РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

АВТОМЕТРИЯ

2009, том 45, № 1

УДК 519.234

НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ПЛОТНОСТИ ВЕРОЯТНОСТИ И ЕЕ ПРОИЗВОДНЫХ. *C*-ПОДХОД

В. Г. Алексеев

Институт физики атмосферы им. А. М. Обухова РАН,
119017, Москва, Пыжевский пер., 3
E-mail: aleks.v.g@mail.ru

В рамках *C*-подхода рассмотрены непараметрические (ядерные) оценки плотности вероятности $f(x)$ по выборке конечного объема. Вводится в рассмотрение параметр гладкости β оцениваемой плотности вероятности. В случае $\beta > 2$ указана возможность улучшения сходимости оценки плотности вероятности $f_n(x)$ к функции $f(x)$ за счет использования знакопеременных весовых функций (весовых функций высших порядков). С аналогичных позиций очень коротко рассмотрены оценки производных плотности вероятности.

Ключевые слова: плотность распределения вероятностей, непараметрическая (ядерная) оценка, равномерная сходимость оценки с ростом объема выборки.

Введение. Предлагаемая работа является продолжением [1], в которой рассматривались непараметрические оценки плотности вероятности и ее производных до третьего порядка включительно. Основное их отличие состоит том, что оценки плотности вероятности представляются здесь в рамках уже не L_2-подхода, а *C*-подхода. Напомним, что сходимость последовательности функций $f_n(x)$ к функции $f(x)$ в рамках *C*-подхода означает ее равномерную сходимость в пределах заданного отрезка $[a, b]$ [2–7]. В дальнейшем оцениваемую плотность вероятности будем считать заданной на всей бесконечной прямой, поэтому отрезок $[a, b]$ преобразуется здесь в множество $\mathsf{R} = (-\infty, \infty)$.

Еще одно (менее существенное) отличие состоит в том, что для оцениваемой плотности вероятности используется другое обозначение. Дело в том, что плотность вероятности случайной величины X является производной от функции распределения, которая практически всегда в математической статистике обозначается символом $F(x)$. В работе [1] функция распределения $F(x)$ никоим образом не использовалась, поэтому обозначение $p(x)$ для оцениваемой плотности вероятности не могло привести к недоразумению. Поскольку в данной работе функция распределения $F(x)$ используется в последующих выкладках, обозначение $p(x)$ для плотности вероятности становится неудобным. Плотность вероятности – производную от функции распределения $F(x)$ – будем обозначать далее $f(x)$.

Что же касается возможностей практического применения предлагаемых методов статистического оценивания, то достаточно сослаться на обзор [8] и монографии [9–13].

Основные результаты. Теоремы о сходимости оценок плотности вероятности и ее производной. Итак, пусть

$$X_1, X_2, \dots, X_n \tag{1}$$

– выборка из n независимых наблюдений случайной величины X с функцией распределения

$$F(x) = \int_{-\infty}^{x} f(y)\,dy.$$

Через $F_n(x)$ обозначим эмпирическую функцию распределения, построенную на основании выборки. Относительно плотности вероятности $f(x)$ будем предполагать, что она дифференцируема r раз ($r = 0, 1, \dots$) и ее r-я производная является функцией из класса $\mathrm{Lip}\,\alpha$, где $0 < \alpha \le 1$. Последнее означает, что для любых точек x' и x'' на прямой x выполнено неравенство

$$\left| f^{(r)}(x') - f^{(r)}(x'') \right| \le M \left| x' - x'' \right|^{\alpha}, \tag{2}$$

где M – некоторая постоянная. Величину $\beta = r + \alpha$ будем называть параметром гладкости плотности вероятности $f(x)$. Условимся также, что символ «~» обозначает далее пропорциональность двух величин, а угловые скобки $\langle \cdot \rangle$ служат символом математического ожидания.

