

# Байесовские блоки

Разбиение на байесовские блоки (Bayesian Blocks) - широко используемый простой непараметрический метод временного анализа кривых блеска, который находит оптимальную сегментацию данных на интервале наблюдений.

Реализация алгоритма описывается в работе Scargle et al. 2013, в которой рассматривается задача обнаружения и характеристики локальной изменчивости во временных рядах и других видах последовательных данных. Цель состоит в том, чтобы выявить и описать статистически значимые вариации, одновременно подавляя неизбежные искажения, возникающие из-за наблюдательных ошибок. Структура алгоритма позволяет использовать его как в ретроспективном режиме (анализировать данные после их регистрации), так и в режиме реального времени, реагируя на первое значительное отклонение сигнала от фонового уровня.

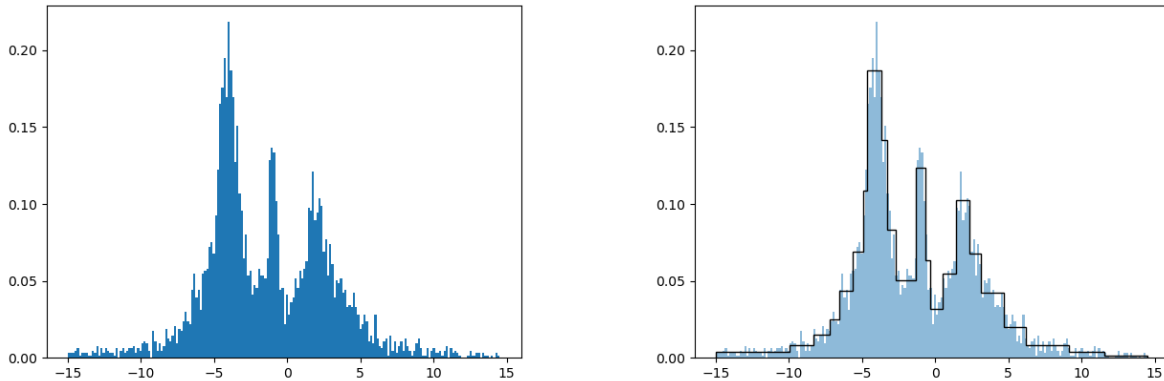


Рис. 1: Пример применения алгоритма разбиения на Байесовские блоки некоторых синтезированных данных

Сейчас мы переходим к краткому описанию алгоритма. Задача состоит в том, чтобы найти оптимальную сегментацию путём максимизации некоторого критерия, называемого *функцией качества* (fitness function). Как мы увидим ниже, важно, чтобы полная функция качества разбиения была аддитивной по блокам. Она должна быть суммой функций качества отдельных блоков, каждая из которых вычисляется только на данных внутри соответствующего блока.

$$F[P] = \sum_{k=1}^{N_{\text{blocks}}} f(B_k) \quad (1)$$

где  $F$  — полное значение функции качества разбиения  $P$ , а  $f(B_k)$  — функция качества  $k$ -го блока.

Если имеется  $N$  ячеек (бинов), то число всех возможных разбиений данных равно  $2^N$ . При  $N = 100$  задача оптимизации перебором становится практически неразрешимой. Однако благодаря аддитивности функции качества по блокам можно использовать алгоритм динамического программирования, который очень похож на математическое доказательство по индукции. Мы последовательно добавляем по одному бину, и для решения задачи для  $n + 1$  бинов используем сохранённую информацию, полученную на предыдущих шагах ( $\leq n$ ).

## Описание алгоритма

Пусть  $P^{\text{opt}}(R)$  обозначает оптимальное разбиение первых  $R$  бинов. В начальном случае  $R = 1$  единственное возможное разбиение является оптимальным — один блок, состоящий только из первого бина. После завершения шага  $R$ , когда найдено оптимальное разбиение  $P^{\text{opt}}(R)$ , значение оптимальной функции качества сохраняется в массиве `best`, а положение последней точки изменения оптимального разбиения сохраняется в массиве `last` (точка изменения — это индекс бина, с которого начинается блок).

