Алгоритмы робастного оценивания спектральных характеристик сигналов и их использование.

Любомищенко Н.С.
6057/12

Содержание

Классические методы оценивания спектральной плотности

- **Периодограммы**

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt$$

$$X_T(f) = \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-i2\pi ft} dt$$

$$S(f) = \lim_{T\to\infty} \frac{1}{T} E[|X_T(f)|^2]$$

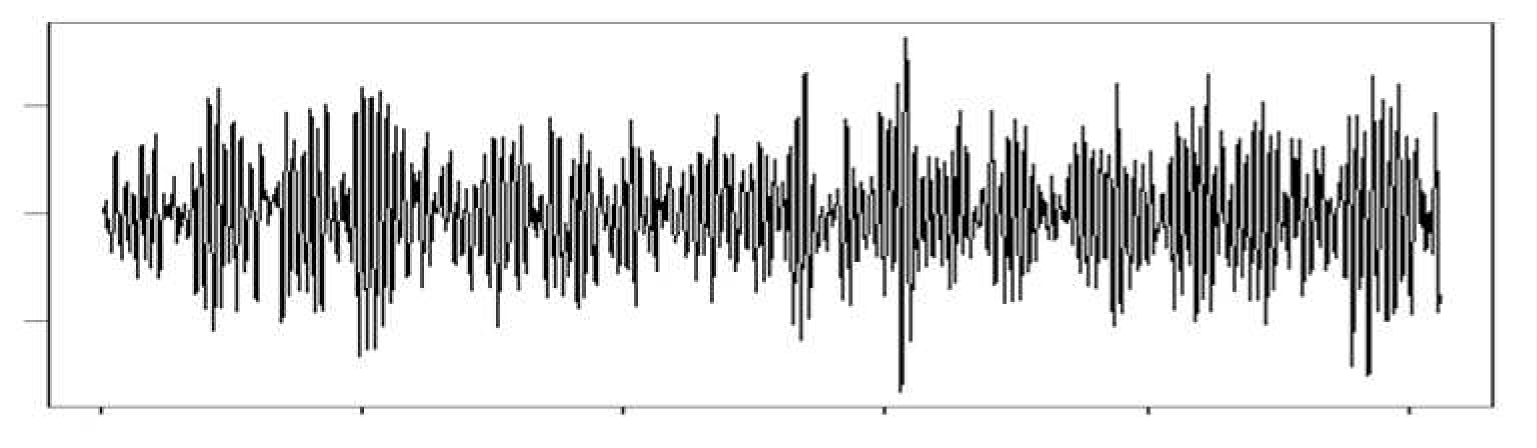
- **Коррелограммы** (теорема Винера-Хинчина)

$$S(f) = \int_{-\infty}^{\infty} R(t) \mathrm{e}^{-i2\pi ft} \, dt$$

Процесс авторегрессии

- *AR(p)*: $$x(t)=\sum_{i=1}^{p}\alpha_i x(t-i)+\varepsilon_t$$
- *AR(2)*: $$x(t)=\alpha_1 x(t-1)+\alpha_2 x(t-2)+\varepsilon_t$$

$$\varepsilon_t \sim N(0,\sigma)$$



Классические методы оценивания спектральной плотности: метод Юла-Уолкера

- Оценка автоковариационной функции:

$$\hat{r}_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X}_t)(X_{t-k} - \bar{X}_t)}{n-k}, 0 \le k \le l$$

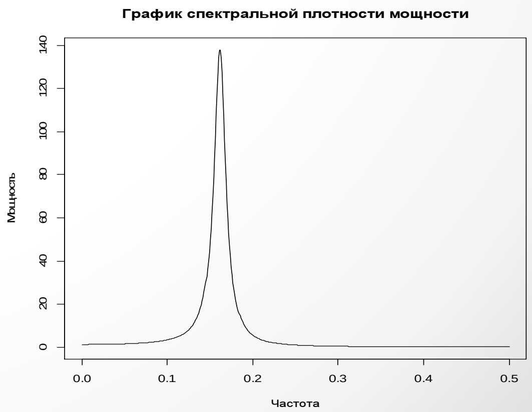
- Система уравнений Юла-Уолкера:

$$r_0 = \hat{\alpha}_1 r_1 + \hat{\alpha}_2 r_2 + \hat{\alpha}_3 r_3 + ... + \hat{\alpha}_l r_l + \hat{\sigma}_\varepsilon^2$$

$$\begin{cases} r_1 = \hat{\alpha}_1 r_0 + \hat{\alpha}_2 r_1 + \hat{\alpha}_3 r_2 + ... + \hat{\alpha}_l r_{l-1} \\ r_2 = \hat{\alpha}_1 r_1 + \hat{\alpha}_2 r_0 + \hat{\alpha}_3 r_1 + ... + \hat{\alpha}_l r_{l-2} \\ r_3 = \hat{\alpha}_1 r_2 + \hat{\alpha}_2 r_1 + \hat{\alpha}_3 r_0 + ... + \hat{\alpha}_l r_{l-3} \\ .... \\ r_l = \hat{\alpha}_1 r_{l-1} + \hat{\alpha}_2 r_{l-2} + \hat{\alpha}_3 r_{l-3} + ... + \hat{\alpha}_l r_0 \end{cases}$$

