

# Эконометрический ликбез: непараметрические и полупараметрические методы

## Непараметрическая регрессия\*

Станислав Анатольев<sup>†</sup>

Российская экономическая школа, Москва, Россия

Настоящее эссе повествует о принципах и методологии непараметрического оценивания регрессии среднего. Акцент делается на методах ядерного сглаживания, но дается и обзор неядерных методов.

### 1 Введение

Пусть имеется случайная выборка  $\{(x_i, y_i)\}_{i=1}^n$  из популяции пар  $(x, y)$ . Нас интересует оценивание регрессии среднего  $g(x) = \mathbb{E}[y|x]$  в предположении, что она существует для всех  $x$  носителя и является гладкой. Для этого чаще всего пользуются *параметрическими* методами, когда предполагается, что регрессионная функция имеет известную функциональную форму и конечное число неизвестных параметров. Оценивание этих параметров автоматически дает оценки для  $g(x)$ . Естественно, неверная спецификация функциональной формы может привести к серьезным искажениям при оценивании и инференции, причем часто непредсказуемым (см. Крил, 2008).

В настоящем эссе мы даем обзор *непараметрического* оценивания регрессии среднего, то есть такого, при котором избегают параметрических предположений о функциональной форме. Как и большая часть соответствующей литературы, мы делаем акцент на методах ядерного сглаживания. В отличие от остальной литературы, однако, мы рассматриваем с самого начала регрессию среднего, избегая предварительного разговора об оценивании плотности. Другие источники информации на данную тему включают обзоры Härdle & Linton (1994) и Расин (2008), а также монографии Härdle (1990), Pagan & Ullah (1999) и Li & Racine (2007).

### 2 Построение непараметрической оценки

Мы вначале предполагаем, что регрессор  $x$  – единственный. Впоследствии мы обсудим и случай многопеременной регрессии.

#### 2.1 Дискретный регрессор

Прежде всего рассмотрим случай дискретного регрессора. Пусть носитель  $x$  сосредоточен в  $a_{(1)}, \dots, a_{(k)}$ , где  $a_{(1)} < \dots < a_{(k)}$ , и  $k$  конечно (если носитель – бесконечное, но счетное множество, мало что меняется в анализе). Зафиксируем  $a_{(j)}$ ,  $j = 1, \dots, k$ . Заметим, что

$$g(a_{(j)}) = \mathbb{E}[y|x = a_{(j)}] = \frac{\mathbb{E}[y \mathbb{I}\{x = a_{(j)}\}]}{\mathbb{E}[\mathbb{I}\{x = a_{(j)}\}]}$$

\*Работа основана на лекциях, читаемых автором в РЭШ. Цитировать как: Анатольев, Станислав (2009) «Непараметрическая регрессия», Квантиль, №7, стр. 37–52. Citation: Anatolyev, Stanislav (2009) “Nonparametric regression,” Quantile, No.7, pp. 37–52.

<sup>†</sup>Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1721(3). Электронная почта: [sanatoly@nes.ru](mailto:sanatoly@nes.ru)

из-за справедливости следующих равенств:

$$\begin{aligned}\mathbb{E} [\mathbb{I} \{x = a_{(j)}\}] &= \mathbb{P}\{x_i = a_{(j)}\}, \\ \mathbb{E} [y \mathbb{I} \{x = a_{(j)}\}] &= \mathbb{E} [y|x = a_{(j)}] \mathbb{P}\{x = a_{(j)}\}.\end{aligned}$$

Согласно принципу аналогий можно построить оценку  $\hat{g}(a_{(j)})$  как

$$\hat{g}(a_{(j)}) = \frac{\sum_{i=1}^n y_i \mathbb{I} \{x_i = a_{(j)}\}}{\sum_{i=1}^n \mathbb{I} \{x_i = a_{(j)}\}}, \quad (1)$$

что можно интерпретировать как среднее по наблюдениям, попадающим в вертикальное сечение  $x = a_{(j)}$ .

## 2.2 Непрерывный регрессор

Если регрессор непрерывно распределен, описанный метод не работает, так как в произвольном сечении  $x = a$  нечего усреднять (ибо туда попадет максимум одно наблюдение), хотя и  $f(a) \neq 0$ , где  $f(a)$  – значение плотности регрессора  $f(x)$  в  $a$ . Поэтому необходимо привлечь информацию откуда-то еще. Если регрессионная кривая непрерывна, наблюдения, попадающие в окрестность  $a$ , являются наиболее информативными о значении регрессии в  $a$ .

Выберем положительное  $h$  и назовем его *шириной окна* (хотя впоследствии необязательно будет задействовано окно в буквальном смысле). Обобщим формулу (1) следующим образом:

$$\hat{g}(a) = \frac{\sum_{i=1}^n y_i \mathbb{I} \{a - h \leq x_i \leq a + h\}}{\sum_{i=1}^n \mathbb{I} \{a - h \leq x_i \leq a + h\}}, \quad (2)$$

т.е. мы усредняем  $y$  по наблюдениям, попадающим в окно  $[a - h, a + h]$ . При изменении  $a$   $\hat{g}(a)$  описывает оцененную регрессионную кривую. Отметим, что последняя имеет разрывы из-за попадания в окно новых наблюдений и выпадения из него старых.

Информация от наблюдений, попавших в окно  $[a - h, a + h]$ , используется одинаково. То есть те наблюдения, которые попали в окно, участвуют в оценивании с равным весом, в то время как наблюдения, не попавшие в окно, вообще в нем не участвуют. Разумной представляется идея сделать веса зависимыми от расстояния от  $x_i$  до  $a$ , а также, возможно, использовать информацию во всех наблюдениях. Введем с этой целью симметричную *ядерную функцию*  $K(u)$ , интегрирующуюся в единицу, т.е.  $\int K(u) du = 1$ , где интеграл  $\int$  берется по всей области определения, которой может быть отрезок, обычно  $[-1, 1]$ , или вся числовая ось. Популярными ядрами являются:

$$\text{Равномерное:} \quad K(u) = \frac{1}{2} \mathbb{I} \{|u| \leq 1\},$$

$$\text{Треугольное:} \quad K(u) = (1 - |u|) \mathbb{I} \{|u| \leq 1\},$$

$$\text{Епанечниково:} \quad K(u) = \frac{3}{4} (1 - u^2) \mathbb{I} \{|u| \leq 1\},$$

$$\text{Гауссово (нормальное):} \quad K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$

Область определения первых трех ядер – отрезок  $[-1, 1]$ , в то время как последнее имеет бесконечный носитель. Следовательно, при использовании равномерного, треугольного или Епанечникова ядра оценка будет использовать информацию в ограниченном окне в окрестности  $a$ , а оценка, использующая гауссово ядро, будет использовать информацию из всех наблюдений. Заметим также, что в принципе ядро не обязано быть всюду неотрицательным.

