Отношение сигнал/шум в астрофизических наблюдениях

Лекция 13

Обнаружение периодических сигналов и статистическая значимость. Методы Фурье и Ломба–Скаргла, вероятность ложной тревоги

КазНУ имени аль-Фараби

Цель и план лекции

Цель

Изучение методов поиска периодических сигналов в равномерных и неравномерных временных рядах и способов определения статистической значимости найденных пиков.

1. Введение
2. Дискретное преобразование Фурье
3. Периодограмма Ломба–Скаргла
4. Окно наблюдений
5. Значимость пика
6. Другие методы поиска периодичности
7. Значимость нескольких сигналов и астросейсмология

1 Введение

- Периодичность является важным источником информации: пульсации звёзд позволяют изучать их внутреннее строение, периоды двойных систем дают массы компонентов, вращение проявляется в модуляции блеска.

2 Дискретное преобразование Фурье

- Для равномерного ряда длиной N с шагом Δt применяют дискретное преобразование Фурье.

$$P(f) = (1/N)\,|\,\Sigma\, x_j\, exp(-2\pi i\, f\, t_j)\,|^2.$$

- Для белого шума значения P распределены экспоненциально. Предельная частота Найквиста равна $f_N = 1/(2\Delta t)$, а разрешение по частоте – примерно 1/T, где T – длина ряда.

2 Дискретное преобразование Фурье (продолжение)

Утечка спектра

Конечная длина записи приводит к появлению боковых лепестков. Функции окон Ханна и других уменьшают их, увеличивая ширину пика.

Алиасинг

Частоты выше предельной отображаются в рабочую область.

Статистическая неустойчивость

Значения P не становятся более точными при увеличении длины ряда, поэтому применяется усреднение сегментов (метод Уэлча).

Нормировка

Для амплитуды синусоиды с амплитудой A значение P(f) составляет приблизительно $NA^2/4$ при нормировке на единичную дисперсию шума.

3 Периодограмма Ломба–Скаргла

- Метод применяется для неравномерно распределённых данных. Периодограмма определена как:

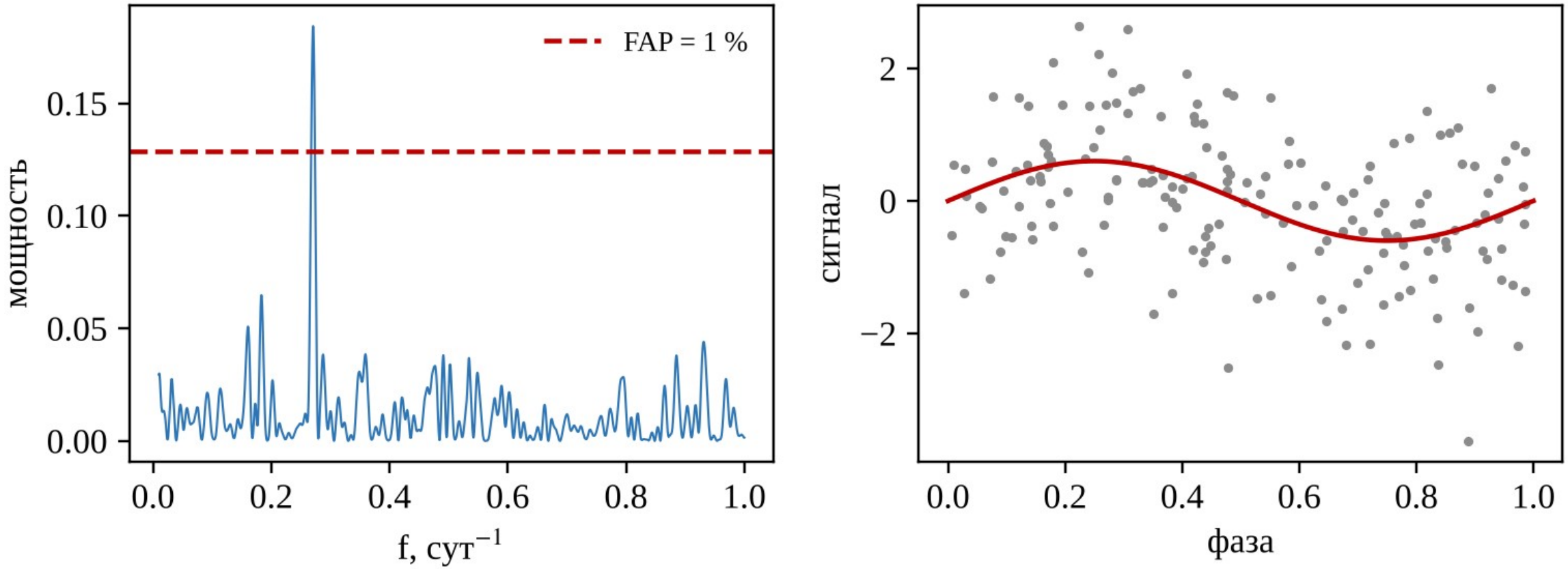
$$P(f) = \tfrac{1}{2}\left\{ [\Sigma\, y_j \cos 2\pi f(t_j - \tau)]^2 / \Sigma \cos^2 2\pi f(t_j - \tau) + [\Sigma\, y_j \sin 2\pi f(t_j - \tau)]^2 / \Sigma \sin^2 2\pi f(t_j - \tau) \right\},$$

