Исследования по задаче “размазанной” разреженности

Алиева Наталия Дмитриевна, гр. 622

Санкт-Петербургский государственный университет
Математико-механический факультет
Статистическое моделирование

Научный руководитель: д.ф.-м.н., профессор Ермаков М. С.

Рецензент: д.ф.-м.н., профессор Шевляков Г. Л.

Санкт-Петербург, 2017г.

Мотивация

- Обработка информации: часто на вход подаются некачественные данные, необходимо максимально точно восстановить исходную картину.
- Умение работать с разреженными данными и получать на выходе более качественный материал для дальнейшей обработки.
- Множество областей применения.

Постановка задачи

Рассмотрим отрезок $[0, 1]$ и модель регрессии:

$$Y_i = f(X_i) + \xi_i, i = 1, \ldots, n$$

где:

- функция f принимает значения 0 или 1
- $X_i, Y_i \in \mathbb{R}$, ξ_i — н.о.р случайные величины $\sim \mathcal{N}(0, \sigma^2)$, где σ^2 — дисперсия.

Пример

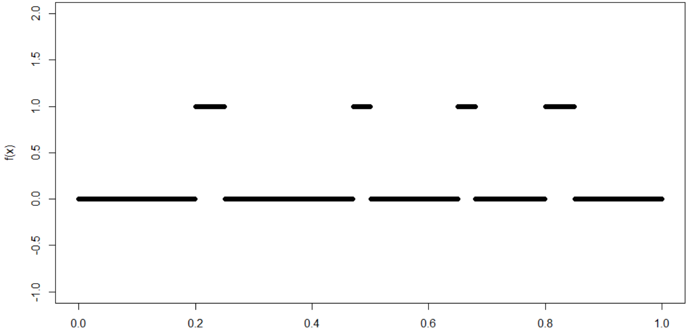

Рис. : Пример скачкообразной функции

Работы по обнаружению моментов разладки

- В. Спокойный (1998),
- Arias-Castro E., Donoho D. L. (2005),
- Arias-Castro E., Meng Wang (2013).

Пример

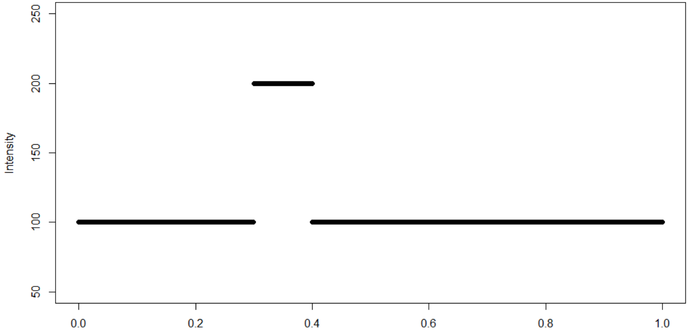

Рис. : Интенсивность пуассоновского процесса

Модификация для пуассоновских процессов

Рассмотрим $X(t), t \in [0,1]$ — неоднородный пуассоновский процесс. Интенсивность $\lambda(t)$ имеет вид:

$$\lambda(t) = \begin{cases} an, & t \in [a_0, a_1], \\ n, & t \notin [a_0, a_1], \end{cases}$$

где $a > 1$, $0 < a_0 < a_1 < 1$.

Полученные оценки

Нижняя граница

Ширина оцениваемого интервала (квадрата): $Cn^{-1}\log n$, где $C > 0$.

