Универсальные ядерные оценки в непараметрической регрессии

Ю. Ю. Линке
Институт математики им. С.Л. Соболева

Вторая конференция Математических центров России
Москва, 7 – 11 ноября 2022

Введение

Пусть наблюдения $X_1, \ldots, X_n$ имеют структуру

$$X_i = f(z_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

где $\{z_i\}$ (регрессоры, точки дизайна) известны, $\{\varepsilon_i\}$ — ненаблюдаемые случайные ошибки. Задача состоит в оценивании неизвестной непрерывной регрессионной функции $f(t)$.

Можно ли построить равномерно состоятельные оценки для неизвестной регрессионной функции без использования традиционных условий зависимости элементов дизайна (при близких к минимальным и наглядных условиях на точки дизайна)?

Linke Yu. Yu., Borisov I. S., Ruzankin P. S., Kutsenko V. A., Yarovaya E. B., Shalnova S.A. (2022) Universal local linear kernel estimators in nonparametric regression // **Mathematics**. 10(15).

Borisov I. S., Linke Yu. Yu., Ruzankin P.S. (2021) Universal weighted kernel-type estimators for some class of regression models // **Metrika**. 84(2), 141–166.

Ядерные оценки. $X_i = f(z_i) + \varepsilon_i, \quad i = 1, \dots, n.$

Выберем $h > 0$ и назовем его шириной окна. Простейший вариант оценки для f:

$$f^*_{SA}(t) = \frac{\sum_{i=1}^n X_i \cdot \mathbb{I}(t-h \le z_i \le t+h)}{\sum_{i=1}^n \mathbb{I}(t-h \le z_i \le t+h)}.$$

$$f^*_{SA}(t) - f(t) = \frac{\sum (f(z_i) - f(t))\mathbb{I}(\dots)}{\sum \mathbb{I}(\dots)} + \frac{\sum \varepsilon_i \mathbb{I}(\dots)}{\sum \mathbb{I}(\dots)}, \qquad h = h_n \to 0, \quad nh \to \infty \text{ при } n \to \infty.$$

Недостатки $f^*_{SA}(t)$:

- функция $f^*_{SA}(t)$ имеет разрывы;
- все наблюдений X_i, для которых $z_i \in [t-h, t+h]$, учитываются в оценке в равной степени.

Пусть функция $K(t)$ (ядро сглаживания) является плотностью симметричного распределения (обычно, с носителем $[-1, 1]$ или $\mathbb{R}$).

Оценка Надарая–Ватсона: $$f^*_{NW}(t) = \frac{\sum_{i=1}^n X_i K_h(t - z_i)}{\sum_{i=1}^n K_h(t - z_i)}, \qquad K_h(u) = h^{-1} K(u/h).$$

Популярными ядрами являются:
Епанечниково $K(u) = \frac{3}{4}(1-u^2)\mathbb{I}\{|u| \le 1\}$, Гауссово (нормальное) $K(u) = \frac{1}{\sqrt{2\pi}} \exp\{-u^2/2\}$, треугольное $K(u) = (1 - |u|)\mathbb{I}\{|u| \le 1\}$.

Оценки Надарая–Ватсона относят к локально-постоянными, поскольку

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

Классический вариант локально–линейных оценок $f^*_{LL}(t)$ определяется соотношением

$$f^*_{LL}(t) = (1,0)\arg\min_{(\theta_0,\theta_1)'} \sum_{i=1}^n (X_i - \theta_0 - \theta_1(z_i - t))^2 K_h(t - z_i).$$

(в малой окрестности точки t выполнено $f(z) \approx f(t) + f'(t)(z - t) \equiv \theta_0 - \theta_1(z - t)$)
Классические локально–полиномиальные оценки $f^*_{LP}(t)$ порядка p

$$f^*_{LP}(t) = (1,0,\ldots,0)\arg\min_{(\theta_0,\theta_1,\ldots,\theta_p)'} \sum_{i=1}^n (X_i - \theta_0 - \ldots - \theta_p(z_i - t)^p)^2 K_h(t - z_i).$$

Оценки Пристли–Чжао и Гассера–Мюллера

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

$$f^*_{GM}(t) = \sum_{i=1}^n \left[\int_{s_{i-1}}^{s_i} K_h(t - z_i)\right] X_i \text{ при } s_i = \frac{z_i + z_{i-1}}{2}$$

(здесь точки дизайна неслучайны, одномерны и упорядочены).

Пусть наблюдения $X_1, \ldots, X_n$ имеют структуру

$$X_i = f(z_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

где элементы дизайна $\{z_i\}$ известны, $\{\varepsilon_i\}$ — ненаблюдаемые случайные ошибки, непрерывная неизвестная регрессионная функция $f(t)$ подлежит оцениванию.

• Наиболее популярные процедуры оценивания в непараметрической регрессии — оценки ядерного типа, среди которых можно выделить оценки Надарая–Ватсона, Пристли–Чжао, Гассера–Мюллера, полиномиальные оценки (**Müller**, 1988; **Härdle**, 1990; **Wand, Jones**, 1995; **Fan, Gijbels**, 1996; **Fan, Yao**, 2003; **Györfi, Kohler, Krzyzak, Walk**, 2002; **Young**, 2017; **Wakefield**, 2013; **Panik**, 2009; **Klemela**, 2014; **Fahrmeir, Kneib, Lang, Marx**, 2013).

• Публикации в этой области можно условно разделить на две группы (в зависимости от условий на элементы дизайна $\{z_i\}$). К одной относятся работы с фиксированным дизайном, к другой — со случайным.

• В случае фиксированного дизайна в подавляющем большинстве работ предполагаются те или иные условия регулярности вида

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

(см., например, **Zhou, Zhu**, 2020; **Tang, Xi, Wu, Wang**, 2018; **Eagleson, Muller**, 1997; **Gu, Roussas, Tran**, 2007; **Beran, Feng**, 2001; **Hardle, Luckhaus**, 1984, **Wu, Chu**, 1994; **Hansen**, 2008; **Benhenni** et. al., 2010; **Ahmad, Lin**, 1984, **Georgiev**, 1989, 1990 и др.). Более общее условие $\max_{i \le n}(z_i - z_{i-1}) \to 0$ используются в работах **Wu, Wang, Balakrishnan** (2020), **He** (2019), **Yang, Yang** (2016), при этом исследуется поточечная сходимость.

