Универсальные локально-линейные ядерные оценки для производной регрессионной функции

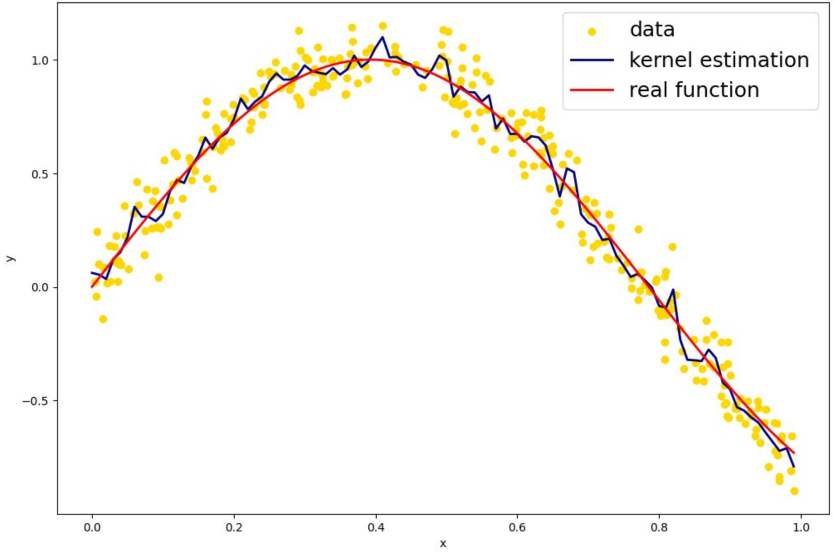
Докладчик: Петренко Сергей, НГУ

Научный руководитель: Линке Ю.Ю. д-р физ.-мат. наук, ИМ СО РАН

Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации № 075-15-2025-348

mathematics & mechanics

ПОСТАНОВКА ЗАДАЧИ РЕГРЕССИИ



Пусть наблюдения $\{(z_i, X_i);\ i = 1, ..., n\}$ имеют структуру:

$$X_i = f(z_i) + \varepsilon_i$$

где $\{z_i\}$ - регрессоры, которые известны, $\{\varepsilon_i\}$ - ненаблюдаемые случайные ошибки (центрированные случайные величины).

Задача состоит в оценивании гладкой регрессионной функции или её производных:

$$f : [0, 1] \to \mathbb{R}$$

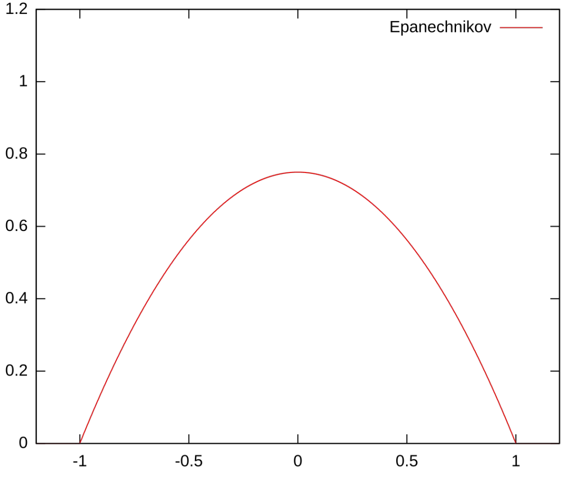
БОЛЬШОЕ ЧИСЛО РАБОТ ПОСВЯЩЕНО ОЦЕНИВАНИЮ ПРОИЗВОДНОЙ:

- Liu, Yang (2024), Prasangika, Tang, Yao (2023), Liu, Li (2023), Comte, Marie (2023),
- Bouzebda, Chaouch, Biha (2022), Calonico, Cattaneo Farrell (2022),
- Liu, De Brabanter (2020), Cattaneo, Farrell (2020), Xie, Sun, Liu (2020),
- Wang, Yu, Lin, Tong (2019), Bercu, Capderou, Durrieu (2019), Dai, Tong, Genton (2016), Song (2016),
- Wang, Lin (2015), Benhenni, Degras (2014), Zheng, Gallagher, Kulasekera (2013),
- De Brabanter, K., De Brabanter, J., De Moor, Gijbels (2013), Martins-Filho, Saraiva (2012),
- Chen, Zhang (2010), Lu, Linton (2007), Newell, Einbeck (2007),
- Delecroix, Rosa (2007), Boente, Rodriguez (2006), Horng (2006), Qin, Tsao (2005),
- Gijbels, Goderniaux (2004), Beran, Feng (2002), Jiang, Mack (2001), Zhou, Wolfe (2000),
- Fan, Gasser, Gijbels, Brockmann, Engel (1997), Masry, Fan (1997), Fan, Gijbels (1995),
- Mack, Muller (1989), Muller, Schmitt (1987), Hardle , Gasser (1985), Gasser, Muller (1984) и др.

В этих работах рассматриваются различные условия на параметры модели: регрессоры, ошибки и тд. Нас же интересуют условия на регрессоры. Мы хотим построить оценки при более общих условиях на регрессоры, чем известные ранее.

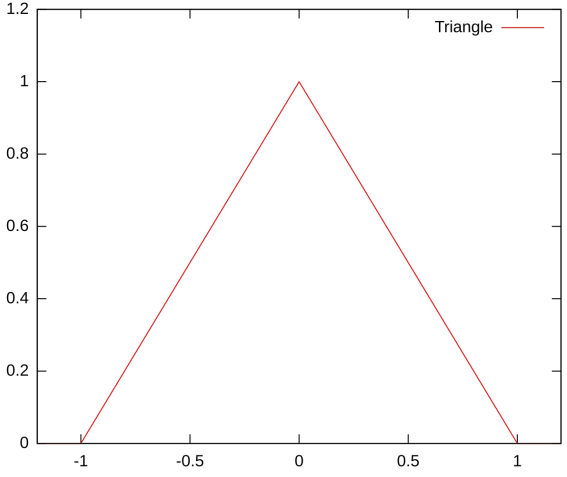
ЯДЕРНЫЕ ФУНКЦИИ

Ядро Епанечникова



$$K(t) = \frac{3}{4}(1 - t^2)I(|t| \leq 1)$$

Треугольное ядро



$$K(t) = (1 - |t|)I(|t| \leq 1)$$

ЯДЕРНЫЕ ОЦЕНКИ

Пусть $K(t)$ - ядро сглаживания - плотность симметричного распределения (обычно с носителем [-1; 1]).

