V. M. Buldakov, G. M. Koshkin, On Recursive Estimates of Probability Density and Regression Line,
Probl. Peredachi Inf., 1977, Volume 13, Issue 1, 58–66

ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Том XIII 1977 Вып. 1

УДК 621.391.1:519.25

О РЕКУРРЕНТНЫХ ОЦЕНКАХ ПЛОТНОСТИ ВЕРОЯТНОСТИ И ЛИНИИ РЕГРЕССИИ

В. М. Булдаков, Г. М. Кошкин

В статье построены рекуррентные оценки плотности вероятности и линии регрессии на основе непараметрических оценок плотности вероятности типа Розенблатта – Парзена [4, 5]; исследованы их асимптотические (при $n\to\infty$) свойства, в частности доказана асимптотическая нормальность этих оценок; приведены результаты статистического моделирования обычных и рекуррентных оценок линии регрессии.

1. В последние годы широкое распространение получили последовательные методы решения различных задач математической статистики [1–3], что в значительной степени связано с внедрением как в производство, так и в научные исследования цифровых вычислительных машин (ЦВМ). Последовательные процедуры обладают следующими преимуществами перед обычными: они, как правило, проще в реализации на ЦВМ, экономят машинную память, и, что весьма немаловажно, работу процедуры можно прервать в любой момент и при этом получить готовый результат. В данной работе построены рекуррентные оценки плотности вероятности парзеновского типа (см. [4–7]), а на их основе – оценки условной плотности вероятности и линии регрессии; рассмотрены вопросы оптимизации и асимптотического поведения этих оценок. Статистическое моделирование на ЦВМ позволило провести наглядное сравнение обычных и рекуррентных оценок линии регрессии.

2. Пусть $X_i=(x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)})$, $i=1, 2, \dots, n$ – выборка k-мерных независимых векторов, полученная из генеральной совокупности, характеризуемой непрерывной плотностью вероятности $f(X)=f(x_1, x_2, \dots, x_k)$. В качестве непараметрической оценки многомерной плотности вероятности возьмем согласно [6, 7]

$$f_n(X)=\frac{1}{nh_n^k}\sum_{j=1}^{n}\prod_{l=1}^{k}K\left(\frac{x_l-x_l^{(j)}}{h}\right), \tag{1}$$

где ядро $K(u)$ удовлетворяет условиям:

$$\begin{aligned}(2)\qquad &1)\ 0\leqslant K(u)<\infty,\\ &2)\ K(u)=K(-u),\\ &3)\ \int_{-\infty}^{\infty}K(u)\,du=1,\\ &4)\ \int_{-\infty}K^2(u)\,du=L<\infty,\end{aligned}$$

5) $K(u_1) \leqslant K(u_2)$ для $|u_1| \leqslant |u_2|$,

6) $\int\limits_{-\infty}^{\infty} u^2 K(u)\, du = 1;$

а $\{h_n\}$ — последовательность положительных чисел, причем

$$nh_n^k \to \infty, \quad h_n \to 0 \quad \text{при } n \to \infty. \tag{3}$$

На основании оценки (1) определим ее рекуррентный аналог в точке $X = (x_1, x_2, \ldots, x_k)$ следующим образом:

$$f_{[n]}(X) = f_{[n-1]}(X) - \frac{1}{n}\left[f_{[n-1]}(X) - \frac{1}{h^k_{[n]}} \prod_{l=1}^{k} K\left(\frac{x_l - x_l^{(n)}}{h_{[n]}}\right)\right]. \tag{4}$$

Отметим, что процедура (4) для $k=1$, по-видимому, впервые была предложена в работе [8].

Преобразуем выражение (4) к удобному для исследований виду: умножим обе части (4) на n, заменим n на j и просуммируем по j от 1 до n. Тогда

$$f_{[n]}(X) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h^k_{[j]}} \prod_{l=1}^{k} K\left(\frac{x_l - x_l^{(j)}}{h_{[j]}}\right), \tag{5}$$

и теперь, исходя из представления (5), будем исследовать свойства рекуррентной оценки (4).

