

АНАЛИЗ АСИМПТОТИЧЕСКИХ СВОЙСТВ МНОГОМЕРНОЙ НЕПАРАМЕТРИЧЕСКОЙ РЕГРЕССИИ*

Исследуются асимптотические свойства многомерной непараметрической регрессии, синтез которой основан на оценках плотности вероятности Розенблатта–Парзена. Устанавливается их количественная зависимость от вида ядерной функции и особенности исходных статистических данных.

Ключевые слова: непараметрическая регрессия, восстановление зависимостей, асимптотические свойства, ядерная функция.

Непараметрическая регрессия, основанная на оценках плотности вероятности типа Розенблатта–Парзена [1; 2], использовалась при восстановлении однозначных стохастических зависимостей по статистическим данным наблюдений их переменных [3]. Для этой регрессии был разработан ряд модификаций в условиях малого [4], большого [5] объема статистических данных и при обработке неоднородной информации [6]. Непараметрическая регрессия является основным элементом структуры гибридных моделей, решающих правил оценивания состояния статических систем и временных процессов [7]. Уже установлены условия асимптотической сходимости непараметрической регрессии и ее модификаций, однако полученные результаты не позволяют определить количественную зависимость показателей эффективности исследуемых статистик от их параметров и особенностей априорной информации. Эта проблема усугубляется тем, что непараметрическая регрессия является оценкой условного математического ожидания. Нелинейный характер ее зависимости от непараметрических оценок плотностей вероятности случайных величин создает трудности при количественном анализе ее асимптотических свойств.

Исследование аппроксимационных свойств непараметрической регрессии значительно упрощается, если априорная информация содержит сведения о виде плотности вероятности аргументов восстанавливаемой зависимости. Такие сведения могут быть получены в результате предварительной обработки исходных статистических данных либо при их формировании в процессе активного эксперимента с исследуемым объектом. Цель данной статьи состоит в анализе асимптотических свойств многомерной непараметрической регрессии в зависимости от вида ядерной функции при известной плотности вероятности аргументов восстанавливаемой зависимости.

Непараметрическая регрессия и ее асимптотические свойства. Пусть $V = (x^i, y^i, i = \overline{1, n})$ – выборка, составленная из n независимых наблюдений случайной величины $(x = (x_v, v = \overline{1, k}), y)$ с плотностью вероятности $p(x, y)$. Обозначим плотность вероятности

случайной величины x через $p(x)$, а кривую регрессии y по x – через $\phi(x)$:

$$\phi(x) = \int_{-\infty}^{\infty} y p\left(\frac{y}{x}\right) dy.$$

Будем считать, что плотность вероятности $p(x)$ известна. В качестве приближения по исходным статистическим данным кривой регрессии примем статистику

$$\bar{\phi}_1(x) = \frac{1}{n p(x) \prod_{v=1}^k c_v} \sum_{i=1}^n \prod_{v=1}^k \Phi\left(\frac{x_v - x_v^i}{c_v}\right), \quad (1)$$

где ядерные функции $\Phi(u_v)$ удовлетворяет условиям H :

$$\begin{aligned} \Phi(u_v) &= \Phi(-u_v), \quad 0 \leq \Phi(u_v) < \infty, \\ \int \Phi(u_v) du_v &= 1, \quad \int u_v^2 \Phi(u_v) du_v = 1, \\ \int u_v^m \Phi(u_v) du_v &< \infty, \quad 0 \leq m < \infty; \quad v = \overline{1, k}, \end{aligned}$$

а их коэффициенты размытости $c_v = c_v(n) \rightarrow 0$ с ростом n . Здесь и далее бесконечные пределы интегрирования опускаются.

При синтезе статистики $\bar{\phi}_1(x)$ используется непараметрическая оценка $\bar{p}(x, y)$ многомерной плотности вероятности $p(x, y)$ типа Розенблатта–Парзена [1; 2].

Для получения аналитически значимых результатов при исследовании свойств непараметрической регрессии будем считать, что интервалы изменения значений компонент $x_v, v = \overline{1, k}$, вектора x одинаковы. В этих условиях появляется возможность полагать одинаковыми значения коэффициентов размытости $c_v = c, v = \overline{1, k}$, ядерных функций. Тогда непараметрическая регрессия (1) запишется в виде

$$\bar{\phi}(x) = \frac{1}{nc^k p(x)} \sum_{i=1}^n \prod_{v=1}^k \Phi\left(\frac{x_v - x_v^i}{c}\right). \quad (2)$$

Асимптотические свойства $\bar{\phi}(x)$ определяются следующим утверждением.