$K_h(u) = h^{-1}K(\frac{u}{h})$ - Плотность симметричного распределения на [-h;h]. Нормированное ядро по h (ширине окна).

Оценка Надарая-Ватсона (1964):

$$f^*_{NW}(t) = \arg\min_{\theta} \sum_{i=1}^{n} (X_i - \theta)^2 K_h(t - z_i)$$

$$f^*_{NW}(t) = \frac{\sum_{i=1}^{n} X_i K_h(t-z_i)}{\sum_{i=1}^{n} K_h(t-z_i)}$$

ЛОКАЛЬНО-ЛИНЕЙНЫЕ ОЦЕНКИ

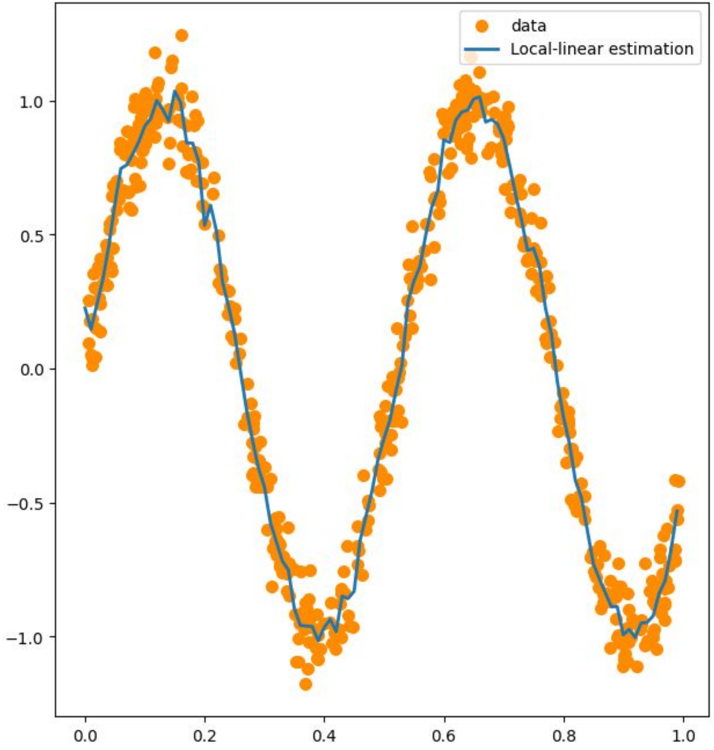
$$f^*_{NW}(t) = \arg\min_{\theta} \sum_{i=1}^{n} (X_i - \theta)^2 K_h(t - z_i)$$

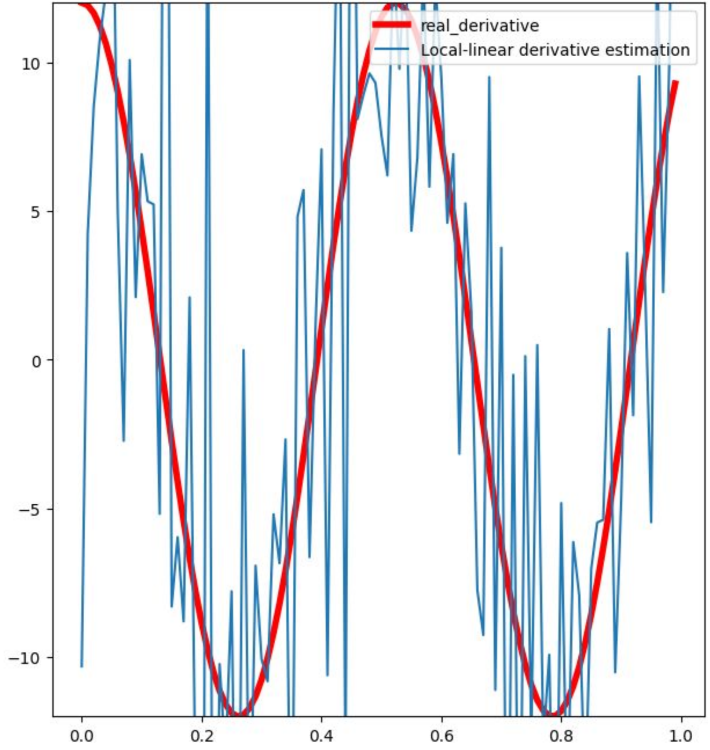
+

$$f(z_i) \approx f(t) + f'(t)(z_i - t)$$

$$f^*_{LL}(t) = \arg\min_{(\theta_0,\theta_1)} \sum_{i=1}^{n} (X_i - \theta_0 - \theta_1(z_i - t))^2 K_h(t - z_i)$$

