М. А. МИРЗАХМЕДОВ, доц.,
Ш. А. ХАШИМОВ, асп.
Ташкентский университет

ОБ ОЦЕНКЕ ПЛОТНОСТИ ВЕРОЯТНОСТИ

1. Пусть $x_1, \ldots, x_n$ — выборка из n независимых наблюдений над случайной величиной ξ с плотностью вероятности $f(x)$. Построение широкого класса оценок для неизвестной функции распределения $F(x)$ и ее плотности $f(x)$, которые имели бы определенные статистические свойства (несмещенность, состоятельность, эффективность и т. д.), представляет большой интерес и является одной из основных задач непараметрической статистики.

Для оценки $f(x)$ Парзен [6] рассмотрел класс статистик вида

$$f_n(x) = \frac{1}{h(n)} \int_{-\infty}^{\infty} K\left(\frac{x-y}{h(n)}\right) dF_n(x) = \frac{1}{nh(n)} \sum_{j=1}^{n} K\left(\frac{x-x_j}{h(n)}\right), \quad (1)$$

где «весовая» функция $K(x)$ — плотность вероятности, удовлетворяющая условиям

$$\sup_x K(x) < \infty, \lim_{|x|\to\infty} |x| K(x) = 0,$$

$\{h(n)\}$ — последовательность неотрицательных чисел,

$$\lim_{n\to\infty} h(n) = 0.$$

Различные свойства оценки $f_n(x)$, а также многомерный аналог (1) были исследованы многими авторами [3—12].

Оценку для $f(x)$ можно построить и при помощи ортонормированных систем функций [5]. В работе [8] Шварц рассматривает для плотности вероятности $f(x)$ оценку вида

$$f_n(x) = \sum_{j=0}^{q(n)} a_{jn} \varphi_j(x),$$

где

$$a_{jn} = \frac{1}{n} \sum_{k=1}^{n} \varphi_j(x_k),$$

$\varphi_i(x)$ — ортонормированные функции Эрмита, $\{q(n)\}$ — последовательность целых положительных чисел и $\lim_{n\to\infty} q(n) = 0$.

Различные интересные свойства многомерных полиномов Эрми та, разложение функции $f(y_1, \ldots, y_n) \in L_2$ в ряд по многомерны полиномам Эрмита рассмотрены в работе [2]. В [9] был изучен слу чай, когда $x_1, \ldots, x_n$ образуют стационарный марковский процесс и приведены условия, при которых оценка для плотности вероят ности перехода, построенная при помощи класса статистик вид (1), является состоятельной.

2. Пусть $x_1, \ldots, x_n, \ldots$ — стационарный марковский процес (выполнена гипотеза D_0), стационарное и начальное распределени совпадают ([1], стр. 202). Обозначим через $p(x)$ и $g(y_1, y_2)$ соответ ственно начальную и двумерную совместную плотности вероятност значений процесса, определенные относительно мер Лебега. Есл $p(x)$ строго положительна на R, то $t(x, y_1, y_2) = \frac{g(y_1, y_2)}{p(x)}$ является плотностью вероятности перехода.

В этой заметке мы построим оценки $p_n(x)$, $g_n(y_1, y_2)$, $t_n(x, y_1, y_2)$ соответственно для $p(x)$, $g(y_1, y_2)$, $t(x, y_1, y_2)$ при помощи орто нормированных функций Эрмита. Качество оценки (например p_n) будем измерять величинами

$$M[p_n(x) - p(x)]^2, \; M\int_{-\infty}^{+\infty}[p_n(x) - p(x)]^2\,dx.$$

Пусть $p(x)$, $g(y) \equiv g(y_1, y_2) \in L_2$,

$$p(x) = \sum_{j=0}^{\infty} a_j \varphi_j(x), \; g(y) = \sum_{i,j=0}^{\infty} a_{ij}\varphi_{ij}(y), \tag{2}$$

где

$$a_j = \int_{-\infty}^{+\infty} \varphi_j(x)\,p(x)\,dx, \; a_{ij} = \int_{R_2} \psi_{ij}(y)\,g(y)\,dy$$

и $\varphi_j(x)$, $\varphi_{ij}(y)$, $\psi_{ij}(y)$ — ортонормированные одномерные и дву мерные функции Эрмита [2].

Известно, (см., например, [87]), что

$$|\varphi_j(x)| < c_1, \; |\varphi_{ij}(y)| < c_2, \; |\psi_{ij}(y)| < c_3, \tag{3}$$

где c_1, c_2, c_3 — постоянные, не зависящие от x, y, i, j.