Ниже приводятся понятия, которые понадобятся в дальнейшем.

Определение 1. Будем говорить, что некоторая ограниченная непрерывная функция $p(X)$ удовлетворяет условию R [9], если для любого вектора $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$

$$p(X+\alpha) = p(X) + \sum_{l=1}^{k} \alpha_l \frac{\partial p(X)}{\partial x_l} + \frac{1}{2}\left[\sum_{l,m=1}^{k} \alpha_l \alpha_m \frac{\partial^2 p(X)}{\partial x_l\, \partial x_m}\right] + q(X, \alpha)\left[\sum_{l=1}^{k} \alpha_l^2\right]^{1+a/2}, \tag{6}$$

где $(X+\alpha) = (x_1+\alpha_1, x_2+\alpha_2, \ldots, x_k+\alpha_k)$; $\frac{\partial p(X)}{\partial x_l}$, $\frac{\partial^2 p(X)}{\partial x_l\, \partial x_m}$, $q(\cdot)$ — ограниченные функции, $0 < a < 1$.

Определение 2. Интегральным среднеквадратическим отклонением (и.с.к.о.) оценки плотности $f_n(X)$ от ее истинного значения назовем функционал вида

$$u^2(f_n) = \int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} \mathbf{M}[f_n(X) - f(X)]^2\, dx_1\, dx_2 \ldots dx_k.$$

Определение 3. Под асимптотической относительной эффективностью (а.о.э.) рекуррентной оценки (4) по отношению к оценке (1) будем понимать предел отношения их и.с.к.о., когда объем выборки неограниченно увеличивается, т. е.

$$e(f_{[\cdot]}, f_{\cdot}) = \lim_{n\to\infty} \frac{u^2(f_{[n]})}{u^2(f_n)}. \tag{7}$$

Теорема 1. *Пусть $K(u)$ удовлетворяет условиям* (2); $h_{[n]}=An^{-\alpha}$, $0<k\alpha<1$, *A — некоторая положительная константа* *; *истинная плотность вероятности $f(X)$ удовлетворяет условию R. Тогда при $n\to\infty$*

1) *среднеквадратическое отклонение*

$$\mathbf{M}[f_{[n]}(X)-f(X)]^2\sim$$

$$\sim\frac{f(X)L^k}{(1+k\alpha)nh^k_{[n]}}+\frac{h^4_{[n]}}{4(1-2\alpha)^2}\left[\sum_{l=1}^{k}\frac{\partial^2 f(X)}{\partial x_l^2}\right]^2;$$

2) $[\mathbf{D}f_{[n]}(X)]^{-1/2}[f_{[n]}(X)-\mathbf{M}f_{[n]}(X)]$ *распределено асимптотически нормально с нулевым средним и единичной дисперсией.*

Доказательство. Из условий (2) и (6) следует, что смещение

$$\mathbf{M}[f_{[n]}(X)-f(X)]=\frac{1}{2n}\sum_{j=1}^{n}h^2_{[j]}\left[\sum_{l=1}^{k}\frac{\partial^2 f(X)}{\partial x_l^2}\right]+o\left(\frac{1}{n}\sum_{j=1}^{n}h^2_{[j]}\right), \tag{8}$$

а дисперсия

$$\mathbf{D}f_{[n]}(X)=\frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{h^k_{[j]}}f(X)L^k+o\left(\frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{h^k_{[j]}}\right). \tag{9}$$

Теперь, учитывая (см. [11], стр. 15), что

$$\sum_{j=1}^{n}j^p=\frac{(n+1)^{p+1}}{p+1}+o(n^{p+1}),\quad p>-1,$$

и $h_{[j]}=Aj^{-\alpha}$ согласно условию теоремы, получаем, что в выражении (8)

$$\frac{1}{n}\sum_{j=1}^{n}h^2_{[j]}=\frac{h^2_{[n]}}{1-2\alpha}+o(h^2_{[n]}), \tag{10}$$