• В работах, имеющих дело со случайным дизайном, рассматриваются или н.о.р. величины, или, как правило, стационарно связанные наблюдения, удовлетворяющие тем или иным известным формам зависимости. Например, используются различные варианты условий перемешивания, схемы скользящих средних, ассоциированных случайных величин, марковские или мартингальные свойства и др. (**Roussas**, 1990, 1991; **Györfi** et. al., 2002; **Masry**, 2005; **Hansen**, 2008; **Honda**, 2010; **Laib, Louani**, 2010; **Kulik, Lorek**, 2011; **Kulik, Wichelhaus**, 2011; **Hardle, Luckhaus**, 1984; **Mack, Silvermann**, 1982; **Muller**, 1997; **Chu, Deng**, 2003; **Linton, Jacho-Chavez**, 2010; **Li, Yang, Hu**, 2016; **Hong, Linton**, 2016; **Shen, Xie**, 2013; **Миллионщиков**, 2005; **Jiang, Mack**, 2001 и др.). Нестационарные последовательности регрессоров с теми или иными формами зависимости (марковские цепи, авторегрессия, частичные суммы скользящих средних) рассматриваются, например, **Linton, Wang**, 2016; **Chan, Wang**, 2014; **Gao** et. al., 2015; **Chen, Gao, Li**, 2012; **Karlsen, Myklebust, Tjostheim**, 2007; **Wang, Phillips**, 2009; **Chen, Li, Zhang**, 2010).

• Задача равномерной аппроксимации оценок ядерного типа изучалась многими авторами (**Nadaraya**, 1970; **Devroye**, 1979; **Mack, Silvermann**, 1982; **Liero**, 1989; **Ioannides**, 1993; **Liang, Jing**, 2005; **Einmahl, Mason**, 2005; **Hansen**, 2008, **Shen, Xie**, 2013; **Wang, Chan** , 2014; **Gao, Kanaya, Li, Tjostheim**, 2015; **Li, Yang, Hu**, 2016 и др.).

Можно ли построить равномерно состоятельные оценки для неизвестной регрессионной функции без использования традиционных условий зависимости элементов дизайна (при близких к минимальным и наглядных условиях на точки дизайна)?

Универсальные локально-постоянные оценки

- пусть $X_i = f(z_i) + \varepsilon_i$, $i = 1, \ldots, n$, неизвестная функция $f(t)$, $t \in [0,1]$, непрерывна;
- $\{z_i;\ i = 1, \ldots, n\}$ — это набор наблюдаемых случайных величин с, вообще говоря, неизвестными распределениями со значениями на $[0,1]$, не обязательно независимых или одинаково распределенных; случайные величины $\{z_i\}$ могут зависеть от n (т.е. данная схема включает в себя модели с фиксированным дизайном);
- при всех $n \geq 1$ с.в. $\{\varepsilon_i;\ i = 1, \ldots, n\}$ с вероятностью 1 при всех $i, j \leq n$, $i \neq j$, удовлетворяют условиям $\mathbb{E}_{\mathcal{F}_n}\varepsilon_i = 0$, $\sup_i \mathbb{E}_{\mathcal{F}_n}\varepsilon_i^2 \leq \sigma^2$, $\mathbb{E}_{\mathcal{F}_n}\varepsilon_i\varepsilon_j = 0$, где $\sigma^2 > 0$ неизвестна и не зависит от n, $\mathbb{E}_{\mathcal{F}_n}$ — у.м.о. относительно σ-алгебры $\mathcal{F}_n$, порожд. $\{z_i,\ i = 1, \ldots, n\}$.
- функция $K(t)$ есть плотность симметричного распределения с носителем $[-1,1]$, $K(t)$ определена на $\mathbb{R}$, удовлетворяет условию Липшица с константой $L \geq 1$ и $K(\pm 1) = 0$.

Обозначим через $z_{n:1} \leq \ldots \leq z_{n:n}$ элементы вариационного ряда, построенного по выборке $\{z_i;\ i \leq n\}$. Положим $z_{n:0} = 0$, $z_{n:n+1} = 1$, $\Delta z_{ni} = z_{n:i} - z_{n:i-1}$, $i = 1, \ldots, n+1$. Отклики и погрешности, относящиеся к $z_{n:i}$, обозначим X_{ni} и ε_{ni}: $X_{ni} = f(z_{n:i}) + \varepsilon_{ni}$, $i = \overline{1, n}$.

Оценку $f^*_{n,h}(t)$ для $f(t)$ определим равенством

$$f^*_{ULC}(t) = \frac{\sum_{i=1}^{n} X_{ni} K_h(t - z_{n:i})\Delta z_{ni}}{\sum_{i=1}^{n} K_h(t - z_{n:i})\Delta z_{ni}}, \qquad K_h(t) = h^{-1}K(h^{-1}t).$$

Заметим, что $f^*_{ULC}(t) = \arg\min_{\theta} \sum_{i=1}^{n} (X_{ni} - \theta)^2 K_h(t - z_{n:i})\Delta z_{ni}$.

Условие на элементы дизайна состоит в следующем:

(D) *Имеет место предельное соотношение* $\delta_n = \max\limits_{1 \le i \le n+1} \Delta z_{ni} \xrightarrow{p} 0.$

Например, условие (D) выполнено в следующих случаях:

- для любого регулярного дизайна;
- если $\{z_i\}$ н.о.р. и отрезок $[0,1]$ является носителем распределения z_1 (в случае существования отделенной от нуля на $[0,1]$ плотности распределения $\delta_n = O\left(\frac{\log n}{n}\right)$ п.н.);
- если $\{z_i\}$ — стационарная последовательность с условием α-перемешивания и маргинальным распределением с носителем $[0,1]$.
- зависимость с.в. $\{z_i\}$ в условии (D) может быть более сильной (см. пример далее).

- Дизайн может быть как фиксированным и не обязательно регулярным, так и случайным, при этом не обязательно удовлетворяющим условиям слабой зависимости.
- Относительно дизайна требуется лишь некоторое условие плотного заполнения области определения регрессионной функции, что по сути является необходимым для восстановления функции на области задания элементов дизайна.

Оценки Надарая–Ватсона и новые оценки:

$$f^*_{NW}(t) = \frac{\sum_{i=1}^n X_i K_h(t - z_i)}{\sum_{i=1}^n K_h(t - z_i)}, \qquad f^*_{ULC}(t) = \frac{\sum_{i=1}^n X_{ni} K_h(t - z_{n:i}) \Delta z_{ni}}{\sum_{i=1}^n K_h(t - z_{n:i}) \Delta z_{ni}}.$$

Пример. Пусть с.в. $\{z_i;\ i \geq 1\}$ определяются равенством

$$z_i = \nu_i U_{1i} + (1 - \nu_i) U_{2i},$$

• $\{U_{1i}\}$ и $\{U_{2i}\}$ независимы в совокупности и равномерно распределены на $[0, 1/2]$ и $[1/2, 1]$ соответственно,
• $\{\nu_i\}$ не зависит от $\{U_{1i}\}$, $\{U_{2i}\}$ и состоит из бернуллиевских с.в. с вер. успеха $1/2$,
• зависимость между с.в. ν_i для любого i определяется соотношениями
$\nu_{2i-1} = \nu_1, \quad \nu_{2i} = 1 - \nu_1.$
Отметим, что $\{z_i;\ i \geq 1\}$ есть стационарная п-ть равномерно распределенных на $[0, 1]$ с.в.,

$$\forall\ m, n \qquad \mathbb{P}(z_{2m} \leq 1/2,\ z_{2n-1} \leq 1/2) = 0.$$

Таким образом, наиболее популярные условия слабой зависимости случайных величин (перемешивания, асимптотическая некоррелируемость, скользящие средние и др.) в данном примере не выполнены, поскольку указанные условия как правило влекут соотношение $\lim_{|n-m|\to\infty} \mathrm{Cov}(z_n, z_m) = 0$. При этом последовательность $\{z_i;\ i \geq 1\}$ удовлетворяет (D).