Чтобы получить оптимальное разбиение  $P^{\text{opt}}(R + 1)$ , рассматриваются все возможные разбиения, в которых последний блок начинается с бина  $r$  ( $1 \leq r \leq R + 1$ ). Пусть функция качества этого последнего блока равна  $F(r)$ . При условии аддитивности функция качества  $P^{\text{opt}}(R + 1)$  равна сумме  $F(r)$  и функции качества  $P^{\text{opt}}(r - 1)$ , сохранённой на предыдущих шагах в массиве `best`:

$$A(r) = F(r) + \begin{cases} 0, & r = 1 \\ \text{best}(r - 1), & 2 \leq r \leq R + 1 \end{cases} \quad (2)$$

Максимизируя функцию  $A(r)$ , находим оптимальное значение:

$$r^{\text{opt}} = \text{argmax}[A(r)]$$

После выполнения всех  $N$  шагов можно получить окончательное оптимальное разбиение, используя информацию, содержащуюся в массиве `last`, где на каждом шаге сохранялся индекс  $r^{\text{opt}}$ . Индексы точек изменения выбираются из `last` в обратном порядке:

$$i_1 = \text{last}(N), \quad i_2 = \text{last}(i_1 - 1), \quad i_3 = \text{last}(i_2 - 1), \quad \dots \quad (3)$$

Красота этого алгоритма заключается в том, что он выполняет полный поиск оптимума в пространстве  $2^N$  разбиений за время всего лишь порядка  $O(N^2)$ .

## Функция качества (кусочно-постоянная модель)

Теперь необходимо определить функцию качества для некоторой модели. Для бинированных данных здесь будет рассмотрена *кусочно-постоянная модель*. Излучение, накопленное детектором, описывается распределением Пуассона. Поэтому функция правдоподобия для  $n$ -го бина равна:

$$L_n = \frac{(\lambda W_n)^{N_n} e^{-\lambda W_n}}{N_n!} \quad (4)$$

где  $N_n$  — обнаруженное число событий в бине  $n$ ,  $\lambda$  — истинная скорость событий (число событий в единицу времени), а  $W_n$  — ширина бина во временных единицах. В кусочно-постоянной модели  $\lambda$  остаётся постоянной внутри блока. Тогда функция правдоподобия блока ( $k$ ) равна:

$$L^{(k)}(\lambda) = \prod_n L_n = \lambda^{N^{(k)}} e^{-\lambda W^{(k)}} \quad (5)$$

где  $N^{(k)} = \sum_n N_n$  и  $W^{(k)} = \sum_n W_n$  — соответственно общее число событий и ширина блока. Здесь мы опускаем постоянный множитель, который зависит только от полной выборки и не влияет на функцию качества при суммировании по всем блокам. Логарифм правдоподобия:

$$\ln L^{(k)}(\lambda) = N^{(k)} \ln \lambda - \lambda W^{(k)} \quad (6)$$

Максимум достигается при  $\lambda_{\max} = N^{(k)}/W^{(k)}$ , что соответствует среднему числу событий в блоке. Максимальное значение логарифма правдоподобия и является функцией качества блока:

$$f^{(k)} = N^{(k)} \ln \frac{N^{(k)}}{W^{(k)}} - N^{(k)} \quad (7)$$

## Число блоков

В описании алгоритма и функции качества мы нигде не учитывали, сколько блоков содержится в разбиении. Необходимо неявно включить в функцию качества параметр, соответствующий числу блоков модели. Для этого умножим правдоподобие на априорное распределение числа блоков:

$$P(N_{\text{blocks}}) = p_0 \gamma^{-N_{\text{blocks}}} \quad (8)$$

где  $p_0$  — нормировочная константа,  $0 \leq N_{\text{blocks}} \leq N$ , а  $\gamma \geq 1$ . При таком выборе априорного распределения гораздо более вероятно, что  $N_{\text{blocks}} \ll N$ , как мы и ожидаем заранее. Тогда апостериорная вероятность (полная функция качества разбиения  $F[P]$ ) равна:

$$F[P](\vec{\lambda}, N_{\text{blocks}}) = \prod_{k=1}^{N_{\text{blocks}}} [L^{(k)}(\lambda_k)] \cdot P(N_{\text{blocks}}) \quad (9)$$