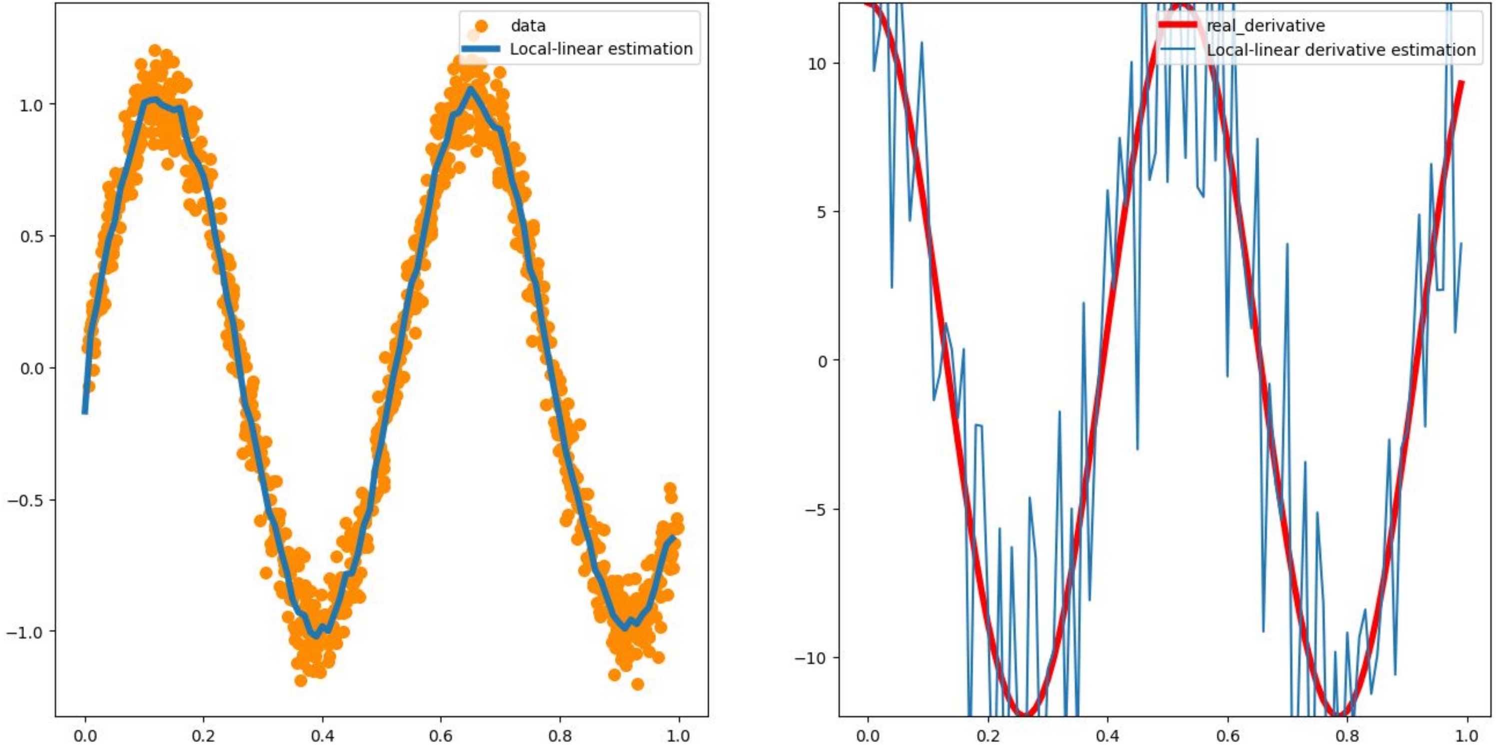
h = 0.01, n = 500





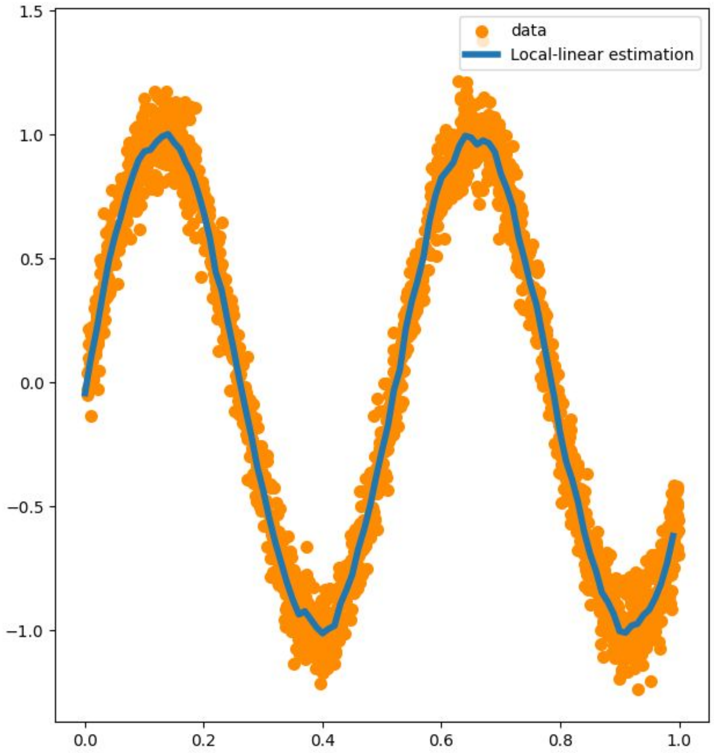
Сходимость локально линейных оценок в зависимости от количества данных. Функция (слева), производная(справа).