а в выражении (9)

$$\frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{h^k_{[j]}}=\frac{(1+k\alpha)^{-1}}{nh^k_{[n]}}+o\left(\frac{1}{nh^k_{[n]}}\right). \tag{11}$$

Так как среднеквадратическое отклонение

$$\mathbf{M}[f_{[n]}(X)-f(X)]^2=\mathbf{D}f_{[n]}(X)+[\mathbf{M}(f_{[n]}(X)-f(X))]^2,$$

то из соотношений (10) и (11) сразу вытекает справедливость первого утверждения теоремы.

Доказательство второго утверждения теоремы опускается: методикой п. 4 [7] нетрудно показать, что для суммы независимых неодинаково распределенных случайных величин, задаваемых в соответствии с (5), при $n\to\infty$ выполняется условие Ляпунова ([12], стр. 129), которое является достаточным для асимптотической нормальности оценки $f_{[n]}(X)$.

Один из результатов теоремы 1, заключающийся в том, что порядок и.с.к.о. оценок (1) и (4) одинаков, позволяет сформулировать такие следствия.

* Выбор $h_{[n]}$ обусловлен тем, что оптимальные h_n^0 для оценки (1) находятся в классе $An^{-\alpha}$ [6, 10].

Следствие 1. *Оптимальные $h^0_{[n]}$, минимизирующие асимптотические и.с.к.о. рекуррентной оценки* (4), *выражаются через оптимальные h^0_n для оценки* (1), *найденные в* [6]

$$h^0_{[n]}=\left[\frac{k+2}{2(k+4)}\right]^{1/(k+4)} h_n^0, \tag{12}$$

где

$$h_n^0=\left[kL^k\left[n\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\left[\sum_{l=1}^{k}\frac{\partial^2 f(x)}{\partial x_l^2}\right]^2 dx_1 dx_2\ldots dx_k\right]^{-1}\right]^{1/(k+4)}. \tag{13}$$

Оптимальная форма ядра [6, 10]

$$K_0(u)=\begin{cases}\dfrac{3}{4\sqrt{5}}-\dfrac{3u^2}{20\sqrt{5}} & \text{при } |u|\leqslant\sqrt{5},\\ 0 \quad \text{при } |u|>\sqrt{5}.\end{cases} \tag{14}$$

для оценок (1) *и* (4) *совпадает.*

Следствие 2. *Для оптимальных $h^0_{[n]}$ и h_n^0, которые определяются согласно* (12) *и* (13), *а.о.э. в смысле* (7)

$$e(f_{[\cdot]}, f_\cdot)=2^{-4/(k+4)}\left(\frac{k+4}{k+2}\right)^{2(k+2)/(k+4)}>1.$$

Из таблицы, в которой приведены значения а.о.э. для некоторых значений k, видно, что рекуррентные оценки практически не проигрывают обычным в эффективности, поэтому в ряде случаев их использование может

k	1	2	3	4
$e(f_{[\cdot]}, f_\cdot)$	1,0620	1,0819	1,0881	1,0884

оказаться полезным. Например, при необходимости оценить плотность вероятности в конечном числе точек $X_1, X_2, \ldots, X_c$, когда по каким-либо причинам для каждой точки X_i, $i=1, 2, \ldots, c$ проще получить свою выборку, чем хранить выборку объема n в памяти ЦВМ (отметим, что иногда объем памяти ЦВМ оказывается недостаточным для хранения выборки такого объема), следует отдавать предпочтение рекуррентным оценкам.