Тогда:

$$\begin{aligned} \ln F[P] &= \sum_{k=1}^{N_{\text{blocks}}} \ln L^{(k)} - N_{\text{blocks}} \ln \gamma = \\ &= \sum_{k=1}^{N_{\text{blocks}}} (\ln L^{(k)} - \ln \gamma) \end{aligned} \quad (10)$$

Максимизация по  $\vec{\lambda}$  приводит к окончательной функции качества разбиения для кусочно-постоянной модели:

$$\ln F[P] = \sum_{k=1}^{N_{\text{blocks}}} f^{(k)} \quad (11)$$

$$f^{(k)} = \left( N^{(k)} \ln \frac{N^{(k)}}{W^{(k)}} - N^{(k)} \right) - \text{ncp\_prior} \quad (12)$$

где  $f^{(k)}$  — функция качества  $k$ -го блока,  $N^{(k)} = \sum_n N_n$  и  $W^{(k)} = \sum_n W_n$  — соответственно общее число событий и ширина блока.  $\text{psr\_prior} \equiv \log \gamma$  — параметр от 0 до  $+\infty$ , который управляет числом блоков: значение 0 соответствует  $N_{\text{blocks}} = N$ , тогда как большое значение соответствует одному или нескольким блокам, покрывающим все данные.

## Кусочно-линейная модель

Вывод функции качества для кусочно-линейной модели проводится аналогично, необходимо лишь ввести новый параметр наклона  $a$ . В этой модели данные могут изменяться внутри блока:

$$\lambda \rightarrow \lambda(1 + a(n - n_i)) = \lambda(1 + a\Delta n) \quad (13)$$

где  $n$  — номер бина,  $n_i$  — номер первого бина блока.

Логарифм функции правдоподобия  $k$ -го блока равен

$$\ln L^{(k)} = N^{(k)} \ln \lambda + \sum_{n=n_i}^{n_f} N_n \ln[(1 + a\Delta n)W_n] - \sum_{n=n_i}^{n_f} \lambda(1 + a\Delta n)W_n \quad (14)$$

При решении кусочно-постоянной модели мы ранее опускали второй член, но теперь он необходим из-за зависимости от нового параметра.

Из условия  $d/d\lambda \ln L^{(k)} = 0$  можно найти  $\lambda_{\text{max}}$ :

$$\lambda_{\text{max}} = \frac{N^{(k)}}{\sum_n (1 + a\Delta n)W_n} \quad (15)$$

И тогда

$$\ln L^{(k)}(\lambda_{\text{max}}, a) = N^{(k)} \left( \ln \frac{N^{(k)}}{\sum_n (1 + a\Delta n)W_n} - 1 \right) + \sum_n N_n \ln[(1 + a\Delta n)W_n] \quad (16)$$

Максимальное значение этого выражения определяет функцию качества  $k$ -го блока (без учёта вычитания  $\text{psr\_prior}$ ). Если  $a = 0$ , мы получаем уже известную функцию для кусочно-постоянной модели (7).

Существует нижняя граница для параметра  $a$ , возникающая из требования, чтобы математическое ожидание числа событий  $\lambda(1 + a\Delta n)$  было положительным:

$$a > -1/N$$

Реализацию алгоритма разложения на байесовские блоки с двумя моделями из данной работы можно найти в репозитории:

[https://github.com/fedsvir/Bayesian\\_block.git](https://github.com/fedsvir/Bayesian_block.git)

## 1 ССЫЛКИ

[1] Scargle, J et al. (2013), ApJ, Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations, 764, 2, 167. doi:10.1088/0004-637X/764/2/167