- Одномерный случай:

$$C = \frac{\ln(1+a)}{(2\ln 2 - 1)(a-1)}.$$

- Двумерный случай:

$$C = \sqrt{\frac{\ln(1+a)}{(2\ln 2 - 1)(a-1)}}.$$

Методы обнаружения моментов разладки

- Исчерпывающий поиск (Braun et al., 2000; Jackson et al., 2005),
- Пошаговый отбор (Scott & Knott, 1974; Vostrikova, 1981),
- Метод обнаружения скачков с помощью диадических интервалов (Arias-Castro E., Donoho D. L., 2005)
- Адаптивный метод оценивания функций регрессии (В. Спокойный, 1998)
- Алгоритм отсеивания и ранжирования SaRa (Hao N., Niu Y. S., Zhang H., 2013)
- Алгоритм SMUCE (Munk A., Frick K., Sieling H., 2010)

Методы обнаружения моментов разладки

- Исчерпывающий поиск (Braun et al., 2000; Jackson et al., 2005),
- Пошаговый отбор (Scott & Knott, 1974; Vostrikova, 1981),
- Метод обнаружения скачков с помощью диадических интервалов (Arias-Castro E., Donoho D. L., 2005)
- Адаптивный метод оценивания функций регрессии (В. Спокойный, 1998)
- Алгоритм отсеивания и ранжирования SaRa (Hao N., Niu Y. S., Zhang H., 2013)
- Алгоритм SMUCE (Munk A., Frick K., Sieling H., 2010)

Пример

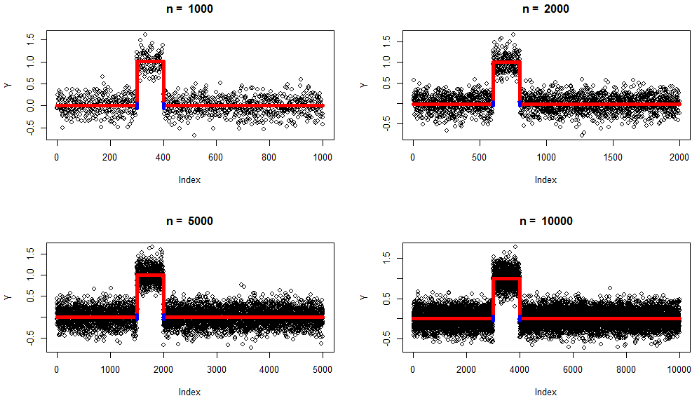

Рис. : Пример обнаружения одного скачка внутри интервала с помощью алгоритма Мунка для гауссовского шума

Оценка высоты скачка

Пусть $Y_1, \dots Y_n$ — входные данные.
Пусть a — высота скачка:

$$\hat{a} = \frac{1}{m}\sum_{i=1}^{m} Y_i,$$

где m — число точек, попавших в оцененный интервал.

Пример

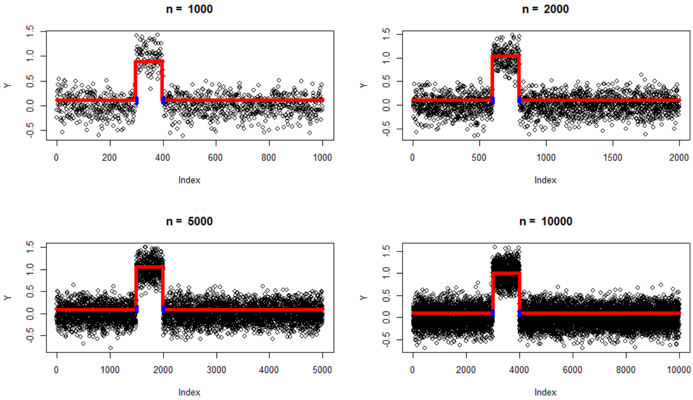

Рис. : Пример обнаружения одного скачка внутри интервала с помощью SaRa для гауссовского шума

Пример

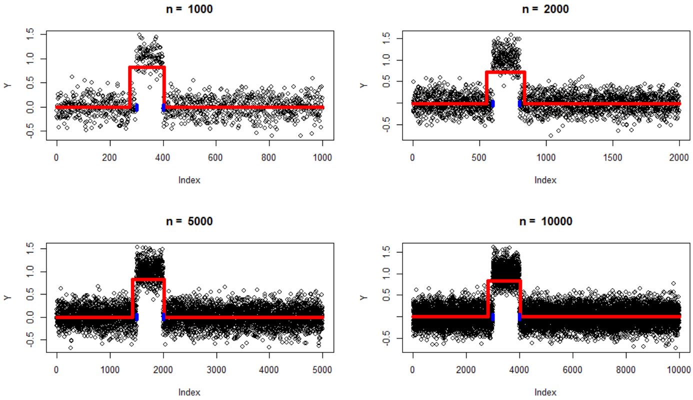

Рис. : Пример обнаружения одного скачка внутри интервала с помощью алгоритма В. Г. Спокойного для гауссовского шума