3. При оценивании условных плотностей и линии регрессии ограничимся рассмотрением выборки из двумерного распределения ($k=2$) ввиду того, что все предположения и выводы для случая $k>2$ по существу не изменяются. Итак, пусть дана выборка независимых реализаций (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$ двумерной случайной величины из генеральной совокупности с совместной плотностью $f(x, y)$, равномерно непрерывной в (x, y), и маргинальной плотностью $g(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$. Так как условная плотность вероятности $f(y|x)=f(x, y)/g(x)$, то в качестве приемлемой оценки $f(y|x)$ возьмем

$$f_{[n]}(y|x)=f_{[n]}(x, y)/g_{[n]}(x), \tag{15}$$

где $f_{[n]}(x, y)$ и $g_{[n]}(x)$ задаются рекуррентными соотношениями типа (4).

Назовем определенную вышеуказанным способом $f_{[n]}(y|x)$ рекуррентной оценкой условной плотности вероятности $f(y|x)$. Оценку (15) можно упростить. В самом деле, из выражения (5) следует, что справедливо представление

$$f_{[n]}(y|x)=C_{[n]}(x, y)/d_{[n]}(x),$$

где

$$C_{[n]}(x,y)=C_{[n-1]}(x,y)+\frac{1}{h^2_{[n]}}K\left(\frac{x-x_n}{h_{[n]}}\right)K\left(\frac{y-y_n}{h_{[n]}}\right),$$

$$d_{[n]}(x)=d_{[n-1]}(x)+\frac{1}{h_{[n]}}K\left(\frac{x-x_n}{h_{[n]}}\right).$$

Теорема 2. *Пусть ядро* $K(u)$ *удовлетворяет условиям* (2); $h_{[n]}=An^{-\alpha}$, $0<2\alpha<1$, $A>0$; *плотности вероятности* $f(x, y)$ *и* $g(x)$ *принадлежат классу* R, *причем* $g(x)\neq 0$. *Тогда при* $n\to\infty$

1) *среднеквадратическое отклонение* $\mathrm{M}[f_{[n]}(y|x)-f(y|x)]^2\sim t_{[n]}(x,y)+ +[b_{[n]}(x,y)]^2$, *где*

$$t_{[n]}(x,y)=\frac{f(y|x)L^2}{(1+2\alpha)nh^2_{[n]}}, \tag{16}$$

$$b_{[n]}(x,y)=\frac{h^2_{[n]}}{2(1-2\alpha)}\left\{\frac{1}{g(x)}\left[\frac{\partial^2 f(x,y)}{\partial x^2}+\frac{\partial^2 f(x,y)}{\partial y^2}\right]-\right.$$
$$\left.-\frac{f(x,y)}{g(x)^2}\frac{\partial^2 g(x)}{\partial x^2}\right\}; \tag{17}$$

2) $[f_{[n]}(y|x)-\mathrm{M}f_{[n]}(y|x)]$ *распределено асимптотически нормально с нулевым средним и дисперсией* (16);

3) *оптимальные* $h^0_{[n]}$, *минимизирующие асимптотическое и.с.к.о. рекуррентной оценки* (15), *имеют вид*:

$$h^0_{[n]}=\left[2L^2\left[3n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[g(x)^{-1}\left(\frac{\partial^2 f(x,y)}{\partial x^2}+\frac{\partial^2 f(x,y)}{\partial y^2}\right)-\right.\right.\right.$$
$$\left.\left.\left.-g(x)^{-2}f(x,y)\frac{\partial^2 g(x)}{\partial x^2}\right]^2 dx\,dy\right]^{-1}\right]^{1/6}; \tag{18}$$

4) *разность между* $f_{[n]}(y|x)$ *и* $f(y|x)$ *может быть сделана имеющей порядок* $O(n^{-1/3})$ *с вероятностью, сколь угодно близкой к единице, если положить* $h_{[n]}=An^{-1/6}$, $A>0$.