Оценку плотности вероятности $f(x)$ по выборке (1) будем искать в виде

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n} \right),$$

где h_n – некоторая последовательность положительных чисел такая, что

$$\lim_{n \to \infty} \left(h_n + \frac{\ln \ln n}{n h_n^2} \right) = 0,$$

а $K(x)$, $-\infty < x < \infty$, – некоторая четная непрерывная функция, удовлетворяющая условиям $K(x) = 0$, если $|x| \ge 1$, и

$$G_0(K) = \int_0^1 K(x)\,dx = \frac{1}{2}. \tag{3}$$

Кроме того, предположим, что функция $K(x)$ имеет непрерывную производную на отрезке $[-1, 1]$.

Впоследствии будем накладывать на функцию $K(x)$ дополнительные условия, выражаемые через функционалы вида

$$G_\nu(K) = \int_0^1 x^\nu K(x)\,dx, \quad \nu \ge 0.$$

Теорема 1. Пусть задан параметр гладкости $\beta > 0$. Выберем последовательность h_n так, чтобы она удовлетворяла условию

$$h_n \sim \left(\frac{\ln\ln n}{n}\right)^{1/[2(1+\beta)]}. \tag{4}$$

От функции $K(x)$ потребуем, чтобы она удовлетворяла условию (3) и в случае $2l < \beta \le 2(l+1)$, где l – целое положительное число, дополнительному условию

$$G_2(K) = G_4(K) = \ldots = G_{2l}(K) = 0. \tag{5}$$

Тогда с вероятностью 1

$$\sup_x \left| f_n(x) - f(x) \right| = O\left(\left(\frac{\ln\ln n}{n}\right)^{\beta/[2(1+\beta)]}\right) \quad \text{при } n \to \infty. \tag{6}$$

Д о к а з а т е л ь с т в о. Легко видеть, что

$$\sup_x \left| f_n(x) - f(x) \right| \le \varepsilon_n + \delta_n, \tag{7}$$

где

$$\varepsilon_n = \sup_x \left| \langle f_n(x) \rangle - f(x) \right|; \qquad \delta_n = \sup_x \left| f_n(x) - \langle f_n(x) \rangle \right|. \tag{8}$$

Начнем с рассмотрения величины ε_n. Представим отдельно два случая: $r \ge 1$ и $r = 0$.

Если $r \ge 1$, то, пользуясь разложением функции $f(x)$ по формуле Тейлора с остаточным членом в форме Лагранжа, находим

$$\langle f_n(x) \rangle - f(x) = h_n^{-1} \int_{-\infty}^{\infty} K\left(\frac{y-x}{h_n}\right) [f(y) - f(x)]\,dy =$$

$$= h_n^{-1} \int_{-\infty}^{\infty} K\left(\frac{y-x}{h_n}\right) \frac{f^{(r)}(\xi)}{r!} (y-x)^r\,dy. \tag{9}$$

Здесь точка $\xi = \xi(y)$ лежит где-то внутри интервала, ограниченного точками x и y. Так как (в силу четности функции $K(x)$ и предположения (5))

$$\int_{-\infty}^{\infty} K\left(\frac{y-x}{h_n}\right) (y-x)^r\,dy = 0,$$

то соотношение (9) можно переписать в виде

$$\langle f_n(x)\rangle - f(x) = h_n^{-1} \int_{-\infty}^{\infty} K\left(\frac{y-x}{h_n}\right) \frac{f^{(r)}(\xi) - f^{(r)}(x)}{r!} (y-x)^r dy.$$

Отсюда в соответствии с (2) и (8) следует, что

$$\varepsilon_n \le Mh_n^{-1} \int_{-\infty}^{\infty} \left| K\left(\frac{u}{h_n}\right)\right| \frac{|u|^\beta}{r!} du = \frac{2Mh_n^\beta}{r!} G_\beta(|K|). \tag{10}$$