• По этой схеме можно конструировать различные удовлетворяющие (D) п-ти зависимых равномерно распределенных на $[0, 1]$ с.в. $\{z_i\}$ (не обязательно стационарные) исходя из выбора $\{\nu_i\}$ с условиями $\nu_{j_k} = 1$ и $\nu_{l_k} = 0$ для неограниченного числа индексов $\{j_k\}$ и $\{l_k\}$.

Например, нестационарная п-ть $\{z_i\}$, не удовлетворяющие ЗБЧ:
• $\nu_j = 1 - \nu_1$ при $j = 2^{2k-1}, \ldots, 2^{2k} - 1$ и $\nu_j = \nu_1$ при $j = 2^{2k}, \ldots, 2^{2k+1} - 1$, где $k = 1, 2, \ldots$ (т.е. мы разыгрываем, в какой из отрезков бросаем наудачу первую точку, и далее чередуем количество бросаний на каждый из двух указанных отрезков следующим образом: 2, 4, 8, 16 и т.д.).

Теорема. *Для любого фиксированного $h \in (0,1)$ с вероятностью 1 выполнено*

$$\sup_{t\in[0,1]} |f^*_{ULC}(t) - f(t)| \le \omega_f(h) + \zeta_n(h), \tag{1}$$

где $\omega_f(h) = \sup_{t,s\in[0,1]:|t-s|\le h} |f(t) - f(s)|$, а с.в. $\zeta_n(h)$ такова, что

$$\mathbb{P}(\zeta_n(h) > y) \le C_0\sigma^2 L^2 y^{-2} h^{-2} \mathbb{E}\delta_n + \mathbb{P}(\delta_n > h/(8L)).$$

Если выполнено (D), то $\zeta_n(h) = O_p(h^{-1}(\mathbb{E}\delta_n)^{1/2})$.[1] В качестве $h = h_n$ можно взять решение уравнения $h^{-1}(\mathbb{E}\delta_n)^{1/2} = \omega_f(h)$ (фактически так выбранный размер окна уравнивает порядок малости (по h) обоих слагаемых в правой части (1)).

Следствие. *Пусть выполнено условие (D) и $\mathcal{C}$ есть произвольное подмножество равностепенно непрерывных функций в $C[0,1]$. Тогда $\gamma_n(\mathcal{C}) = \sup_{f\in\mathcal{C}} \sup_{t\in[0,1]} |f^*_{ULC}(t) - f(t)| \xrightarrow{p} 0$, где $h = h_n$ есть решение уравнения $h_n^{-1}(\mathbb{E}\delta_n)^{1/2} = \omega_{\mathcal{C}}(h_n)$ при $\omega_{\mathcal{C}}(h) = \sup_{f\in\mathcal{C}} \omega_f(h)$. Кроме того, выполнено $\gamma_n(\mathcal{C}) = O_p(\omega_{\mathcal{C}}(h_n))$.*

Если $\mathbb{E}\delta_n = O(1/n)$ и $\mathcal{C}$ состоит из гельд. функций f с пок. $\alpha \in (0,1]$ и унив. конст., то $h_n = O\left(n^{-\frac{1}{2(1+\alpha)}}\right)$ и $\omega_{\mathcal{C}}(h_n) = O\left(n^{-\frac{\alpha}{2(1+\alpha)}}\right)$. В частности, $\gamma_n(\mathcal{C}) = O_p(n^{-\frac{1}{4}})$ при $\alpha = 1$.

[1] $\zeta_n = O_p(\eta_n)$, если для всех чисел $M > 0$ выполнено $\limsup \mathbb{P}(|\zeta_n|/\eta_n > M) \le \beta(M)$, где $\{\eta_n\}$ – положительные (возможно, случайные) величины, $\lim_{M\to\infty} \beta(M) = 0$ ($\beta(M)$ может зависеть от K и σ^2).

Сравнение с оценками Н.-В. в случае н.о.р. величин

Пусть функция $f(t)$ дважды непрерывно дифференцируема на $[0,1]$, $\{z_i\}$ н.о.р., $\{\varepsilon_i\}$ н.о.р. и не зависят от $\{z_i\}$. Ф.р. $F(t)$ с.в. z_1 имеет непрерывно дифференцируемую положительную на $(0,1)$ плотность $p(t)$. Оценки Надарая–Ватсона и новые оценки:

$$f^*_{NW}(t) = \frac{\sum_{i=1}^n X_i K_h(t-z_i)}{\sum_{i=1}^n K_h(t-z_i)}, \qquad f^*_{ULC}(t) = \frac{\sum_{i=1}^n X_{ni} K_h(t-z_{n:i})\Delta z_{ni}}{\sum_{i=1}^n K_h(t-z_{n:i})\Delta z_{ni}}.$$

Лемма 1 (Rosenblatt, 1969). Если $n\to\infty$ и $h\to 0$ так, что $h^3 n\to\infty$, то $\forall t\in(0,1)$

$$\mathrm{Bias}\, f^*_{NW}(t) = \frac{h^2\kappa_2}{2p(t)}\left(f''(t)p(t)+2f'(t)p'(t)\right)+o(h^2), \qquad \mathbb{D} f^*_{NW}(t)\sim\frac{\sigma^2}{hnp(t)}\|K\|^2,$$

где $\kappa_2 = \int\limits_{-1}^{1} u^2K(u)du$, $\|K\|^2 = \int\limits_{-1}^{1} K^2(u)du$.