Пример

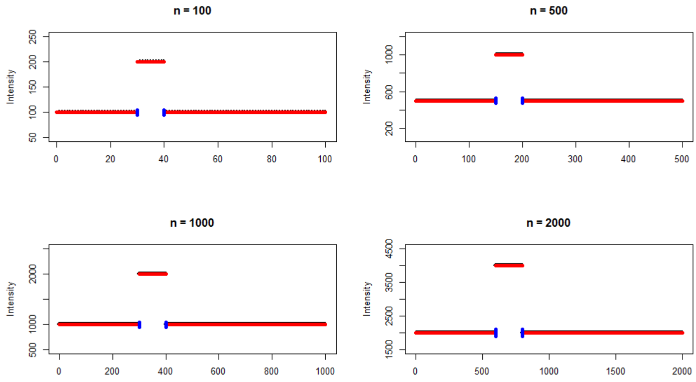

Рис. : Пример обнаружения одного скачка внутри интервала с помощью алгоритма Мунка для интенсивности пуассоновского процесса

Пример

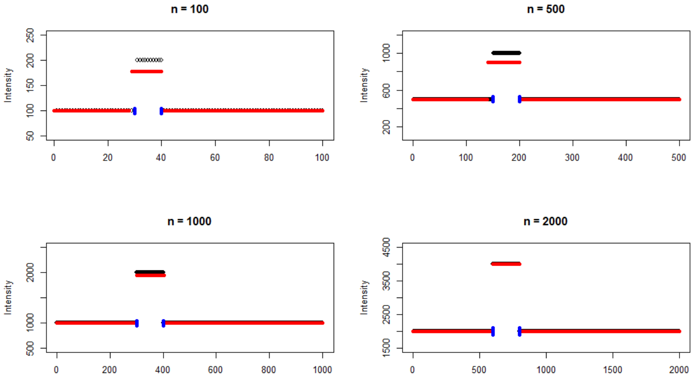

Рис. : Пример обнаружения одного скачка внутри интервала с помощью SaRa для интенсивности пуассоновского процесса

Пример

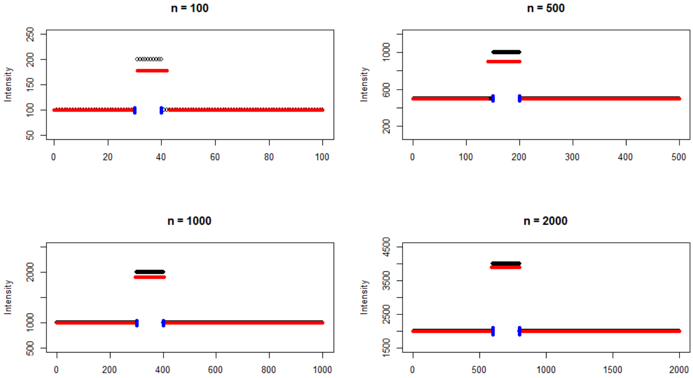

Рис. : Пример обнаружения одного скачка внутри интервала с помощью алгоритма В. Г. Спокойного для интенсивности пуассоновского процесса

Результаты

- Найдены нижные границы ширины прямоугольного скачка для пуассоновских процессов.
- Результат продолжен на двумерный случай и для регрессионной модели с гауссовским шумом, и для пуассоновских процессов.
- Проведено сравнение 3–х наиболее известных алгоритмов обнаружения моментов разладки.