Далее, введем обозначение

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right).$$

Теперь обобщим формулу (2) и получим

$$\hat{g}(a) = \frac{\sum_{i=1}^n y_i K_h(x_i - a)}{\sum_{i=1}^n K_h(x_i - a)}. \quad (3)$$

Оценка (3) называется *оценкой Надарайа–Уотсона* для регрессии среднего, в честь Nadaraya (1965) и Watson (1964). Заметим, что нормализация делением на  $h$  в определении  $K_h(u)$  не влияет на численное значение оценки и сделано лишь для удобства, в чем мы убедимся позднее.

Заметим, что при использовании трех последних ядер (треугольного, Епанечникова и гауссова), так же как и многих других, оцененная регрессионная кривая непрерывна, так как новые и старые наблюдения вводятся в формулу и выводятся из нее непрерывным образом по мере того как  $a$  меняется. При этом велика роль параметра ширины окна. Если  $h$  слишком велика, оценка задействует слишком много нерелевантной информации, что увеличивает смещение и приводит к явлению *сверхсглаживания*. Сверхсглаженная кривая слишком «линейная», в то время как ей положено быть более извилистой и более пристально отслеживать рисунок наблюдений. Если же  $h$  слишком мала, оценка задействует маловато точек, что увеличивает дисперсию и приводит к явлению *недосглаживания*. Недосглаженная кривая слишком извилистая, так как она слишком пристально отслеживает отдельные наблюдения.

### 3 Асимптотические свойства

В случае дискретного регрессора легко получить, используя ЗБЧ и ЦПТ, что

$$\sqrt{n} (\hat{g}(a_{(j)}) - g(a_{(j)})) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\mathbb{V}[y|x = a_{(j)}]}{\mathbb{P}\{x = a_{(j)}\}} \right). \quad (4)$$

Интерпретация выражения для асимптотической дисперсии следующая. Качество оценивания положительно связано с частотой попадания точек в вертикальное сечение  $x = a_{(j)}$  и отрицательно связано со степенью разброса точек вдоль него. Заметим, что скорость сходимости оценки – параметрическая,  $\sqrt{n}$ . Действительно, задачу можно трактовать как параметрическую, ибо вектор параметров  $(a_{(1)}, \dots, a_{(k)})'$  конечномерен.

В случае непрерывно распределенного регрессора необходимо, чтобы ширина окна асимптотически падала до нуля, иначе смещение из-за нерелевантности используемой информации из соседних к  $a$  наблюдений испортит состоятельность оценки. Таким образом, необходимо установить правило  $h \rightarrow 0$  по мере того как  $n \rightarrow \infty$ . С другой стороны, ширина окна не должна падать слишком быстро, иначе дисперсия из-за малого количества участвующих в оценивании точек не будет падать, что также испортит состоятельность оценки. Точнее, поскольку дисперсия обратно пропорциональна эффективному количеству участвующих наблюдений, которое в свою очередь прямо пропорционально  $nh$ , необходимо установить правило  $nh \rightarrow \infty$  по мере того как  $h \rightarrow 0$  и  $n \rightarrow \infty$ .

Обозначим

$$\sigma_K^2 = \int u^2 K(u) du$$

и

$$R_K = \int K(u)^2 du.$$

Эти две константы зависят только от выбранного ядра. Подразумевается, что обе величины конечны. Далее, установим дополнительное требование к скорости падения ширины окна:

$$\lambda = \lim_{n \rightarrow \infty} \sqrt{nh^5},$$

предполагая  $\lambda < \infty$ . Заметим, что  $\lambda$  может равняться, а может и не равняться нулю. Мы также предполагаем непрерывность и ограниченность  $g(x)$ ,  $g'(x)$ ,  $g''(x)$ ,  $f(x)$  и  $f'(x)$  всюду на области определения.

Рассмотрим разницу между оценкой и оцениваемой величиной:

$$\hat{g}(a) - g(a) = \frac{\hat{q}_1(a) + \hat{q}_2(a)}{\hat{f}(a)},$$

где

$$\begin{aligned}\hat{q}_1(a) &= \frac{1}{n} \sum_{i=1}^n e_i K_h(x_i - a), \\ \hat{q}_2(a) &= \frac{1}{n} \sum_{i=1}^n (g(x_i) - g(a)) K_h(x_i - a), \\ \hat{f}(a) &= \frac{1}{n} \sum_{i=1}^n K_h(x_i - a),\end{aligned}$$

и за  $e_i$  обозначены, как обычно, регрессионные ошибки в точках выборки:  $e_i = y_i - g(x_i)$ .

Начнем со знаменателя

$$\hat{f}(a) = \frac{1}{n} \sum_{i=1}^n K_h(x_i - a).$$

Величина  $\hat{f}(a)$ , называемая *оценкой плотности Надарайа–Уотсона*, и правда оценивает плотность регрессора  $f(x)$  в точке  $a$ , отсюда и обозначения. Действительно, рассмотрим

$$\begin{aligned}\mathbb{E}[\hat{f}(a) - f(a)] &= \mathbb{E}[K_h(x - a)] - f(a) \\ &= \frac{1}{h} \int K\left(\frac{x - a}{h}\right) f(x) dx - f(a) \\ &= \int K(u) f(a + hu) du - f(a) \\ &= \int K(u) f(a + hu) du - f(a).\end{aligned}$$

Разложим  $f(a + hu)$  до первого порядка:

$$\begin{aligned}\mathbb{E}[\hat{f}(a) - f(a)] &= \int K(u) (f(a) + f'(a) hu + o(h)) du - f(a) \\ &= f(a) \int K(u) du + f'(a) h \int u K(u) du + o(h) - f(a) \\ &= o(h),\end{aligned}$$

так как ядро интегрируется в единицу и симметрично. Использую ту же технологию,

$$\begin{aligned}\mathbb{V}[\hat{f}(a)] &= \frac{1}{n} \mathbb{V}[K_h(x - a)] \\ &= \frac{1}{n} \mathbb{E}[K_h(x - a)^2] - \frac{1}{n} \mathbb{E}[K_h(x - a)]^2 \\ &= \frac{1}{n} \int \left(\frac{1}{h} K\left(\frac{x - a}{h}\right)\right)^2 f(x) dx - \frac{1}{n} \left(\int \frac{1}{h} K\left(\frac{x - a}{h}\right) f(x) dx\right)^2 \\ &= \frac{1}{n h} \int K(u)^2 f(a + hu) du - \frac{1}{n} \left(\int K(u) f(a + hu) du\right)^2 \\ &= \frac{1}{n h} \int K(u)^2 (f(a) + o(1)) du - \frac{1}{n} O(1) \\ &= O\left(\frac{1}{n h}\right).\end{aligned}$$

Поскольку  $h \rightarrow 0$  и  $nh \rightarrow \infty$ , имеем  $\mathbb{E}[\hat{f}(a) - f(a)] \rightarrow 0$  и  $\mathbb{V}[\hat{f}(a)] \rightarrow 0$ , откуда следует, что действительно  $\hat{f}(a) \xrightarrow{P} f(a)$ .

