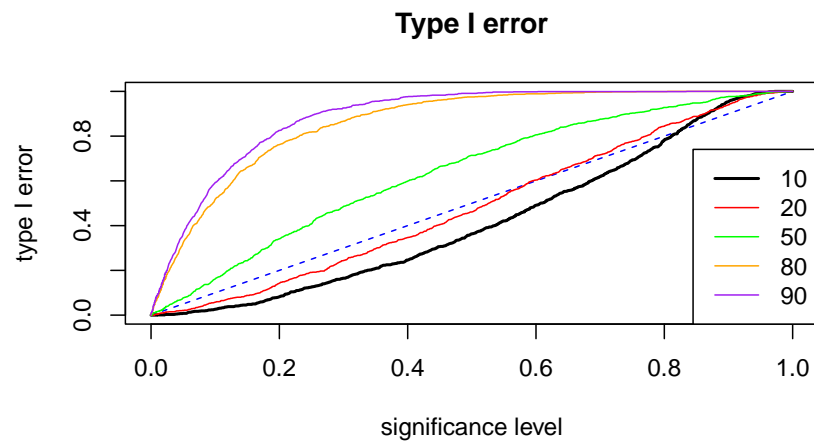


(б) Мощность

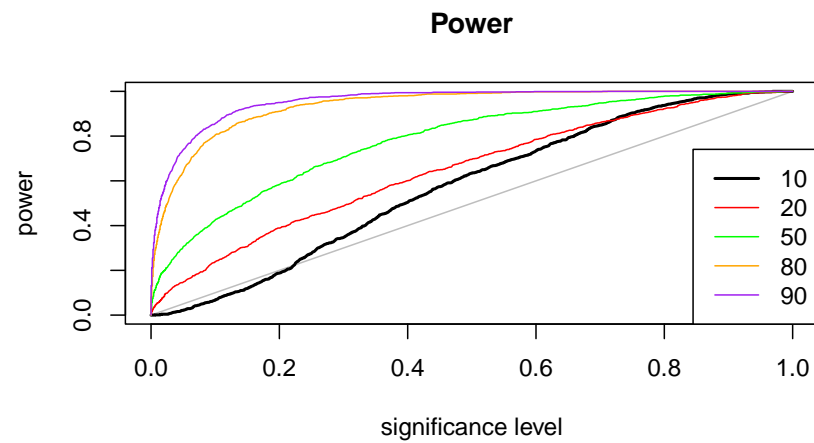


(в) ROC-кривая

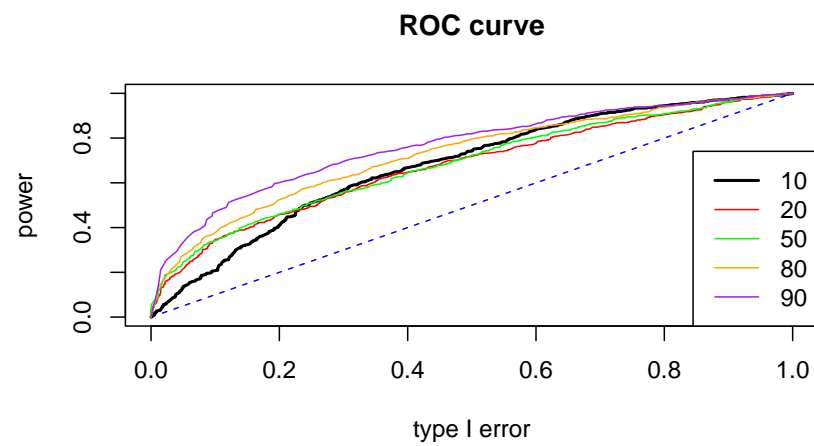
Рис. 3.5. Анализ метода, случай периодической компоненты (сигнал и параметры красного шума известны точно)



