Отсутствие гладкости при непараметрическом оценивании*

Виктория Зинде-Уолш†

Университет МакГилл, Монреаль, Канада

Непараметрические оценки широко используются в статистике и эконометрике, при этом большая часть асимптотических результатов опирается на предположения о гладкости функции распределения; эти предположения могут не выполняться на практике. Нарушение предположений о гладкости распределения может иметь такие последствия, как неправильный выбор ширины окна, наличие серьезного асимптотического смещения и ошибок при тестировании. Подход, заключающийся в оптимальном комбинировании нескольких непараметрических оценок, каждая со своей парой ядро/ширина окна, может автоматически достичь (неизвестной заранее) оптимальной скорости сходимости. Этот метод был применен к ядерной оценке плотности распределения, к оценке усредненных производных, а также к сглаженной оценке максимальных очков в модели бинарного выбора. В крайнем случае, когда функция плотности не существует, ядерная оценка «оценивает» несуществующую функцию; тем не менее, ее предельное распределение можно полностью охарактеризовать с помощью обобщенных (в смысле обобщенных функций) Гауссовых процессов. Проверка гипотез о существовании плотности и ее гладкости – вопрос пока что малоисследованный; обсуждаются некоторые предварительные результаты.

1 Введение

Непараметрические оценки широко используются в статистике и эконометрике, при этом большая часть асимптотических результатов опирается на предположения о гладкости функции распределения. Однако зачастую на практике эти предположения могут не выполняться. Вмешательство в экономику государства и других регулирующих органов в форме, например, трансфертов или различных ставок процентов и налогов ведет к нарушению гладкости в функциях плотности распределения или даже в самом распределении соответствующих экономических процессов; подобное влияние на наблюдаемые переменные могут оказывать и особенности естественных или технологических процессов (например, необходимость полностью разгрузить электрический генератор при продаже электроэнергии). Нарушение предположений о гладкости распределения может иметь важные последствия для статистического оценивания и проверки гипотез, такие как неправильный выбор ширины окна, который приведет к замедленной (по сравнению с оптимальной) скорости сходимости, наличию серьезного асимптотического смещения и, возможно, к ошибкам при тестировании.

В последнее время были разработаны новые методы, цель которых – предохранить исследователя от ошибок, проистекающих либо от недостаточной гладкости, либо от неправильных предположений о гладкости. При условии существования функции плотности, но, возможно, не непрерывной, или непрерывной, но с неизвестным (и, возможно, недостаточным для стандартной асимптотики) количеством производных, этот подход заключается в

*Автор благодарит за поддержку Social Sciences and Humanities Research Council of Canada и Fonds Québéqois de la Recherche sur la Société et la Culture. Цитировать как: Зинде-Уолш, Виктория (2008) «Отсутствие гладкости при непараметрическом оценивании», Квантиль, №4, стр. 57–69. Citation: Zinde-Walsh, Victoria (2008) "Consequences of lack of smoothness in nonparametric estimation," Quantile, No.4, pp. 57–69.

†Адрес: Department of Economics, 855 Sherbrooke Street West, Montreal, Quebec, H3A 2T7, Canada. Электронная почта: *victoria.zinde-walsh@mcgill.ca*

оптимальном комбинировании нескольких непараметрических оценок, каждая со своей парой ядро/ширина окна; оказывается что такая комбинация может автоматически достичь (неизвестной заранее) оптимальной скорости сходимости, в частности, за счет снижения смещения (Kotlyarova & Zinde-Walsh, 2006). Этот метод был применен к ядерной оценке плотности рапределения (Kotlyarova & Zinde-Walsh, 2007), к оценке усредненных производных (например, для индексовой модели) (Schafgans & Zinde-Walsh, 2007), а также к сглаженной оценке максимальных очков в модели бинарного выбора (Kotlyarova & Zinde-Walsh, 2004). Во всех исследованных случаях методы Монте Карло подтверждают надежность комбинированного оценивания, в то время как общеизвестные способы оценивания могут приводить к существенным ошибкам для различных (даже гладких, но с несколькими пиками, например, смешанных нормальных) распределений. Подобная стратегия комбинирования (взвешенных) оценок рекомендуется и в других непараметрических оценках, как, например, в оценке параметра в моделях с долгой памятью (Guggenberger & Sun, 2006).

В случае, когда функция плотности не определена (распределение имеет сингулярности), ядерная оценка оценивает несуществующую функцию плотности; тем не менее ее предельное распределение можно полностью охарактеризовать с помощью обобщенных (в смысле обобщенных функций) Гауссовых процессов, введенных Гельфандом и Виленкиным (1961). Эти асимптотические результаты (Zinde-Walsh, 2008) дают возможность построения доверительных интервалов и тестирования. Интересно, что сингулярность в функции распределения не всегда ведет к отрицательным последствиям для оценки: например, сходимость условного ожидания ускоряется в точках сингулярности.