Теперь рассмотрим первую часть числителя

$$\hat{q}_1(a) = \frac{1}{n} \sum_{i=1}^n e_i K_h(x_i - a).$$

Это среднее регрессионных ошибок, взвешенных ядром. Как и в параметрическом анализе, такое среднее должно дать асимптотическую нормальность. Разница в том, что дисперсия отдельного слагаемого здесь не постоянна, а зависит от асимптотически падающего  $h$ . Следовательно, необходимо рассчитать предел этой дисперсии. Представим  $\sqrt{nh}\hat{q}_1(a)$  как

$$\sqrt{nh}\hat{q}_1(a) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{e_i}{\sqrt{h}} K\left(\frac{x_i - a}{h}\right).$$

Дисперсия отдельного слагаемого равна

$$\begin{aligned} \mathbb{V}\left[\frac{e}{\sqrt{h}} K\left(\frac{x-a}{h}\right)\right] &= \mathbb{E}\left[\frac{\sigma^2(x)}{h} K\left(\frac{x-a}{h}\right)^2\right] \\ &= \frac{1}{h} \int \sigma^2(x) K\left(\frac{x-a}{h}\right)^2 f(x) dx \\ &= \int \sigma^2(a+hu) K(u)^2 f(a+hu) du \\ &= \int \sigma^2(a) K(u)^2 f(a) du + o(1) \\ &= \sigma^2(a) f(a) R_K + o(1). \end{aligned}$$

По ЦПТ Линдберга–Леви,  $\sqrt{nh}\hat{q}_1(a) \xrightarrow{d} \mathcal{N}(0, \sigma^2(a) f(a) R_K)$ .

Наконец, рассмотрим вторую часть числителя

$$\hat{q}_2(a) = \frac{1}{n} \sum_{i=1}^n (g(x_i) - g(a)) K_h(x_i - a).$$

Это среднее взвешенных ядром отклонений значений регрессионной функции в точках наблюдений от ее значения в точке  $a$ , где она оценивается. Эти отклонения рожают смещение. Конечно же,  $\hat{q}_2(a)$  обладает и дисперсией, но она мала по сравнению с дисперсией, рожаемой в  $\hat{q}_1(a)$  регрессионными ошибками.

Рассмотрим математическое ожидание  $\hat{q}_2(a)$ :

$$\begin{aligned} \mathbb{E}[\hat{q}_2(a)] &= \mathbb{E}[(g(x) - g(a)) K_h(x - a)] \\ &= \frac{1}{h} \int (g(x) - g(a)) K\left(\frac{x-a}{h}\right) f(x) dx \\ &= \int (g(a+hu) - g(a)) K(u) f(a+hu) du \\ &= \int \left(g'(a) hu + \frac{g''(a)}{2} (hu)^2 + o(h^2)\right) K(u) (f(a) + f'(a) hu + o(h)) du \\ &= g'(a) f(a) h \int u K(u) du \\ &\quad + \left(g'(a) f'(a) + \frac{g''(a)}{2} f(a)\right) h^2 \int u^2 K(u) du + o(h^2) \\ &= h^2 f(a) B(a) \sigma_K^2 + o(h^2), \end{aligned}$$

где

$$B(a) = \frac{g'(a)f'(a)}{f(a)} + \frac{g''(a)}{2}.$$

Легко также получить, что

$$\mathbb{V}[\hat{q}_2(a)] = o\left(\frac{1}{nh}\right).$$

Эти два результата приводят к тому, что

$$\sqrt{nh}\hat{q}_2(a) \xrightarrow{p} \lambda f(a) B(a) \sigma_K^2.$$

Все выведенное выше в совокупности дает

$$\begin{aligned} \sqrt{nh}(\hat{g}(a) - g(a)) &= \frac{\sqrt{nh}\hat{q}_1(a) + \sqrt{nh}\hat{q}_2(a)}{\hat{f}(a)} \\ &\xrightarrow{d} \frac{\mathcal{N}(0, \sigma^2(a) f(a) R_K) + \lambda f(a) B(a) \sigma_K^2}{f(a)} \\ &\sim \mathcal{N}\left(\lambda B(a) \sigma_K^2, \frac{\sigma^2(a)}{f(a)} R_K\right). \end{aligned}$$

Отметим две интересные черты этого асимптотического результата. Во-первых, непараметрическая скорость сходимости  $\sqrt{nh}$  меньше, чем параметрическая  $\sqrt{n}$ , поскольку асимптотически  $h \rightarrow 0$ . Этот факт отражает меньшую точность оценивания бесконечномерных объектов, чем точность оценивания конечномерных. Во-вторых, хотя асимптотическое распределение и нормальное, оно нецентрированное. Асимптотическое смещение отражает тот факт, что информация, используемая при оценивании, не является всецело релевантной.

Формула для асимптотической дисперсии похожа на свой аналог в случае дискретного регрессора: скедастическая функция в числителе и «вероятностная масса» в знаменателе. Асимптотическое смещение зависит от множества характеристик форм регрессионной и плотностной функций, зашитых в формуле для  $B(a)$ . Одна часть асимптотического смещения пропорциональна  $g'(a)f'(a)$  и отражает смещение, возникающее, когда регрессионная кривая имеет наклон, а наблюдения падают несимметрично слева и справа от  $a$ , в результате чего точки слева и точки справа создают неодинаковое и взаимно не компенсирующееся смещение вверх и вниз. Вторая часть асимптотического смещения пропорциональна  $g''(a)$  и отражает смещение, возникающее, когда регрессионная кривая локально нелинейна, в результате чего ординаты точек слева и точек справа от  $a$  несимметрично распределены выше и ниже  $g(a)$ , даже если их абсциссы распределены симметрично слева и справа от  $a$ . Отметим, что и дисперсия, и смещение в общем случае обратно пропорциональны плотности  $f(a)$ , что отражает тот факт, что точность оценивания, и в смысле дисперсии, и в смысле смещения, низка при оценивании  $g(a)$  около тех границ носителя  $x$ , где плотность убывает в ноль.

Полученный асимптотический результат также означает, что оптимальная скорость падения ширины окна  $h \propto n^{-1/5}$ , так как в этом случае  $\lambda > 0$ , и асимптотические смещение и дисперсия уравновешены. С другой стороны, если положить  $h = o(n^{-1/5})$ , можно добиться того, что  $\lambda$  будет равно нулю, и асимптотическое смещение исчезнет. Конечно, это удобно с точки зрения реализации, т.к. в таком случае нет необходимости оценивать компоненты  $B(a)$ , но подобные действия скрывают истинную картину, связанную со смещением оценки, и могут привести к плохому качеству асимптотического приближения.

