Отношение сигнал/шум в астрофизических наблюдениях

Лекция 14

Систематический и коррелированный шум. Белый шум, красный шум, инструментальные и систематические эффекты и их влияние на оценку S/N

КазНУ имени аль-Фараби

Цель и план лекции

Цель

Изучение свойств белого и коррелированного шумов, методов их диагностики и способов корректировки оценки S/N, а также подходов к учёту инструментальных и систематических эффектов.

1. Введение
2. Классификация шумов по спектру
3. Влияние корреляции на S/N
4. Модель Понта и фактор бета
5. Методы анализа коррелированного шума
6. Систематические эффекты инструмента
7. Рекомендации по анализу
8. Оценка шума в многоканальных данных

1 Введение

- Большинство простых формул оценки S/N предполагают, что шум независим от отсчёта к отсчёту.

2 Классификация шумов по спектру

Свойства стационарного шума описываются автокорреляционной функцией или спектральной плотностью мощности S(f).

Белый шум

S(f) постоянна, отсчёты независимы.

Фликкер-шум, или шум 1/f

$S(f) \propto 1/f^{\alpha}$ при $\alpha \approx 1$. Присутствует во многих электронных системах и естественных процессах.

Красный шум

$S(f) \propto 1/f^{2}$; большая мощность приходится на низкие частоты. На практике красным называют любой шум, для которого мощность растёт на низких частотах.

В радиоастрономии шум 1/f возникает из-за нестабильности усиления, поэтому применяются переключение на эталонный источник и пространственная модуляция.

2 Классификация шумов по спектру (продолжение)

Коррелированный процесс первого порядка

Значение процесса зависит от предыдущего значения: $x_i = \varphi\, x_{i-1} + \varepsilon_i$; $S(f)$ имеет плато и спад.

В радиоастрономии шум 1/f возникает из-за нестабильности усиления, поэтому применяются переключение на эталонный источник и пространственная модуляция.