Очевидно, было бы желательно иметь тест для определения степени гладкости. Имеются тесты для определения фиксированных точек на сингулярность; эти тесты базируются на использовании специальных ядер; они получили распространение, начиная с работы Müller (1992). Обший тест на сингулярность был предложен в Frigyesi & Hössjer (1998), однако для практического применения этот тест малопригоден, так как даже для выбора между вполне обычным абсолютно непрерывным распределением (χ^2) и сингулярным, равномерным на равноотстоящих друг от друга ста точках в интервале (0,1), требуется огромная (несколько тысяч) выборка. Новый тест (Zinde-Walsh & Galbraith, 2007), основанный на асимптотической теории в Zinde-Walsh (2008), позволяет различать классы функций с различными характеристиками гладкости: тестовая статистика имеет стандартное нормальное асимптотическое распределение для любой непрерывной функции плотности, но стремится к бесконечности при наличии сингулярности. Тесты также позволяют различать классы функций с разным количеством непрерывных производных.

В следующем разделе приведены примеры, иллюстрирующие причины, по которым негладкие и сингулярные распределения встречаются на практике. Раздел 3 обьясняет общие свойства комбинированной оценки, особенности ее практического построения и результаты для таких моделей, как индексовая модель, модель бинарного выбора, и просто для оценки плотности вероятности. В разделе 4 приводятся результаты для случая, когда (в каких-то точках) функция распределения, возможно, не существует.

2 Примеры сингулярных распределений и негладких функций плотности

Многие прикладные исследователи сталкиваются с проблемой сингулярности в распределениях наблюдаемых величин; это обшеизвестные случаи в биологических и медицинских статистиках (например, пики риска смертности после инфаркта или младенца после родов); в экономических исследованиях также встречаются точечные веса, например, в проработанных неделях в связи с определенными ограничениями на право получения пособия по безработице (см. Green & Riddell, 1997). Также встречаются и негладкие или даже не непрерывные функции плотности.

Приведенные ниже примеры иллюстрируют причины, по которым распределения данных, таких как предложение труда, доходы от заработка, потребление и совместное распределение предложения труда мужчин и женщин, могут быть негладкими, например в результате одновременного трансферта. Более реалистичные формы пособий, сложные правила налогообложения приводят к менее очевидным нарушениям гладкости распределений.

Предположим, что семья предлагает рабочую силу в обьеме y, где $0 \leq y \leq \overline{y}$, и $\overline{y}$ имеет в популяции функцию распределения $F_{\overline{y}}(\cdot)$ (по предположению, достаточно гладкую) и плотность $f_{\overline{y}}(\cdot)$. Пусть y^* – величина предложения рабочей силы (равного заработанному доходу), а c – величина располагаемого дохода (равного потреблению). Семья максимизирует потребление c. Если $c(y) = y$, то $y^* = \overline{y}$. Теперь предположим, что трансферт в размере t предлагается при доходе, не превышающем y_t; тогда

$$c(y) = \begin{cases} y + t, & \text{если } y \leq y_t, \\ y, & \text{если } y > y_t. \end{cases} \tag{1}$$

Рассмотрим два случая.

Случай 1. Непредусмотренный трансферт. Предложение труда по-прежнему $y^* = \overline{y}$, но $c(y^*)$ определяется (1); функция распределения располагаемого дохода такова, что функция плотности разрывна:

$$f_c(c) = \begin{cases} f_{\overline{y}}(c - t), & \text{если } c \leq y_t, \\ f_{\overline{y}}(c - t) + f_{\overline{y}}(c), & \text{если } y_t < c \leq y_t + t, \\ f_{\overline{y}}(c), & \text{если } c > y_t + t. \end{cases}$$

Случай 2. Предусмотренный трансферт. В таком случае предложение рабочей силы меняется, чтобы максимизировать c в (1):

$$y^* = \begin{cases} \overline{y}, & \text{если } \overline{y} \leq y_t \text{ или } \overline{y} > y_t + t, \\ y_t, & \text{если } y_t < \overline{y} \leq y_t + t. \end{cases}$$

Тогда функция распределения предложения труда (и заработка) для семьи принимает вид

$$F_{y^*}(y^*) = F_{\overline{y}}(y^* - t)\mathbb{I}_{[y^* < y_t]} + [F_{\overline{y}}(y_t) - F_{\overline{y}}(y_t - t)]I(y^* = y_t) + F_{\overline{y}}(y^*)\mathbb{I}_{[y^* > y_t]},$$

где индикатор $\mathbb{I}_{[a]} = 1$, когда a верно, 0 в противном случае. Это распределение сингулярно (с точечным весом). Если рассмотреть при этих условиях совместное распределение предложения рабочей силы мужчин и женщин, веса могут быть как в отдельных точках, так и на одномерных подмножествах в зависимости от того, как принимаются совместные решения в семьях.

Дополнительный пример можно найти в олигополистических ценах на запасы электричества, где весовые точки обусловлены тем, что продажа осуществляется в количествах, соответствующих полной разгрузке генератора.

3 Проблемы недостаточной гладкости и комбинированные оценки

В этой части статьи мы исследуем вопрос, как наилучшим образом строить ядерные непараметрические оценки, когда функция плотности существует, но ее гладкость неизвестна.

3.1 Некоторые общие результаты

Допустим, мы имеем дело с ядерной непараметрической оценкой; такая оценка требует выбора функции ядра K и ширины окна h; в результате такого выбора получается оценка $b(h, K)$. Выбирая разные окна h_t, $t = 1, ..., m$, и ядра K_s, $s = 1, ..., l$; получаем разные оценки $b(t, s)$.

Обычно асимптотические результаты для такой оценки (в зависимости от свойств ядра и окна) выражаются в одной из следующих форм.