Лемма 2. Пусть $\inf_{t\in[0,1]} p(t) > 0$ и $n\to\infty$ и $h\to 0$ так, что $(\log n)^{-1}h\sqrt{n}\to\infty$, $h^{-2}\mathbb{E}\delta_n\to 0$ и $h^{-3}\mathbb{E}\delta_n^2\to 0$. Тогда при любом $t\in(0,1)$

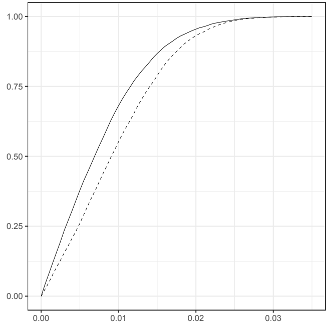
$$\mathrm{Bias}\, f^*_{ULC}(t) = \frac{h^2\kappa_2}{2}f''(t)+o(h^2), \qquad \mathbb{D} f^*_{ULC}(t)\sim\frac{2\sigma^2}{hnp(t)}\|K\|^2.$$

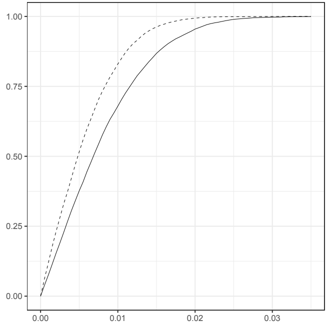
Вывод: Если стандартное отклонение погрешностей σ не сильно велико и

$$\left|f''(t)p(t)+2f'(t)p'(t)\right| > \left|f''(t)p(t)\right|, \qquad (*)$$

то оценка $f^*_{ULC}(t)$ может быть лучше, чем оценка Надарая–Ватсона $f^*_{NW}(t)$.

Приведенные графики иллюстрируют влияние $(*)$ на точность аппроксимации.





$p(t) = 0.5 + t$ $\qquad$ $p(t) = 1.5 - t$

Графики изображают эмпирические функции распределения для $|f^*_{ULC}(0.5) - f(0.5)|$ (сплошная линия) и $|f^*_{NW}(0.5) - f(0.5)|$ (пунктирная линия), построенные по 10000 симуляционных прогонок, при этом $f(t) = t^2$, $h = 0.15$, $K(\cdot)$ – ядро Епанечникова, погрешности нормально распределены со средним 0 и $\sigma = 0.1$, размер выборки $n = 1000$, плотность дизайна $p(t) = 0.5 + t$ или $p(t) = 1.5 - t$.

Для плотности $p(t) = 0.5 + t$ неравенство $(*)$ выполнено и оценка $f^*_{ULC}(t)$ оказывается лучше, нежели $f^*_{NW}(t)$. Если $p(t) = 1.5 - t$, то ситуация обратная.

Универсальные локально–линейные оценки

Для любого $h \in (0,1)$ введем в рассмотрение класс оценок

$$f^*_{ULL}(t) := I(\delta_n \le c_* h) \sum_{i=1}^{n} \frac{w_{n2}(t) - (t - z_{n:i}) w_{n1}(t)}{w_{n0}(t) w_{n2}(t) - w^2_{n1}(t)} X_{ni} K_h(t - z_{n:i}) \Delta z_{ni},$$

где $I(\cdot)$ — индикаторная функция,

$$c_* \equiv c_*(K) := \frac{\kappa_2 - \kappa_1^2}{96L(6L + \kappa_2 + \kappa_1/2)} < \frac{1}{864L};$$

$$w_{nj}(t) := \sum_{i=1}^{n} (t - z_{n:i})^j K_h(t - z_{n:i}) \Delta z_{ni}, \qquad \kappa_j = \int_{-1}^{1} |u|^j K(u) du, \qquad j = 0, 1, 2.$$

- Ядерная оценка $f^*_{ULL}(t)$, без индикаторного множителя, является первой координатой двумерной точки (θ_0^*, θ_1^*), на которой достигается следующий минимум:

$$\min_{\theta_0, \theta_1} \sum_{i=1}^{n} \left(X_{ni} - (\theta_0 + \theta_1(t - z_{n:i}))\right)^2 K_h(t - z_{n:i}) \Delta z_{ni}.$$

- Качественное отличие оценок $f^*_{ULC}(t)$ и $f^*_{ULL}(t)$ наблюдается в окрестностях граничных точек 0 и 1: для оценки $f^*_{ULC}(t)$ в h-окрестностях указанных точек порядок малости смещения h, а для $f^*_{ULL}(t)$ этот порядок h^2. Такая связь между оценками представляется весьма естественной ввиду известной взаимосвязи в граничных точках между классическими оценками Надарая–Ватсона и классическими локально–линейными оценками.

Многомерные регрессоры

$$X_i = f(\mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

- $f : [0,1]^k \to \mathbb{R}$ неизвестна и непрерывна;
- $\{\mathbf{z}_i\}$ — набор наблюдаемых k-мерных векторов в схеме серий.

$(\mathbf{D}_k)$ Для каждого n существует такое разбиение множества $[0,1]^k$ на n измеримых по Жордану подмножеств $\{\mathcal{P}_i,\ i = 1, \ldots, n\}$, что каждый элемент разбиения содержит только одну точку из $\{\mathbf{z}_i\}$ (занумеруем $\mathcal{P}_i$ так, чтобы $\mathbf{z}_i \in \mathcal{P}_i$) и $\max\limits_{i \le n} d(\mathcal{P}_i) \xrightarrow{p} 0$, где $d(A) = \sup\limits_{\mathbf{x},\mathbf{y} \in A} \|\mathbf{x} - \mathbf{y}\|$ — диаметр множества, $\|\cdot\|$ — супремальная норма в $\mathbb{R}^k$.

$$f^*_{ULC}(\mathbf{t}) = \frac{\sum_{i=1}^n X_i K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i)}{\sum_{i=1}^n K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i)},$$

где $\Lambda_k(\cdot)$ есть мера Лебега в $\mathbb{R}^k$, $K_h(\mathbf{s}) = h^{-k}K(h^{-1}\mathbf{s})$, $K(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^k$ — ядерная функция.

Многомерные точки дизайна. Компьютерное моделирование

Генерируется 10000 независимых выборок вида $X_i = f(\mathbf{z}_i) + \varepsilon_i, \quad i = 1, \ldots, n$, при этом

- $k = 2$ (размерность точек дизайна), $\{\mathbf{z}_i\}$ независимы и имеют равномерное распределение на единичном квадрате $[0,1]^2$,
- $f(\mathbf{z}) = \max\{0, \sin(4z_2 - 2)\} + 0.5$,
- $\{\varepsilon_i\}$ н.о.р. с распределением $\mathcal{N}(0, 0.01)$.

Для каждой выборки вычисляется $f^*_{ULC}(\mathbf{t})$ с двумя вариантами разбиений и $f^*_{NW}(\mathbf{t})$.

