Исследование статистических свойств метода Монте-Карло SSA

Бояров Андрей Александрович, гр. 522

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н., доц. Голяндина Н.Э.
Рецензент: аспирант Шлемов А.Ю.

Санкт-Петербург
2012г.

Постановка задачи

- $F_N = (f_0, \ldots, f_{N-1})$ — исходный ряд, $F_N = R_N + S_N$,
  $R_N = (r_0, \ldots, r_{N-1})$ — красный шум,
  $r_i = u + \gamma(r_{i-1} - u) + \delta\varepsilon_i$, $0 < \gamma < 1, \delta > 0$, $\varepsilon_i \sim N(0,1)$
  S_N — сигнал
- $H_0 : S_N = 0$, $H_1 : S_N \neq 0$
- $S_N = (s_0, \ldots, s_{N-1})$, $s_i = A\sin(2\pi\omega i + \varphi)$, $i = 0, \ldots, N-1$
- Критерий: Монте-Карло SSA из статьи Allen, Smith (1996)
- Проблемы: Недостаточная формализация, построение новых корректных критериев, сравнение по мощности

Гусеница-SSA

Исходный ряд: $F_N = (f_0, \ldots, f_{N-1})$

1. L — длина окна,
   $K = N - L + 1$, $X_i = (f_{i-1}, \ldots, f_{i+L-2})^{\mathrm{T}}, i = 1, \ldots, K$,
   $\mathbf{X} = [X_1 : \ldots : X_K]$ — L-траекторная матрица F_N
2. $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$, $\lambda_1 \geqslant \ldots \geqslant \lambda_d > 0$ — собственные числа $\mathbf{S}$,
   $U_1, \ldots, U_d$ — собственные вектора $\mathbf{S}$, $V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i}$,
   $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$,
   $i = 1, \ldots, d$.
   $\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$,
   $(\sqrt{\lambda_i}, U_i, V_i)$ — i-ая собственная тройка
3. Группировка собственных троек
4. Диагональное усреднение

Результат: $F_N = F_N^{(1)} + \ldots + F_N^{(m)}$

- $F_N = (f_0, \ldots, f_{N-1})$, $H_0 : S_N = 0$
- Входные данные: α — уровень значимости, L, $W \in \mathbf{R}^L$, $\mathbf{X}$ — L-траекторная матрица, G — число повторов Монте-Карло
- Статистика критерия: $t = \|\mathbf{X}^{\mathrm{T}} W\|^2$
- Использование м. Монте-Карло:
  - $\xi_N^i = (\xi_0^i, \ldots, \xi_{N-1}^i)$ — AR(1), $\mathbf{\Xi}_i$ — его L-траек. матрица
  - $p_i = \|\mathbf{\Xi}_i^{\mathrm{T}} W\|^2$ — проекция на W, $i = 1, \ldots, G$
- Доверительная область:
  - $((\alpha/2)$-квантиль, $(1-\alpha/2)$-квантиль) по $(p_1, \ldots, p_G)$
  - $(-\infty, (1-\alpha)$-квантиль) по $(p_1, \ldots, p_G)$
- Ситуации:
  - Реальная: параметры красного шума оцениваются по F_N
  - Модельная: параметры красного шума известны

- Выбираем систему векторов $W_1, \ldots, W_H$
- Статистика критерия $\bar{t}$:
  - $\bar{t}^{\mathrm{sing}} = (t_1^{\mathrm{sing}}, \ldots, t_H^{\mathrm{sing}})$, где $t_k^{\mathrm{sing}} = \|\mathbf{X}^{\mathrm{T}} W_k\|^2$
  - $\bar{t}^{\mathrm{min}} = (t_1^{\mathrm{min}}, \ldots, t_{H-1}^{\mathrm{min}})$, где $t_k^{\mathrm{min}} = \min(t_k^{\mathrm{sing}}, t_{k+1}^{\mathrm{sing}})$
  - $\bar{t}^{\mathrm{sum}} = (t_1^{\mathrm{sum}}, \ldots, t_{H-1}^{\mathrm{sum}})$, где $t_k^{\mathrm{sum}} = (t_k^{\mathrm{sing}} + t_{k+1}^{\mathrm{sing}})$
- Распределение $\bar{t}$ по м. Монте-Карло.