A1(1). Преобладание асимптотического смещения:

$$r(t,s)(b(t,s)-\beta)-A(t,s)\xrightarrow{p}0;$$

A1(2). Предельное нормальное распределение (со смещением):

$$r(t,s)(b(t,s)-\beta)\xrightarrow{d}N(A(t,s),\Gamma(t,s));$$

A1(3). Предельное нормальное распределение (без смещения): A1(2) с $A(t,s)=0$.

Выбор K и h обычно основывается на предположении о гладкости функции плотности. Когда нет уверенности в гладкости, выбор ядра и ширины окна, основывающийся на возможно ложных предположениях, может привести к ошибке.

Для разработки комбинированной оценки Kotlyarova & Zinde-Walsh (2006) рассматривают совместное предельное распределение нескольких оценок с разными K и h, про которые известно только то, что реализуется один из случаев в A1; для интересующих нас оценок обычно можно доказать следующее: предельное совместное распределение $b_j(=b(t,s))-$ нормальное, но, возможно, вырожденное:

$$\begin{pmatrix} r_1(b_1-\beta)\\ \vdots \\ r_J(b_J-\beta)\end{pmatrix}\sim N\left(\begin{pmatrix}A_1\\ \vdots\\ A_J\end{pmatrix},\begin{pmatrix}\Gamma_{11} & \cdots & \Gamma_{1J}\\ \vdots & \ddots & \vdots\\ \Gamma_{1J} & \cdots & \Gamma_{JJ}\end{pmatrix}\right).$$

Когда $\frac{r_i}{r_j}\to\infty$ (или к 0), $\Gamma_{ij}=0$, но также Γ_{ij} может равнятся нулю и когда $\frac{r_i}{r_j}\to d\neq 0,\infty$, так что во всех этих случаях соответствующие оценки распределены асимптотически независимо. Когда b_i слишком сглажено (случай A1(1)), $\Gamma_{ii}=0$ (вырожденное распределение); когда недостаточно сглажено (случай A1(3)), $A=0$ (нет смещения). Совместное распределение показывает, что разные оценки могут привносить дополнительную информацию; таким образом, линейная комбинация оценок может снизить величину среднеквадратичной ошибки (СКО); это может быть особенно полезным, когда из-за отсутствия уверенности в гладкости неизвестна оптимальная скорость сходимости непараметрической оценки.

Комбинированная оценка минимизирует оценку асимптотической СКО (АСКО) и может автоматически приобрести заранее неизвестную оптимальную скорость сходимости.

Как правило, комбинированные оценки в статистических и эконометрических исследованиях построены на выпуклых комбинациях (с неотрицательными коэффициентами). Juditsky & Nemirovki (2000) рассматривают оценку, построенную на основе набора заданных оценок регрессионной функции, дающую приближение, наилучшее среди выпуклых комбинаций заданных оценок; Yang (2000) в одной из статей рассматривает выпуклую комбинацию оценок для функции регрессии, а в другой – оценок для функции плотности; Fan & Ullah (1999) рассматривают выпуклую комбинацию параметрической и непараметрической регрессионных оценок с целью предохранения от неправильно определенной функции регрессии. Однако, в недавней статье об оценивании параметра в моделях с долгой памятью Guggenberger & Sun (2006) предлагают, подобно нам, использовать и отрицательные коэффициенты для уменьшения смещения оценки.

Чтобы сформулировать общий результат, определим вектор b, включающий все оценки (векторов) параметров, каждая из которых получена на основе какой-то пары ядро/окно:

$$b=\begin{pmatrix}b_1\\ \vdots\\ b_J\end{pmatrix},$$

и соответственно вектор $a:\ a'\iota = 1$; если $a \to \overline{a}$, то $a'b \sim N(\sum_j \overline{a}_j \tilde{A}_j, \sum_{i,j} \overline{a}_i \overline{a}_j \tilde{\Gamma}_{ij})$. Тогда

$$ACKO(a'b) = \sum_{i,j} a_i a_j (\tilde{A}_i' \tilde{A}_j + tr\tilde{\Gamma}_{ij}) = a'Da = a_I' D^I a_I + a_{II}' D^{II} a_{II},$$

где матрица D^I обратима и включает не более, чем одну оценку типа A1(1) (со смещением); D^{II} соответствует оценкам типа A1(1); $rk(D^{II}) = 1$, $\dim D^{II} \geq 2$.

Определим

$$a_n = \arg\min_a \ ACKO(a'b), \text{ где } a'\iota = 1; \tag{2}$$

а через D_{11} в D^I обозначим подматрицу, соответствующую наиболее быстрой скорости сходимости.

ТЕОРЕМА. Предположим, что существует предельное распределение при $n \to \infty$.
(a) Если $D \equiv D^I$, имеем $a_n = a^I_{\lim} + o(1)$, где

$$a^I_{\lim} = \left(\left(\frac{1}{\iota' D_{11}^{-1} \iota} D_{11}^{-1} \iota \right)', 0, ..., 0 \right)'. \tag{3}$$

(b) Если $D \equiv D^{II}$, имеем $\inf\limits_{a^{II}_{\lim} \in \Omega} \left\| a_n - a^{II}_{\lim} \right\| = o(1)$, где Ω состоит из векторов $a^{II}_{\lim}$, для которых

$$a^{II\prime}_{\lim} D^{II} a^{II}_{\lim} = 0;\ a^{II\prime}_{\lim} \iota = 1. \tag{4}$$

При общем D вектор коэффициентов a стремится к оптимальной комбинации $a^I_{\lim}$ и $a^{II}_{\lim}$. Таким образом, в случае (a) скорость сходимости совпадет с той, которая доминирует среди различных оценок, входящих в комбинацию, в случае (б) и в общем случае смещения могут взаимно сокращаться, приводя к лучшей сходимости оценки.