Выведенный выше асимптотический результат означает, что приближенно

$$\hat{g}(a) \sim \mathcal{N}\left(g(a) + \frac{\sqrt{nh^5} B(a) \sigma_K^2}{\sqrt{nh}}, \frac{\sigma^2(a) R_K}{f(a) nh}\right).$$

Как обычно, асимптотическое распределение можно использовать для тестирования статистических гипотез о  $g(a)$  и построения доверительных интервалов для этой величины. Доверительный интервал, например, выглядит так:

$$\hat{g}(a) - h^2 \hat{B}(a) \sigma_K^2 \mp z_{1-\alpha/2} \sqrt{\frac{\hat{\sigma}^2(a) R_K}{\hat{f}(a) nh}},$$

где  $z_{1-\alpha/2}$  –  $(1 - \alpha/2)$ -квантиль стандартного нормального распределения, а  $\hat{f}(a)$ ,  $\hat{\sigma}^2(a)$  и  $\hat{B}(a)$  – непараметрические оценки соответствующих функций в точке  $a$ . Заметим, что если построить доверительные интервалы для  $g(a)$  при всех значениях  $a$  (на решетке) с вероятностью покрытия  $1 - \alpha$ , возникнет *поточечный доверительный коридор* для  $g(x)$ . Естественно, неверно говорить, что истинная регрессионная кривая находится внутри этого коридора с вероятностью  $1 - \alpha$ .

## 4 Выбор ширины окна

### 4.1 Правило подстановки

Из выведенного выше результата следует, что асимптотическая среднеквадратическая ошибка оценивания равна

$$AMSE(a) = h^4 B(a)^2 \sigma_K^4 + \frac{\sigma^2(a) R_K}{f(a) nh}.$$

Если минимизировать эту величину по  $h$ , то возникнет *правило подстановки* для (локально) оптимальной ширины окна

$$h^*(a) = \left( \frac{\sigma^2(a) R_K}{4f(a) B(a)^2 \sigma_K^4} \right)^{1/5} n^{-1/5}.$$

Заметим, что скорость падения равна оптимальной, выведенной выше.

Практическое применение правила подстановки заключается в (непараметрическом!) оценивании  $f(a)$ ,  $f'(a)$ ,  $g'(a)$ ,  $g''(a)$ ,  $\sigma^2(a)$ , вычислении  $R_K$ ,  $\sigma_K^4$  и подстановке полученных результатов в формулу для  $h^*(a)$ . Эта трудоемкая процедура дает численное значение оптимальной ширины окна всего для одного значения  $a$ , так что ее приходится повторить для всех  $a$ . Конечно, это очень нелегко, и исследователи чувствовали бы себя комфортнее с одним-единственным значением *глобально оптимальной ширины окна*  $h^*$ , общей для всех  $a$ .

Глобально оптимальную ширину окна легко вывести, минимизируя критерий интегрированной асимптотической среднеквадратической ошибки оценивания

$$IAMSE(a) = \int AMSE(a) da = h^4 \sigma_K^4 \int B(a)^2 da + \frac{R_K}{nh} \int \frac{\sigma^2(a)}{f(a)} da.$$

Естественно, можно ввести какую-либо взвешивающую схему, если есть причины не использовать равномерное взвешивание для разных  $a$ . При равномерном взвешивании

$$h^* = \left( \frac{\int \sigma^2(a) / f(a) da R_K}{4 \int B(a)^2 da \sigma_K^4} \right)^{1/5} n^{-1/5}.$$

Данная стратегия, однако, не лишает исследователя необходимости оценивать  $f(a)$ ,  $f'(a)$ ,  $g'(a)$ ,  $g''(a)$ ,  $\sigma^2(a)$ , то есть процедура настолько же трудоемкая, как и прежде. Статистик Бернард Сильверман предложил универсальную формулу, основанную на вышеописанной процедуре, но специфичной для определенного частного случая, например, в предположении

нормальной плотности  $f$ .<sup>1</sup> Вот эта универсальная формула, называемая *правилом Сильвермана*:

$$h^S = 1.364 \left( \frac{R_K}{\sigma_K^4} \right)^{1/5} \hat{\sigma}_x n^{-1/5},$$

где  $\hat{\sigma}_x^2$  – выборочная дисперсия регрессора. В частности, для гауссова ядра формула выглядит как  $h^S = 1.06 \hat{\sigma}_x n^{-1/5}$ .

На практике правило Сильвермана обычно обеспечивает приемлемые результаты, кроме, возможно, тех случаев, когда оценивание происходит вблизи краев носителя  $x$ . Иногда, впрочем, вид оцененной регрессионной кривой неудовлетворителен, и подобную ширину окна используют лишь как стартовое значение при поиске более приемлемой.

## 4.2 Кросс-валидация

Радикально иное правило выбора глобальной ширины окна – это *кросс-валидация*. Она основана на качестве подгонки, а не на асимптотических свойствах оценки. Можно обычную среднеквадратическую ошибку, оцененную в точках выборки,

$$MSE(h) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{g}(x_i))^2,$$

устремить к нулю, что приводит к идеальной подгонке, если положить  $h$  равным своему наименьшему значению, когда каждое наблюдение объясняется только им же самим. Естественно, это не что иное как экстремальная степень недосглаживания, и оно нас не устраивает. От источника проблемы можно легко избавиться, если запретить объяснять наблюдение самим собой. Исходя из этого, разумным критерием будет *функция кросс-валидации*

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{g}^{-i}(x_i))^2,$$

где  $\hat{g}^{-i}(x_i)$  – это оценка Надарайа–Уотсона в точке  $x_i$ , которая (оценка) использует все наблюдения, за исключением  $i$ -го, т.е.

$$\hat{g}^{-i}(x_i) = \frac{\sum_{j=1, j \neq i}^n y_j K_h(x_j - x_i)}{\sum_{j=1, j \neq i}^n K_h(x_j - x_i)}.$$

Оптимальная в смысле кросс-валидации ширина окна  $h^{CV}$  минимизирует  $CV(h)$ . К сожалению, при практическом применении данная ширина окна часто приводит к сильному недосглаживанию. Следовательно, в этих случаях ее можно использовать в качестве начального приближения для приемлемой ширины окна, и окончательное решение опять же остается за визуальным анализом.

## 5 Многопеременная ядерная регрессия

До сих пор регрессор  $x$  был единственным. Как обобщить оценку Надарайа–Уотсона на постановку с множественными регрессорами?

Обозначим через  $d$  размерность  $x$ . У ядра теперь будет  $d$ -мерный аргумент, отображающий расстояние между  $a$  и каждым  $x_i$  в  $d$ -мерном пространстве в скалярный вес:  $K : \mathbb{R}^d \rightarrow \mathbb{R}$ .