- где величина τ определена условием:

$$\tan(4\pi f \tau) = \Sigma \sin(4\pi f\, t_j) / \Sigma \cos(4\pi f\, t_j).$$

- Метод эквивалентен подгонке синусоиды методом наименьших квадратов на каждой частоте.

3 Периодограмма Ломба–Скаргла (продолжение)



Периодограмма Ломба–Скаргла и свёрнутая по фазе кривая

4 Окно наблюдений

- Наблюдаемая периодограмма является свёрткой истинного спектра со спектральным окном.

5 Значимость пика

- Для нормированной периодограммы при белом гауссовом шуме вероятность того, что значение z превышается в одной частоте, равна exp(−z).

$$FAP = 1 - (1 - e^{-z})^{M}.$$

- Величина M приближённо равна числу разрешённых частот в рассматриваемой полосе, порядка f_{max} Т. Хорн и Балиунас (1986) предложили оценку эффективного числа M. Более точным приближением является формула Балуева (2008), основанная на теории экстремальных значений.

5 Значимость пика (продолжение)

- Метод бутстрепа является надёжным для произвольной выборки: значения данных случайным образом перемешиваются между моментами времени, для каждого такого набора определяется максимум периодограммы, и формируется распределение максимумов.

5 Периодограмма Ломба–Скаргла

Листинг 13.1.

```
import numpy as np
from scipy.signal import lombscargle

rng = np.random.default_rng(13)
N = 180
t = np.sort(rng.uniform(0, 120, N))
P0, A, s = 3.7, 0.6, 1.0
y = A * np.sin(2 * np.pi * t / P0) + rng.normal(0, s, N)
y -= y.mean()

f = np.linspace(0.01, 1.0, 8000)
p = lombscargle(t, y, 2 * np.pi * f, normalize=True)
k = p.argmax()

Xc = np.vstack([np.sin(2 * np.pi * f[k] * t),
                np.cos(2 * np.pi * f[k] * t)]).T
coef = np.linalg.lstsq(Xc, y, rcond=None)[0]
print(f"Лучший период: {1 / f[k]:.3f} (задан {P0})")
print(f"Мощность пика: {p[k]:.3f}")
print(f"Амплитуда: {np.hypot(*coef):.3f} (задана {A})")
print(f"Ожидаемое S/N: {A / s * np.sqrt(N / 2):.1f}")
```

5 Результат: Периодограмма Ломба–Скаргла

```
Лучший период: 3.693 (задан 3.7)
Мощность пика: 0.102
Амплитуда: 0.494 (задана 0.6)
Ожидаемое S/N: 5.7
```

5 Вероятность ложной тревоги

Листинг 13.2.

```
def fap_analytic(pw, N, M):
    z = -(N - 3) / 2 * np.log(1 - pw)
    return 1 - (1 - np.exp(-z))**M

M = (f[-1] - f[0]) * t.max()
print(f"Оценка числа независимых частот: {M:.0f}")
print(f"FAP (аналитически) = {fap_analytic(p[k], N, M):.2e}")

nb = 1000
peaks = []
fc = f[::4]
for _ in range(nb):
    yb = rng.permutation(y)
    peaks.append(lombscargle(t, yb, 2 * np.pi * fc, normalize=True).max())
peaks = np.array(peaks)
print(f"Порог FAP=1% (бутстреп): {np.quantile(peaks, 0.99):.3f}")
print(f"Найденных превышений: {(peaks >= p[k]).sum()} из {nb}")
```

5 Результат: Вероятность ложной тревоги

```
Оценка числа независимых частот: 118
FAP (аналитически) = 9.02e-03
Порог FAP=1% (бутстреп): 0.109
Найденных превышений: 22 из 1000
```

Обе оценки дают вероятность ложной тревоги порядка 1–2 %: пик заметен, но при заранее заданном пороге 1 % его значимость остаётся пограничной.

6 Другие методы поиска периодичности

Минимизация разброса фазы (PDM)

Данные сворачиваются для пробных периодов, и минимизируется разброс в фазовых бинах; это эффективно для несинусоидальных кривых.