* Работа выполнена при поддержке гранта ФЦП «Научные и научно-педагогические кадры инновационной России, на 2009–2013 гг.», ГК № 02.740.11.0621.

Теорема. Пусть $\varphi(x)$, $p(x, y)$, $p(x) \neq 0$ и первые две их производные по каждой компоненте x_v , $v = \overline{1, k}$, ограничены и непрерывны; ядерные функции $\Phi(u_v)$ удовлетворяют условиям H ; последовательности $c = c(n)$ коэффициентов размытости ядерных функций такие, что при $n \rightarrow \infty$ значения $c \rightarrow 0$, а $nc^k \rightarrow \infty$. Тогда непараметрическая оценка регрессии $\bar{\varphi}(x)$ обладает свойствами асимптотической несмещенности и состоятельности.

Доказательство. По определению имеем

$$\begin{aligned} M(\bar{\varphi}(x)) &= \frac{1}{nc^k p(x)} \times \\ &\times \sum_{i=1}^n \int \dots \int y^i \prod_{v=1}^k \Phi\left(\frac{x_v - x_v^i}{c}\right) p(y^i, x_1^i, \dots, x_k^i) dy^i dx_1^i \dots dx_k^i = \\ &= \frac{1}{c^k p(x)} \int \dots \int y \prod_{v=1}^k \Phi\left(\frac{x_v - t_v}{c}\right) p(y, t_1, \dots, t_k) dy dt_1 \dots dt_k = \\ &= \frac{1}{c^k p(x)} \int \dots \int \varphi(t) \prod_{v=1}^k \Phi\left(\frac{x_v - t_v}{c}\right) p(t_1, \dots, t_k) dt_1 \dots dt_k, \quad (3) \end{aligned}$$

где M – знак математического ожидания. При выполнении данных преобразований учитывается, что элементы статистической выборки V являются значениями одних и тех же случайных величин (t, y) с плотностью вероятности $p(y, t_1, \dots, t_k)$.

Проведем в выражении (3) замену переменных $(x_v - t_v)c^{-1} = u_v$ и разложим функции $\varphi(x_v - ct_v, v = \overline{1, k})$, $p(x_v - ct_v, v = \overline{1, k})$ в ряд Тейлора в точке x . Тогда с учетом свойств H ядерных функций при достаточно больших значениях n получим асимптотическое выражение смещения непараметрической регрессии

$$\begin{aligned} M(\bar{\varphi}(x) - \varphi(x)) &\sim \frac{c^2}{2p(x)} \sum_{v=1}^k (\varphi(x)p(x))^{(2)} + \\ &+ \frac{c^4}{4p(x)} \left(\sum_{v=1}^k \sum_{r=1}^k \varphi_v^{(2)}(x) p_r^{(2)}(x) + \right. \\ &\left. + \sum_{v=1}^k \varphi_v^{(2)}(x) p_v^{(2)}(x) \int u_v^4 \Phi(u_v) du_v \right) + O(c^6), \quad (4) \end{aligned}$$

где $\varphi_v^{(2)}(x)$, $p_v^{(2)}(x)$, $(\varphi(x)p(x))^{(2)}$ – вторые производные функций $\varphi(x)$, $p(x)$ и их произведения по компоненте x_v ; символом $O(c^6)$ обозначены слабые степени малости порядка c^6 . Отсюда, из условия $c \rightarrow 0$ при $n \rightarrow \infty$, следует свойство асимптотической несмещенности непараметрической регрессии $\bar{\varphi}(x)$.