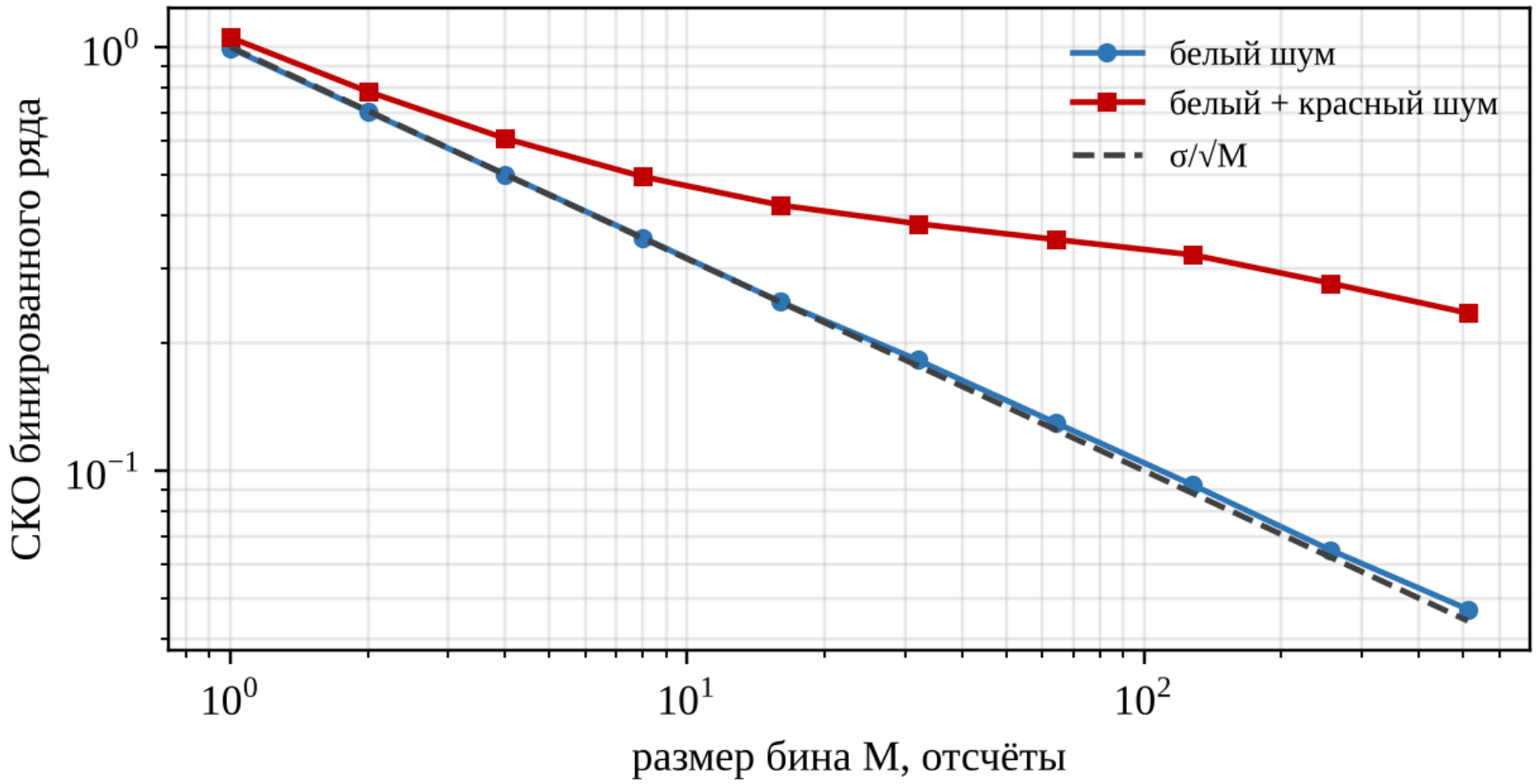
3 Влияние корреляции на S/N

- При наличии корреляций дисперсия среднего отличается от белого значения. Для стационарного процесса с автокорреляцией ρ_k:

$$\sigma_{\bar{x}}^2 = (\sigma^2/N)\,[\,1 + 2\,\Sigma_{k=1}^{N-1}\,(1 - k/N)\,\rho_k\,].$$

- Если ρ_k положительна, шум среднего больше ожидаемого. Эффективное число независимых отсчётов приблизительно равно $N_{eff} \approx N/(1 + 2\Sigma\rho_k)$.
- Диагностика в простейшем случае заключается в построении зависимости СКО бинированных данных от размера бина (time-averaging).

3 Влияние корреляции на S/N (продолжение)



Зависимость СКО бинированного ряда от размера бина для белого и коррелированного шума

4 Модель Понта и фактор бета

- Понт, Цукер и Келос (2006) предложили простое описание шума в наземной транзитной фотометрии, включив белый компонент σ_w и красный компонент σ_r:

$$\sigma_N^2 = \sigma_w^2 / N + \sigma_r^2.$$

- Отсюда следует, что S/N транзита, содержащего N отсчётов, составляет:

$$S/N = \delta / \sqrt{(\sigma_w^2/N + \sigma_r^2)}.$$

- Когда σ_r велико, увеличение числа точек внутри транзита перестаёт приносить выигрыш.

5 Методы анализа коррелированного шума

Вейвлет-анализ

Метод Картера и Винна (2009) использует разложение по вейвлетам для приближённого вычисления правдоподобия при шуме со спектром $1/f^{\gamma}$.

Гауссовы процессы

Ковариационная матрица C задаётся функцией ядра, а логарифм правдоподобия равен:

$$\ln L = -\tfrac{1}{2}\, r^{T} C^{-1} r - \tfrac{1}{2} \ln \det C - (N/2) \ln 2\pi.$$

Методы удаления систематических эффектов

К ним относятся SysRem, TFA и использование вспомогательных параметров.

Остаточный сдвиг (prayer-bead)

Циклический сдвиг остатков относительно модели сохраняет корреляцию, что позволяет оценить неопределённость параметров.

Важно выделить различия между шумом и сигналом: звёздная грануляция может являться помехой при поиске планет, но информативным сигналом для астросейсмологии.

5 Методы анализа коррелированного шума (продолжение)

Инъекции сигналов

Искусственный сигнал вводится в реальный ряд, и определяется доля его обнаружения, что неявно учитывает реальные свойства шума.

Важно выделить различия между шумом и сигналом: звёздная грануляция может являться помехой при поиске планет, но информативным сигналом для астросейсмологии.

6 Систематические эффекты инструмента

- Систематические эффекты удобно разделять на аддитивные и мультипликативные.
- сравнивать измерения одного объекта различными приборами;
- наблюдать постоянные стандартные объекты;
- выполнять тесты с нулевым сигналом, например разность половин данных (null test);
- изменять шаги обработки и контролировать устойчивость результата.
- Погрешность итогового результата записывают в виде суммы статистической и систематической составляющих, причём систематическая обычно не уменьшается с ростом объёма данных.

6 Оценка белого и красного шума и фактора бета

Листинг 14.1.

```
import numpy as np

rng = np.random.default_rng(14)
N = 50000
phi, sr = 0.995, 0.35
e = rng.normal(0, 1, N)
red = np.zeros(N)
for i in range(1, N):
    red[i] = phi * red[i - 1] + e[i]
red *= sr / red.std()
x = rng.normal(0, 1, N) + red

def binned_std(a, M):
    n = a.size // M
    return a[:n * M].reshape(n, M).mean(1).std()

print("   M   СКО бина   при белом  бета")
s1 = x.std()
for M in (1, 10, 50, 200, 1000):
    b = binned_std(x, M)
    w = s1 / np.sqrt(M) * np.sqrt((N / M) / (N / M - 1))
    print(f"{M:5} {b:9.4f} {w:9.4f} {b / w:6.2f}")
```

6 Результат: Оценка белого и красного шума и фактора бета

```
   M   СКО бина   при белом  бета
    1    1.0564    1.0564   1.00
   10    0.4640    0.3341   1.39
   50    0.3615    0.1495   2.42
  200    0.3112    0.0748   4.16
 1000    0.1956    0.0337   5.80
```