Если же $r = 0,$ то в соответствии с предположением (2)

$$\left|\langle f_n(x)\rangle - f(x)\right| = \left| h_n^{-1} \int_{-\infty}^{\infty} K\left(\frac{y-x}{h_n}\right)[f(y) - f(x)]dy \right| \le$$

$$\le Mh_n^{-1} \int_{-\infty}^{\infty} \left| K\left(\frac{u}{h_n}\right)\right| |u|^\beta du$$

и, следовательно,

$$\varepsilon_n \le 2Mh_n^\beta G_\beta(|K|).$$

Если учесть, что по принятому в алгебре определению $0! = 1,$ то отсюда следует, что формула (10) верна не только при $r \ge 1,$ но и при $r = 0.$

Что же касается величины δ_n, то, как показано в работе [14],

$$\delta_n \le \mu h_n^{-1} \sup_x \left| F_n(x) - F(x)\right|, \tag{11}$$

где

$$\mu = \int_{-1}^{1} |K'(x)| dx.$$

Рассмотрим теперь множитель $\sup_x \left| F_n(x) - F(x)\right|$ в правой части неравенства (11). Согласно результатам работ [15–17] с вероятностью 1

$$\sup_x \left| F_n(x) - F(x)\right| = O\left(\sqrt{\frac{\ln\ln n}{n}}\right), \tag{12}$$

какой бы ни была функция распределения $F(x)$.

Из соотношений (11) и (12) очевидным образом следует, что с вероятностью 1

$$\delta_n = O\left(h_n^{-1}\sqrt{\frac{\ln\ln n}{n}}\right). \tag{13}$$

Сопоставляя теперь соотношения (4), (7), (10) и (13), приходим к утверждению теоремы.

В качестве небольшого комментария к теореме 1 укажем на то обстоятельство, что ядра (весовые функции) $K(x)$, удовлетворяющие условиям (5) при $\beta > 2$, знакопеременны. Их использование диктуется стремлением уменьшить смещение оценки $f_n(x)$ и в конечном счете улучшить скорость ее сходимости к функции $f(x)$ в метрике пространства непрерывных функций C. Нетрудно показать, что, ограничиваясь применением лишь неотрицательных весовых функций $K(x)$, при всех $\beta \geq 2$ невозможно получить ничего лучшего, чем

$$\sup_x \left| f_n(x) - f(x) \right| = O\left(\left(\frac{\ln\ln n}{n} \right)^{1/3} \right). \tag{14}$$

Легко видеть, что скорость сходимости (6) оказывается более высокой, чем (14), если $\beta > 2$.

Равномерно сходящиеся оценки производных плотности вероятности можно построить с помощью аналогичных приемов. Коротко рассмотрим оценку первой производной функции $f(x)$. Предположим, что $\beta = r + \alpha > 1$. Оценку величины $f'(x)$ по выборке (1) будем искать в виде

$$Df_n(x) = (nh_n^2)^{-1} \sum_{i=1}^{n} N\left(\frac{X_i - x}{h_n} \right),$$

где h_n – некоторая последовательность положительных чисел такая, что

$$\lim_{n \to \infty} \left(h_n + \frac{\ln\ln n}{nh_n^4} \right) = 0,$$

а $N(x)$, $-\infty < x < \infty$, – некоторая нечетная непрерывная функция, удовлетворяющая условиям $N(x) = 0$, если $|x| \geq 1$, и

$$G_1(N) = \int_0^1 x N(x)\, dx = \frac{1}{2}. \tag{15}$$

Кроме того, будем предполагать, что функция $N(x)$ имеет непрерывную производную на отрезке $[-1, 1]$.

Теорема 2. Пусть задан параметр гладкости $\beta > 1$. Выберем последовательность h_n так, чтобы она удовлетворяла условию (4).