  Матрица проекций $\mathbf{P} = (p_{ij})_{i,j=1}^{H,G}$, где P_k — ее строка:
  - $P_k^{\mathrm{sing}} = (p_{k1}^{\mathrm{sing}}, \ldots, p_{kG}^{\mathrm{sing}})^{\mathrm{T}}$, где $p_{ki}^{\mathrm{sing}} = \|\boldsymbol{\Xi}_i^{\mathrm{T}} W_k\|^2$
  - $P_k^{\mathrm{min}} = (p_{k1}^{\mathrm{min}}, \ldots, p_{kG}^{\mathrm{min}})^{\mathrm{T}}$, где $p_{ki}^{\mathrm{min}} = \min(p_{ki}^{\mathrm{sing}}, p_{(k+1)i}^{\mathrm{sing}})$
  - $P_k^{\mathrm{sum}} = (p_{k1}^{\mathrm{sum}}, \ldots, p_{kG}^{\mathrm{sum}})^{\mathrm{T}}$, где $p_{ki}^{\mathrm{sum}} = (p_{ki}^{\mathrm{sing}} + p_{(k+1)i}^{\mathrm{sing}})$
- Проблема: построение критерия: $\alpha_I \approx \alpha$

Методы выбора системы векторов

- С.век.AR(1) (Allen, Smith (1996)):
  $W_k = W_k(\gamma)$ — k-ый собственный вектор $\mathbf{S}_R = \mathbf{\Gamma}$,
  где $\Gamma_{ij} = \gamma^{|i-j|}$
- sin,cos: $W_1, \ldots, W_H$ — чередующиеся синусы и косинусы
  с шагом по частоте: $1/(2L)$, $1/(4L)$, $1/(8L)$
- С.в., порожденные рядом (Allen, Smith (1996)):
  - $W_k = U_k$
  - t_k — собственные числа F_N
  - p_{ki} — собственные числа реализации красного шума

Проблема множественных сравнений

- Односторонний критерий
- Одномерная доверительная область для каждого W_k
- Попадание в критическую область хотя бы на одном W_k
- Параметры: $\gamma = 0.72$, $\delta = 1$, $u = 0$, $N = 200$, $L = 40$, $G = 1000$, $M = 1000$ — число повторов метода

Оценка вероятности ошибки первого рода, $\alpha = 0.05$				
$\bar{t}$	С.век.AR(1)	sin,cos($1/2L$)	sin,cos($1/4L$)	sin,cos($1/8L$)
$\bar{t}^{\mathrm{sing}}$	0.543	0.503	0.573	0.606
$\bar{t}^{\mathrm{min}}$	0.539	0.533	0.610	0.608
$\bar{t}^{\mathrm{sum}}$	0.477	0.477	0.563	0.582

- Критерий нельзя применять, $\alpha_I >> \alpha$

Теоретический эллипсоид в нормальной модели

- Модель: $\bar{s} \in \mathcal{N}(\mu, \Sigma)$
- Преобразование: $\bar{t} \mapsto \bar{s}$
  - $\bar{s} = \log(\bar{t})$
- Статистика критерия: $\rho^2_{\mathrm{Mahalanobis}}(\bar{s}, \hat{\mu}) \sim T^2_m(G-1)$, где $\hat{\mu}$ — вектор средних, m — размерность
- Аналог двустороннего критерия

Оценка вероятности ошибки первого рода				
Теоретический эллипсоид, $\alpha = 0.1$				
$\bar{t}$	С.век.AR(1)	sin,cos($1/2L$)	sin,cos($1/4L$)	sin,cos($1/8L$)
$\bar{t}^{\mathrm{sing}}$	0.114	0.107	0.202	0.260
$\bar{t}^{\mathrm{min}}$	0.143	0.165	0.176	0.259
$\bar{t}^{\mathrm{sum}}$	0.147	0.149	0.155	0.206

- Критерий неприменим, $\alpha_I > \alpha$: $\bar{s} = \log(\bar{t}) \notin \mathcal{N}(\mu, \Sigma)$

Доверительные области, построенные по выборке

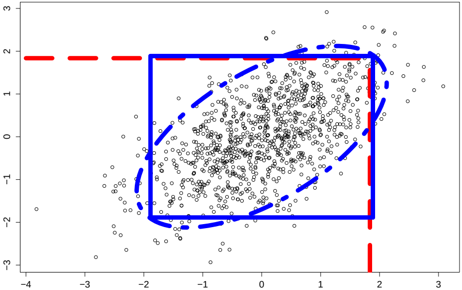
- Обозначения: $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_H)$ — вектор средних,