Для доказательства состоятельности оценки $\bar{\varphi}(x)$ вычислим ее дисперсию

$$D(\bar{\varphi}(x)) = M(\bar{\varphi}(x) - \varphi(x))^2 - (M(\bar{\varphi}(x) - \varphi(x)))^2. \quad (5)$$

Исследуем асимптотические свойства среднеквадратического отклонения:

$$\begin{aligned} M(\bar{\varphi}(x) - \varphi(x))^2 &= \\ &= M(\bar{\varphi}^2(x)) - 2\varphi(x)M(\bar{\varphi}(x)) + \varphi^2(x). \quad (6) \end{aligned}$$

Следуя ранее использованной технологии исследований, проведем преобразования:

$$\begin{aligned} M(\bar{\varphi}^2(x)) &= \frac{1}{n^2 c^{2k} p^2(x)} \left[\sum_{i=1}^n M\left(\left(y^i\right)^2 \prod_{v=1}^k \Phi^2\left(\frac{x_v - x_v^i}{c}\right)\right) + \right. \\ &+ \sum_{i=1}^n \sum_{j=1}^n M\left(y^i \prod_{v=1}^k \Phi\left(\frac{x_v - x_v^i}{c}\right) y^j \prod_{v=1}^k \Phi\left(\frac{x_v - x_v^j}{c}\right)\right) \left. \right] = \\ &= \frac{1}{n^2 c^{2k} p^2(x)} \left[n \int \dots \int \varphi^2(t_1, \dots, t_k) \times \right. \\ &\times \prod_{v=1}^k \Phi^2\left(\frac{x_v - t_v}{c}\right) p(t_1, \dots, t_k) dt_1 \dots dt_k + \\ &+ n(n-1) \left(\int \dots \int \varphi(t_1, \dots, t_k) \times \right. \\ &\times \prod_{v=1}^k \Phi\left(\frac{x_v - t_v}{c}\right) p(t_1, \dots, t_k) dt_1 \dots dt_k \left. \right)^2 \left. \right]. \end{aligned}$$

Пренебрегая величинами малости $O\left(\frac{1}{n}\right)$,

$O\left(\frac{1}{nc^{k-2}}\right)$, $O(c^6)$, найдем значение асимптотического выражения

$$\begin{aligned} M(\bar{\varphi}^2(x)) &\sim \varphi^2(x) + \frac{1}{nc^k p(x)} \varphi^2(x) \times \\ &\times \prod_{v=1}^k \int \Phi^2(u_v) du_v + \frac{c^4}{4p^2(x)} \left(\sum_{v=1}^k (\varphi(x)p(x))^{(2)} \right)^2 + \\ &+ c^2 \frac{\varphi(x)}{p(x)} \sum_{v=1}^k (\varphi(x)p(x))^{(2)} + c^4 \frac{\varphi(x)}{2p(x)} \times \\ &\times \left(\sum_{v=1}^k \sum_{r=1}^k \varphi_v^{(2)}(x) p_r^{(2)}(x) + \sum_{v=1}^k \varphi_v^{(2)}(x) p_v^{(2)}(x) \int u_v^4 \Phi(u_v) du_v \right). \quad (7) \end{aligned}$$

Подставим выражения (4) и (7) в (6) и при достаточно больших n получим

$$\begin{aligned} M(\bar{\varphi}(x) - \varphi(x))^2 &\sim \frac{\varphi^2(x)}{nc^k p(x)} \prod_{v=1}^k \int \Phi^2(u_v) du_v + \\ &+ \frac{c^4}{4p^2(x)} \left(\sum_{v=1}^k (\varphi(x)p(x))^{(2)} \right)^2. \quad (8) \end{aligned}$$

Отсюда, с учетом соотношений (4) и (8), из условия $c \rightarrow 0$, $nc^k \rightarrow \infty$ при $n \rightarrow \infty$ следует свойство состоятельности непараметрической регрессии $\bar{\varphi}(x)$.

При $k = 1$ результат (8) совпадает с утверждением работы [8], что подтверждает корректность выполненных преобразований.

Анализ аппроксимационных свойств статистики $\bar{\varphi}(x)$. На основании полученных аналитических результатов установим количественную зависимость аппроксимационных свойств $\bar{\varphi}(x)$ от вида ядерной функции и особенностей исходных статистических данных.

Определим минимальное значение W_2 выражения

$$\frac{\prod_{v=1}^k \int \Phi^2(u_v) du_v}{nc^k} \int \dots \int \varphi^2(x) p^{-1}(x) dx_1 \dots dx_k + \frac{c^4}{4} \int \dots \int \left(p^{-1}(x) \sum_{v=1}^k (\varphi(x) p(x))^{(2)} \right)^2 dx_1 \dots dx_k, \quad (9)$$

которое получено путем интегрирования результата (8).