Примеры:

1. Две одинаково быстро сходящиеся оценки, b_1 и b_2, обе недостаточно сглажены. Тогда

$$a_{\lim} = \left(\frac{\Gamma_{22} - \Gamma_{12}}{\Gamma_{11} + \Gamma_{22} - 2\Gamma_{12}}, \frac{\Gamma_{11} - \Gamma_{12}}{\Gamma_{11} + \Gamma_{22} - 2\Gamma_{12}}, 0, ..., 0 \right)'$$

и у комбинации АСКО ниже по сравнению с каждой индивидуальной оценкой.

2. Для недостаточно сглаженной b_1 и слишком сглаженной b_2

$$a_{\lim} = \left(\frac{A_2^2}{\Gamma_{11} + A_2^2}, \frac{\Gamma_{11}}{\Gamma_{11} + A_2^2}, 0, ..., 0 \right)',$$

$$r_1(a'_{\lim} b - \beta) \xrightarrow{d} N \left(\frac{\Gamma_{11} A_2}{\Gamma_{11} + A_2^2}, \left(\frac{A_2^2}{\Gamma_{11} + A_2^2} \right)^2 \Gamma_{11} \right),$$

и у комбинации АСКО ниже по сравнению с каждой индивидуальной.

3. Две оценки, b_1 и b_2, слишком сглаженные. Тогда

$$a_{\lim} = \left(\frac{A_2}{A_2 - A_1}, -\frac{A_1}{A_2 - A_1}, 0, ..., 0 \right)',$$

и комбинация сходится быстрее, чем каждая из оценок.

При применении оптимальной комбинации используется оценка АСКО, $\widehat{ACKO(a'b_n)}$. Комбинированная оценка определяется как

$$b_{comb} = \sum_j a_{combj} b_j(n),$$

где

$$a_{comb} = \arg\min_{a_n} \ \widehat{ACKO}\left(\sum_j a_{nj} b_j(n)\right).$$

Можно показать, что при наличии состоятельной оценки для смещения и для дисперсии

$$a_{comb} = a_{\lim} + o_p(1).$$

Трудность состоит в получении состоятельных оценок, которые не основываются на предположениях о гладкости; оценки дисперсии могут быть получены разными методами, включая бутстрап. Для оценивания смещения предлагается использовать недостаточно сглаженные оценки, $b_i, i = 1, ..., L$; тогда возможная оценка смещения задается как $b_j - \frac{1}{L}\sum_{i=1}^{L} b_i$ (несостоятельная); к ней можно применить бутстрап и увеличивать L. Для получения недостаточно сглаженных оценок можно использовать кросс-валидацию или другую предварительную оценку ширины окна, а затем его сузить; конечно, построение этих оценок зависит от конкретной задачи.

Как показывают конкретные приложения, описанные в последующих разделах, применение комбинированной оценки дает преимущества над индивидуальным выбором пары ядро/ширина окна, хотя ни для одной из них теоретическая оптимальность метода комбинированной оценки пока не доказана из-за трудностей, связанных с определением оценки смещения и выводом ее предельных свойств.

3.2 Комбинированные оценки для плотности распределения

Этот подраздел основывается на статье Kotlyarova & Zinde-Walsh (2007).

Предположим, что (X_i), $i = 1, \ldots, n$, – независимая выборка из распределения X, и что функция плотности $f(x)$ непрерывна в точке x. Мы не делаем никаких предположений о гладкости этой функции. Ядро K удовлетворяет стандартным условиям $\int K(z)dz = 1$, $\int |K(z)|dz < \infty$, $|z||K(z)| \to 0$ при $|z| \to \infty$, $\sup |K(z)| < \infty$, $\int K(z)^{2+\delta}dz < \infty$ для какой-то $\delta > 0$, но K не обязано быть симметричным или неотрицательным. Ширина окна такова, что при $n \to \infty$, $h_n \to 0$ и $h_n n \to \infty$.

Ядерная оценка плотности распределения есть

$$\hat{f}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right). \tag{5}$$

При предположениях выше оценка плотности состоятельна для СКО и имеет предельное нормальное распределение (Parzen, 1962). Известно также, что если существуют две непрерывные производные, оптимальная ширина окна для ядра второго порядка – $h_n = cn^{-\frac{1}{5}}$, и тогда скорость сходимости оценки плотности равна $O(n^{-2/5})$ (см., например, Pagan & Ullah, 1999). При существовании производных более высокого порядка сходимость может быть ускорена за счет выбора ядра более высокого порядка.

Без предположений о гладкости смещение ядерной оценки плотности равно

$$B(h_n, K, x) = \mathbb{E}[\hat{f}(x) - f(x)] = \int K(z)\left[f(x + zh_n) - f(x)\right]dz. \tag{6}$$

В этом случае без дальнейших предположений известны только общие черты предельного распределения (как в предыдущем разделе): для маленькой ширины окна нет смещения и распределение нормальное, для большой распределение вырожденное и доминирует смещение; оптимальная ширина не может быть определена без дальнейших предположений.