- Оценка спектральной плотности мощности:

$$\hat{S}(f) = \frac{\hat{\sigma}_e^2}{\left|1 - \sum_{j=1}^{l} \hat{\alpha}_j \exp(i2\pi f j)\right|^2}$$



Робастность в статистике

«robust» с лат.: сильный, грубый, надежный

- Робастность – устойчивость статистических процедур к возможным отклонениям от принятых моделей распределений.
- Неробастная процедура: **МНК**

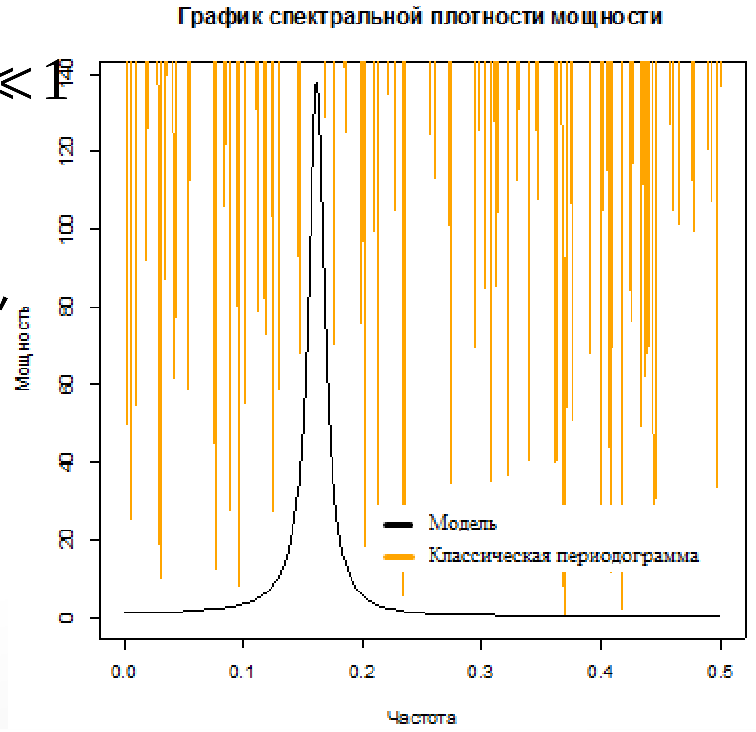
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$x_1, x_2, \ldots x_n \sim (1-\varepsilon)N(x;0,1) + \varepsilon C(x;\theta),\ \ 0 < \varepsilon \ll 1$$

- Робастная процедура: **МНМ**

[Tukey(1960)],[Huber(1964,1981)], [Hampel(1986)]

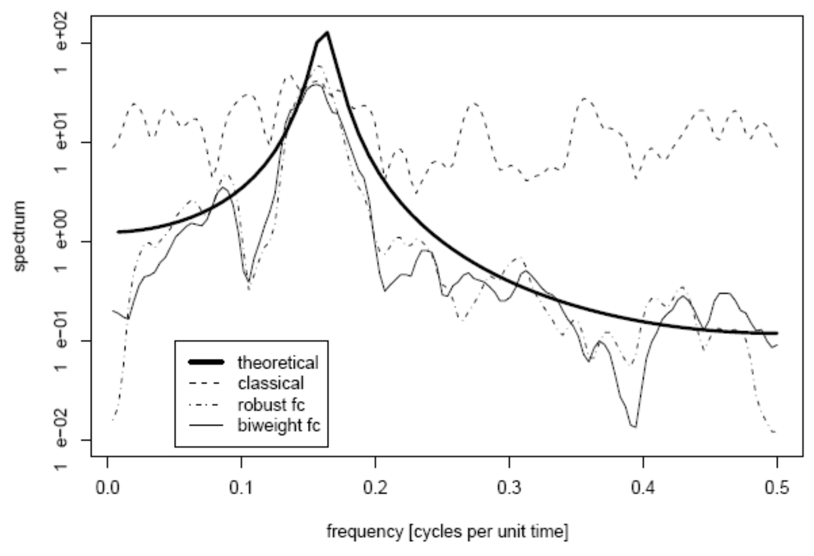
$$med\{x_i\} = \begin{cases} x(k+1), & n = 2k+1 \\ \dfrac{x(k+1)+x(k)}{2}, & n = 2k \end{cases}$$