Доказательство. Первые утверждения теорем 1 и 2 доказываются одинаковой методикой, поэтому в данном случае доказательство не приводится. Из теоремы 1 работы [9] следует, что асимптотические распределения статистик $T_{[n]}=f_{[n]}(y|x)-\mathrm{M}f_{[n]}(y|x)$ и $T'_{[n]}=g(x)^{-1}[f_{[n]}(x,y)-\mathrm{M}f_{[n]}(x,y)]$ совпадают. Применение теоремы 1 данной работы к статистике $T'_{[n]}$ позволяет немедленно доказать второе утверждение теоремы. Оптимальные $h^0_{[n]}$, определяемые выражением (20), нетрудно получить методикой п. 2 [6]. Последнее утверждение теоремы очевидно.

Рассмотрим теперь оценивание линии регрессии $r(x)$. Так как

$$r(x)=\int_{-\infty}^{\infty}yf(y|x)dy=g(x)^{-1}\int_{-\infty}^{\infty}yf(x,y)dy=\frac{a(x)}{g(x)},$$

то в качестве оценки $r(x)$ естественно взять

$$r_{[n]}(x)=\int_{-\infty}^{\infty} y f_{[n]}(y|x)\,dy=\frac{a_{[n]}(x)}{g_{[n]}(x)},$$

где $a_{[n]}(x)$ и $g_{[n]}(x)$ задаются соотношениями типа (4). Если

(19) $$f(x, y)\leqslant B(1+x^2+y^2)^{-5/2},\quad B<\infty,$$

то линия регрессии будет определена и непрерывна [9]. Как и в случае оценивания условной плотности вероятности, возможен упрощенный вариант $r_{[n]}(x)$:

$$r_{[n]}(x)=s_{[n]}(x)/d_{[n]}(x),$$

где

$$s_{[n]}(x)=s_{[n-1]}(x)+h_{[n]}^{-1}y_nK(h_{[n]}^{-1}(x-x_n)),$$

$$d_{[n]}(x)=d_{[n-1]}(x)+h_{[n]}^{-1}K(h_{[n]}^{-1}(x-x_n)).$$

Ниже приводится лемма из [13], которая будет использована при доказательстве асимптотической нормальности оценки линии регрессии.

Лемма. *Пусть* (P_n, Q_n), $n=1, 2, \ldots$ *и* (P, Q) — *двумерные случайные величины. Если*

1) (P_n, Q_n) *сходится по распределению к* (P, Q), *когда* $n\to\infty$;

2) d_n, $n=1, 2, \ldots$ *есть последовательность ненулевых констант, таких, что* $\lim\limits_{n\to\infty} d_n=0$;

3) $H(z, u)$ *есть функция вещественных переменных* (z, u), *которая имеет полный дифференциал в* $(0, 0)$ *с*

$$H_1=\frac{\partial H(z,u)}{\partial z}\bigg|_{(0,0)},\quad H_2=\frac{\partial H(z,u)}{\partial u}\bigg|_{(0,0)}$$

причем H_1 *и* H_2 *не равны нулю одновременно, то при* $n\to\infty$ *случайная величина* $W_n=d_n^{-1}[H(d_n P_n, d_nQ_n)-H(0,0)]$ *сходится по распределению к* H_1P+H_2Q.

Далее, если (P, Q) *распределена по нормальному закону со средним* $(0, 0)$ *и невырожденной ковариационной матрицей* $\begin{Vmatrix} A_1 & B_1 \\ B_1 & C_1 \end{Vmatrix}$, *то* W_n *имеет предельное нормальное распределение со средним* 0 *и дисперсией* $H_1^2A_1+2H_1H_2B_1+H_2^2C_1$.

