

Поддерживающие временные ряды в MSSA

Федоров Никита Алексеевич, гр. 14.Б02-мм

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

Научный руководитель: к. ф.-м. н. доц. Голяндина Нина
Эдуардовна

Рецензент: м. н. с. Шлемов Александр Юрьевич



Санкт-Петербург
2018г.

Временной ряд

Вещественный временной ряд длины N : $F = (f_0, \dots, f_{N-1})$, $f_i \in \mathbb{R}$.

Пусть имеется временной ряд $F_1 = S_1 + R_1$ длины N , где

- 1 Сигнал S_1 – детерминированная составляющая ряда.
- 2 Шум R_1 – случайная (мешающая) составляющая ряда.

Задача: выделить сигнал S_1 .

Пусть помимо ряда F_1 имеется временной ряд F_2 длины N .

Идея: использование ряда F_2 при выделении сигнала S_1 может помочь выделить сигнал более точно.

В работе используется:

- Для анализа одного ряда – метод SSA.
- Для анализа двух рядов – метод MSSA.

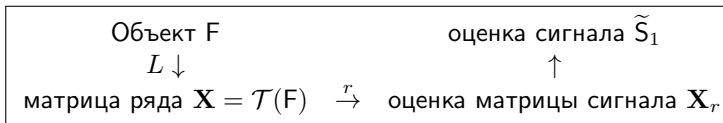
Подробное описание теории методов SSA и MSSA — в работе [Golyandina et al., 2015].

Схема получения оценок сигнала методами SSA и MSSA

Вход:

- Ряд $F_1 = S_1 + R_1$ длины N , длина окна $1 \leq L \leq N$, r
- (MSSA) Ряд F_2 длины N

Схема:



Рекомендации (SSA): $L \sim N/2$, $r = \text{rank } \mathcal{T}(S_1)$

SSA: $F = F_1$, $\mathbf{X} = \mathcal{T}(F_1) = \begin{pmatrix} f_0^{(1)} & f_1^{(1)} & \cdots & f_{K-1}^{(1)} \\ f_1^{(1)} & f_2^{(1)} & \ddots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ f_{L-1}^{(1)} & f_L^{(1)} & \cdots & f_{N-1}^{(1)} \end{pmatrix}$

MSSA: $F = (F_1; F_2)$, $\mathbf{X} = [\mathcal{T}(F_1); \mathcal{T}(F_2)] = [\mathbf{X}_1; \mathbf{X}_2]$

Выход:

- \tilde{S}_{SSA} - оценка S_1 методом SSA
- \tilde{S}_{MSSA} - оценка S_1 методом MSSA

Модель: $F_1 = S_1 + R_1$

$\tilde{S}_{SSA}, \tilde{S}_{MSSA}$ – оценки S_1

Ошибка оценки \tilde{S}_1 сигнала S_1 : $MSE(\tilde{S}; S_1) = \frac{1}{N} \sum_{i=0}^{N-1} (\tilde{s}_i - s_i)^2$,

где $S_1 = (s_0, \dots, s_{N-1})$, а $\tilde{S} = (\tilde{s}_0, \dots, \tilde{s}_{N-1})$

- $MSE_{SSA} = MSE(\tilde{S}_{SSA}, S_1)$ – ошибка оценки сигнала методом SSA
- $MSE_{MSSA} = MSE(\tilde{S}_{MSSA}, S_1)$ – ошибка оценки сигнала методом MSSA

Определение

F_2 поддерживающий для F_1 : $MSE_{MSSA} < MSE_{SSA}$

Модель:

- $F_1 = S_1 + R_1, F_2 = S_2 + R_2$
- R_1, R_2 – два независимых белых гауссовских шума со средними, равными 0, и дисперсиями σ_1^2, σ_2^2 , соответственно.
- $r = r_{\text{SSA}}$ для \tilde{S}_{SSA} , $r = r_{\text{MSSA}}$ для \tilde{S}_{MSSA}

Рассматриваем случаи:

- Константных сигналов: $s_j^{(i)} \equiv c_i$ – константы
- Гармонических сигналов: $s_j^{(i)} = A_i \cos(2\pi(j+1)/T_i)$

Для гармонических сигналов:

- Сигналы одинаковой структуры:
 $T_1 = T_2 \neq 2 \Rightarrow r_{\text{SSA}} = r_{\text{MSSA}} = 2$
- Сигналы разной структуры: $T_1 \neq T_2, T_1, T_2 \neq 2 \Rightarrow r_{\text{SSA}} = 2,$
 $r_{\text{MSSA}} = 4$

Для константных сигналов: сигналы всегда одинаковой структуры:

$$r_{\text{SSA}} = r_{\text{MSSA}} = 1$$

Задачи:

- ① **Исследование структуры:** определить, как структура рядов F_1, F_2 влияет на то, является ли F_2 поддерживающим для F_1 .
- ② **Случай неизвестного сигнала:** выяснить, как, не зная сигнала S_1 , определить, является ли F_2 поддерживающим для F_1 .

Схема исследования:

- ① Исследование структуры:
 - ① Теоретические результаты для константных сигналов
 - ② Численные эксперименты для гармонических сигналов
 - ③ Задача масштабирования F_2 для константных сигналов – аналитически
- ② Случай неизвестного сигнала:
 - ① Идея алгоритма CHOICE определения поддерживающих рядов
 - ② Исследование качества алгоритма CHOICE
 - ③ Приведение рекомендаций по выбору параметров алгоритма CHOICE

Исследование структуры: теоретические результаты

Результаты для константных сигналов:

- c_1, c_2 – значения констант
- σ_1, σ_2 – значения уровня шума
- $\mathbb{E} \text{MSE}_{\text{SSA}}, \mathbb{E} \text{MSE}_{\text{MSSA}}$ – матожидания ошибок оценки сигнала

Утверждение

Для первого порядка $\mathbb{E} \text{MSE}_{\text{SSA}}$ при $L = \left\lceil \frac{N+1}{2} \right\rceil$ верно:

$$\mathbb{E} \text{MSE}_{\text{SSA}}^{(1)} \sim \frac{26\sigma_1^2}{9N} \text{ при } N \rightarrow \infty.$$

Утверждение

Пусть $a = \frac{c_2}{c_1}$ и $b = \frac{\sigma_2}{\sigma_1}$. Для первого порядка $\mathbb{E} \text{MSE}_{\text{MSSA}}$ при

$$L = \left\lceil \frac{N+1}{2} \right\rceil \text{ верно: } \mathbb{E} \text{MSE}_{\text{MSSA}}^{(1)} \sim \frac{\sigma_1^2}{9N} \cdot \frac{15a^4 + 4a^2b^2 + 37a^2 + 26}{a^4 + 2a^2 + 1} \text{ при } N \rightarrow \infty.$$

Исследование структуры: теоретические результаты

Результаты для константных сигналов:

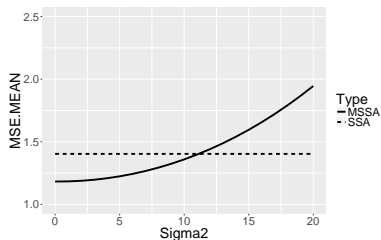
- $\tilde{s}_l^{(1)}$ – первый порядок восстановления сигнала S_1 в точке на l -ой позиции
- $\mathbb{D}\tilde{s}_l^{(1)}$ – дисперсия $\tilde{s}_l^{(1)}$

Утверждение

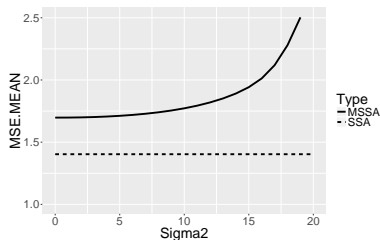
Пусть $L - 1 \leq l \leq K - 1$. При $L < \frac{N}{2}$ $\mathbb{D}\tilde{s}_l^{(1)}$ не зависит от F_2 и задаётся выражением $\mathbb{D}\tilde{s}_l^{(1)} = \frac{\sigma_1^2(1+2L^2)}{3L^3}$.