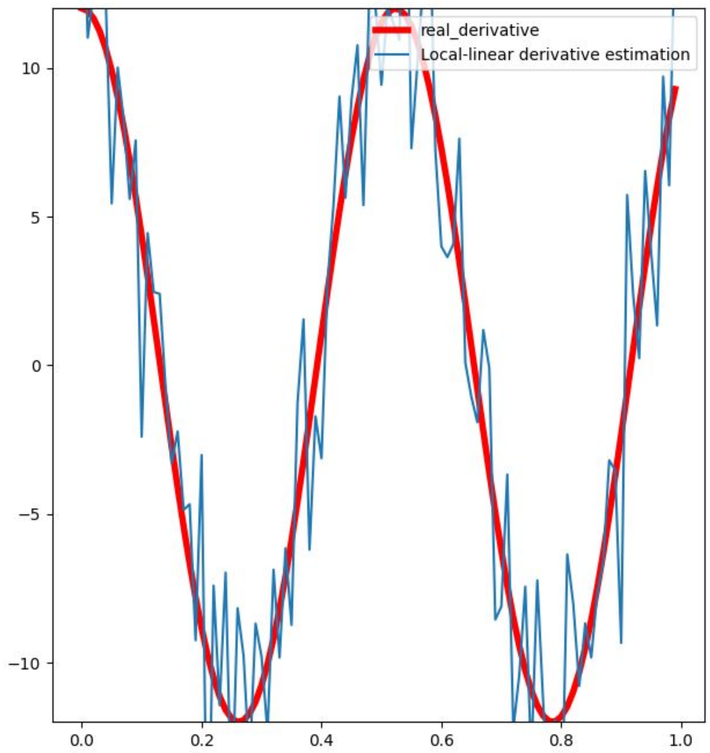
h = 0.01, n = 1000



Сходимость локально линейных оценок в зависимости от количества данных. Функция (слева), производная(справа).