Из условия минимума (9) по коэффициенту размытости c нетрудно получить его оптимальное значение

$$c^* = \left(\frac{k A \prod_{v=1}^k \int \Phi^2(u_v) du_v}{nB} \right)^{1/(k+4)},$$

где

$$A = \int \dots \int \varphi^2(x) p^{-1}(x) dx_1 \dots dx_k;$$

$$B = \int \dots \int \left(p^{-1}(x) \sum_{v=1}^k (\varphi(x) p(x))^{(2)} \right)^2 dx_1 \dots dx_k.$$

Тогда, подставляя c^* в выражение (9), получим

$$W_2 = \left[\left(\frac{A \prod_{v=1}^k \int \Phi^2(u_v) du_v}{n} \right)^4 B^k \right]^{1/(k+4)} \frac{4+k}{4k^{k/(k+4)}}. \quad (10)$$

Дополнительно уменьшить значение W_2 можно за счет минимизации выражения $\int \Phi^2(u_v) du_v$ по виду ядерной функции $\Phi(u_v)$ с учетом выполнения условий H .

Подобная задача решена в работе [2], в которой получена оптимальная форма ядерной функции

$$\Phi_0(u_v) = \begin{cases} \frac{3}{4\sqrt{5}} - \frac{3u_v^2}{20\sqrt{5}} & \forall |u_v| \leq \sqrt{5}, \\ 0 & \forall |u_v| > \sqrt{5}. \end{cases}$$

Для исследования влияния вида ядерных функций на аппроксимационные свойства многомерной непараметрической регрессии определим отношение

$$R_2 = \frac{W_2(\Phi(u))}{W_2(\Phi_0(u))},$$

где $W_2(\Phi_0(u))$, $W_2(\Phi(u))$ – значения критерия W_2 (10) при ядерных функциях $\Phi_0(u)$, $\Phi(u)$. При этом по каждой компоненте u_v , $v = \overline{1, k}$, использовались следующие ядерные функции:

$$\Phi(u_v) = \begin{cases} \frac{1}{\sqrt{6}} - \frac{|u_v|}{6} & \forall |u_v| \leq \sqrt{6} \\ 0 & \forall |u_v| > \sqrt{6}, \end{cases} \quad (11a)$$

$$\Phi(u_v) = (2\pi)^{-1/2} e^{-u_v^2/2}, \quad (11б)$$

$$\Phi(u_v) = \begin{cases} \frac{1}{2\sqrt{3}} & \forall |u_v| \leq \sqrt{3} \\ 0 & \forall |u_v| > \sqrt{3}, \end{cases} \quad (11в)$$

$$\Phi(u_v) = \frac{1}{2} e^{-\sqrt{2}|u_v|}. \quad (11г)$$

Нетрудно убедиться, что в данных условиях при $\Phi(u_v) = \Phi(u)$ и $\Phi_0(u_v) = \Phi_0(u)$, $v = \overline{1, k}$, отношение

$$R_2 = \alpha^{4k/(k+4)},$$

где

$$\alpha = \frac{\int \Phi^2(u) du}{\int \Phi_0^2(u) du}.$$

Значения α приведены в работе [2].

Результаты вычислительных экспериментов при исследовании зависимости отношения R_2 от количества k аргументов восстанавливаемой зависимости и вида ядерных функций представлены в таблице.

Зависимость отношения R_2 от количества k аргументов восстанавливаемой зависимости и вида ядерных функций

$\Phi(u)$	$k = 1$	$k = 3$	$k = 5$	$k = 7$	$k = 9$
(11a)	1,012	1,026	1,034	1,039	1,042
(11б)	1,041	1,089	1,117	1,135	1,148
(11в)	1,061	1,136	1,179	1,208	1,228
(11г)	1,249	1,61	1,853	2,027	2,157

С ростом размерности k вектора аргументов восстанавливаемой зависимости возрастает влияние выбора вида ядерной функции на аппроксимационные свойства многомерной непараметрической регрессии, что проявляется в увеличении значений отношения R_2 . Такая тенденция особо характерна для ядерных функций $\Phi(u)$ и $\Phi_0(u)$, которым свойственны

большие значения α . Например, отношение R_2 для ядерной функции (11a) при $k = 9$ равно 1,042, а для (11г) – 2,157. При этом соответствующие им значения α равны 1,015 (11a) и 1,32 (11г).