От функции $N(x)$ потребуем, чтобы она удовлетворяла условию (15) и в случае $2l - 1 < \beta \leq 2l + 1$, где l – целое число, большее 1, дополнительному условию

$$G_3(N) = G_5(N) = \ldots = G_{2l-1}(N) = 0.$$

Тогда с вероятностью 1

$$\sup_x \left| Df_n(x) - f'(x) \right| = O\left(\left(\frac{\ln\ln n}{n} \right)^{(\beta - 1)/[2(1+\beta)]} \right) \quad \text{при } n \to \infty.$$

Доказательство теоремы опускаем, так как оно реализуется по той же схеме, что и доказательство теоремы 1.

В случае оценивания второй производной плотности вероятности $f(x)$ выбор ядра (весовой функции) снова будет зависеть от предполагаемого значения параметра гладкости β. Если, в частности, $\beta > 4$, то всегда можно достичь улучшения равномерной сходимости оценки функции $f''(x)$ за счет использования весовых функций высших порядков (ср. теоремы 1 и 2).

Кроме того, укажем на работы [18–23], в которых представлены наборы весовых функций, позволяющие строить непараметрические оценки самой плотности вероятности и ее производных. Следует, однако, иметь в виду, что во всех случаях применение весовых функций высших порядков становится целесообразным лишь при достаточно больших (хотя и не астрономических) значениях объема выборки n. Если же объем выборки невелик (например, не превосходит нескольких десятков), то от применения весовых функций высших порядков лучше воздержаться. Аналогичные рекомендации были сформулированы в работах [24, 25], посвященных спектральному и биспектральному оцениванию.

Заключение. Критерии качества статистических оценок плотности вероятности, лежащие в основе L_2-подхода и C-подхода, разумеется, не являются единственно возможными. Широко известен, например, L_1-подход [26], в рамках которого ошибка оценивания плотности вероятности $f(x)$ описывается величиной

$$\left\langle \int_{-\infty}^{\infty} \left| f_n(x) - f(x) \right| dx \right\rangle.$$

Известны также «взвешенные» L_1- и L_2-подходы, в рамках которых ошибка оценивания плотности вероятности описывается соответственно величинами

$$\left\langle \int_{-\infty}^{\infty} \left| f_n(x) - f(x) \right| f(x)\, dx \right\rangle, \quad \left\langle \int_{-\infty}^{\infty} [f_n(x) - f(x)]^2 f(x)\, dx \right\rangle.$$

Краткое обсуждение вопросов, касающихся двух последних подходов, представлено в [27].

Наконец, в основе подхода, избранного авторами работы [28], лежит расстояние Хеллингера

$$H(f_n, f) = \int_{-\infty}^{\infty} \left[\sqrt{f_n(x)} - \sqrt{f(x)} \right]^2 dx.$$

Выбор того или иного критерия качества оценки плотности вероятности остается за читателем. Выбор критерия качества оценки, в свою очередь, однозначно определяет выбор подхода к ее построению. Например, если избранным критерием качества оценки является расстояние Хеллингера, то нужно исключить возможность появления отрицательной оценки плотности вероятности.