<sup>1</sup>На самом деле Сильверман использовал оптимальную ширину окна для оценки плотности Надарайа–Уотсона. См. одну из задач в конце эссе.



Определим  $d \times d$ -мерную симметричную и положительно определенную *матрицу ширины окна*  $H$ . Далее, положим

$$K_H(u) = \frac{1}{\det H} K(H^{-1}u).$$

Оценка Надарайа–Уотсона выглядит по-прежнему как

$$\hat{g}(a) = \frac{\sum_{i=1}^n y_i K_H(x_i - a)}{\sum_{i=1}^n K_H(x_i - a)}.$$

Есть много способов организовать структуру  $K$  и  $H$ . На практике широко используются два способа.

Первый использует разложение  $d$ -мерного пространства на произведение  $d$  одномерных:

$$K_H(u) = \prod_{\ell=1}^d K_{h_{\ell}, \ell}(u_{\ell}) = \prod_{\ell=1}^d \frac{1}{h_{\ell}} K_{\ell}\left(\frac{u_{\ell}}{h_{\ell}}\right),$$

где  $K_{\ell}$  и  $h_{\ell}$  – ядро и ширина окна, соответственно, по  $\ell$ -ой размерности. Такое  $K_H(u)$  называется *ядром-произведением*.  $K_{\ell}$  потенциально могут быть разными по разным координатам, но обычно они одинаковы. Матрица ширины окна равна  $H = \text{diag}\{h_{\ell}\}_{\ell=1}^d$ , и каждая  $h_{\ell}$  выбирается отдельно (например, в простейшем случае – с помощью правила Сильвермана).

Ядро-произведение игнорирует зависимость между регрессорами. Кроме того, выбор  $d$  значений ширины окна не особенно привлекателен. У второго метода нет этих недостатков. Матрица ширины окна имеет следующую структуру:

$$H = hS^{1/2},$$

где  $h$  – единая ширина окна,  $S$  – выборочная дисперсионная матрица регрессоров, а  $S^{1/2}$  – квадратный корень из нее (например, определяемый через разложение Холецки).

К сожалению, точность оценивания при больших значениях  $d$  мала. Это явление называют *проклятием размерности*. Причина в том, что при прочих равных в  $d$ -мерное гиперокно попадает намного меньше наблюдений, чем в его одномерный аналог, и дисперсия оценки быстро возрастает с увеличением  $d$ . В частности, скорость сходимости (при второй схеме) равна  $\sqrt{nh^d}$ , а оптимальная ширина окна –  $O(n^{-1/(d+4)})$ . На практике непараметрическое оценивание обычно возможно только при очень малом количестве регрессоров  $d$  вроде 1, 2, 3, редко выше, и требует больших выборок для достижения приличной надежности оценивания.

## 6 Локальная полиномиальная регрессия

Вернемся к скалярному  $x$ . Заметим, что оценку Надарайа–Уотсона можно представить в виде решения задачи минимизации взвешенной суммы квадратов:

$$\hat{g}(a) = \arg \min_{\beta_0} \sum_{i=1}^n (y_i - \beta_0)^2 K_h(x_i - a).$$

Это означает, что оценивание Надарайа–Уотсона – это локальная (в том смысле, что взвешивание основано на локальности наблюдений к  $a$ ) регрессия на константе. Нет причин останавливаться на регрессии на константе. Естественной является идея расширить ее на линейную регрессию не только на константе, но также и на  $x$ . Результатом будет *локальная линейная регрессия*:

$$\hat{g}_1(a) = (1, 0) \arg \min_{(\beta_0, \beta_1)} \sum_{i=1}^n (y_i - \beta_0 - \beta_1(x_i - a))^2 K_h(x_i - a).$$

Вектор  $(1, 0)$  отбирает первый элемент вектора  $2 \times 1$ . В качестве сопутствующего результата второй элемент этого вектора дает локально линейную оценку наклона регрессионной кривой в  $a$ , то есть ее первой производной  $g'(a)$ .

Оценка локальной линейной регрессии имеет то преимущество перед оценкой Надарайа–Уотсона, что она учитывает наклонность регрессионной кривой, что имеет значение, если наблюдения распределены неравномерно вокруг  $a$ , и таким образом уменьшает смещение. Это проявляется в асимптотических свойствах оценки, идентичных свойствам оценки Надарайа–Уотсона, кроме того момента, что теперь

$$B(a) = \frac{g''(a)}{2}.$$

Локальную полиномиальную регрессию можно обобщить далее и получить *локальную полиномиальную регрессию порядка  $p$* :

$$\hat{g}_p(a) = (1, 0, \dots, 0) \arg \min_{(\beta_0, \dots, \beta_p)} \sum_{i=1}^n (y_i - \beta_0 - \dots - \beta_p (x_i - a)^p)^2 K_h(x_i - a).$$

Данная оценка учитывает не только наклонность, но и свойства кривизны регрессионной кривой в точке  $a$ , и в качестве сопутствующего результата дает оценки производных  $g(x)$  до  $p$ -го порядка в точке  $a$ . Удобно, что оценку можно записать в явном виде как оценку взвешенного метода наименьших квадратов:

$$\hat{g}_p(a) = (1, 0, \dots, 0) (X'WX)^{-1} X'WY,$$

где  $Y = (y_1, \dots, y_n)'$ ,  $X = (X_1, \dots, X_n)'$ ,  $X_i = (1, x_i - a, \dots, (x_i - a)^p)'$ , и, наконец,  $W = \text{diag} \{K_h(x_i - a)\}_{i=1}^n$ . Если  $p > 1$ , асимптотическое смещение  $B(a)$  равно нулю. Это означает, что оптимальная ширина окна и результирующая скорость сходимости оценки меняются.

При практическом применении существуют серьезные недостатки использования локальной полиномиальной регрессии. Использование информации в локальных непараметрических методах очень ограничено, а увеличение  $p$  означает уменьшение степеней свободы. В качестве экстремального примера рассмотрим равномерное ядро и узкое окно, настолько узкое, что в него попадает лишь два наблюдения. Оценка Надарайа–Уотсона усредняет ординаты этих двух наблюдений, локальная линейная регрессия соединяет их прямой и берет ординату пересечения с вертикальной прямой  $x = a$ , а локальная полиномиальная регрессия при  $p > 1$  попросту не существует.

На практике стоит ограничиваться малыми  $p$ , вроде 0, 1 или 2, не больше.

## 7 Временные ряды

Если вместо случайной выборки у нас стационарный временный ряд, основные методы, приведенные выше, в целом работают. Интересно, что, в отличие от параметрических задач, в асимптотической дисперсии не возникают автоковариации  $y_t$  в  $x_t = a$ . Причина в том, что эффект последних асимптотически исчезает на фоне дисперсии  $y_t$  в  $x_t = a$  (см. Bierens, 1994). Таким образом, можно использовать те же самые формулы.