В качестве оценок для неизвестных плотностей $p(x)$, $g(y)$ рас смотрим класс статистик вида

$$p_n(x) = \sum_{j=0}^{q(n)} a_{jn}\varphi_j(x), \; g_n(y) = \sum_{i=0}^{q_1(n)}\sum_{j=0}^{q_2(n)} a_{ijn}\varphi_{ij}(y),$$

где

$$a_{jn} = \frac{1}{n}\sum_{k=1}^{n}\varphi_j(x_k), \; a_{ijn} = \frac{1}{n}\sum_{k=1}^{n}\psi_{ij}(y_k), \; y_k = (x_k, x_{k+1}).$$

Сперва изучим свойства статистики $p_n(x)$. Очевидно, что $Ma_{jn} = a_j$.

Рассмотрим среднеквадратичную сходимость $p_n(x)$ к $p(x)$

$$\sigma_n^2 = \boldsymbol{M}[p_n(x) - p(x)]^2 = Dp_n(x) + [p(x) - p_q(x)]^2,$$

где

$$p_q(x) = \sum_{j=0}^{q(n)} a_j \varphi_j(x),$$

$$Dp_n(x) = D\sum_{j=0}^{q(n)} a_{jn}\varphi_j(x) = \frac{1}{n^2}\sum_{k=1}^{n} D\left(\sum_{j=0}^{q(n)} \varphi_j(x_k)\varphi_j(x)\right) +$$

$$+ \frac{2}{n^2}\sum_{1\leqslant m<k\leqslant n} \operatorname{cov}\left(\sum_{j=0}^{q(n)} \varphi_j(x_m)\varphi_j(x), \sum_{j=0}^{q(n)} \varphi_j(x_k)\varphi_j(x)\right) = A_n + B_n.$$

В силу одинаковой распределенности

$$\sum_{k=1}^{n} D\left(\sum_{j=0}^{q(n)} \varphi_j(x_k)\varphi_j(x)\right) = n\sum_{j=0}^{q(n)} D\varphi_j(x_1)\varphi_j(x) +$$

$$+ 2n\sum_{0\leqslant i<j\leqslant q(n)} \operatorname{cov}(\varphi_i(x_1)\varphi_i(x), \varphi_j(x_1)\varphi_j(x)).$$

В силу (3) имеем

$$n\sum_{j=0}^{q(n)} D\varphi_j(x_1)\varphi_j(x) < nc_1^2 \sum_{j=0}^{q(n)} D\varphi_j(x_1) < nc_1^4(1 + q(n))$$

$$2n\sum_{i<j} |\operatorname{cov}(\varphi_i(x_1)\varphi_i(x), \varphi_j(x_1)\varphi_j(x))| < 2nc_1^2 \times$$

$$\times \sum_{i<j} |\operatorname{cov}|\varphi_i(x_1), \varphi_j(x_1))| < 2nc_1^4 q(n)(1 + q(n)),$$

так как по неравенству Буняковского — Шварца

$$|\boldsymbol{M}\varphi_i(x_1)\varphi_j(x_1)| \leqslant \sqrt{\boldsymbol{M}\varphi_i^2(x_1)\boldsymbol{M}\varphi_j^2(x_1)} < c_1^2.$$

Итак,

$$A_n < \frac{c_1^4}{n}(1 + q(n))^2.$$

Теперь оценим B_n

$$B_n \leqslant \frac{2}{n^2}\sum_{i,j=0}^{q(n)}\sum_{m<k} |\operatorname{cov}(\varphi_i(x_m)\varphi_i(x), \varphi_i(x_k)\varphi_j(x)| <$$

$$< \frac{2c_1^2}{n^2}\sum_{i,j=0}^{q(n)}\sum_{m<k} |\operatorname{cov}(\varphi_i(x_m), \varphi_j(x_k))|.$$

Поскольку выполнена гипотеза (D_0), то, применяя лемму 7.1 ([1], стр. 203), получим

$$|\operatorname{cov}(\varphi_i(x_1), \varphi_j(x_k))| \leqslant 2\gamma^{\frac{1}{2}}\rho^{\frac{k-1}{2}}\{\boldsymbol{M}\varphi_i^2(x_1)\boldsymbol{M}\varphi_j^2(x_1)\}^{\frac{1}{2}},$$

где $\gamma > 0, 0 < \rho < 1$.