Теорема 3. *Пусть ядро* $K(u)$ *удовлетворяет условиям* (2); *для* $f(x, y)$ *имеет место ограничение* (19); $\int_{-\infty}^{\infty} y^2 f(x,y)\,dy<\infty$; $a(x)$ *и* $g(x)$ *удовлетворяют условию* R, *причем* $g(x)\neq 0$; $h_{[n]}=An^{-\alpha}$, $0<\alpha<1$. *Тогда при* $n\to\infty$

1) $[r_{[n]}(x)-r(x)]$ *распределено асимптотически нормально с нулевым средним и дисперсией*

$$\frac{(1+\alpha)^{-1}L}{nh_{[n]}g(x)}\left[\int_{-\infty}^{\infty} y^2 f(y|x)\,dy-r(x)^2\right];$$

2) *разность между* $r_{[n]}(x)$ *и* $r(x)$ *может быть сделана имеющей порядок* $O(n^{-2/5})$ *с вероятностью, сколь угодно близкой к единице, если положить* $h_{[n]}=An^{-1/5}$, $A>0$.

Доказательство. В соответствии с леммой обозначим

$$P_n=\sqrt{nh_{[n]}}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_{[j]}}y_jK\left(\frac{x-x_j}{h_{[j]}}\right)-a(x)\right],$$

$$Q_n=\sqrt{nh_{[n]}}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_{[j]}}K\left(\frac{x-x_j}{h_{[j]}}\right)-g(x)\right],$$

$$d_n=\frac{1}{\sqrt{nh_{[n]}}},\quad H(z,u)=\frac{z+a(x)}{u+g(x)}.$$

Применяя те же приемы, что и при выводе дисперсии рекуррентной оценки плотности вероятности в теореме 1, получим, что при $n\to\infty$ ковариа-

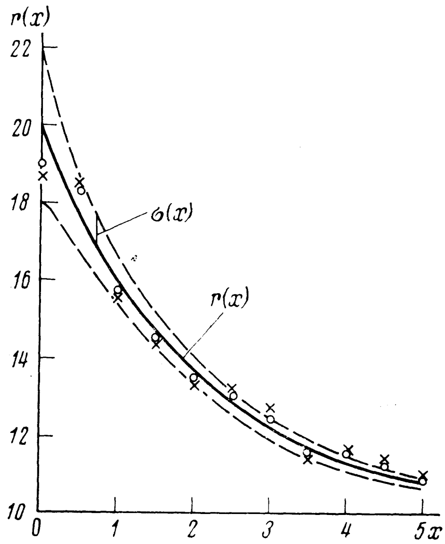

Рис. 1

Рис. 1. × – значения рекуррентной оценки, ● – значения нерекуррентной оценки

Рис. 2. Зависимость ошибки $M(n)$ от объема выборки n в случае оценивания линии регрессии экспоненциального типа с помощью рекуррентной (1) и нерекуррентной (2) оценок

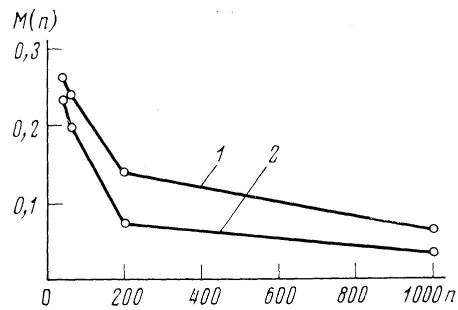

Рис. 2

ционная матрица случайной величины (P_n, Q_n) стремится к невырожденной ковариационной матрице

$$\frac{L}{1+\alpha}\left\|\begin{matrix}\int\limits_{-\infty}^{\infty}y^2f(x,y)\,dy & a(x)\\ a(x) & g(x)\end{matrix}\right\|. \tag{20}$$

Далее, использование методики п. 4 [7] и соотношений (10) и (11) позволяет показать, что для двумерной случайной величины (P_n, Q_n) при $n\to\infty$ выполняется условие Ляпунова. Таким образом, из многомерной центральной предельной теоремы ([14], стр. 269) следует, что (P_n, Q_n) сходится по распределению к случайной величине (P, Q), которая распределена нормально со средним (0, 0) и невырожденной ковариационной матрицей (20). Итак, согласно лемме имеем