Kotlyarova & Zinde-Walsh (2007) исследовали построение и поведение комбинированной оценки в этих условиях. В экспериментах Монте Карло были рассмотрены распределения с нормальной плотностью, смешанные нормальные распределения и распределения с недифференцируемой функцией плотности. Для этого набора ни один из обычно применяемых методов оценивания не привел к результатам, удовлетворительным для всех случаев. Так, например, ядро четвертого порядка с кросс-валидацией доминировало для нормального и смешанного распределений, но привело к существенной ошибке для распределения с недифференцируемой функцией плотности, где доминировало ядро второго порядка. Комбинированная оценка во всех случаях дала результаты лучше, чем неправильно выбранная обычная оценка; во многих случаях (особенно в негладком случае) комбинированная оценка приводит к меньшей СКО, чем наилучшая из обычных оценок.

3.3 Оценивание усредненных производных

Усредненные производные дают удобный способ непараметрического оценивания коэффициентов в индексовых моделях, примерами которых являются линейная регрессия, Тобит-модель, модель бинарного выбора и т.п., которые могут быть представлены в виде

$$y = g(x\beta) + u.$$

При дифференцируемой функции условного ожидания можно, например, получить (в предположении что функция плотности определена):

$$\delta_0 = \mathbb{E}\left[\frac{\partial g}{\partial x}(x) f(x)\right] = \mathbb{E}\left[\frac{\partial g}{\partial\,(x\beta)}(x) f(x)\right]\beta.$$

Из этого равенства следует, что оценка вектора параметров может быть с точностью до постоянного множителя получена на основе непараметрической оценки δ_0 (усредненной взвешанной функцией плотности производной).

Оценка для δ_0 была предложена в известной статье Powell, Stock & Stoker (1989):

$$\hat{\delta}_N(K,h) = \frac{-2}{N}\sum_{i=1}^{N}\hat{f}'_{(K,h)}(x_i)y_i, \tag{7}$$

с

$$\hat{f}'_{(K,h)}(x_i) = \frac{1}{N-1}\sum_{j\neq i}^{N}\left(\frac{1}{h}\right)^{k+1} K'\left(\frac{x_i - x_j}{h}\right),$$

где ядро K и ширина окна h такие, что $h \to 0$ при $N \to \infty$. При достаточной дифференцируемости (когда непрерывны $k+2$ производные, где k – количество переменных) скорость сходимости параметрическая, несмотря на использование непараметрической оценки, и асимптотическое распределение нормальное с асимптотической дисперсией $\Sigma_{2\delta}$.

В обшем случае асимптотическая дисперсия может быть представлена разложением

$$\mathbb{V}[\hat{\delta}_N(K,h)] = \Sigma_{1\delta}(K)N^{-2}h^{-(k+2)} + \Sigma_{2\delta}N^{-1} + O(N^{-2}), \tag{8}$$

и если $f(x)$ недостаточна гладкая, то может доминировать $\Sigma_{1\delta}(K)N^{-2}h^{-(k+2)}$, в результате чего скорость сходимости зависит от h. Более того, при недостаточной гладкости смещение может преобладать, приводя к вырожденному асимптотическому распределению. Таким образом, в общем случае при неизвестной степени гладкости распределения можно только указать общие черты, отмеченные в первом подразделе.

Powell & Stoker (1996) вывели оптимальную ширину окна для минимизации СКО, когда предположения о гладкости выполнены. В общем случае Scafgans & Zinde-Walsh (2007) показали, что оптимальная ширина

$$h^{opt} = cN^{-2/(2\overline{v}+k+2)} \tag{9}$$

зависит от (неизвестной) степени гладкости, которая определяет и оптимальный выбор порядка ядра, и дает наилучшую скорость сходимости (параметрическую или непараметрическую).

При неизвестной степени гладкости эта скорость априори неизвестна, и Scafgans & Zinde-Walsh (2007) выводят совместное асимптотическое распределение для оценок с разными ядрами и окнами, на основе которого оправдано применение комбинированной оценки. Пример в этой статье показывает, что при определенных условиях комбинированная оценка может достигать параметрической скорости сходимости, даже если каждая из составляющих оценок сходится медленнее, чем с параметрической скоростью.

Эксперимент Монте Карло использует тобит-модель с различными распределениями для (двух) переменных; эти распределения построены как смешанные нормальные и, хотя формально каждое из них бесконечно гладкое, некоторые производные настолько велики, что соответствующие смешанные распределения мало похожи на гладкие. Как и при оценивании функции плотности в предыдущем подразделе, ни один из индивидуальных методов оценивания не дает хороших результатов во всех случаях, для каждой индивидуальной оценки есть распределения, где СКО очень велики по сравнению с другими. В то же время комбинированная оценка дает стабильные результаты, хоть и не лучше, чем лучшая для этого случая индивидуальная оценка, но ненамного хуже, и существенно лучше, чем плохо работающая оценка.

3.4 Сглаженная оценка, основанная на максимизации очков, для модели бинарного выбора

Самая устойчивая оценка в модели бинарного выбора

$$y_i = \operatorname{sgn}(x_i'\beta + u_i), \quad i = 1, ..., n,$$

где

$$\operatorname{sgn}(z) = \begin{cases} 1, & \text{если } z \geq 0, \\ -1, & \text{если } z < 0, \end{cases}$$

и $\mathbb{M}[u_i] = 0$, была предложена в Manski (1975). Оценка (основанная на максимизации очков) базировалась на максимизации совпадений между y_i и $\operatorname{sgn}(x_i'\hat{b})$. Однако скорость сходимости этой оценки $O(n^{-1/3})$, и предельный процесс сложно описать.