Типичное приложение непараметрических методов во временных рядах – непараметрическая авторегрессия

$$y_t = g(y_{t-1}, y_{t-2}, \dots, y_{t-k}) + \sigma(y_{t-1}, y_{t-2}, \dots, y_{t-k}) \eta_t$$

где  $\mathbb{E}[\eta_t | y_{t-1}, y_{t-2}, \dots] = 0$  и  $\mathbb{V}[\eta_t | y_{t-1}, y_{t-2}, \dots] = 1$ .

Оценку Надарайа–Уотсона  $\hat{g}(a_1, \dots, a_k)$  авторегрессионной функции  $g(y_{t-1}, \dots, y_{t-k})$  в  $(y_{t-1}, \dots, y_{t-k}) = (a_1, \dots, a_k)$  можно построить по знакомой схеме, а оценку автоскедастичной функции  $\sigma^2(y_{t-1}, \dots, y_{t-k})$  как

$$\hat{\sigma}^2(y_{t-1}, \dots, y_{t-k}) = \hat{\delta}(a_1, \dots, a_k) - \hat{g}(a_1, \dots, a_k)^2,$$

где  $\hat{\delta}(a_1, \dots, a_k)$  – оценка Надарайа–Уотсона непараметрической регрессии  $y_t^2$  на  $(y_{t-1}, \dots, y_{t-k})$  в точке  $(y_{t-1}, \dots, y_{t-k}) = (a_1, \dots, a_k)$ . Если порядок авторегрессии  $k$  неизвестен, его можно оценить совместно с регрессионной функцией (см., например, Tschernig & Yang, 2000). Обзор непараметрических методов в контексте временных рядов содержится в Heiler (2001).

## 8 Другие непараметрические методы

Непараметрический метод может быть одного из двух типов: локальный или глобальный. Оценка Надарайа–Уотсона, локальная линейная и локальная полиномиальная регрессии принадлежат классу локальных методов, так как оценивание  $g(x)$  в точке  $a$  использует информацию в наблюдениях, находящихся вблизи  $a$ . Глобальные непараметрические методы вместо этого пытаются подогнать всю кривую ко всем точкам выборки одновременно. При этом влияние одного наблюдения не так ограничено, и значения его абсциссы и ординаты влияют не только на оцененную регрессию поблизости этой точки, но и на положение и форму всей оцененной регрессионной кривой.

Независимо от того, локальный метод или глобальный, всегда наличествует *параметр сглаживания*, контролирующий степень последнего. Выбор этого параметра осуществляется исследователем. Параметр сглаживания в уже рассмотренных методах – это ширина окна.

### 8.1 Метод ближайших соседей

Другим локальным непараметрическим методом является *оценка  $k$  ближайших соседей*

$$\hat{g}_{NN}(a) = \frac{1}{k} \sum_{i=1}^n y_i 1_i,$$

где  $1_i = \mathbb{I}\{x_i \text{ является одним из } k \text{ ближайших соседей к } a\}$ , и о соседстве судится по тому, насколько близки абсциссы точек выборки к  $a$ . Здесь  $k$  является параметром сглаживания, и для состоятельности необходимы условия  $k \rightarrow \infty$  и  $k/n \rightarrow 0$  по мере того как  $n \rightarrow \infty$ . Другой версией является *симметризованная оценка  $k$  ближайших соседей*. Предполагая, что  $k$  четно,

$$\hat{g}_{SNN}(a) = \frac{1}{k} \sum_{i=1}^n y_i 1_i,$$

где  $1_i = \mathbb{I}\{x_i \text{ является одним из левых } k/2 \text{ или правых } k/2 \text{ ближайших соседей к } a\}$ . Асимптотические свойства этих оценок схожи со свойствами оценки Надарайа–Уотсона.

Одним из преимуществ методов ближайших соседей перед ядерными методами является существование оценки при любом раскладе, так как у любой точки есть соседи в любой выборке, в то время как в окно могут не попасть наблюдения совсем. Кроме того, количество усредняемых величин контролируется напрямую. Конечно же, обе идеи можно скомбинировать.

## 8.2 Разложение в серии (решето)

Оценивание сериями, или решето, – один из глобальных методов, основанный на разложении гладкой функции по базису в функциональном пространстве. Положим, существует, предпочтительно ортогональный, упорядоченный базис  $\{\psi_j(x)\}_{j=0}^{\infty}$ , такой, что

$$g(x) = \sum_{j=0}^{\infty} \gamma_j \psi_j(x).$$

Термин «упорядоченный» означает, что  $\psi_0(x)$  наиболее важный член,  $\psi_1(x)$  менее важный,  $\psi_2(x)$  еще менее важный и т.д. Например, базисом могут быть полиномы Эрмита или синусы и косинусы Фурье. Выберем параметр отсечения  $J$ , такой, что асимптотически  $J \rightarrow \infty$  и  $J/n \rightarrow 0$  по мере того как  $n \rightarrow \infty$ . Тогда

$$\hat{g}_S(x) = \sum_{j=0}^J \hat{\gamma}_j \psi_j(x),$$

где  $(\hat{\gamma}_0, \dots, \hat{\gamma}_J)'$  – вектор МНК-оценок в «линейной регрессии»  $y$  на  $(\psi_0(x), \dots, \psi_J(x))'$ . Этот метод также иногда называют полунепараметрическим, так как он непараметрический по сути, но параметрический по технической реализации. См. обзор Chen (2007), а также статью Кристенсен (2009) в настоящем номере журнала.

## 8.3 Искусственные нейронные сети

Идея искусственных нейронных сетей похожа на идею решета, т.е. аппроксимацию нелинейной функции линейной комбинацией неких базовых функций, в таком количестве, в каком необходимо для достижения хорошего приближения. В качестве базисных обычно используются логистические функции:

$$\hat{g}_{ANN}(x) = \hat{\phi}_{0,0} + \hat{\phi}_{1,0}x + \sum_{j=1}^J \frac{\hat{\gamma}_j}{1 + \exp(\hat{\phi}_{0,j} + \hat{\phi}_{1,j}x)},$$

где коэффициенты оцениваются нелинейным методом наименьших квадратов вместо МНК, используемого в решете. Расширенные версии задействуют интересные иерархические структуры, см. Franses & van Dijk (2000, глава 5).