Следствие: наибольшее влияние F_2 на MSE_{MSSA} при $L = \left\lceil \frac{N+1}{2} \right\rceil$.

Исследование структуры: гармонические сигналы



(a) Сигналы одинаковой структуры ($T_1 = T_2 = 10$)



(b) Сигналы разной структуры ($T_1 = 10, T_2 = 7$)

Рис. 1: Зависимость $\hat{\mathbb{E}} \text{MSE}_{\text{MSSA}}$ от σ_2 и сравнение с $\hat{\mathbb{E}} \text{MSE}_{\text{SSA}}$ для гармонических сигналов ($A_1 = 30, A_2 = 20, \sigma_1 = 5, N = 100$)

Пороговое значение: для сигналов одинаковой структуры $\exists \sigma_2^{(0)}$:

- $\sigma_2 < \sigma_2^{(0)} \Rightarrow \hat{\mathbb{E}} \text{MSE}_{\text{MSSA}} < \hat{\mathbb{E}} \text{MSE}_{\text{SSA}}$
- $\sigma_2 > \sigma_2^{(0)} \Rightarrow \hat{\mathbb{E}} \text{MSE}_{\text{MSSA}} > \hat{\mathbb{E}} \text{MSE}_{\text{SSA}}$

Идея: $F_2 \rightarrow \alpha F_2$, $\alpha > 0$.

Введём: $U(\alpha) = \mathbb{E} \text{MSE}_{\text{MSSA}} / \mathbb{E} \text{MSE}_{\text{SSA}}$

Утверждение

Опишем поведение $U(\alpha)$ в зависимости от $\text{SNR}(F_1)/\text{SNR}(F_2)$:

- ❶ При $\text{SNR}(F_1)/\text{SNR}(F_2) < \frac{7}{8}$ $U(\alpha)$ убывает на $(0, +\infty)$. При всех значениях α ряд αF_2 в среднем является поддерживающим для F_1 .
- ❷ При $\frac{7}{8} \leq \text{SNR}(F_1)/\text{SNR}(F_2)$ $U(\alpha)$, убывая, достигает своего минимума в точке α_0 , после чего, возрастая, сходится к некоторому значению $U_{+\infty}$:
 - В случае $\text{SNR}(F_1)/\text{SNR}(F_2) < \frac{11}{4}$ значение $U_{+\infty} < 1$ и ряд αF_2 также в среднем остаётся поддерживающим для F_1 при всех значениях α .
 - В случае $\text{SNR}(F_1)/\text{SNR}(F_2) \geq \frac{11}{4}$ значение $U_{+\infty} \geq 1$ и ряд αF_2 в среднем уже не будет поддерживающим для F_1 при достаточно больших значениях α .

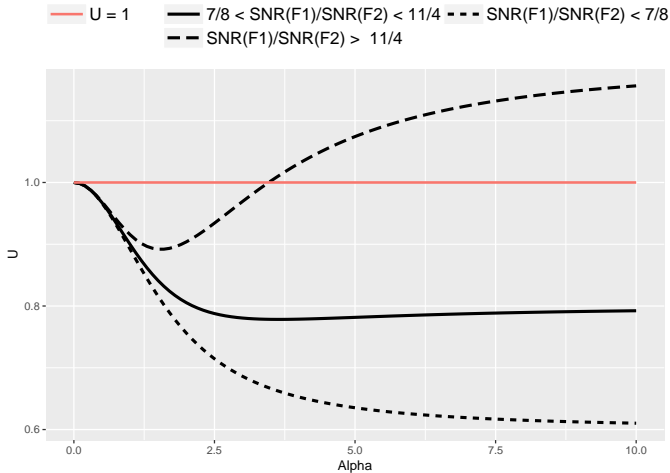


Рис. 2: Зависимость поведения отношения ошибок оценки сигнала $U(\alpha)$ от масштабирующего параметра α

Случай неизвестного сигнала: идея

По определению: $MSE_{SSA} \leq MSE_{MSSA} \rightarrow$ Ряд F_2
(не) поддерживающий для F_1 .