(a) Ошибка первого рода

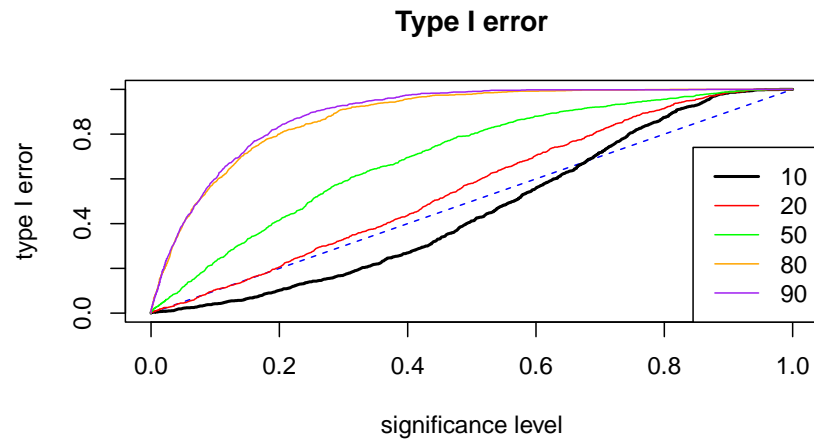


(б) Мощность

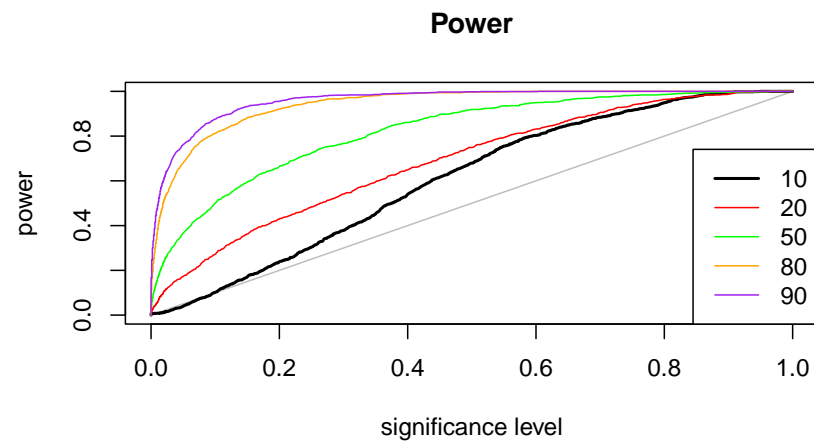


(в) ROC-кривая

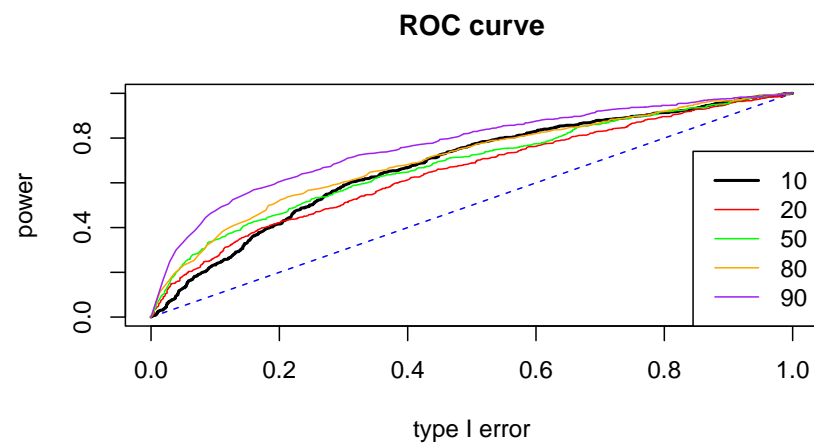
Рис. 3.6. Анализ метода, случай периодической компоненты (оцененные параметры красного шума)