Horowitz (1992) предложил сглаженую оценку

$$\widehat{b} = \underset{|b_1|=1}{\arg\max}\, \frac{1}{n} \sum y_i \cdot G\left(\frac{x_i'b}{\sigma_n}\right), \tag{10}$$

где G – сглаживающая функция, производная от которой совпадает с ядром K (определенным выше). Здесь, как и в уже рассмотренных оценках, оптимальный выбор ядра и ширины окна определяется степенью гладкости распределения.

Общие результаты (Kotlyarova & Zinde-Walsh, 2004) относительно асимптотического поведения отдельных индивидуальных оценок, их совместного асимптотического распределения

Таблица 1: СКО для оценок параметра

Размер выборки	Оценка	Гладкая модель	Негладкая модель
2000	Хоровиц	0,00026	0,00156
	$comb4$	0,00021	0,00127
	$comb24$	0,00020	0,00132
	$comb334$	0,00023	0,00131
4000	Хоровиц	0,00022	0,00130
	$comb4$	0,00012	0,00090
	$comb24$	0,00011	0,00091
	$comb334$	0,00013	0,00088

и поведения комбинированной оценки приложимы к этому случаю. В экспериментах Монте Карло в Kotlyarova & Zinde-Walsh (2004) получены результаты, сходные с приведенными выше; интересно, что здесь поведение комбинированной оценки стабильно и зачастую лучше, чем для индивидуальной оценки, притом что для индивидуальной оценки использовалась предложенная Хоровицем методика поправки смещения, а в комбинации использовались неисправленные оценки. Этот факт вновь иллюстрирует преимущество комбинирования для избавления оценки от смещения.

В таблице 1 приведены некоторые из результатов оценивания параметра в модели бинарного выбора с гетероскедастичной ошибкой в случае гладкого распределения и негладкого распределения; «Хоровиц» обозначает оценку, оптимальную по методике Хоровица, $comb4$, $comb24$, $comb334$ – различные комбинированные оценки; критерием является СКО.

Следует еще отметить, что комбинированные оценки стабильны еще и в том, что СКО комбинаций, основанных на разных ядрах и окнах, близки друг к другу.

4 Что известно для случая, когда функция плотности не определена?

Как видно из примеров, предположения о существовании функции плотности могут не осуществляться на практике. Возникает вопрос о поведении непараметрических оценок в этом случае, а также вопрос о возможности строить заключения о том, существует ли функция плотности (и ее производные) или нет.

4.1 Как ведут себя непараметрические оценки, когда функция плотности не определена

Если функция плотности не определена, исследователь, не зная об этом или не зная, что делать в таком случае, может строить непараметрические оценки. Таким образом, вопрос о предельном поведении таких оценок имеет смысл и для этого случая.

Результаты для ядерной оценки условного распределения для (очень специального) случая, когда распределение в окрестности точки сингулярности чисто фрактальное, исследовано в Lu (1999); автор показал, что здесь сходимость ядерной оценки быстрее, чем при условии существования плотности.

Общий подход к изучению ядерных оценок предложен в Zinde-Walsh (2008). Он основан на теории обобщенных функций. Классические книги в этой области – Гельфанд & Шилов (1958), Гельфанд & Виленкин (1961), а также главы в Соболев (1974).

Для сингулярного распределения обычная функция плотности не существует, однако всегда можно определить соответствующую обобщенную функцию как обобщенную производную обычной функции (распределения). Рассмотрим для примера одномерный случай $k = 1$. Функция распределения $F(x)$ может быть рассмотрена как функционал на пространстве

D «хороших» функций, бесконечно дифференцируемых и имеющих ограниченную область определения. Определим соответствующий функционал его действием на $\psi \in D$:

$$(F, \psi) = \int F(x)\psi(x)dx, \tag{11}$$

где правая часть, очевидно, определена и задает значение функционала. Теперь определим (обобщенную) функцию плотности (как обобщенную производную от функции распределения), тоже как линейный непрерывный функционал на D:

$$(f, \psi) = (F', \psi) = -(F, \psi') = -\int F(x)\psi'(x)dx. \tag{12}$$

Поскольку правая часть этого выражения полностью определена, то и нужный функционал, определяющий f, определен. Если обычная функция f существует, то это равенство получается интегрированием по частям. Если в распределении есть весовая точка x_0, то f пропорциональна в этой точке δ-функции Дирака:

$$(f(x), \psi(x)) = \int \delta(x - x_0)\psi(x)dx = \psi(x_0).$$

Производные (обобщенные) для функции плотности определяются аналогично:

$$(f', \psi) = -(f, \psi') = (F, \psi'') = \int F(x)\psi''(x)dx.$$

Определим также усредненную ядром K (см. Соболев, 1974, стр. 104–108) функцию $f_{hK}(x)$:

$$f_{hK}(x) = \frac{1}{h}\int f(\tilde{x})K\left(\frac{\tilde{x} - x}{h}\right)d\tilde{x} = -\frac{1}{h^2}\int F(\tilde{x})K'\left(\frac{\tilde{x} - x}{h}\right)d\tilde{x}. \tag{13}$$