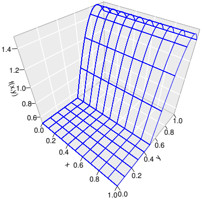
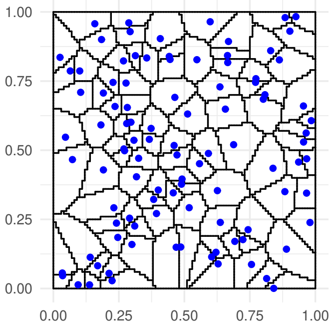
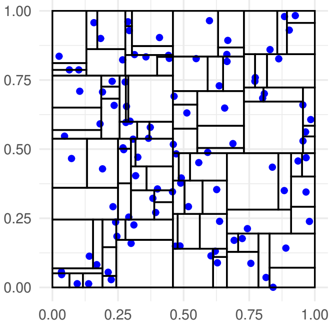


График функции f

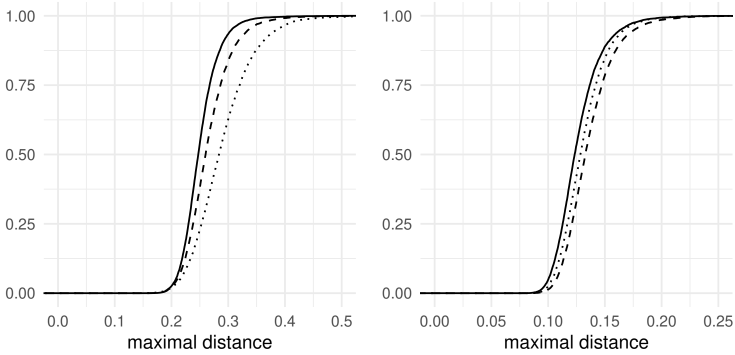


Разбиение методом Вороного



Разбиение покоординатно-медианными сечениями

На рисунке приведены графики эмпирических функций распределения для трех величин:
$\max\limits_{\mathbf{t}\in\mathcal{P}} |f^*_{ULC}(\mathbf{t}) - f(\mathbf{t})|$ с разбиением Вороного (сплошная линия),
$\max\limits_{\mathbf{t}\in\mathcal{P}} |f^*_{ULC}(\mathbf{t}) - f(\mathbf{t})|$ с разбиением покоорд.-медиан. сечениями (пунктирная линия),
$\max\limits_{\mathbf{t}\in\mathcal{P}} |f^*_{NW}(\mathbf{t}) - f(\mathbf{t})|$ (точечная линия), $f^*_{NW}(\mathbf{t}) = \frac{\sum_{i=1}^n X_i K_h(\mathbf{t}-\mathbf{z}_i)}{\sum_{i=1}^n K_h(\mathbf{t}-\mathbf{z}_i)}$ — оценка Надарая–Ватсона.
Максимум вычислен по сетке 1000×1000. Каждый из графиков эмпирических функций распределения построен по 10000 прогонкам.



$n = 100,\ h = 0.3$ $\qquad$ $n = 1000,\ h = 0.1$

Оценивание функций среднего и ковариации случайного процесса

Рассмотрим набор $f_1(t), \dots, f_n(t)$ независимых реализаций непрерывного случайного процесса $f(t)$, определенного на $[0, 1]$. Задача состоит в оценивании функций среднего $\mu(t) = \mathbb{E} f(t)$ и ковариации $\psi(t, s) = \mathrm{cov}\{f(t), f(s)\}$ по парам наблюдений $\{(z_{ij}, X_{ij}),\ i = 1, \dots n,\ j = 1, \dots, m_i\}$ со структурой

$$X_{ij} = f_i(z_{ij}) + \varepsilon_{ij}, \qquad i = 1, \dots, n, \quad j = 1, \dots, m_i.$$

• При широких ограничениях на $\{z_{ij}\}$ можно строить оценки для μ и ψ как в случае разреженного дизайна (количество точек дизайна для каждой из траекторий равномерно ограничено конечной константой), так и плотного (количество наблюдений для каждой траектории растет с ростом n).

• Новые оценки обладают свойством универсальности относительно структуры дизайна: он может быть как фиксированным и не обязательно регулярным, так и случайным, при этом не обязательно состоящим из слабо зависимых случайных величин.

• Относительно разреженного дизайна мы требуем лишь, чтобы вся совокупность точек дизайна образовывала измельчающееся разбиение отрезка $[0, 1]$, а для плотного дизайна подобное условие должно быть выполнено для элементов дизайна каждой из траекторий.

Rice, Wu (2001); James, Hastie (2001); Lin, Carroll (2000), Wang (2003), Yao, Müller, Wang (2005a, 2005b); Muller (2005); Ramsay, Silverman (2005), Yao, Lee (2006); Wu, Zhang (2006); Hall, Müller, Wang (2006); Gervini (2006); Zhang, Chen (2007); Yao (2007); Degras (2008); Li, Hsing (2010), Cai, Yuan (2011); Bunea, Ivanescu, Wegkamp (2011); Cai, Yuan (2011); Ma, Yang, Carroll (2012); Cao, Yang, Todem (2012); Kim, Zhao (2013); Song, Liu, Shao, Yang (2014); Zheng, Yang, Hardle (2014); Cuevas (2014); Hsing, Eubank (2015); Cao, Wang, Li, Yang (2016); Wang, Chiou, Muller (2016); Kokoszka, Reimherr (2017); Zhang, Wang (2016, 2018); Zhou, Lin, Liang (2018); Wang J. Cao, Wang L., Yang (2020); Lin, Wang (2022).

[1] Linke Yu., Borisov I., Ruzankin P., Kutsenko V., Yarovaya E., Shalnova S. (2022) Universal local linear kernel estimators in nonparametric regression // **Mathematics**. 10(15).
[2] Borisov I. S., Linke Yu. Yu., Ruzankin P.S. (2021) Universal weighted kernel-type estimators for some class of regression models // **Metrika**. 84(2), 141–166.