$$H_1=1/g(x),\qquad H_2=-r(x)/g(x),\qquad H(0,0)=r(x),$$

$$W_n = \sqrt{nh_{[n]}}[\ddot{r}_{[n]}(x) - r(x)], \quad A_1 = \frac{L}{1+\alpha}\int_{-\infty}^{\infty} y^2 f(x,y)\,dy,$$

$$B_1 = L(1+\alpha)^{-1}a(x), \qquad C_1 = L(1+\alpha)^{-1}g(x),$$

и, следовательно, $\sqrt{nh_{[n]}}\,[r_{[n]}(x) - r(x)]$ распределено асимптотически нормально с нулевым средним и дисперсией

$$\frac{L}{(1+\alpha)g(x)}\left[\int_{-\infty}^{\infty} y^2 f(y|x)\,dy - r(x)^2\right],$$

что и требовалось доказать.

Громоздкие выражения для оптимальных $\overset{0}{h}_{[n]}$ и смещения оценки $r_{[n]}(x)$ не выписываются, так как они не имеют принципиальных расхождений с выражениями (18) и (17).

Рис. 3. × – значения рекуррентной оценки, ● – значения нерекуррентной оценки

Рис. 4. Зависимость ошибки $M(n)$ от объема выборки в случае оценивания линии регрессии параболического типа рекуррентной (1) и нерекуррентной (2) оценками

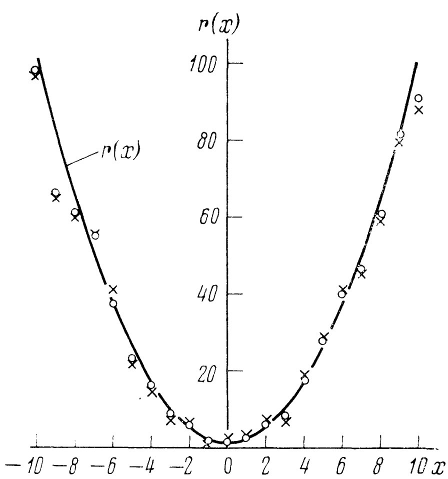

Рис. 3

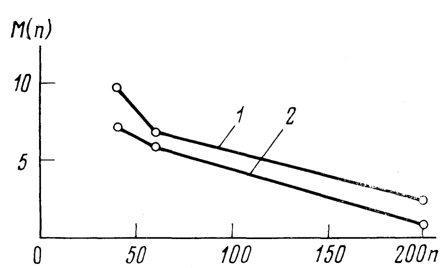

Рис. 4

Следует отметить, что результаты п.п. 2 и 3 можно обобщить для выборки из стационарного процесса со слабой зависимостью [15]. Некоторые результаты и принципы такого обобщения приводятся в [16–18, 10].

4. Методика проведения статистического эксперимента при моделировании оценок линии регрессии в основном совпадает с [19]. С помощью ЦВМ была получена совокупность независимых наблюдений (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$ с плотностью вероятности $f(x, y) = f(y|x)g(x)$, где x — независимая, а y — зависимая переменная, ошибка измерения которой подчинена нормальному закону с дисперсией

(21) $$\sigma^2(x) = [1+\exp(-0{,}5x)]^2.$$

В качестве моделей линии регрессии были выбраны

(22) $$r_1(x) = 10 + 10\exp(-0{,}5x),$$

(23) $$r_2(x) = 1 + x^2.$$

Реализации $x_1, x_2, \ldots, x_n$ независимой переменной получили из равномерного распределения на интервале $[z_1, z_2]$, т. е. $g(x) = (z_2 - z_1)^{-1}$, а $y_1, y_2, \ldots$

..., y_n при условии, что $X=x$ получили из нормального распределения

$$f(y|x)=\frac{1}{\sqrt{2\pi}\sigma(x)}\exp\left(-\frac{(y-r_i(x))^2}{2\sigma^2}\right),$$

где $r_i(x)$, $i=1$, 2 задаются соотношениями (22) и (23), а $\sigma^2(x)$ — (21). Для модели (22) взяли оптимальную форму ядра (14), $[z_1, z_2]=[0, 5]$, а для (23) было выбрано

$$K(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right),\ [z_1, z_2]=[-10, 10].$$