На основе анализа асимптотических свойств многомерной непараметрической регрессии определена количественная зависимость ее аппроксимационных свойств от вида ядерной функции и особенностей статистических данных. В условиях их достаточно большого объема вид ядерной функции оказывает значимое влияние на среднеквадратическое отклонение непараметрической регрессии от условного математического ожидания, которое возрастает с увеличением размерности аргументов восстанавливаемой зависимости. Данная закономерность особо проявляется с увеличением степени отличия ядерной функции от оптимального ядра Епанечникова.

Библиографические ссылки

1. Parzen E. On Estimation of a Probability Density Function and Mode // Annu. Math. Statistic. 1962. Vol. 33. P. 1065–1076.
2. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и ее применения. 1969. Т. 14, вып. 1. С. 156–161.
3. Надабая Э. А. Замечания о непараметрических оценках плотности вероятности и кривой регрессии // Теория вероятностей и ее применения. 1970. Т. 15, вып. 1. С. 139–142.
4. Лапко В. А., Бадмаев Р. В. Синтез и анализ нелинейных непараметрических коллективов решающих правил в задачах восстановления стохастических зависимостей, основанных на последовательных процедурах формирования решений // Вестник КрасГАУ. 2006. № 10. С. 53–62.
5. Лапко А. В., Лапко В. А., Варочкин С. С. Коллектив непараметрических регрессий, основанных на принципе декомпозиции обучающей выборки // Вестник СибГАУ. 2009. Вып. 1 (22). Ч. 2. С. 38–40.
6. Лапко А. В., Лапко В. А., Суханов А. В. Непараметрические восстановления стохастических зависимостей в условиях пропуска данных // Проблемы автоматизации и управления : науч.-техн. журн. Нац. акад. наук Кыргыз. Республики. 2005. С. 74–81.
7. Лапко В. А. Непараметрические коллективы решающих правил. Новосибирск : Наука, 2002.
8. Обучающиеся системы обработки информации и принятия решений / А. В. Лапко, С. В. Ченцов, С. И. Крохов, Л. А. Фельдман. Новосибирск : Наука, 1996.

A. V. Lapko, V. A. Lapko

THE ANALYSIS OF ASYMPTOTIC PROPERTIES OF MULTIDIMENSIONAL NONPARAMETRIC REGRESSION

Asymptotic properties of multidimensional nonparametric regression, synthesis of which is grounded on estimations of a probability density of Rosenblatt-Parzen, are explored. Their quantitative dependence on flavor of kernel function and features of initial statistical data is determined.

Keywords: nonparametric regression, restoring of dependences, asymptotic properties, kernel function.

© Лапко А. В., Лапко В. А., 2012

УДК 631.331

Г. Н. Лимаренко, М. В. Шевчугов, А. Н. Щепин

К РАСЧЕТУ ДИНАМИКИ УСТАНОВИВШЕГОСЯ ДВИЖЕНИЯ ВЫХОДНОГО ЗВЕНА В ВОЛНОВОЙ РЕЕЧНОЙ ПЕРЕДАЧЕ

Приведены результаты теоретических исследований и методов расчета в поступательном приводе с волновой реечной передачей неравномерности установившегося движения его выходного звена – роликовой рейки, взаимодействующей с многокулачковым механизмом.

Ключевые слова: толкатель, диаграмма, фазовый сдвиг, приведенный момент инерции, динамическая ошибка скорости.

Волновая реечная передача (ВРП) с роликовой рейкой [1] основана на взаимодействии многокулачкового вала с толкателями, контактирующими с выходным звеном – рейкой на линейных участках диаграмм их относительного движения. Количество толкателей в механизме преобразования движения выбирается

таким, чтобы в зацеплении с роликами рейки одновременно находилось не менее двух толкателей. Это теоретически обеспечивает линейность функции положения выходного звена, малые изменения характеристики приведенной жесткости и равномерное распределение движущих сил по промежуточным звеньям.