[3] Линке Ю. Ю. К вопросу о нечувствительности оценок Надарая–Ватсона относительно корреляции элементов дизайна. // **Теория вероятн. и ее примен.** (в печати)
[4] Linke Yu. Yu. (2022) Kernel estimators for the mean function of a stochastic process under sparse design conditions. // **Siberian Adv. Math.** 32(4), 269–276.
[5] Linke Yu. Yu., Borisov I. S. (2022) Insensitivity of Nadaraya–Watson estimators to design correlation // **Communications in Statistics – Theory and Methods**, 51, 6909–6918.
[6] Линке Ю.Ю., Борисов И.С. Универсальные непараметрические ядерные оценки для функций среднего и ковариации случайного процесса. **Теория вероятн. и ее примен.** (сдана в печать)

[7] Linke Yu. Yu. (2019) Asymptotic properties of one-step M-estimators // **Communications in Statistics – Theory and Methods** 48, 4096–4118.
[8] Linke Yu. Yu. (2017) Asymptotic normality of one-step *M*-estimators based on non-identically distributed observations // **Statist. Probab. Lett.** 129, 216–221.
[9] Линке Ю. Ю. (2017) Асимптотические свойства одношаговых взвешенных *M*-оценок с приложениями к задачам регрессии // **Теория вероятн. и ее примен.** 62(3), 468–498.
[10] Линке Ю. Ю., Борисов И. С. (2018) Построение явных оценок в задачах нелинейной регрессии // **Теория вероятн. и ее примен.** 63(1) 29–56.
[11] Linke Yu. Yu., Borisov I. S. (2017) Constructing initial estimators in one-step estimation procedures of nonlinear regression // **Statist. Probab. Lett.** 120(1), 87–94.
[12] Linke Yu. Yu., Borisov I. S. (2019) Toward the notion of intrinsically linear models in nonlinear regression // **Siberian Adv. Math.** 29(3), 210–216.

Взаимосвязь с задачами нелинейной регрессии.

Пусть наблюдения $X_1, \ldots, X_n$ представимы в виде

$$X_i = f(\theta, z_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

где функция f известна. Требуется оценить параметр θ.

• В задачах нелинейной регрессии асимптотически оптимальные оценки как правило задаются неявно в виде решений тех или иных уравнений.

• Для задач нелинейной регрессии весьма типична ситуация, когда имеется несколько корней того или иного уравнения, определяющего оценку (см., например,
Small C.G., Yang Z. *Multiple Roots of Estimating Functions.* (1999);
Small C.G., Wang J. *Numerical Methods for Nonlinear Estimating Equations.* (2003).

Указанный факт является главной проблемой, затрудняющей использование приближенных численных методов поиска таких оценок. Дело в том, что при неудачном выборе начального приближения параметра итерационные процедуры обнаруживают лишь корень, ближайший к этой стартовой точке, а не к параметру. Более того, если корней несколько, то как среди всех корней выбрать состоятельный корень? Данная проблематика обсуждается, например, в монографиях **Heyde**, 1997; **Small, Wang**, 2003.

Например, пусть требуется оценить параметр θ по выборке $X_1, \ldots, X_n$, имеющей структуру

$$X_i = f(\theta, z_i) + \varepsilon_i, \qquad \mathbb{E}\varepsilon_i = 0, \quad \mathbb{E}\varepsilon_i^2 = \sigma^2 / w_i(\theta), \quad i = 1, \ldots, n,$$

где f и $\{w_i\}$ — известные функции, $\{z_i\}$ известны, погрешности $\{\varepsilon_i\}$ независимы. Оценка **квазиправдоподобия** (**Heyde**, 1997) определяется здесь уравнением

$$\psi_n(t) := \sum_{i=1}^n w_i(t) f'(t, z_i)\big(X_i - f(t, z_i)\big) = 0.$$

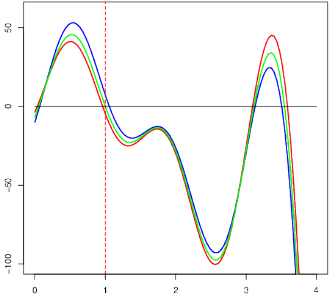


Рис.: Графики функции $\psi_n(t)$ для некоторых f и $\{w_i\}$ и трех независимых реализаций выборки $\{X_i\}$ ($\theta = 1$, точные параметры этой модели приведены далее, см. слайд 28, модель 4).

Какой из корней приближает неизвестный параметр?

Помимо нелинейной регрессии, имеется немало разделов статистического оценивания, также связанных с поиском корней тех или иных уравнений. При этом уравнения лишь в исключительных случаях разрешаются в явном виде и найти состоятельные корни уравнения или их приближения бывает технически сложно, особенно в случае существования нескольких корней исходного уравнения.

Van der Vaart A.W. (2000) *Asymptotic Statistics.*
Small C.G., Yang Z. Multiple Roots of Estimating Functions. *Can. J. Stat.*, 1999;
Small C.G., Wang J. (2003). *Numerical Methods for Nonlinear Estimating Equations.*
Heyde C.C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation.*

Выход: одношаговые оценки (one-step estimators).

• Идея одношагового оценивания, восходящая к работам Р. Фишера, заключается в следующем: в качестве стартовой точки итерационной процедуры ньютоновского типа используется не произвольная точка, а предварительная состоятельная оценка, сходящаяся к параметру с нужной скоростью. В этом случае достаточно лишь одного шага итерационной процедуры, чтобы получить явную оценку (так называемую «одношаговую»), имеющую ту же асимптотическую точность, как и искомая статистика, определяемая соответствующим уравнением.

• Например, одношаговую оценку θ_n^{**} для параметра θ в вышеприведенном примере, связанном с уравнением $\psi_n(t) = 0$, можно задать равенством

$$\theta_n^{**} = \theta_n^* - \psi_n(\theta_n^*)/\psi_n'(\theta_n^*).$$

Одношаговые оценки. Библиографические ссылки

Одношаговые процедуры являются весьма популярным современным инструментом статистического оценивания в различных задачах, связанных с поиском корней тех или иных уравнений, и в последние годы интерес к одношаговому оцениванию в современной статистической литературе только нарастает.
Antoniadis, Fan (2001), **Zhao** (2000, 2001), **Bianco, Boente** (2002),
Welsh, Ronchetti (2002), **Fan, Li** (2002), **Xie, Yang** (2003),
Fan, Yao, Cai (2003), **Qian, Correa** (2003), **Yang** (2004),
Li, Marron (2005), **Giloni, Simonoff** (2005), **Jurečková, Picek** (2006),
Bianco, Boente, Martinez (2006), **Fan, Lin, Zhou** (2006),
Cai, Fan, H.Zhou, Y. Zhou (2007), **Linton, Xiao** (2007),
Li, Zheng (2007), **Zou, Li** (2008), **Johansen, Nielsen** (2009),
Chen, Li, Zhang (2010), **Bergesioa, Yohaia** (2011),
Bradic, Fan, Wang (2011), **Karunamuni, Wu** (2011), **Jureckova** (2012),
Cai, Chen, Fang (2012), **Li, Calder, Cressie** (2012), **Jureckova, Sen, Picek** (2012),
Acitas, Kasap, Senoglu, Arslan (2013), **Chen, Lin** (2013),
Hall, Ma (2014), **Fan, Xue, Zou** (2014),
Johansen, Nielsen (2016), **Wang, Wang** (2016), **Taddy** (2016),
Lipsitz, Fitzmaurice, Sinha, Hevelone, Hu, Nguyen (2017), **Jiao, Nielsen** (2017),
Ning, Liu (2017), **Huo, Huang, Ni** (2017),
Dattner, Gugushvili (2018), **Morgan** (2018),
Huang, Huo (2019), **Bradic, Guo** (2019),
Bassett, Deride (2020), **Duchesne, Micheaux, Tatsinkou** (2020),
Eisenach, Bunea, Ning, Dinicu (2020), **Fang, Zhao, Ahmed, Qu** (2020) и др.