Робастные подходы к оценке спектральной плотности

- **Двухэтапные процедуры**:
  1. Предварительная робастная фильтрация – чистка процесса от «засоряющих» данных
  2. Применение классических методов к очищенным данным

[Kleiner, Martin (1979)],
[Martin, Thomson (1982)],
[Челпанов и пр. (1985)],
[Tatum and Hurvich (1993)],
[Dutter, Spangl (2008)].



Постановка задачи:

поиск перспективных подходов к робастному оцениванию спектральной плотности

1. Робастный вариант периодограмм
2. Робастный вариант коррелограмм
3. Робастный вариант метода Юла-Уолкера

Используемые робастные оценки

- **Параметра положения:** выборочная медиана (**med**) как робастный аналог выборочного среднего
- **Параметра масштаба:**
  - медиана абсолютных отклонений (**MAD**)

[Rousseeuw, 1984]

$$MAD(X_t) = med(|\, X_t - med(X_t)\,|), \quad \text{eff} = 37\%$$

  - **FQn** оценка [Смирнов & Шевляков, 2010]

$$FQn(X_t) = MAD(X_t) * (1 - \frac{Z_0 - n/\sqrt{2}}{Z_2}), \quad \text{eff} = 81\%,$$

$$Z_k = \sum u_i^k e^{-u_i^2/2} \qquad u_i = \frac{X_i - med(X_t)}{MAD(X_t)}$$

как робастные аналоги среднеквадратического отклонения

Робастные периодограммы: медианное преобразование Фурье

- Оценка спектральной плотности:

$$\hat{S}(f) = \frac{1}{T}\left|\tilde{X}_T^c(f) + i\tilde{X}_T^s(f)\right|^2$$

- Медианное преобразование Фурье:

$$\tilde{X}_T^c(f) = T \arg\min_z \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)\cos(2\pi ft) - z|\,dt$$

$$\tilde{X}_T^s(f) = T \arg\min_z \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)\sin(2\pi ft) - z|\,dt$$

Робастные периодограммы: медианное преобразование Фурье

- МНК:

$$X_T^c(f) = T \arg\min_z \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)\cos(2\pi ft) - z|^2\, dt$$

$$X_T^s(f) = T \arg\min_z \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)\sin(2\pi ft) - z|^2\, dt$$

- МНМ:

$$\tilde{X}_T^c(f) = T \arg\min_z \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)\cos(2\pi ft) - z|\, dt$$

$$\tilde{X}_T^s(f) = T \arg\min_z \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)\sin(2\pi ft) - z|\, dt$$

$$X_T^c(f) = T\, ave\{x_1\cos(2Pi\ f),\ldots,x_N\cos(2Pi\, Nf)\} \qquad \tilde{X}_T^c(f) = T\, med\{x_1\cos(2Pi\ f),\ldots,x_N\cos(2Pi\, Nf)\}$$

$$X_T^s(f) = T\, ave\{x_1\sin(2Pi\ f),\ldots,x_N\sin(2Pi\, Nf)\} \qquad \tilde{X}_T^s(f) = T\, med\{x_1\sin(2Pi\ f),\ldots,x_N\sin(2Pi\, Nf)\}$$

Робастные коррелограммы: процедура Блекмэна-Тьюки

$$\hat{S}_x(f) = \int_{-T/2}^{T/2} \hat{R}(t)\, e^{-i2\pi f\tau}\, dt$$

- Робастные оценки коэффициентов корреляции:
  - **MAD** оценка: $$\hat{\rho}_{MAD}(k) = \frac{MAD^2(U) - MAD^2(V)}{MAD^2(U) + MAD^2(V)}$$
  - **FQn** оценка: $$\hat{\rho}_{FQn}(k) = \frac{FQn^2(U) - FQn^2(V)}{FQn^2(U) + FQn^2(V)}$$

$$U = \frac{X_t - med(X_t)}{\sqrt{2}MAD(X_t)} + \frac{X_{t-k} - med(X_t)}{\sqrt{2}MAD(X_t)} \qquad V = \frac{X_t - med(X_t)}{\sqrt{2}MAD(X_t)} - \frac{X_{t-k} - med(X_t)}{\sqrt{2}MAD(X_t)}$$