  $\bar{\sigma} = (\sigma_1, \ldots, \sigma_H)$, где σ_k — стандартное отклонение по P_k

- Метод:
  1. $t = f(\bar{t}) \in \mathbf{R}^1$
  2. Монте-Карло для построения доверительной области

- Преобразование (f):
  - Доверительный эллипсоид: расстояние Махаланобиса
  - Доверительный «куб»: $\max(|t_i - \hat{\mu}_i|/\sigma_i)$, $i = 1, \ldots, H$
  - Доверительный «полукуб»: $\max((t_i - \hat{\mu}_i)/\sigma_i)$, $i = 1, \ldots, H$

Двумерный случай



- Доверительный «полукуб», $\alpha = 0.05$
- Доверительный «куб», $\alpha = 0.1$
- Доверительный эллипсоид, $\alpha = 0.1$

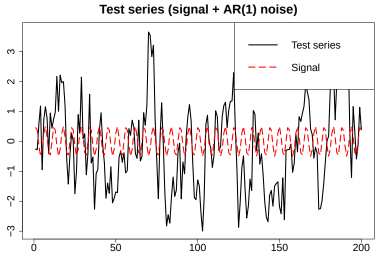
Проверка применимости критериев

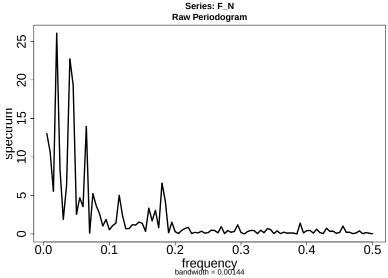
Оценка вероятности ошибки первого рода				
Полукубы, $\alpha = 0.05$				
$\bar{t}$	С.век.AR(1)	sin,cos($1/2L$)	sin,cos($1/4L$)	sin,cos($1/8L$)
$\bar{t}^{\mathrm{sing}}$	0.048	0.048	0.044	0.048
$\bar{t}^{\mathrm{min}}$	0.043	0.044	0.044	0.048
$\bar{t}^{\mathrm{sum}}$	0.045	0.051	0.044	0.045

Трудоемкости методов:

- sin,cos($1/8L$) — большая
- sin,cos($1/4L$) — средняя
- С.в.AR(1) — меньше
- sin,cos($1/2L$) — меньше

- Параметры метода: $N = 200$, $L = 40$, $G = 1000$, $M = 1000$
- Параметры красного шума: $\gamma = 0.72$, $\delta = 1$, $u = 0$
- Параметры сигнала: $A = 0.5$, $T = 5.5$, $\varphi = 0$





Сравнение критериев по мощности

Оценка мощности $\hat{\beta}$, сигнал $s(x) = 0.5\sin(2\pi x/5.5)$				
$\bar{t}^{\mathrm{sing}}$, $\alpha = 0.1$				
Метод	С.век.AR(1)	sin,cos($1/2L$)	sin,cos($1/4L$)	С.ч.
Куб	0.651	0.656	0.691	0.249
Эллипс.	0.586	0.577	0.596	0.225

- Лучше «куб» $\Rightarrow$ «полукуб» как аналог одностороннего критерия

Полукубы, $\alpha = 0.05$				
$\bar{t}$	С.век.AR(1)	sin,cos($1/2L$)	sin,cos($1/4L$)	С.ч.
$\bar{t}^{\mathrm{sing}}$	0.719	0.709	0.744	0.214
$\bar{t}^{\mathrm{min}}$	0.710	0.745	0.756	0.225
$\bar{t}^{\mathrm{sum}}$	0.743	0.733	0.743	0.222

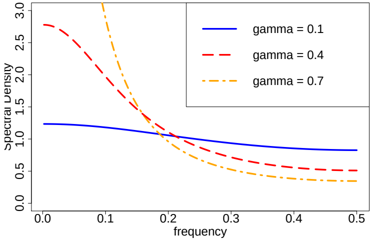
- Критерий «С.ч.» маломощен, остальные одинаковы