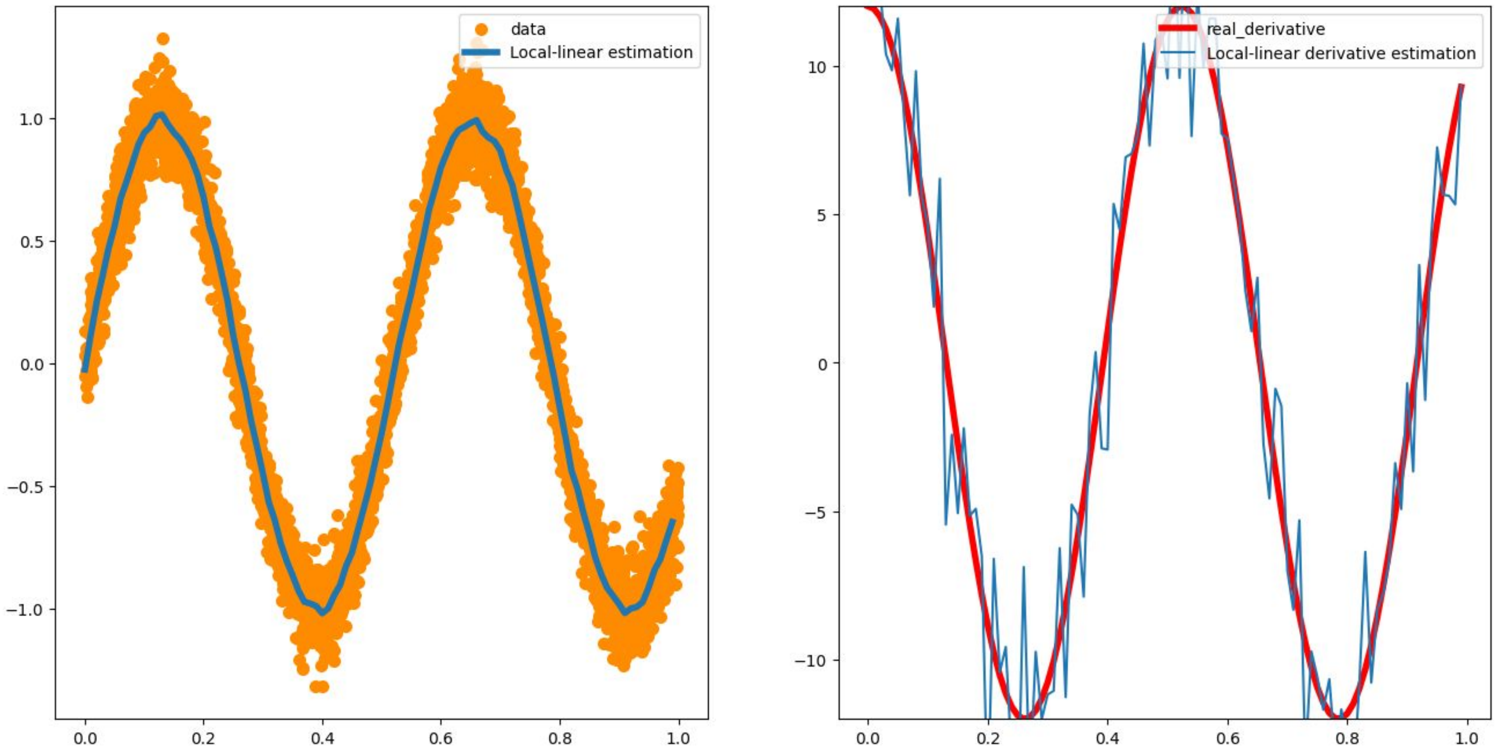
h = 0.01, n = 2000





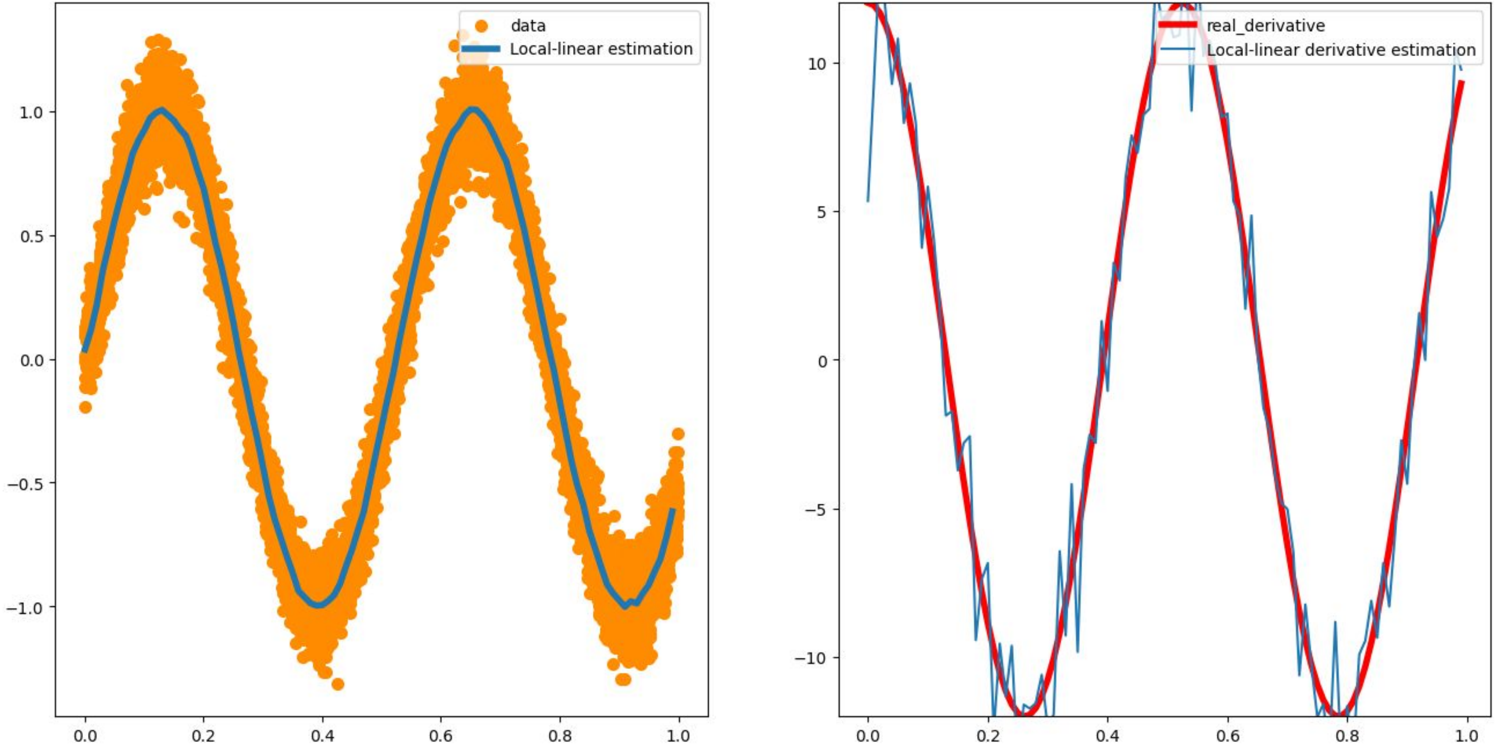
Сходимость локально линейных оценок в зависимости от количества данных. Функция (слева), производная(справа).

h = 0.01, n = 4000



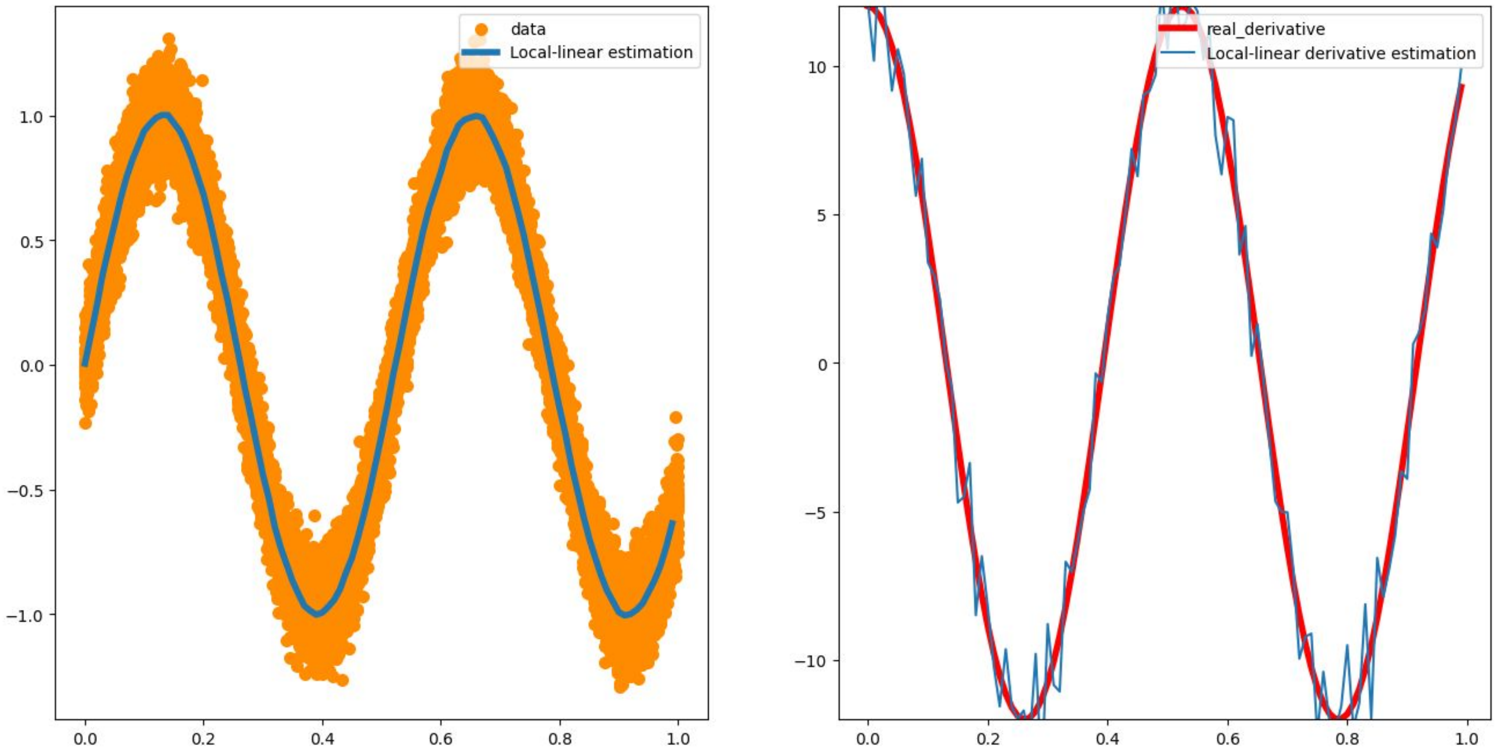
Сходимость локально линейных оценок в зависимости от количества данных. Функция (слева), производная(справа).