Одношаговые процедуры. Библиографические ссылки

Le Cam (1956), **Bickel** (1975), **Janssen, Jureckova, Veraverbeke** (1985), **Poetscher, Prucha** (1986), **Jureckova, Portnoy** (1987), **Williams** (1987), **Robinson** (1988), **White** (1989), **Marx, Smith** (1990), **Jureckova, Sen** (1990), **Tableman** (1990), **Dollinger, Staudte** (1991), **Simpson, Ruppert, Carroll** (1992), **Müller** (1994), **Field, Wiens** (1994), **Jurečková, Malý** (1995), **Simpson, Chang** (1997), **Simpson, Yohai** (1998), **Markatou** (1999), **Lopuhaa** (1999), **Fan, Chen** (1999).

Одношаговые оценки находят широкое применение и в монографической литературе. **Zacks** (1971), **Serfling** (1980), **Rousseeuw, Leroy** (1987), **Rieder** (1994), **Fan, Gijbels** (1996), **Jureckova, Sen** (1996), **Lehmann, Casella** (1998), **Van der Vaart** (2000), **Small, Wang** (2003), **Seber, Lee** (2003), **Fan, Yao** (2003), **Maronna, Martin, Yohai** (2006), **Jureckova, Picek** (2006), **Hampel, Ronchetti, Rousseeuw, Stahel** (2011), **Jureckova, Sen, Picek** (2012), **Wakefield** (2013), **Shults, Hilbe** (2014), **Morgan** (2018), **Jureckova, Picek, Schindler** (2019), **Fan, Li, Zhang, Zou** (2020), **Боровков** (2021) и др.

Проблема в нелинейной регрессии: построение предварительных оценок.

Во всех работах, связанных с одношаговым оцениванием в нелинейной регрессии, существование предварительной оценки лишь постулируется.

Идея построения явных оценок в задачах нелинейной регрессии

Пусть наблюдения $X_1, \ldots, X_n$ представимы в виде $X_i = f(\theta, z_i) + \varepsilon_i$, где функция f известна, $\{z_i\}$ — наблюдаемые с. в. в схеме серий с неизвестными распределениями, не обязательно независимые или одинаково распределенные, $\{\varepsilon_i\}$ — ненаблюдаемые с. в., которые могут зависеть от n. Требуется оценить параметр $\theta \in \Theta = (a, b) \subset \mathbb{R}$.

- Идея построения оценок состоит в использовании сумм специальным образом взвешенных наблюдений со структурой интегральных сумм Римана.
- При каждом фиксированном $n \geq 1$ упорядочим $z_1, \ldots, z_n$ по возрастанию: $z_{n:1} \leq \ldots \leq z_{n:n}$. Далее перенумеруем соответствующим образом отклики и погрешности, которые обозначим через X_{ni} и ε_{ni}. В этом случае $X_{ni} = f(\theta, z_{n:i}) + \varepsilon_{ni}$, $i = 1, \ldots, n$.

(D$_0$) Пусть $z_i \in [c, d]$ п.н. и $\max\limits_{1 \leq i \leq n+1} \Delta z_{ni} \xrightarrow{p} 0$, где $\Delta z_{ni} = z_{n:i} - z_{n:i-1}$, $z_{n:0} = c$, $z_{n:n+1} = d$. При всех $t \in \Theta$ функция $f(t, z)$ интегрируема по Риману по z на отрезке $[c, d]$ и интеграл Римана $T(t) = \int_c^d f(t, z)dz$ — строго монотонная и непрерывная функция.

Оценку параметра θ определим равенством $\theta_n^* = T^{-1}\left(\sum_{i=1}^n \Delta z_{ni} X_{ni}\right)$,

поскольку в широких условиях $\sum_{i=1}^n \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} 0$, а потому

$$\sum_{i=1}^n \Delta z_{ni} X_{ni} = \sum_{i=1}^n \Delta z_{ni} f(\theta, z_{n:i}) + \sum_{i=1}^n \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} T(\theta) \quad \text{и} \quad \theta_n^* \xrightarrow{p} T^{-1}(T(\theta)) \equiv \theta.$$

Иногда при построении оценки следует “отбросить” некоторое количество наблюдений X_{ni}, чтобы обеспечить выполнение нужных нам условий. Например,

• если $\{z_i\}$ «плотно» заполняют лишь часть области своего задания (не обязательно односвязную, точки дизайна (или их часть) могут разбиваться на кластеры с условием измельчения в каждом);

• если функция $T(t) = \int_A f(t,z)dz$ не является строго монотонной и непрерывной (A — множество плотного заполнения точками дизайна), но существует подмножество $A_o \subset A$ такое, что функция $T_o(t) = \int_{A_o} f(t,z)dz$ строго монотонна и непрерывна.

$$X_{ni} = f(\theta, z_{n:i}) + \varepsilon_i, \quad \Delta z_{ni} = z_{n:i} - z_{n:i-1}, \quad T(t) = \int_c^d f(t,z)dz, \quad \theta_n^* = T^{-1}\Big(\sum_{i=1}^n \Delta z_{ni} X_{ni}\Big).$$

Приведем несколько примеров известных регрессионных моделей и соответствующих функций $T(\cdot)$, определяющих θ_n^*. При этом мы не задаемся вопросом явного представления обратной функции $T^{-1}(\cdot)$, поскольку ее значения легко могут быть вычислены с любой наперед заданной точностью. Более того, в общем случае не требуется чтобы и интеграл Римана $T(\cdot)$ вычислялся в элементарных функциях.

Примеры. Пусть $z_i \in [c,d]$, $c \geq 0$, $d < \infty$ и $\theta > 0$.