Заметим что для ядерной оценки функции плотности

$$\mathbb{E}[\widehat{f(x)}] = f_{hK}(x).$$

В этих обозначениях мы видим, что если считать, что ядерная оценка оценивает обобщенную функцию плотности, то разность $f_{hK}(x) - f(x)$ представляет собой (обобщенное) смещение этой оценки. Для ядра порядка $l + 1$ и $\psi \in D$ смещение выражается как

$$(f_{hK}(x), \psi) - (f(x), \psi) = (-1)^l \int F(\tilde{x})\frac{h^l}{l!}\frac{\partial^{l+1}\psi}{\partial x^{l_i+1}}(\tilde{x})d\tilde{x}\int K(w)w^l dw + R(h) = O(h^l).$$

Таким образом, можно получить скорость сходимости обобщенного смещения. В Zinde-Walsh (2008) получено полное описание предельного распределения для оценки обобщенной функции плотности в форме обобщенного нормального процесса, для которого ковариационный функционал C для $\psi_1, \psi_2 \in D$ имеет вид

$$(C, (\psi_1, \psi_2)) = \int \psi_1(x)\psi_2(x)f(x)dx\int K(w)^2 dw = E(\psi_1(x)\psi_2(x))\int K(w)^2 dw. \tag{14}$$

В Zinde-Walsh (2008) получены также результаты для широко используемой оценки условного ожидания без предположения о существовании функции плотности. Подобный результат в частном случае фрактального распределения был также получен в Lu (1999).

Допустим, что для модели

$$y_i = m(x_i) + u_i$$

используется непараметрическая оценка Надарая–Уотсона:

$$\widehat{m(x)} = \frac{\sum\limits_{j=1}^{n} y_j K(\frac{x_j-x}{h})}{\sum\limits_{j=1}^{n} K(\frac{x_j-x}{h})} = \frac{\frac{1}{nh^k}\sum\limits_{j=1}^{n} y_j K(\frac{x_j-x}{h})}{\widehat{f(x)}}.$$

При довольно широких предположениях и без ограничений на распределение доказано, что

$$(nh^k)^{\frac{1}{2}}\widehat{f(x)}\left(\widehat{m(x)} - m(x)\right)$$

сходится к обобщенному нормальному процессу с ковариационным фунционалом

$$(C_x, (\psi_1, \psi_2)) = \mathbb{E}\left[\psi_1(x)\psi_2(x)\sigma^2(x)\right]\int K(w)^2 dw, \tag{15}$$

где $\sigma^2(x)$ – дисперсия.

Этот результат может быть использован для проверки гипотез. При дополнительных условиях (достаточно общих) получены результаты относительно сходомости и самой оценки $\widehat{m(x)}$; эти результаты показывают, что для оценивания условного ожидания сходимость ускоряется в точках сингулярности:

$$(nh^{k+2\eta})^{\frac{1}{2}}\left(\widehat{m(x)} - m(x)\right) \approx O_p(1),$$

где $\eta > 0$.

4.2 Как проверить свойства распределения: существование и гладкость плотности или сингулярность?

Как правило, проверка гипотез относительно сингулярности применяется для определения отдельных точек разрыва или негладкости функции. Müller (1992) предложил метод для определения точек разрыва в функции условного ожидания (и ее производных); этот метод базируется на использовании разности статистик, построенных с помошью несимметричных ядер. Подобный метод использован в Müller & Wang (1994) для определения точки разрыва в функции риска.

Frigyesi & Hössjer (1998) предложили общий метод для различения класса сингулярных распределений (которые, наряду с разрывами функции распределения, могут включать и более сложные, например фрактальные, сингулярности) от некоторого подкласса абсолютно гладких распределений. Их статистика основывается на функционале от ядерной оценки функции плотности; статистика ограничена в подклассе и расходится к бесконечности для сингулярных распределений. Проблема в том, что поведение этой статистики для представляющих интерес классов распределений изучено мало, ее значения могут быть очень велики даже в больших выборках (например, для χ^2), так что их трудно отличить от сингулярных.

Вообще задача проверки гипотезы о том, принадлежит ли распределение классу сингулярных или абсолютно непрерывных, неразрешима, как показано в Donoho (1988) и Lalley & Nobel (2003). Возможно только отличать друг от друга некоторые подклассы, и работа Frigyesi & Hossjer (1998) – шаг в этом направлении.

Результаты в Zinde-Walsh (2008) указывают на возможность построения более удобного функционала, чем у Frigyesi & Hössjer (1998) для построения статистики для проверки гипотез о степени гладкости. Такая статистика (Zinde-Walsh & Galbraith, 2007)

$$z = \frac{(nh)^{\frac{1}{2}}\left((\widehat{f(x)}, \psi(x)) - \mathbb{E}[\psi]\right)}{\left(\mathbb{E}[\psi^2]\int K(w)^2 dw\right)^{\frac{1}{2}}}$$

со специально подобранными функциями ψ (построенными на основе самого распределения) базируется на обобщенном смещении ядерной оценки, в то же время подобно статистике Frigyesi & Hössjer (1998) она построена так, что расходится к бесконечности для сингулярных распределений. Преимущество заключается в том, что известно асимптотическое распределение этой статистики в классе распределений с гладкой функцией плотности:

$$z \xrightarrow{d} N(0,1).$$

В подклассах распределений с сингулярностью можно указать скорость расхождения статистики, которая определяется расхождением функционала смещения обобщенной ядерной оценки.