Изменение параметров сигнала

Оценка мощности $\hat{\beta}$ для различных сигналов					
Полукубы, $\alpha = 0.05$					
$\bar{t}$	$T = 5.5$ $\varphi = 0$	$T = 5.5$ $\varphi = \pi/4$	$T = 5$ $\varphi = 0$	$T = 5$ $\varphi = \pi/2$	$T = 5$ $\varphi = \pi/4$
$\bar{t}^{\mathrm{sing}}$	0.709	0.715	0.845	0.847	0.872
$\bar{t}^{\mathrm{min}}$	0.745	0.758	0.631	0.635	0.632
$\bar{t}^{\mathrm{sum}}$	0.733	0.737	0.787	0.798	0.793

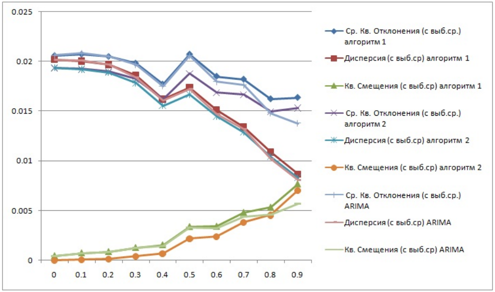
- Вычисления для sin,cos$(1/2L)$
- Выводы:
    - Наибольшая минимальная мощность у $\bar{t}^{\mathrm{sum}}$
    - $T = 5.5$: лучший — $\bar{t}^{\mathrm{min}}$
    - $T = 5$: лучший — $\bar{t}^{\mathrm{sing}}$
    - Изменение φ не влияет на результат

Изменение параметра красного шума



Оценка мощности $\hat{\beta}$, сигнал $s(x) = 0.5\sin(2\pi x/10)$				
Полукубы, $\bar{t}^{\mathrm{sing}}$, $\alpha = 0.05$				
γ	С.век.AR(1)	sin,cos($1/2L$)	sin,cos($1/4L$)	sin,cos($1/8L$)
0.1	0.680	0.703	0.684	0.688
0.4	0.350	0.399	0.372	0.371
0.7	0.219	0.227	0.214	0.219

- Красный шум: $\xi_i = u + \gamma(\xi_{i-1} - u) + \delta\varepsilon_i$, $0 < \gamma < 1, \delta > 0$, $\mathsf{E}\varepsilon_i = 0$, $\mathsf{D}\,\varepsilon_i = 1$
- Алгоритмы оценок γ, δ:
  1. $\hat{c}_l$ — оценка автоковариации, $\hat{\gamma} = \hat{c}_1/\hat{c}_0, \hat{\delta} = \sqrt{(\hat{c}_0^2 - \hat{c}_1^2)/\hat{c}_0}$
  2. Алгоритм из статьи Allen, Smith (1996)
  3. Алгоритм из статьи Gardner, Harvey, Phillips (1980)
- Сравнение алгоритмов



Критерий с оцениванием параметров красного шума

Оценка вероятности ошибки первого рода				
Полукубы, $\bar{t}^{\text{sing}}$, $\alpha = 0.05$				
	С.век.AR(1)	sin,cos(1/$2L$)	sin,cos(1/$4L$)	sin,cos(1/$8L$)
γ	0.048	0.048	0.044	0.048
$\hat{\gamma}$	0.020	0.026	0.031	0.026

Оценка мощности $\hat{\beta}$, сигнал $s(x) = 0.5\sin(2\pi x/5.5)$				
Полукубы, $\bar{t}^{\text{sing}}$, $\alpha = 0.05$				
$\bar{t}$	С.век.AR(1)	sin,cos(1/$2L$)	sin,cos(1/$4L$)	sin,cos(1/$8L$)
γ	0.719	0.709	0.744	0.753
$\hat{\gamma}$	0.560	0.589	0.622	0.636

- $\hat{\alpha}_I$ и $\hat{\beta}$ уменьшились по сравнению с модельной ситуацией

Заключение

1. Формализована задача из Allen, Smith (1996)
2. Предложены различные корректные модификации метода
3. Модификации метода сравнены с помощью мощности
   Лучшие модификации:
   - Полукуб в качестве доверительной области
   - Равномерная решетка по частоте
4. Рассмотрено влияние изменения различных параметров на результаты работы метода
   - $\bar{t}^{\mathrm{sum}}$ наиболее устойчив к изменению периода сигнала
   - Свойства критерия не зависят от фазы сигнала
5. Рассмотрено применение метода в реальных условиях
   - Использование оценок параметров шума уменьшает мощность критерия примерно в 1.5 раза