Метод BLS

Алгоритм Ковача и др. (2002) предназначен для поиска прямоугольных транзитов.

Метод Гаусса–Марковских процессов

Применяется для квазипериодической модуляции вращения.

Вейвлеты

Применяются к сигналам с меняющимся периодом.

Необходимо учитывать, что истинный сигнал и его гармоники могут проявиться на частотах 2f, f/2.

6 Другие методы поиска периодичности (продолжение)

Многогармонический анализ

Помогает определить несинусоидальные сигналы, например для пульсирующих звёзд типа RR Лиры.

Необходимо учитывать, что истинный сигнал и его гармоники могут проявиться на частотах 2f, f/2.

7 Значимость нескольких сигналов и астросейсмология

- Если в данных присутствует несколько периодичностей, применяют последовательную процедуру (prewhitening): находят самый сильный пик, вычитают синусоиду, анализируют остаток.
- В астросейсмологии анализируется спектр мощности, который содержит гребёнку мод.

$$\ln L = -\sum_k [\ln M_k + P_k / M_k],$$

- где M_k – модель спектра, а P_k – наблюдаемая мощность. Такой подход корректно описывает статистику периодограммы.
- В присутствии коррелированного шума FAP, рассчитанное для белого шума, будет заниженным на низких частотах.

7 Значимость нескольких сигналов и астросейсмология (продолжение)

- В поиске пульсаров применяется складывание сигнала по пробному периоду и дисперсионной мере (epoch folding), а в радиоастрономии – быстрое складывание (FFA).
- Упражнение: для ряда с неравномерной выборкой вычислите окно наблюдений и установите положение алиасов основного сигнала; определите, можно ли различить истинный период и его суточный алиас.

8 Разбор примера

1. Пусть серия длиной $T = 120$ сут, рассматриваемая полоса частот $f_{max} = 1$ сут$^{-1}$, тогда приближённо $M \approx 120$.
2. Если для пика $z = 10$, то вероятность превышения в одной частоте равна $e^{-10} \approx 4{,}5 \cdot 10^{-5}$, а FAP $\approx 120 \cdot 4{,}5 \cdot 10^{-5} \approx 5{,}4 \cdot 10^{-3}$.
3. Для пороговой FAP = 1 % требуется $e^{-z} = 0{,}01/120$, откуда $z = \ln(12000) \approx 9{,}4$.
4. Для равномерного разрешения при $T = 120$ сут ширина пика составляет примерно $1/120 \approx 0{,}0083$ сут$^{-1}$, а шаг сетки – около 0,001 сут$^{-1}$.

9 Вопросы для обсуждения на занятии

Окно

Какие следствия имеет наблюдение только в ночное время для поиска периодов около суток?

Значимость

Почему оценка FAP зависит от диапазона частот, просмотренных при поиске?

Метод

В каких случаях PDM или BLS предпочтительнее метода Ломба–Скаргла?

Контрольные вопросы

1. В чём заключаются особенности периодограммы Фурье?
2. Какую задачу решает метод Ломба–Скаргла?
3. Что такое спектральное окно и как оно проявляется?
4. Как выбрать шаг сетки частот?
5. Как определяется вероятность ложной тревоги?
6. Каким образом выполняется бутстреп-оценка значимости?
7. Какие методы подходят для поиска несинусоидальных и транзитных сигналов?
8. Как отличить истинный период от алиаса?

Список литературы

1. Lomb N.R. Least-squares frequency analysis of unequally spaced data // Astrophysics and Space Science. – 1976. – Vol. 39. – P. 447–462.
2. Scargle J.D. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data // The Astrophysical Journal. – 1982. – Vol. 263. – P. 835–853.
3. VanderPlas J.T. Understanding the Lomb–Scargle periodogram // The Astrophysical Journal Supplement Series. – 2018. – Vol. 236. – Art. 16.
4. Baluev R.V. Assessing the statistical significance of periodogram peaks // Monthly Notices of the Royal Astronomical Society. – 2008. – Vol. 385. – P. 1279–1285.
5. Zechmeister M., Kürster M. The generalised Lomb–Scargle periodogram // Astronomy & Astrophysics. – 2009. – Vol. 496. – P. 577–584.
6. Horne J.H., Baliunas S.L. A prescription for period analysis of unevenly sampled time series // The Astrophysical Journal. – 1986. – Vol. 302. – P. 757–763.
7. Kovács G., Zucker S., Mazeh T. A box-fitting algorithm in the search for periodic transits // Astronomy & Astrophysics. – 2002. – Vol. 391. – P. 369–377.
8. Basu S., Chaplin W.J. Asteroseismic Data Analysis: Foundations and Techniques. – Princeton: Princeton University Press, 2017.