h = 0.01, n = 8000



Сходимость локально линейных оценок в зависимости от количества данных. Функция (слева), производная(справа).

h = 0.01, n = 16000



Сходимость локально линейных оценок в зависимости от количества данных. Функция (слева), производная(справа).

УСЛОВИЯ НА РЕГРЕССОРЫ В РАБОТАХ ПРЕДШЕСТВЕННИКОВ

Модели с **фиксированными** и **случайными** регрессорами принято рассматривать отдельно .

- Если $\{z_i\}$ неслучайны, то предполагаются те или иные условия **регулярного плотного заполнения** регрессорами области определения регрессионной функции. Например:

$$z_i = i/n, \qquad z_i = g(i/n) + o(1/n), \qquad \max_{i\le n}(z_i - z_{i-1}) = O(1/n)$$

- Если $\{z_i\}$ случаины, то рассматриваются:
  - **независимые и одинаково распределенные величины**
  - **слабо зависимые величины** (например, используются условия перемешивания, схемы скользящих средних, ассоциированные с. в., марковские или мартингальные свойства, авторегрессия и др.)

mathematics & mechanics

Мотивация

- Можно ли построить оценки для производной неизвестной регрессионной функции при **близких к минимальным и наглядных условиях на регрессоры** (без использования условий слабой зависимости, в случае сильно зависимых величин)?
- Ранее ответ на аналогичный вопрос был получен в задаче оценивания регрессионной функции: для того, чтобы оценить регрессионную функцию, относительно регрессоров достаточно требовать лишь условие плотного заполнения этими величинами области задания регрессионной функции.

НОВЫЙ КЛАСС ОЦЕНОК

Пусть регрессоры $\{z_i;\ i = 1, ..., n\}$ - это набор наблюдаемых случайных величин, с неизвестным распределением на [0;1]; Не обязательно независимых и одинаково распределённых, регрессоры могут зависеть от n. Частный случай: детерминированные регрессоры.

Обозначим через $z_{n:1} \leq \ldots \leq z_{n:n}$ элементы вариационного ряда. $\Delta z_{ni} = z_{n:i} - z_{n:i-1}$. Отклики, относящиеся к $z_{n:i}$ обозначим X_{ni}

Введём в рассмотрение следующий класс оценок для производных:

$$\widehat{f'_{ULL}}(t) = \sum_{i=1}^{n} \frac{(z_{n:i} - t)w_{n0}(t) + w_{n1}(t)}{w_{n0}(t)w_{n2}(t) - w_{n1}^2(t)} X_{ni} K_h(t - z_{n:i}) \Delta z_{ni} \cdot$$

$$w_{nj}(t) = \sum_{i=1}^{n} (t - z_{n:i})^j K_h(t - z_{n:i}) \Delta z_{ni}, \quad j = 0, 1, 2$$

ИНТУИЦИЯ ОЦЕНКИ

$$w_{nj}(t) = \sum_{i=1}^{n} (t - z_{n:i})^j K_h(t - z_{n:i}) \Delta z_{ni}, \quad j = 0, 1, 2$$

$$w_j(t) = \int_{t-h}^{t+h} (t - z)^j K_h(t - z) dz = h^j \int_{-1}^{1} v^j K(v) dv.$$

при малых Δz_{ni} выполнено: $w_{nj}(t) \approx w_j(t)$

УСЛОВИЕ НА РЕГРЕССОРЫ, ГАРАНТИРУЮЩЕЕ ПОТОЧЕЧНУЮ СОСТОЯТЕЛЬНОСТЬ ПО ВЕРОЯТНОСТИ

Соотношение на регрессоры, гарантирующее поточечную состоятельность оценки с ростом n:

$$\delta_n = \max_{1 \le i \le n+1} \Delta z_{ni} \xrightarrow{p} 0$$

- Универсально относительно стохастической природы регрессоров, включает как случайные, так и детерминированные регрессоры
- Зависимость регрессоров может быть более сильной, чем в известных ранее работах
- В оценивании никак не используется характер зависимости регрессоров.

ДОПОЛНИТЕЛЬНЫЕ УСЛОВИЯ

$(\boldsymbol{K})$: Ядерная Функция $K(t), t \in \mathbb{R}$ является плотностью симметричного распределения с носителем [-1, 1], Кроме того $K(t)$ удовлетворяет условию Липшица с константой $L \geq 1$ и $K(\pm 1) = 0$.

$(\boldsymbol{M})$: $f \in C^1[0,1]$

$(\boldsymbol{E})$ При всех $n \geq 1$ случайные погрешности $\{\varepsilon_i;\ i = 1, \ldots, n\}$ с вероятностью 1 при всех $i, j \leq n, \quad i \neq j$ удовлетворяют условиям:

$$\mathbb{E}_{\mathcal{F}_n} \varepsilon_i = 0, \qquad \sup_{i \leq n} \mathbb{E}_{\mathcal{F}_n} \varepsilon_i^2 \leq \sigma^2, \qquad \mathbb{E}_{\mathcal{F}_n} \varepsilon_i \varepsilon_j = 0, \quad i \neq j$$

ОСНОВНОЙ РЕЗУЛЬТАТ

ТЕОРЕМА 1.