Робастный аналог метода Юла-Уолкера

- Система уравнений Юла-Уолкера в матричном виде:

$$r_0^l = \alpha^T r_1^l + \delta_\varepsilon^2$$

$$R(r_0^l)\alpha = r_1^l$$

- Робастные оценки автоковариационной функции:
  - MAD оценка $\hat{r}_{MAD}(k) = \hat{\rho}_{MAD}(k) * MAD(X_t) * MAD(X_{t-k})$
  - FQn оценка $\hat{r}_{FQn}(k) = \hat{\rho}_{FQn}(k) * FQn(X_t) * FQn(X_{t-k})$
  - Оценка спектральной плотности $\hat{S}(f) = \dfrac{\hat{\sigma}_e^2}{\left|1 - \sum_{j=1}^{l} \hat{\alpha}_j \exp(i2\pi f j)\right|^2}$

Модели засорения

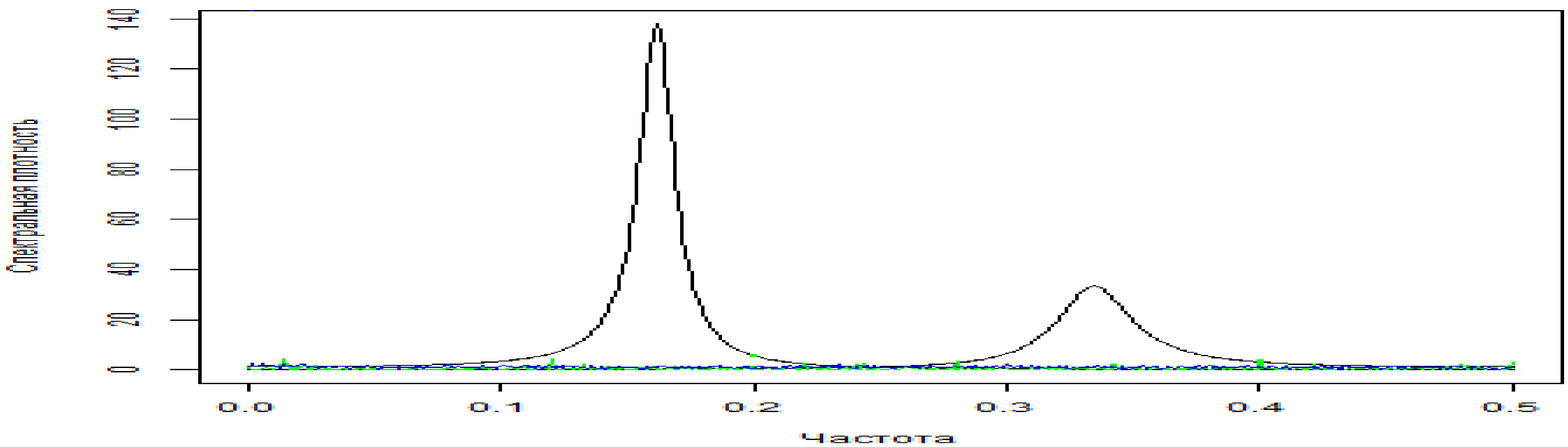
- Амплитудная модель засорения:

$$x(t) = \alpha_1 x(t-1) + \alpha_2 x(t-2) + \varepsilon_t$$

$$\varepsilon_t \sim (1-\xi)N(0,1) + \xi N(0,100)$$

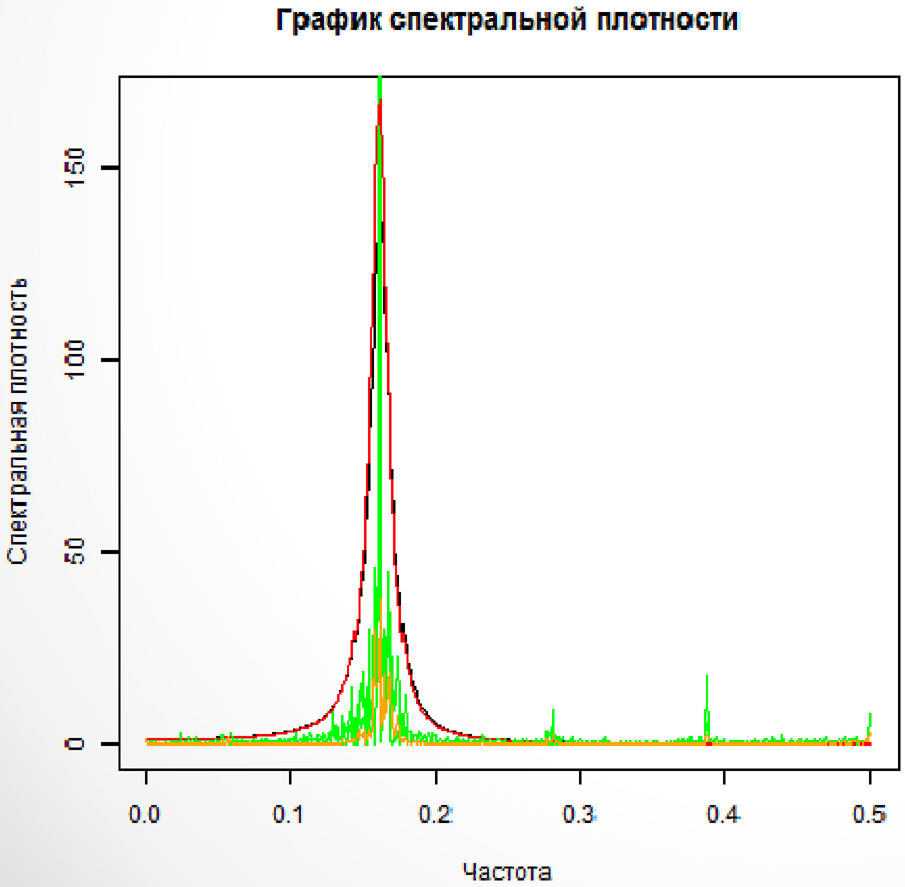
- Модель «разладки» - маскирующий эффект:

$$x(t) = \alpha_1 x(t-1) + \alpha_2 x(t-2) + \varepsilon_t$$



Эксперимент: модель без засорения

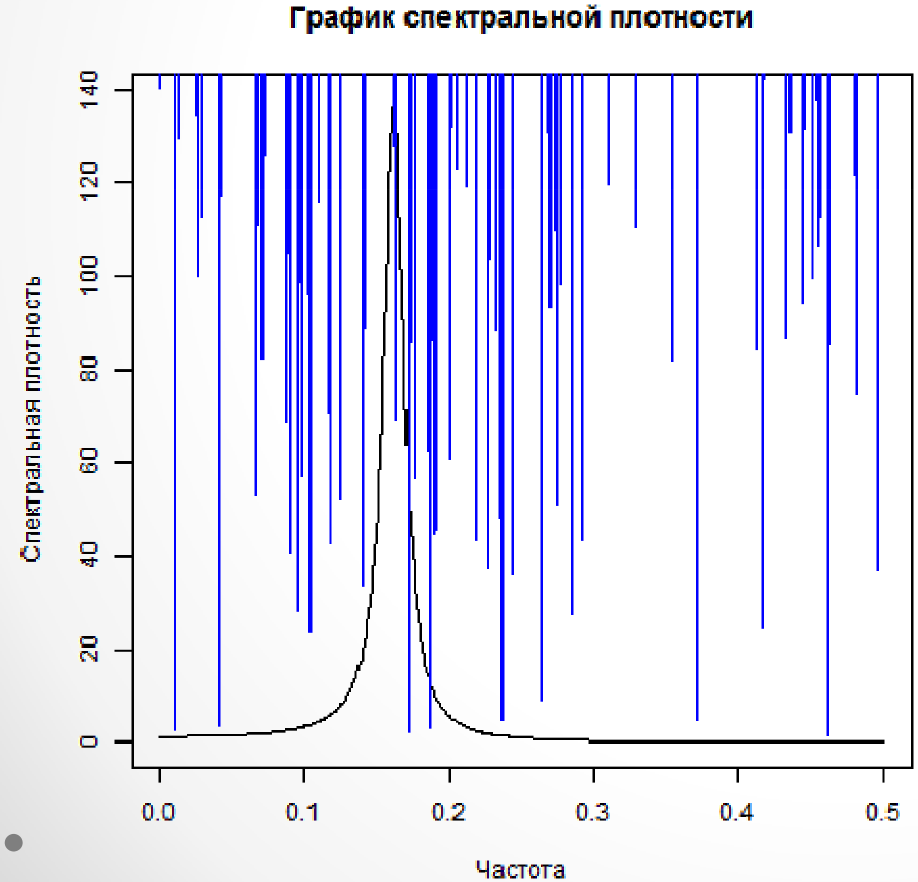
Выборка из 1024 элементов
Число экспериментов Монте-Карло: 10



Легенда:
Медианная периодограмма
Метод Юла-Уолкера с FQn оценкой
Модель
Коррелограмма с MAD оценкой