## 8.4 Сплаины

Другим глобальным методом является оценивание сплайнами. Положим, мы хотим подогнать всю регрессионную кривую, используя обычный критерий подгонки, сумму квадратов ошибок. Это, конечно же, немедленно приведет к интерполяции, т.е. недосглаживанию максимальной степени. Однако мы можем добавить штрафной член, наказывающий за недосглаживание. Интерполирующая кривая слишком извилистая, то есть обладает большой второй производной по абсолютной величине. Эти идеи приводят к следующей оценке сплайнами:

$$\hat{g}_{CS}(x) = \arg \min_{\hat{g}(x)} \sum_{i=1}^n (y_i - \hat{g}(x_i))^2 + \lambda \int (\hat{g}''(u))^2 du,$$

где параметром сглаживания является  $\lambda$ . Если  $\lambda$  мало, решение близко к интерполирующей кривой, в то время как если  $\lambda$  очень большое, решение близко к линейному предиктору с точки зрения критерия наименьших квадратов. Индекс CS расшифровывается как «кубический сплайн», что означает решение в виде кусочно кубического полинома с непрерывно дифференцируемыми переходами в абсциссах точек наблюдений. Серьезное пособие по сплайнам – Wahba (1990).

## 9 Задачи

### 9.1 Оценка плотности Надарайа–Уотсона

Вывести асимптотическое распределение оценки Надарайа–Уотсона плотности скалярной случайной величины  $x$ , имеющей непрерывное распределение, аналогично тому, как выведено асимптотическое распределение оценки Надарайа–Уотсона регрессионной функции, при аналогичных предположениях. Дать интерпретацию зависимости выражений для асимптотического смещения и асимптотической дисперсии от формы плотности.

### 9.2 Несмещенность ядерных оценок

Рассмотрим оценку Надарайа–Уотсона  $\hat{g}(x)$  условного среднего  $g(x)$  для случайной выборки. Показать, что если  $g(x) = c$ , где  $c$  – некоторая константа, то оценка  $\hat{g}(x)$  несмещена. Какова интуиция за этим результатом? Выяснить, при каких обстоятельствах оценка локальной линейной регрессии  $g(x)$  будет несмещена. Будет ли оценка плотности  $f(x)$  несмещена?

### 9.3 Оценивание при ограничении на форму

Фирмы производят продукт, используя технологию  $f(l, k)$ . Функциональная форма  $f$  неизвестна, но известно, что она обладает свойством постоянного эффекта от масштаба. Для фирмы  $i$  наблюдается труд  $l_i$ , капитал  $k_i$  и выпуск  $y_i$ , а порождающий данные процесс принимает форму  $y_i = f(l_i, k_i) + \varepsilon_i$ , где  $\mathbb{E}[\varepsilon_i] = 0$ , и ошибка  $\varepsilon_i$  независима от  $(l_i, k_i)$ . Для случайной выборки  $\{y_i, l_i, k_i\}_{i=1}^n$  предложить непараметрическую оценку  $f(l, k)$ , которая бы тоже обладала свойством постоянного эффекта от масштаба.

### 9.4 Непараметрическая функция риска

Пусть  $z_1, \dots, z_n$  – скалярные независимые одинаково распределенные случайные величины с неизвестной плотностью  $f(\cdot)$  и функцией распределения  $F(\cdot)$ . Предположим, что распределение  $z$  имеет носитель  $\mathbb{R}$ . Возьмем  $t \in \mathbb{R}$ , такое что  $0 < F(t) < 1$ . Целью является оценивание функции риска

$$H(t) = \frac{f(t)}{1 - F(t)}.$$

Предложить непараметрическую оценку  $\hat{F}(t)$  для  $F(t)$ . Обозначим за  $\hat{f}(t)$  оценку Надарайа–Уотсона для  $f(t)$ , и выберем ширину окна  $h$  так, что  $nh^5 \rightarrow 0$ . Предложить оценку  $\hat{H}(t)$  для  $H(t)$ , использующую  $\hat{F}(t)$  и  $\hat{f}(t)$ , и найти ее асимптотическое распределение.

## 10 Решения задач

### 10.1 Оценка плотности Надарайа–Уотсона

Применим знакомую технологию:

$$\begin{aligned} \mathbb{E}[\hat{f}(a) - f(a)] &= \int K(u) \left( f(a) + hu f'(a) + \frac{1}{2}(hu)^2 f''(a) + O(h^3) \right) du - f(a) \\ &= \frac{h^2}{2} f''(a) \sigma_K^2 + O(h^3) \end{aligned}$$

и

$$\begin{aligned} \mathbb{V}[\hat{f}(a)] &= \frac{1}{nh} \int K(u)^2 (f(a) + o(1)) du - \frac{1}{n} O(1) \\ &= \frac{1}{nh} f(a) R_K + o\left(\frac{1}{nh}\right). \end{aligned}$$

Используя ЦПТ Линдберга–Леви, по мере того как  $n \rightarrow \infty$ ,  $h \rightarrow 0$  и  $nh \rightarrow \infty$ , имеем

$$\sqrt{nh} \left( \hat{f}(a) - f(a) \right) \xrightarrow{d} \mathcal{N} \left( \frac{1}{2} \lambda f''(a) \sigma_K^2, f(a) R_K \right),$$

при условии, что  $\lambda \equiv \lim_{n \rightarrow \infty} \sqrt{nh^5}$  существует и конечно.

Асимптотическое смещение пропорционально  $f''(a)$ , значение которой говорит о том, насколько плотность в окрестности  $a$  отличается от оцениваемой плотности в  $a$ . Отметим, что асимптотическое смещение не зависит от  $f(a)$ , т.е. как часто наблюдения попадают в данную область, и от  $f'(a)$ , т.е. отличаются ли плотности слева и справа от  $a$ . Асимптотическая дисперсия пропорциональна  $f(a)$ , плотности в  $a$ , что может показаться странным (бóльшая частота выпадения наблюдений приводит к худшему качеству оценивания). Однако мы оцениваем  $f(a)$ , так что ее большее значение также означает больший разброс оценки вокруг истинной величины, и этот эффект превалирует (эффект частоты дает  $\propto f(a)^{-1}$ , эффект размера дает  $\propto f(a)^2$ ).

## 10.2 Несмещенность ядерных оценок

Математическое ожидание равно

$$\begin{aligned} \mathbb{E}[\hat{g}(a)] &= \mathbb{E} \left[ \mathbb{E} \left[ \frac{\sum_{i=1}^n y_i K_h(x_i - a)}{\sum_{i=1}^n K_h(x_i - a)} \middle| x_1, \dots, x_n \right] \right] \\ &= \mathbb{E} \left[ \frac{\sum_{i=1}^n \mathbb{E}[y_i | x_i] K_h(x_i - a)}{\sum_{i=1}^n K_h(x_i - a)} \right] = \mathbb{E} \left[ \frac{\sum_{i=1}^n c K_h(x_i - a)}{\sum_{i=1}^n K_h(x_i - a)} \right] = c, \end{aligned}$$

т.е. оценка  $\hat{g}(a)$  несмещена для  $c = g(a)$ . Причина проста: все элементы выборки одинаково релевантны при оценивании тривиального условного среднего, так что от участия точек вдали от  $a$  смещение не возникает.