1) если $f(\theta, z_i) = (1+\theta z_i)^r$, $r \neq 0, -1, 1$, то $T(\theta) = \dfrac{(1+d\theta)^{r+1} - (1+c\theta)^{r+1}}{\theta(r+1)}$;
если $r < -1$, то при $c = 0$ и $d = \infty$ выполнено $T^{-1}(t) = t^{-1}|r+1|^{-1}$;

2) если $f(\theta, z_i) = 1/(1+\theta z_i)$, то $T(\theta) = \dfrac{1}{\theta}\log\dfrac{1+d\theta}{1+c\theta}$;

3) если $f(\theta, z_i) = 1/(1+e^{-\theta z_i})$, то $T(\theta) = (d-c) - \dfrac{1}{\theta}\log\dfrac{1+e^{-\theta c}}{1+e^{-\theta d}}$;

4) если $f(\theta, z_i) = 1/(\theta + z_i)$, то $T(\theta) = \log\dfrac{\theta+d}{\theta+c}$ and $T^{-1}(t) = \dfrac{d-ce^t}{e^t-1}$;

5) если $f(\theta, z_i) = e^{-\theta z}$, то $T(\theta) = \theta^{-1}\big(e^{-\theta c} - e^{-\theta d}\big)$; $T^{-1}(t) = 1/t$ при $c = 0$ и $d = \infty$;

6) если $f(\theta, z_i) = \log(1+\theta z_i)$, $T(\theta) = \big((1+d\theta)\log(1+\theta d) - (1+c\theta)\log(1+\theta c)\big)/\theta - (d-c)$;

7) если $f(\theta, z_i) = z^\theta$ $(c, d \leq 1)$, то $T(\theta) = \big(d^{\theta+1} - c^{\theta+1}\big)/(\theta+1)$; $T^{-1}(t) = 1/t - 1$ при $c = 0$ и $d = 1$.

Результаты компьютерного моделирования

Для каждой модели генерируется 1000 независимых выборок объема $n = 100$ вида
$X_i = f_i(\theta) + \varepsilon_i$ при $f_i(\theta) = f(\theta, \mathbf{z}_i)$, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2/w_i(\theta))$, $i = 1, \ldots, n$.

Одношаговые процедуры применяются для приближения оценок квазиправдоподобия, которые являются наилучшими в некотором классе и определяются здесь уравнением

$$\psi_n(t) := \sum_{i=1}^{n} \psi_i(t, X_i) = 0 \qquad \text{при} \quad \psi_i(t, X_i) = w_i(t) f_i'(t)(X_i - f_i(t))$$

Таблица: Метод квазиправдоподобия. Результаты компьютерного моделирования при использовании одношаговых процедур для приближения оценок квазиправдоподобия.

	$f_i(t)$	$\sigma^2/w_i(t)$	$z_i = (z_{1i}, z_{2i})$	θ	θ_n^*	θ_n^{**}
1.	$z_{1i} t + z_{2i} \log(t)$	$(1 + t z_{1i} + z_{1i} \log(t))^{-2}$	$z_{1i} = 0.5 + \frac{i}{n}$, $z_{2i} = 1.5 - \frac{i}{n}$	2	1.895 2.381 1.746	2.067 1.966 2.182
2.	$z_{1i} t + z_{2i} t^{-1/5}$	$1/(z_{1i} t + z_{2i} t^{-1/5})$	$z_{1i} = \frac{i}{n}$, $z_{2i} = 2 - \frac{i}{n}$	3	1.804 3.941 2.074	3.204 2.928 3.017
3.	$z_{1i} t + z_{2i} \cos(t)$	$0.4(1 + t z_{1i} + z_{2i} \cos(t))^{-2}$	$z_{1i} = i/n$, $z_{2i} = 2 - i/n$	1	0.662 0.726 1.589	1.065 1.030 0.973
4.	$z_{1i} t + z_{2i} \cos(t)$	$0.25(t z_{1i} + z_{2i} \cos(t))^{-1}$	$z_{1i} = i/n$, $z_{2i} = 2 - i/n$	1	0.805 1.703	1.032 1.025

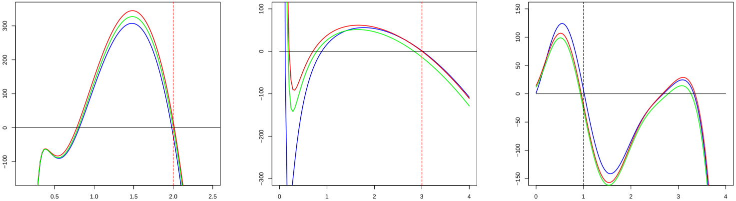


Рис.: Метод квазиправдоподобия. Иллюстрируется поведение функции $\psi_{\boldsymbol{n}}(x) = \sum_{i=1}^{n} w_{\boldsymbol{i}}(x) f_{\boldsymbol{i}}'(x)(X_{\boldsymbol{i}} - f_{\boldsymbol{i}}(x))$ в окрестности истинного значения параметра при трех различных реализациях выборки.

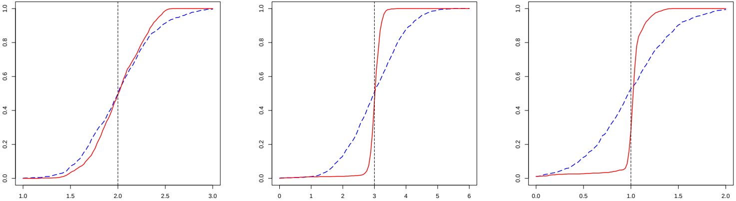


Рис.: Одношаговая аппроксимация оценок квазиправдоподобия. Эмпирические функции распределения для $\theta_{\boldsymbol{n}}^*$ (пунктирная линия) и $\theta_{\boldsymbol{n}}^{**}$, построенные по 1000 независимых выборок.

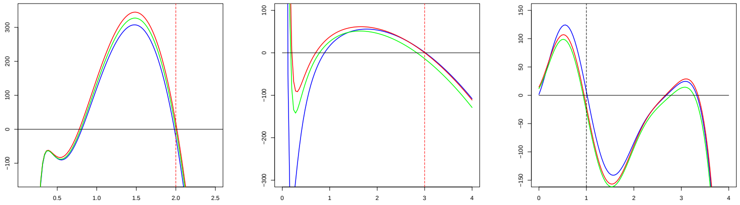


Рис.: Метод квазиправдоподобия. Иллюстрируется поведение функции $\psi_n(x) = \sum_{i=1}^{n} w_i(x) f_i'(x)(X_i - f_i(x))$ в окрестности истинного значения параметра при трех различных реализациях выборки.

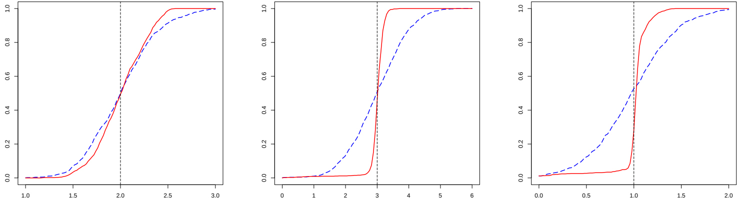


Рис.: Одношаговая аппроксимация оценок квазиправдоподобия. Эмпирические функции распределения для θ_n^* (пунктирная линия) и θ_n^{**}, построенные по 1000 независимых выборок.

СПАСИБО ЗА ВНИМАНИЕ!