Эксперимент: амплитудное засорение

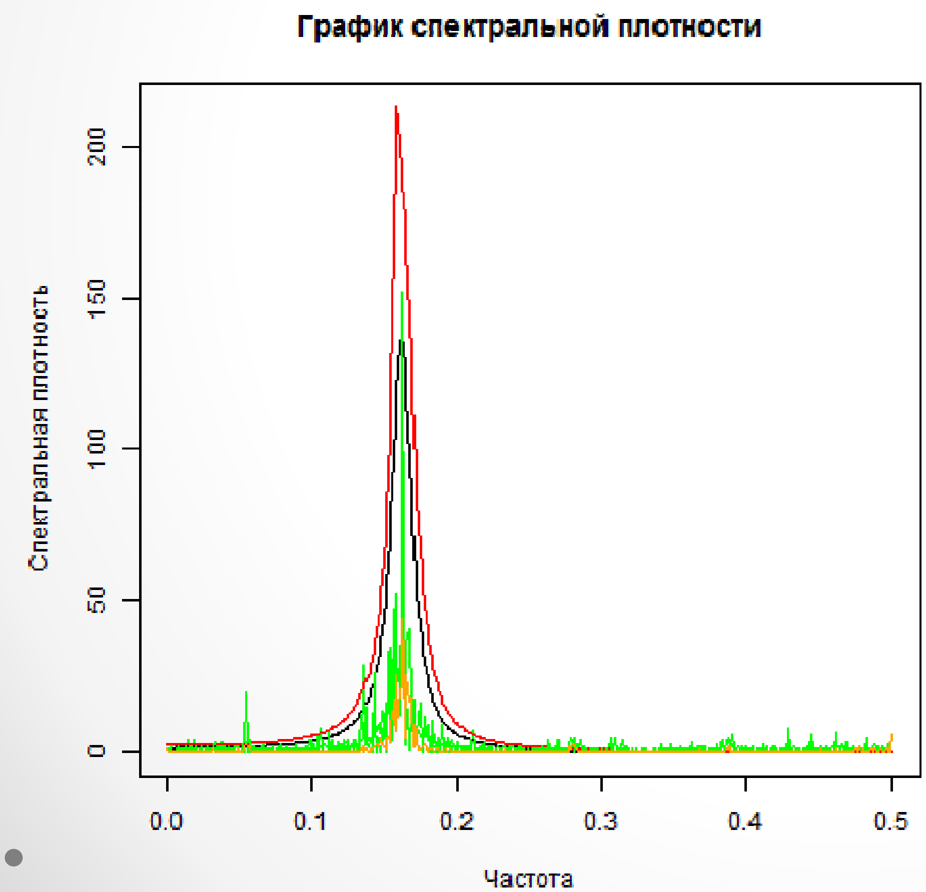
Выборка из 1024 элементов
Число экспериментов Монте-Карло: 10
Выбросы: 10%



Легенда:
Классическая периодограмма
Модель

Эксперимент: амплитудное засорение

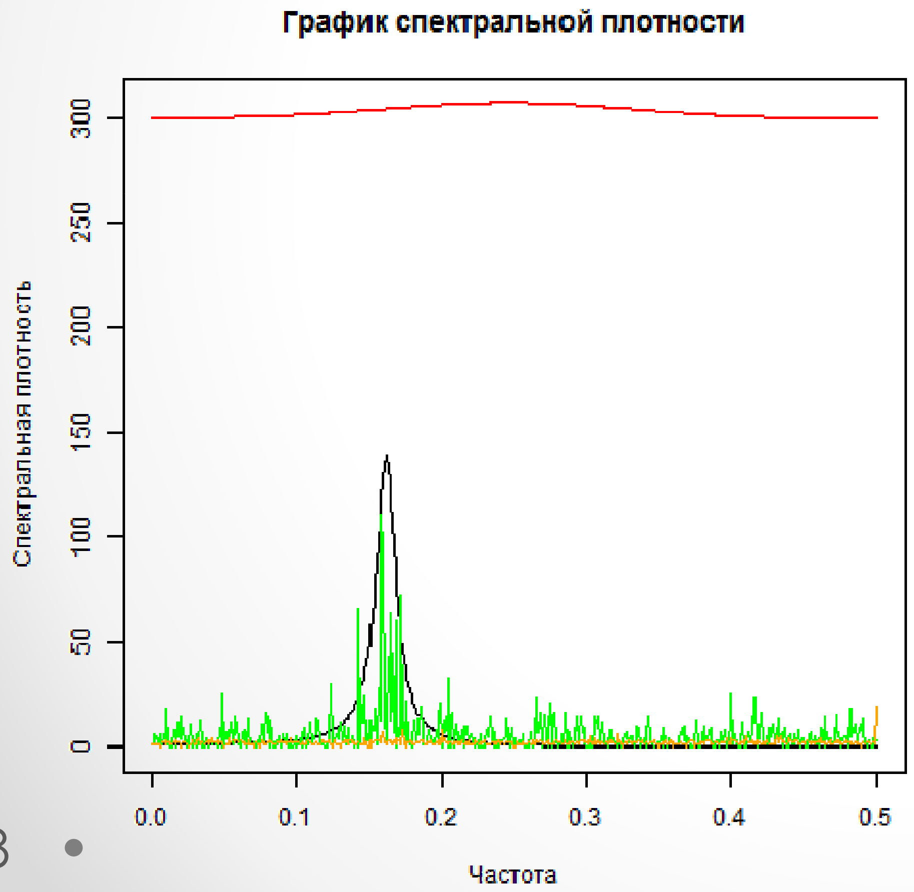
Выборка из 1024 элементов
Число экспериментов Монте-Карло: 10
Выбросы: 10%