Оценка локальной линейной регрессии будет несмещена, если  $g(x) = c + bx$ . Тогда все элементы выборки одинаково релевантны при оценивании, так как оценивается линейная, хоть и локальная, регрессия. Действительно,

$$\hat{g}_1(a) = \bar{y} + \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) K_h(x_i - a)}{\sum_{i=1}^n (x_i - \bar{x})^2 K_h(x_i - a)} (a - \bar{x}),$$

так что

$$\begin{aligned} \mathbb{E}[\hat{g}_1(a)] &= \mathbb{E}[\mathbb{E}[\bar{y} | x_1, \dots, x_n]] \\ &\quad + \mathbb{E} \left[ \mathbb{E} \left[ \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) K_h(x_i - a)}{\sum_{i=1}^n (x_i - \bar{x})^2 K_h(x_i - a)} \middle| x_1, \dots, x_n \right] (a - \bar{x}) \right] \\ &= \mathbb{E}[c + b\bar{x}] \\ &\quad + \mathbb{E} \left[ \mathbb{E} \left[ \frac{\sum_{i=1}^n (c + bx_i - c - b\bar{x})(x_i - \bar{x}) K_h(x_i - a)}{\sum_{i=1}^n (x_i - \bar{x})^2 K_h(x_i - a)} \middle| x_1, \dots, x_n \right] (a - \bar{x}) \right] \\ &= \mathbb{E}[c + b\bar{x} + b(a - \bar{x})] = c + bx. \end{aligned}$$

Что же касается плотности, вряд ли стоит ожидать несмещенности. Действительно,

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_h(x_i - a),$$

так что

$$\mathbb{E}[\hat{f}(a)] = \mathbb{E}[K_h(x_i - a)] = \frac{1}{h} \int K \left( \frac{x_i - a}{h} \right) f(x) dx.$$

Это матожидание сильно зависит от ширины окна и ядра, и вряд ли будет равно  $f(x)$ , кроме как в особых условиях (например, равномерная  $f(x)$ ,  $a$  далеко от границ и т.д.).

### 10.3 Оценивание при ограничении на форму

Технология с постоянным эффектом от масштаба обладает свойством

$$f(l, k) = kf\left(\frac{l}{k}, 1\right).$$

Регрессия для нормированных переменных выглядит как

$$\frac{y_i}{k_i} = f\left(\frac{l_i}{k_i}, 1\right) + \frac{\varepsilon_i}{k_i}.$$

Поэтому можно построить (одномерную!) ядерную оценку для  $f(l, k)$  как

$$\hat{f}(l, k) = k \times \frac{\sum_{i=1}^n \frac{y_i}{k_i} K_h\left(\frac{l_i}{k_i} - \frac{l}{k}\right)}{\sum_{i=1}^n K_h\left(\frac{l_i}{k_i} - \frac{l}{k}\right)}.$$

По сути, мы присваиваем больший вес тем наблюдениям, которые ближе к лучу  $l/k$ .

### 10.4 Непараметрическая функция риска

Простая непараметрическая оценка для  $F(t) \equiv \Pr\{z \leq t\}$  – это выборочная частотность

$$\hat{F}(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}\{z_j \leq t\}.$$

Согласно ЗБЧ, она состоятельна для  $F(t)$ . Согласно ЦПТ, ее скорость сходимости равна  $\sqrt{n}$ .

Мы знаем из разделов 9.1 и 10.1, что

$$\sqrt{nh}(\hat{f}(t) - f(t)) \xrightarrow{d} \mathcal{N}(0, R_K f(t)),$$

используя  $\lambda = 0$ . По принципу аналогий,

$$\hat{H}(t) = \frac{\hat{f}(t)}{1 - \hat{F}(t)}.$$

По теореме Slutsky, эта оценка состоятельна для  $H(t)$ , а также

$$\begin{aligned} \sqrt{nh}(\hat{H}(t) - H(t)) &= \frac{\sqrt{nh}(\hat{f}(t) - f(t))}{1 - \hat{F}(t)} + \sqrt{h}f(t) \frac{\sqrt{n}(\hat{F}(t) - F(t))}{(1 - \hat{F}(t))(1 - F(t))} \\ &\xrightarrow{d} \frac{1}{1 - F(t)} \mathcal{N}(0, R_K f(t)) + 0 \sim \mathcal{N}\left(0, R_K \frac{f(t)}{(1 - F(t))^2}\right). \end{aligned}$$

Причина того, что неопределенность в  $\hat{F}(t)$  не влияет на асимптотическое распределение  $\hat{H}(t)$ , в том, что  $\hat{F}(t)$  сходится быстрее, чем  $\hat{f}(t)$ .

К сожалению,  $\hat{H}(t)$  выглядит не очень аппетитно из-за скачков в  $\hat{F}(t)$ .

## Список литературы

- Крил, М. (2008). Некоторые ловушки параметрической инференции. *Квантиль* 4, 1–6.
- Кристенсен, Д. (2009). Полупараметрическая эконометрика: вводный курс. *Квантиль* 7, 53–83.
- Расин, Дж. (2008). Непараметрическая эконометрика: вводный курс. *Квантиль* 4, 7–56.
- Bierens, H.J. (1994). *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models*. New York: Cambridge University Press.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. Глава 76 в *Handbook of Econometrics* (под редакцией J.J. Heckman & E.E. Leamer), том 6/2. Elsevier Science.
- Franses, P. & D. van Dijk (2000). *Nonlinear Time Series Models in Empirical Finance*. New York: Cambridge University Press.
- Härdle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.
- Härdle, W. & O. Linton (1994). Applied nonparametric methods. Глава 38 в *Handbook of Econometrics* (под редакцией R. Engle & D. McFadden), том 4. Elsevier Science.
- Heiler, S. (2001). Nonparametric time series analysis: nonparametric regression, locally weighted regression, autoregression, and quantile regression. Глава в *A Course in Time Series Analysis* под редакцией D. Peña, G. Tiao & R. Tsay. Wiley.
- Li, Q. & J.S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Nadaraya, E.A. (1965). On nonparametric estimates of density functions and regression curves. *Theory of Applied Probability* 10, 186–190.
- Pagan, A. & A. Ullah (1999). *Nonparametric Econometrics*. New York: Cambridge University Press.
- Tschernig, R. & L. Yang (2000). Nonparametric lag selection for time series. *Journal of Time Series Analysis* 21, 457–487.
- Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.
- Watson, G.S. (1964). Smooth regression analysis. *Sankhya* 26, 359–372.

## Nonparametric regression

Stanislav Anatolyev

*New Economic School, Moscow, Russia*

This essay covers the principles and methodology of nonparametric estimation of a mean regression. The emphasis is put on kernel smoothing, but non-kernel methods are also reviewed.