В обоих случаях положили $h_{[n]}=h_n=n^{-1/5}$, $n=40$, 60, 200, 1000. За меру близости оценки к истинной кривой регрессии приняли

$$M(n)=\frac{1}{m}\sum_{l=1}^{m}(r_{[n]}(x_l)-r(x_l))^2.$$

На рис. 1 представлены вычисленные в точках $x_l=l\cdot 0{,}5$, $l=0, 1, \ldots, 10$ значения рекуррентной и нерекуррентной оценок линии регрессии ($n=40$), задаваемой экспоненциальной формулой (22), а на рис. 2 для этого случая показана зависимость меры близости $M(n)$ от объема выборки n. На рис. 3 и 4 аналогичные результаты представлены для линии регрессии, задаваемой параболой (23).

ЛИТЕРАТУРА

1. *Невельсон М. Б., Хасьминский Р. З.* Стохастическая аппроксимация и рекуррентное оценивание. М., «Наука», 1972.
2. *Цыпкин Я. З.* Адаптация и обучение в автоматических системах. М., «Наука», 1968.
3. *Вазан М.* Стохастическая аппроксимация. М., «Мир», 1972.
4. *Parsen E.* On Estimation of a Probability Density and Mode. Ann. Math. Statist., 1965, 33, *3*, 1065–1076.
5. *Rosenblatt M.* Remarks on Some Nonparametric Estimates of a Density Function. Ann. Math. Statist., 1956, 27, *3*, 832–835.
6. *Епанечников В. А.* Непараметрическая оценка многомерной плотности вероятности. Теор. вероятн. и ее примен., 1969, 14, *1*, 156–162.
7. *Murthy V. K.* Nonparametric Estimation of Multivariate Densities with Applications. Multivariate Analysis, I. New York — London, Academic Press, 1966, 43—56.
8. *Wolverton C. T., Wagner T. J.* Asymptotically Optimal Discriminant Functions for Pattern Classification. IEEE Trans. Inform. Theory, 1969, 15, *2*, 258–266.
9. *Rosenblatt M.* Conditional probability density and regression estimators. Multivariate Analysis, II. New York — London, Academic Press, 1969, 25–31.
10. *Зияева З. Т.* О задаче оптимизации в оценке плотности вероятности. Изв. АН УзбССР, 1973, *1*, 30–35.
11. *Градштейн И. С. и Рыжик И. М.* Таблицы интегралов, сумм, рядов и произведений. М., «Наука», 1971.
12. *Петров В. В.* Суммы независимых случайных величин. М., «Наука», 1972.
13. *Hoeffding W., Robbins H.* The Central Limit Theorem for Dependent Random Variables. Duke Math. J., 1948, 15, *2*, 773—780.
14. *Уилкс С.* Математическая статистика. М., «Наука», 1967.
15. *Ибрагимов И. А., Линник Ю. В.* Независимые и стационарно связанные величины. М., «Наука», 1965.
16. *Булдаков В. М., Кошкин Г. М.* Рекуррентное оценивание условной плотности вероятности и линии регрессии по зависимой выборке. Материалы V научн. конф. по математике и механике, I. Томск, 1975, 89–90.
17. *Кошкин Г. М., Симахин В. А., Тарасенко Ф. П.* Об одной оценке условной функции распределения и регрессии. Материалы IV научн. конф. по математике и механике, I. Томск, 1974, 135–136.
18. *Rosenblatt M.* Density estimates and Markow sequences. Nonparametric techniques in statistical inference. Cambridge, Univer. Press, 1970, 199–210.
19. *Апраушева Н. Н., Конаков В. Д.* Использование непараметрических оценок в регрессионном анализе. Заводск. лаб., 1973, 39, *5*, 566—569.

Поступила в редакцию
20 февраля 1975 г.