(a) Ошибка первого рода

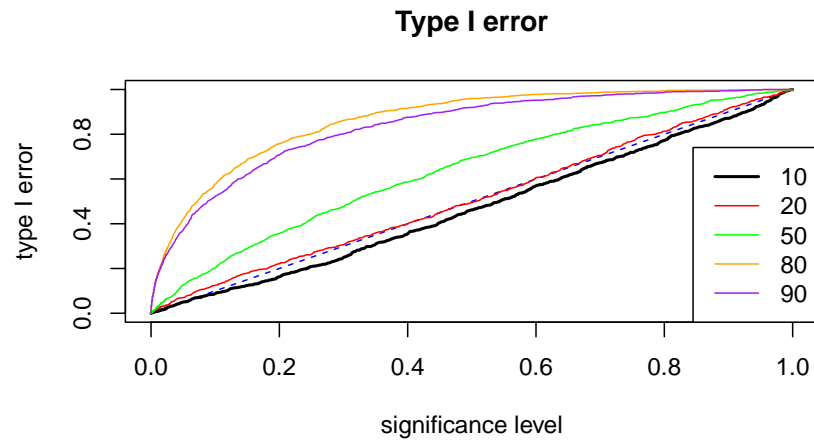


(б) Мощность

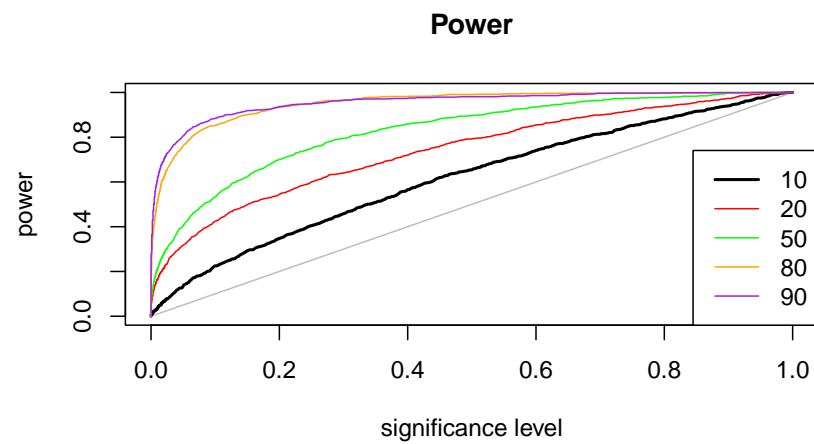


(в) ROC-кривая

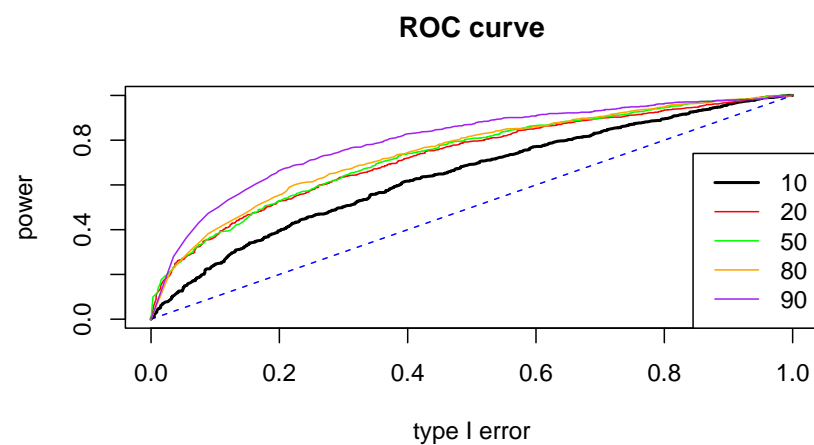
Рис. 3.7. Анализ метода, случай периодической компоненты (оцененный сигнал и параметры красного шума)