Легенда:
Медианная периодограмма
Метод Юла-Уолкера с FQn оценкой
Модель
Коррелограмма с MAD оценкой

Эксперимент: амплитудное засорение

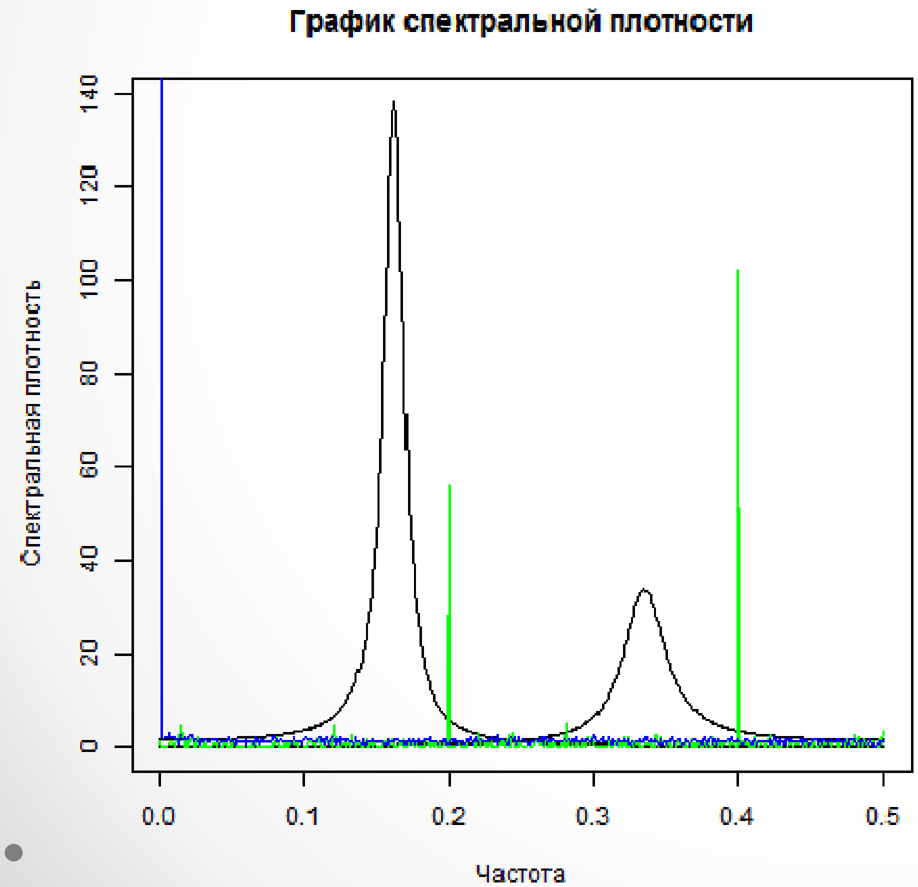
Выборка из 1024 элементов.
Число экспериментов Монте-Карло: 10
Выбросов: 45%



Легенда:
Медианная периодограмма
Метод Юла-Уолкера с FQn оценкой
Модель
Коррелограмма с MAD оценкой