Проблема: в реальных задачах сигнал неизвестен \Leftrightarrow не вычислить MSE_{SSA} , MSE_{MSSA} .

Идея:

- 1 Делить ряд на две части
- 2 Строить прогноз одной части ряда по другой
- 3 Вычислять γ_{SSA} , γ_{MSSA} – ошибки прогноза
- 4 $\gamma_{SSA} \leq \gamma_{MSSA} \rightarrow$ Решение: ряд F_2 (не) поддерживающий для F_1

Алгоритм: CHOICE($k, r_{SSA}, r_{MSSA}, L = L(k)$)

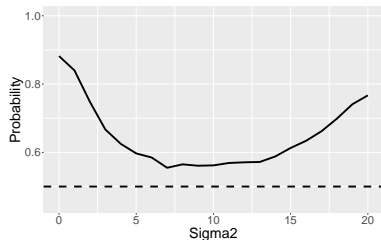
- k – длина отрезка, по которому строится прогноз
- r_{SSA} , r_{MSSA} – значения параметра r для SSA и MSSA прогнозов
- L – значение параметра длины окна для SSA и MSSA прогнозов

Для исследования качества:

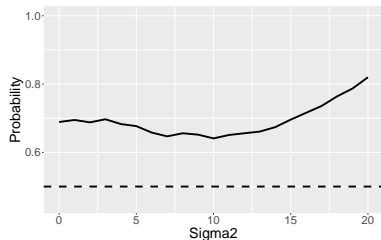
$\mathbb{P}_{\text{CHOICE}} = \mathbb{P}(\mathbb{1}_{\{\gamma_{MSSA} < \gamma_{SSA}\}} = \mathbb{1}_{\{MSE_{MSSA} < MSE_{SSA}\}})$ – вероятность правильного определения, является ли ряд поддерживающим

Случай неизвестного сигнала: качество алгоритма для гармонических сигналов

Пусть известна структура двумерного сигнала $(S_1; S_2)$. При всех значениях σ_2 $\hat{P}_{\text{CHOICE}} > 0.5$.



(a) Сигналы одинаковой структуры ($T_1 = T_2 = 10, r_{\text{SSA}} = r_{\text{MSSA}} = 2$)



(b) Сигналы разной структуры ($T_1 = 10, T_2 = 7, r_{\text{SSA}} = 2, r_{\text{MSSA}} = 4$)

Рис. 3: Зависимость \hat{P}_{CHOICE} от σ_2 для гармонических сигналов для $N = 100, k = \lfloor \frac{3N}{4} \rfloor, L = \lfloor \frac{k+1}{2} \rfloor$

Случай неизвестного сигнала: рекомендации по выбору параметров

Пусть известны ранги сигналов S_1, S_2 .

Проблема: неизвестно, имеют ли сигналы S_1, S_2 одинаковую или разную структуру \Rightarrow неизвестен ранг сигнала $(S_1; S_2)$.

Численное моделирование (гармонические сигналы):

- Параметры r_{SSA}, r_{MSSA} рекомендуется брать равными $\text{rank } S_1 + \text{rank } S_2$
- Параметр k рекомендуется брать достаточно большим ($k \geq \frac{N}{2}$)

В моей работе были получены следующие результаты:

- ① В задаче об исследовании структуры:
 - Были получены теоретические результаты, описывающие поведение MSE_{SSA} , MSE_{MSSA} для константных сигналов.
 - С помощью моделирования были исследованы свойства MSE_{SSA} , MSE_{MSSA} для гармонических сигналов одинаковой и разной структуры.
 - Получены теоретические результаты, описывающие поведение MSE_{MSSA} при масштабировании второго ряда в случае константных сигналов.
- ② В задаче о неизвестном сигнале:
 - Был разработан алгоритм CHOICE определения того, является ли ряд поддерживающим.
 - С помощью моделирования было исследовано качество алгоритма CHOICE в случае гармонических сигналов.
 - Были приведены рекомендации по выбору параметров алгоритма CHOICE.