Вычисление z требует численной оценки интегралов; знаменатель может быть определен из асимптотической формулы или через бутстрап дисперсии оценки. Поскольку асимптотическое распределение z в классе, где функция плотности гладкая, пивотально, к самой статистике можно применить параметрический бутстрап.

Предварительные результаты Монте Карло показывают, что для равномерного, нормального и сингулярного распределений, рассмотренных в Frigyesi & Hössjer (1998) (равномерные веса на 100 равноотстоящих точках в [0,1]), получаются результаты, согласующиеся с теоретическими; кроме того, эта статистика легче идентифицирует свойства распределений (такие, как сингулярность), чем статистика Frigyesi & Hössjer (1998).

Список литературы

Гельфанд, И.М. & Г.Е. Шилов (1958). *Обобщенные функции.* Том 1: Обобщенные функции и действия над ними. Том 2: Пространства основных и обобщенных функций. Москва: Физматгиз.

Гельфанд, И.М. & Н.Я. Виленкин (1961). *Обощенные функции.* Том 4: Некоторые применения гармонического анализа. Оснащенные гильбертовы пространства. Москва: Физматгиз.

Соболев, С.Л. (1974). *Введение в теорию кубатурных формул.* Москва: Наука.

Donoho, D.L. (1988). One-sided inference about functionals of a density. *Annals of Statistics* 16, 1390–1420.

Fan, Y. & A. Ullah (1999). Asymptotic normality of a combined regression estimator. *Journal of Multivariate Analysis* 71, 191–240.

Frigyesi, A. & O. Hössjer (1998). A Test for Singularity. *Statistics and Probability Letters* 40, 215–226.

Gel'fand, I.M. & G.E. Shilov (1964). *Generalized Functions*, vol. 1,2. Academic Press.

Gel'fand, I.M. & N.Y. Vilenkin (1964). *Generalized Functions*, vol. 4. Academic Press.

Green, D.A. & W.C. Riddell (1997). Qualifying for unemployment insurance: An empirical analysis. *Economic Journal* 107, 67–84.

Guggenberger, P. & Y. Sun (2006). Bias-reduced log-periodogram and Whittle estimation of the long-memory parameter vithout variance inflation. *Econometric Theory* 22, 863–912.

Horowitz, J.L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.

Juditsky A. & A. Nemirovski (2000). Functional aggregation for nonparametric regression. *Annals of Statistics* 28, 681–712.

Kotlyarova, Y. & V. Zinde-Walsh (2004). Improving the efficiency of the smoothed maximum score estimator. Working paper, McGill University.

Kotlyarova, Y. & V. Zinde-Walsh (2006). Non- and semi-parametric estimation in models with unknown smoothness. *Economics Letters* 93, 379–386.

Kotlyarova, Y. & V. Zinde-Walsh (2007). Robust kernel estimator for densities of unknown smoothness. *Journal of Nonparametric Statistics* 19, 89–101.

Lalley, S.P. & A. Nobel (2003). Indistinguishability of absolutely continuous and singular distributions. *Statistics and Probability Letters* 62, 145–154.

Li, Q. & J.S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Lu, Z.-Q. (1999). Nonparametric regression with singular design. *Journal of Multivariate Analysis* 70, 177–201.

Manski, C.F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3, 205–228.

Müller & H.-G. (1992). Change-points in nonparametric regression analysis. *Annals of Statistics* 20, 737–761.

Müller, H.-G. & J.-L. Wang (1994). Change-point models for hazard functions. In *Change-point problems*, IMS Lecture Notes 23, 224–241.

Pagan, A. & A. Ullah (1999). *Nonparametric Econometrics*. Cambridge University Press.

Parzen, E. (1962). On estimation of a probability density and mode. *Annals of Mathematical Statistics* 33, 1065–1076.

Schafgans, M. & V. Zinde-Walsh (2007). Robust Average derivative estimation. Working paper, McGill University & CIREQ.

Sobolev, S. (1992). *Cubature formulas and modern analysis*. Gordon and Breach Science Publishers.

Yang, Y. (2000). Combining different procedures for adaptive regression, *Journal of Multivariate analysis* 74, 135–161.

Yang, Y. (2000). Mixing strategies for density estimation. *Annals of Statistics* 28, 75–87.

Zinde-Walsh, V. & J.W. Galbraith (2007). When is a distribution smooth enough for nonparametric estimation? Manuscript, McGill University.

Consequences of lack of smoothness in nonparametric estimation

Victoria Zinde-Walsh

McGill University, Montreal, Canada

Nonparametric estimation is widely used in statistics and econometrics with many asymptotic results relying on smoothness of the underlying distribution, however, there are cases where such assumptions may not hold in practice. Lack of smoothness may have undesirable consequences such as an incorrect choice of window width, large estimation biases and incorrect inference. Optimal combinations of estimators based on different kernel/bandwidth can achieve automatically the best unknown rate of convergence. The combined estimator was successfully applied in density estimation, estimation of average derivatives and for smoothed maximum score in a binary choice model. In the extreme case when density does not exist the estimator “estimates” a non-existent function; nevertheless its limit process can be described in terms of generalized (in terms of generalized functions) Gaussian processes. Inference about existence of density and about its smoothness is not yet well developed; some preliminary results are discussed.