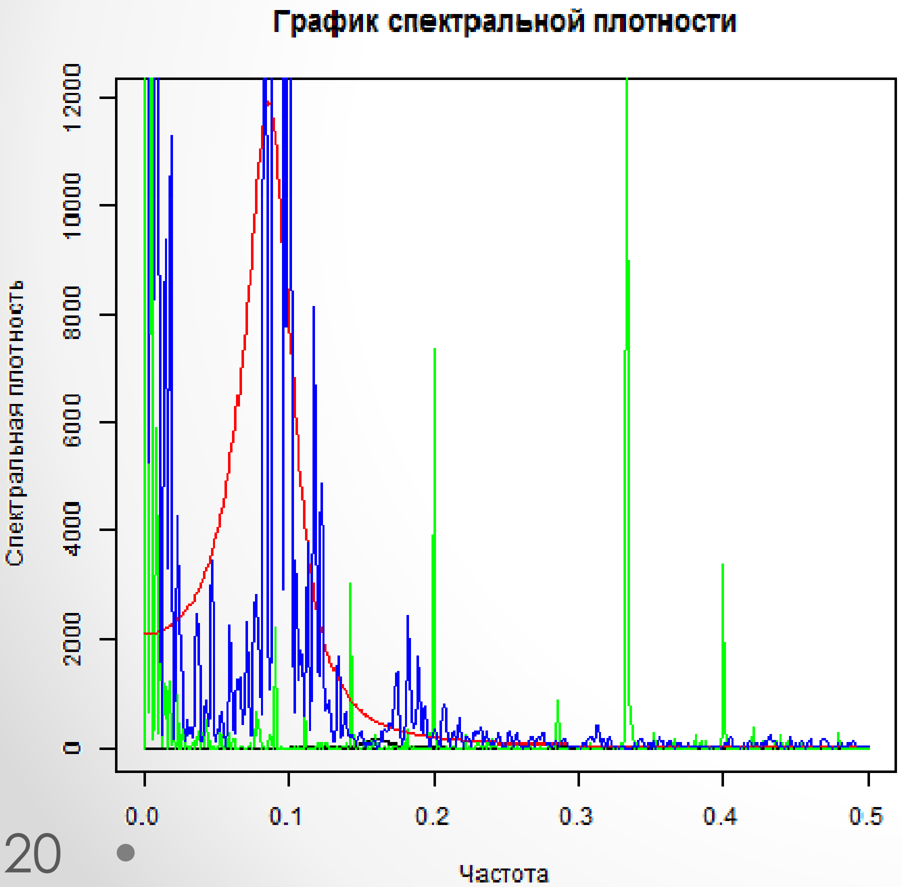
Эксперименты: модель «разладки»

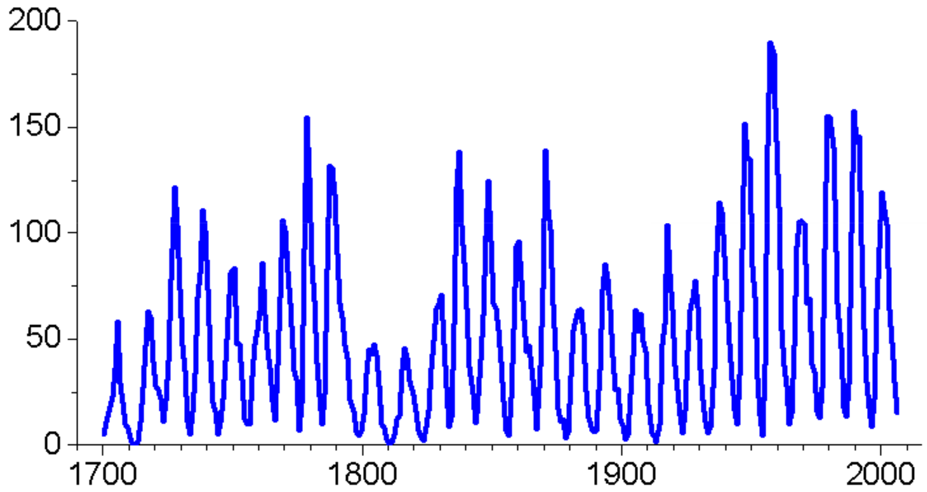
Выборка из 1024 элементов.
Число экспериментов Монте-Карло: 10



Легенда:
Классическая периодограмма
Медианная периодограмма
Модель

Эксперименты с реальными данными: годовые числа Вольфа





Легенда:
Классическая периодограмма
Медианная периодограмма
Метод Юла-Уолкера с FQn оценкой

Выводы

1. Робастный метод Юла-Уолкера оказался лучшим на модели авторегрессии с засорениями
2. Медианная периодограмма наследует свойство устойчивости выборочной медианы к выбросам
3. Робастная процедура Блекмэна-Тьюки ведет себя примерно как робастная периодограмма, но менее устойчива к выбросам

Направления будущих исследований

- Уменьшение смещения оценок
- Сглаживание оценок путем применения временных, спектральных и корреляционных «окон»
- Исследование описанных подходов на более сложных моделях процессов

Спасибо за внимание