Пусть выполнены условия (K), (M), (E) и предельное условие на регрессоры, тогда существует $h = h_n \to 0$ последовательность положительных чисел такая, что при $n \to \infty$:

$$\mathbb{E}\frac{\delta_n}{h^3} \to 0, \quad \mathbb{E}\frac{\delta_n^2}{h^4} \to 0$$

Тогда имеет место предельное соотношение:

$$\forall t \in [0,1]: \quad \widehat{f'}_{ULL}(t) \xrightarrow{\mathrm{P}} f'(t)$$

ОСНОВНОЙ РЕЗУЛЬТАТ

ТЕОРЕМА 2.

Пусть выполнены условия (K), (M), (E) и предельное условие на регрессоры, тогда существует $h = h_n \to 0$ последовательность положительных чисел такая, что при $n \to \infty$:

$$\mathbb{E}\frac{\delta_n}{h^4} \to 0$$

Тогда имеет место предельное соотношение:

$$\sup_{t\in[0,1]} |\widehat{f'}_{ULL}(t) - f'(t)| \xrightarrow{\mathrm{p}} 0$$

АСИМПТОТИЧЕСКАЯ НОРМАЛЬНОСТЬ

Нам понадобится некоторое упрощение условия на случайные ошибки.

(**E1**) При всех $n \geq 1$ случайные погрешности $\{\varepsilon_i;\ i = 1, \ldots, n\}$ представимы в виде

$$\varepsilon_i = \sigma(z_i)\xi_i, \quad i = 1, \ldots, n,$$

где неизвестная функция $\sigma(t)$, $t \in [0, 1]$, непрерывна, $\{\xi_i;\ i = 1, \ldots, n\}$– независимые одинаково распределенные случайные величины с нулевым средний и единичной дисперсией, не зависящие от набора регрессоров $\{z_i;\ i = 1, \ldots, n\}$.

АСИМПТОТИЧЕСКАЯ НОРМАЛЬНОСТЬ

Усиление теоремы Гаека-Шидака.

Теорема 2. Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых одинаково распределенных центрированных случайных величин с конечным вторым моментом $\sigma^2 = \mathbb{E}\xi_1^2$, а случайные величины a_{nk}, где $k = 1, \ldots, n$; $n = 1, 2, \ldots$, не зависящие от ξ_i, где $i = 1, 2, \ldots$, удовлетворяют следующим двум условиям:

$$\max_{1 \le k \le n} |a_{nk}| \xrightarrow{p} 0, \quad \sum_{i=1}^{n} a_{nk}^2 = 1 \quad \text{при всех } n.$$

Тогда имеет место предельное соотношение

$$\sum_{k=1}^{n} a_{nk}\xi_k \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

АСИМПТОТИЧЕСКАЯ НОРМАЛЬНОСТЬ

Теорема 3. Пусть выполнены условия $(M), (E1), (K)$ и при некотором $t \in [0,1]$ и $h \equiv h_n$

$$\frac{\max_{k\le n} |((z_{n:k} - t)w_{n0}(t) + w_{n1}(t))K_h(t - z_{n:k})\sigma(z_{n:i})\Delta z_{nk}|}{(\sum_{i=1}^{n}((z_{n:i} - t)w_{n0}(t) + w_{n1}(t))^2 K_h^2(t - z_{n:i})\sigma^2(z_{n:i})(\Delta z_{ni})^2)^{1/2}} \xrightarrow{\mathrm{P}} 0$$

Тогда имеет место сходимость по распределению:

$$B_{n,h}^{-1}(t)\left(\hat{f}_{n,h}(t) - f'(t) - r_{n,h}(t)\right) \xrightarrow{d} \mathcal{N}(0,1),$$

где:

$$r_{n,h}(t) = \mathbb{E}_{\mathcal{F}_n}(\hat{f}_{n,h}(t) - f'(t))$$

$$B_{n,h}^2(t) = \sum_{i=1}^{n} \beta_{n,i}^2(t) K_h^2(t - z_{n:i})\sigma^2(z_{n:i})(\Delta z_{ni})^2.$$

Новизна результата

Полученный результат позволяет:

- Значительно ослабить известные ограничения на регрессоры в задаче оценивания производной регрессионной функции;
- Оценивать производную функции при отсутствии какой-либо информации о характере зависимости регрессоров;
- В данной задаче непараметрической регрессии в едином подходе рассматривать модели с детерминированными и случайными регрессорами.

Спасибо за внимание

$$\widehat{\sigma^2}_{n,h,\overline{h}}(t) = \frac{\sum_{i=1}^{n}\big(X_{ni} - f^{*}_{n,h}(z_{n:i})\big)^2 \overline{K}_{\overline{h}}(t - z_{n:i})\Delta z_{ni}}{\sum_{i=1}^{n}\overline{K}_{\overline{h}}(t - z_{n:i})\Delta z_{ni}},$$

$$\beta_{n,i}(t) = \frac{(z_{n:i} - t)w_{n0}(t) + w_{n1}(t)}{w_{n0}(t)w_{n2}(t) - w_{n1}^2(t)}$$