СПИСОК ЛИТЕРАТУРЫ

1. **Алексеев В. Г.** Непараметрическое оценивание плотности вероятности и ее производных. L_2-подход // Автометрия. 2007. **43**, № 6. С. 39–47.
2. **Parzen E.** On estimation of a probability density function and mode // Ann. Math. Statist. 1962. **33**, N 3. P. 1065.
3. **Woodroofe M.** On the maximum deviation of the sample density // Ann. Math. Statist. 1967. **38**, N 2. P. 475.
4. **Wegman E. J.** Nonparametric probability density estimation: I. A summary of available methods // Technometrics. 1972. **14**, N 3. P. 533.
5. **Silverman B. W.** Weak and strong uniform consistency of the kernel estimate of a density and its derivatives // Ann. Statist. 1978. **6**, N 1. P. 177.
6. **Хасьминский Р. З.** О границе снизу рисков непараметрических оценок плотности в равномерной метрике // Теория вероятностей и ее применения. 1978. **23**, № 4. С. 824.
7. **Хашимов Ш. А.**, **Убайдуллаев К. Х.** О сильной состоятельности сплайн-оценки плотности распределения // Узбекский матем. журнал. 2000. № 1. С. 60.
8. **Шапиро Е. И.** Непараметрические оценки плотности вероятности в задачах обработки результатов наблюдений // Зарубеж. радиоэлектроника. 1976. № 2. С. 3.
9. **Фукунага К.** Введение в статистическую теорию распознавания образов. М.: Наука, 1979.
10. **Лапко А. В.**, **Ченцов С. В.**, **Крохов С. И.**, **Фельдман Л. А.** Обучающиеся системы обработки информации и принятия решений. Непараметрический подход. Новосибирск: Наука, 1996.
11. **Лапко А. В.**, **Лапко В. А.**, **Соколов М. И.**, **Ченцов С. В.** Непараметрические системы классификации. Новосибирск: Наука, 2000.
12. **Лапко А. В.**, **Лапко В. А.** Непараметрические системы обработки неоднородной информации. Новосибирск: Наука, 2007.
13. **Абсава Р. М.**, **Надарая Э. А.** Некоторые задачи теории непараметрического оценивания функциональных характеристик закона распределения наблюдений. Тбилиси: Изд-во Тбилисского ун-та, 2005.
14. **Schuster E. F.** Estimation of a probability density function and its derivatives // Ann. Math. Statist. 1969. **40**, N 4. P. 1187.
15. **Смирнов Н. В.** Приближение законов распределения случайных величин по эмпирическим данным // Успехи матем. наук. 1944. Вып. 10. С. 179.
16. **Kiefer J.** On large deviations of the empiric D. F. of vector chance variables and a law of the iterated logarithm // Pacif. Journ. Math. 1961. **11**, N 2. P. 649.
17. **Singh R. S.** Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density // Ann. Statist. 1977. **5**, N 2. P. 394.
18. **Алексеев В. Г.** Некоторые практические рекомендации по спектральному анализу гауссовских стационарных случайных процессов // Проблемы передачи информации. 1973. **9**, № 4. С. 42.
19. **Алексеев В. Г.** О вычислении спектров стационарных случайных процессов по выборкам большого объема // Проблемы передачи информации. 1980. **16**, № 1. С. 42.
20. **Алексеев В. Г.** Об оценках производных плотности вероятности // Вычислительная и прикладная математика. Киев: Вища шк., 1981. Вып. 43. С. 139.

21. **Алексеев В. Г.** О вычислении спектральных плотностей случайных процессов по выборкам большого объема // Там же. Вып. 44. С. 32.
22. **Gasser Th.**, **Müller H.-G.** Kernel estimation of regression functions // Lecture Notes in Mathematics. Berlin – Heidelberg – New York: Springer-Verlag, 1979. N 757. P. 23.
23. **Müller H.-G.** Smooth optimum kernel estimators of densities, regression curves and modes // Ann. Statist. 1984. **12**, N 2. P. 766.
24. **Алексеев В. Г.** О непараметрических методах прикладного спектрального анализа // Автометрия. 2007. **43**, № 1. С. 56–64.
25. **Алексеев В. Г.** О непараметрических методах прикладного биспектрального анализа // Автометрия. 2006. **42**, № 1. С. 13–22.
26. **Деврой Л.**, **Дьёрфи Л.** Непараметрическое оценивание плотности. L_1-подход. М.: Мир, 1988.
27. **Rosenblatt M.** Global measures of deviation for kernel and nearest neighbor density estimates // Lecture Notes in Mathematics. Berlin – Heidelberg – New York: Springer-Verlag, 1979. N 757. P. 181.
28. **Eggermont P. P. B.**, **LaRiccia V. N.** Optimal convergence rates for Good's nonparametric maximum likelihood density estimator // Ann. Statist. 1999. **27**, N 5. P. 1600.

Поступила в редакцию 27 февраля 2008 г.