6 Влияние красного шума на S/N транзита

Листинг 14.2.

```
depth = 0.3
Nin = 200
s_w = 1.0
for s_r in (0.0, 0.02, 0.05, 0.1):
    sn_w = depth / (s_w / np.sqrt(Nin))
    sn_c = depth / np.sqrt(s_w**2 / Nin + s_r**2)
    print(f"sigma_r={s_r:.2f}: S/N белый {sn_w:.2f}, "
          f"с красным {sn_c:.2f}")

print(f"N_eff/N при phi={phi}: {(1 - phi) / (1 + phi):.4f}")
```

```
sigma_r=0.00: S/N белый 4.24, с красным 4.24
sigma_r=0.02: S/N белый 4.24, с красным 4.08
sigma_r=0.05: S/N белый 4.24, с красным 3.46
sigma_r=0.10: S/N белый 4.24, с красным 2.45
N_eff/N при phi=0.995: 0.0025
```

7 Рекомендации по анализу

- Перед вычислением значимости следует проверить остатки: построить автокорреляционную функцию и спектр мощности, определить зависимость шума от размера бина и сравнить распределение с нормальным.

8 Оценка шума в многоканальных данных

- Если данные содержат много каналов или объектов, коррелированный шум проявляется как общая для них мода.
- Ковариационная матрица данных C определяет оптимальный вес: значение S/N сигнала с шаблоном s в случае коррелированного шума равно:

$$(S/N)^2 = s^T C^{-1} s.$$

- Эта формула обобщает случай белого шума. Для диагонализуемых в частотной области стационарных шумов она совпадает с интегралом вида $\Sigma|\tilde{s}(f)|^2/S(f)$, применяемым в гравитационно-волновых исследованиях.
- В радиоастрономии и космологии ковариационная матрица реликтового излучения оценивается по большому числу смоделированных карт.

8 Оценка шума в многоканальных данных (продолжение)

- На практике коррелированный шум может быть вызван самой обработкой. Сглаживание, интерполяция и выравнивание спектров вводят ковариации между соседними элементами.
- Упражнение: для готовой кривой блеска постройте автокорреляционную функцию и зависимость шума от размера бина; сравните значения S/N транзита, полученные при предположении белого шума и с учётом β-фактора.

9 Разбор примера

1. Для наблюдения транзита пусть $\sigma_w = 1$ ммаг, $\sigma_r = 0{,}3$ ммаг, а $N = 100$ отсчётов.

2. Белая оценка шума на транзите равна $1/\sqrt{100} = 0{,}1$ ммаг. Реальный шум равен $\sqrt{(0{,}1^2 + 0{,}3^2)} = \sqrt{0{,}1} \approx 0{,}316$ ммаг. Значит, при глубине 1 ммаг получаем $S/N \approx 3{,}2$, а не 10.

3. Коэффициент β равен 3,16.

4. Для процесса с $\varphi = 0{,}9$ эффективное число отсчётов составляет $(1 - 0{,}9)/(1 + 0{,}9) \approx 0{,}053$ от общего числа, то есть S/N меньше в $\sqrt{(1/0{,}053)} \approx 4{,}4$ раза, чем при расчёте для белого шума.

10 Вопросы для обсуждения на занятии

Диагностика

Какие графики вы включили бы в статью для демонстрации отсутствия коррелированного шума?

Метод

Когда достаточно применить β-фактор, а когда необходимо использовать гауссовы процессы?

Значимость

Как коррелированный шум влияет на вероятность ложного обнаружения транзита?

Контрольные вопросы

1. Чем белый шум отличается от красного?
2. Что утверждает теорема Винера–Хинчина?
3. Как влияет корреляция на дисперсию среднего значения?
4. Как определить наличие коррелированного шума по бинированным данным?
5. Запишите модель шума Понта и объясните её смысл.
6. Какова роль гауссовых процессов в анализе временных рядов?
7. Какие тесты позволяют выявить систематические эффекты?
8. В чём состоит идея слепого анализа?

Список литературы

1. Pont F., Zucker S., Queloz D. The effect of red noise on planetary transit detection // Monthly Notices of the Royal Astronomical Society. – 2006. – Vol. 373. – P. 231–242.
2. Carter J.A., Winn J.N. Parameter estimation from time-series data with correlated errors: A wavelet-based method and its application to transit light curves // The Astrophysical Journal. – 2009. – Vol. 704. – P. 51–67.
3. Foreman-Mackey D., Agol E., Ambikasaran S., Angus R. Fast and scalable Gaussian process modeling with applications to astronomical time series // The Astronomical Journal. – 2017. – Vol. 154. – Art. 220.
4. Rasmussen C.E., Williams C.K.I. Gaussian Processes for Machine Learning. – Cambridge, MA: MIT Press, 2006.
5. Ivezić Ž., Connolly A.J., VanderPlas J.T., Gray A. Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. – Updated ed. – Princeton: Princeton University Press, 2020.