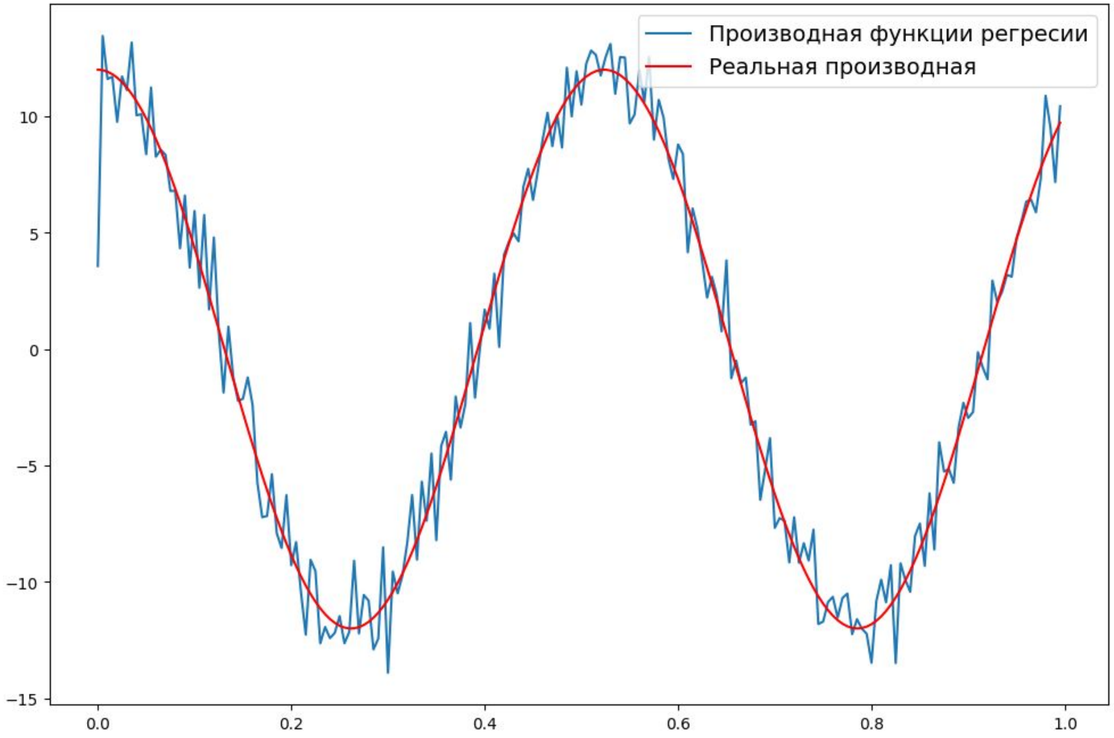
$$r_{n,h}(t) = \sum_{i=1}^{n}\beta_{n,i}(t)\big(f(z_{n:i}) - f(t) - f'(t)(z_{n:i} - t)\big)K_h(t - z_{n:i})\Delta z_{ni}$$

$$\beta_{n,i}(t) = \frac{(z_{n:i} - t)w_{n0}(t) + w_{n1}(t)}{w_{n0}(t)w_{n2}(t) - w_{n1}^2(t)}$$

$$r_{n,h}(t) = \sum_{i=1}^{n} \beta_{n,i}(t)\big(f(z_{n:i}) - f(t) - f'(t)(z_{n:i} - t)\big)K_h(t - z_{n:i})\Delta z_{ni}$$

$$\nu_{n,h}(t) = \sum_{i=1}^{n} \beta_{n,i}(t)K_h(t - z_{n:i})\Delta z_{ni}\varepsilon_{ni}$$

$$\boxed{\widehat{f'}_{n,h}(t) = f'(t) + r_{n,h}(t) + \nu_{n,h}(t)}$$



Оценка производной функции sin(12x), полученная дифференцированием оценки Надарая-Ватсона при h = 0.005, n = 80000.

КРИТИКА ДИФФЕРЕНЦИРОВАНИЯ ЯДЕРНЫХ ОЦЕНОК ДЛЯ ФУНКЦИИ

De Brabanter, K., De Brabanter, J., De Moor, B., Gijbels, I. (2013) Derivative estimation with local polynomial fitting // J. Mach. Learn. Res.

Было бы заманчиво дифференцировать ядерные оценки для регрессионной функции для получения оценок для производных регрессионной функции, но это может привести к неправильным оценкам для производных, поскольку такая процедура может работать только в случае, если очень точно оценена сама функция. Иначе прямое дифференцирование может привести к накоплению ошибок, которые увеличиваются с ростом производной.

Dai W., Tong T., Genton M.G. (2016) Optimal estimation of derivatives in nonparametric regression // J. Mach. Learn. Res.

Даже хорошая оценка для регрессионной функции может не гарантировать получение хороших оценок для производных.

К ВОПРОСУ О СКОРОСТИ СХОДИМОСТИ

Рассмотрим случай **фиксированных регрессоров** с **равномерной решёткой**. То есть:

$$z_i = i/n, \quad \delta_n = 1/n$$

- Локально-линейные: $nh^3 \to \infty$
- Сплайны: $nh^3 \to \infty$
- Универсально локально-линейные: $nh^3 \to \infty$

$$f^*_{NW}(t) = \frac{\sum_{i=1}^n X_i K_h(t-z_i)}{\sum_{i=1}^n K_h(t-z_i)}$$

Асимптотическая сложность построения оценки в точке t: $O(n)$
Сложность построения оценки на всей сетке: $O(n/h)$

$$f^*_{ULL}(t) = \sum_{i=1}^{n} \frac{W_{n0}(t)(z_{n:i}-t) - W_{n1}(t)}{W_{n2}(t)W_{n0}(t) - W_{n1}(t)^2} X_{ni} K_h(t - z_{n:i}) \Delta z_{ni},$$

$$W_{nj}(t) = \sum_{i=1}^{n} (z_{n:i} - t)^j K_h(t - z_{n:i}) \Delta z_{ni}, \quad j = 0, 1, 2.$$

Асимптотическая сложность построения оценки в точке t: $O(n)$
Сложность построения оценки на всей сетке: $O(n/h)$