(a) Ошибка первого рода

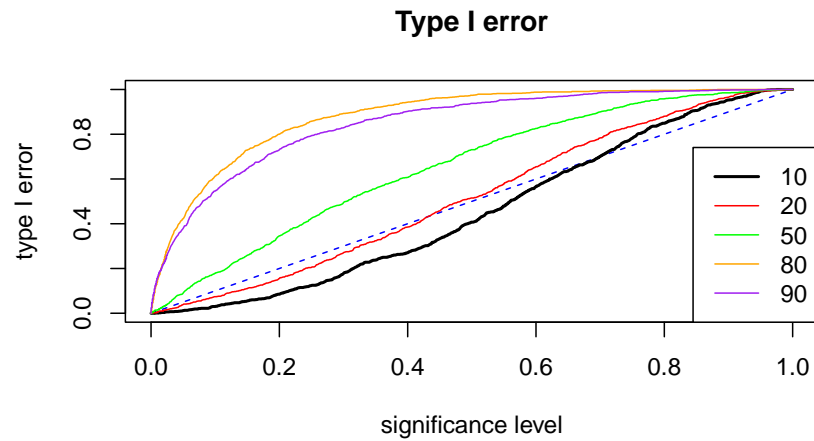


(б) Мощность

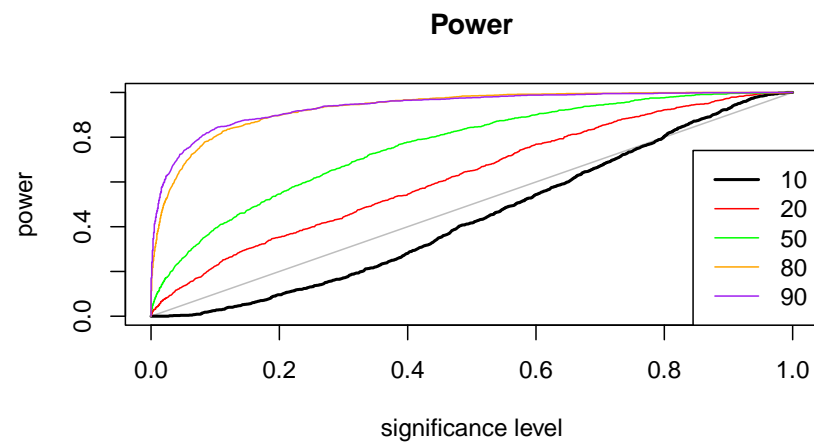


(в) ROC-кривая

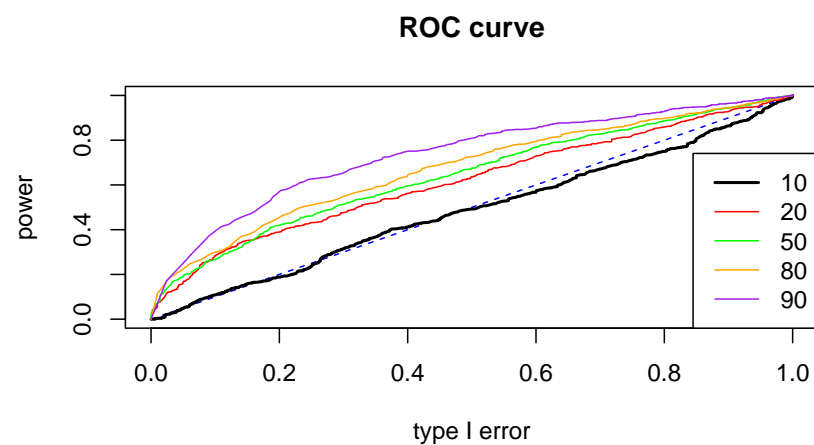
Рис. 3.8. Анализ метода, случай тренда (сигнал и параметры красного шума известны точно)



(a) Ошибка первого рода



(б) Мощность



(в) ROC-кривая

Рис. 3.9. Анализ метода, случай тренда (оцененные параметры красного шума)

Заключение

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.