Тогда

$$\frac{2c_1^2}{n^2}\sum_{i,j=0}^{q(n)}\sum_{m<k}|\operatorname{cov}(\varphi_i(x_m),\ \varphi_j(x_k))|\leqslant$$

$$\leqslant\frac{2c_1^2}{n^2}\sum_{i,j=0}^{q(n)}\sum_{m<k}2\gamma^{\frac{1}{2}}\rho^{\frac{k-m}{2}}\{M\varphi_i^2(x_m)\,M\varphi_j^2(x_k)\}^{\frac{1}{2}}\leqslant$$

$$\leqslant\frac{4c_1^4}{n^2}\sum_{i,j=0}^{q(n)}\sum_{m<k}\gamma^{\frac{1}{2}}\rho^{\frac{k-m}{2}}=\frac{4c_1^4}{n^2}\gamma^{\frac{1}{2}}(1+q(n))^2\sum_{k=1}^{n-1}(n-k)\rho^{\frac{k}{2}}.$$

Итак,

$$B_n<4c_1^4\gamma^{\frac{1}{2}}\rho^{\frac{1}{2}}(1-\rho^{\frac{1}{2}})^{-1}\frac{(1+q(n))^2}{n}.$$

Следовательно, для среднеквадратичной сходимости получим оценку

$$\sigma_n^2<[p(x)-p_q(x)]^2+\frac{(1+q(n))^2}{n}[c_1^4+4c_1^4\gamma^{\frac{1}{2}}\rho^{\frac{1}{2}}(1-\rho^{\frac{1}{2}})^{-1}]=$$

$$=[p(x)-p_q(x)]^2+N_1\frac{(1+q(n))^2}{n}. \qquad (4)$$

Таким образом, нами доказано следующее утверждение.

Теорема 1. Пусть $q(n)=0(\sqrt{n})$ и $p(x)$ удовлетворяет условиям теоремы работы [2] при $n=1$. Тогда при сделанных выше предположениях оценка $p_n(x)$ является состоятельной оценкой для плотности вероятности $p(x)$.

Чтобы получить оценку скорости сходимости, сформулируем одну лемму, нужную нам и для дальнейшего, доказательство которой получается использованием некоторых дифференциальных соотношений для многомерных полиномов Эрмита [2].

Л е м м а. Пусть функция

$$\sum_{p_1=0}^{r_1}c_{r_1}^{p_1}\sum_{p_2=0}^{r_2}c_{r_2}^{p_2}(-1)^{-(p_1+p_2)}2^{-\frac{p_1+p_2}{2}}h_{p_1p_2}\left(\frac{y_1}{\sqrt{2}},\ \frac{y_2}{\sqrt{2}}\right)\frac{\partial^{r-p_1-p_2}}{\partial y_1^{r_1-p_1}\partial y_2^{r_2-p_2}}\otimes$$

$$\times\, g(y_1,\ y_2)$$

интегрируема с квадратом. Тогда для коэффициентов разложения (2) верна оценка

$$|a_{ij}|<\frac{c(r)}{(2\pi)^{\frac{1}{2}}2^{\frac{r}{2}}i^{\frac{r_1}{2}}j^{\frac{r_2}{2}}},$$

где

$$c(r)=\left\|h_{p_1p_2}\left(\frac{y_1}{\sqrt{2}},\ \frac{y_2}{\sqrt{2}}\right)\cdot\frac{\partial^{r-p_1-p_2}}{\partial y_1^{r_1-p_1}\partial y_2^{r_2-p_2}}g(y_1y_2)\right\|_{L_2},$$

r_1, r_2 — порядок произвольной функции $q(y_1, y_2)$ по аргументам y_1, y_2 соответственно, $r = r_1 + r_2$, $h_{p_1 p_2}(y_1 y_2)$ — двумерный полином Эрмита.

Теперь легко показать, что если справедлива теорема 1 и выполняются условия леммы для начальной (одномерной) плотности $p(x)$ при $r \geqslant 3$ и $q(n) \sim \alpha n^{1/r}$, когда $n \to \infty$, то

$$\sigma_n^2 = O\left(\frac{1}{n^{\frac{r-2}{r}}}\right). \tag{5}$$

Действительно, используя лемму, получим

$$\sum_{j=q(n)+1}^{\infty} |a_j \varphi_j(x)| < c_1 \sum_{j=q(n)+1}^{\infty} |a_j| < c_1 \sum_{j=q(n)+1}^{\infty} \frac{c(r)}{(2j)^{\frac{r}{2}}} <$$

$$< \frac{c_1 c(r)}{2^{\frac{r}{2}} \left(\frac{r}{2} - 1\right) \cdot (1 + q(n))^{\frac{r-2}{2}}}.$$

Тогда (4) можно переписать в виде

$$\sigma_n^2 < \frac{c_1^2 c^2(r)}{2^r \left(\frac{r}{2} - 1\right)^2 (1 + q(n))^{r-2}} + N_1 \frac{(1 + q(n))^2}{n}.$$

Отсюда получается (5).

Точно так же, как теорема 1, доказывается следующая теорема.

Теорема 2. Пусть $q_1(n)\, q_2(n) = o(\sqrt{n})$ и $g(y)$ удовлетворяет условиям теоремы работы [2] при $n = 2$. Тогда, при сделанных выше предположениях, оценка $g_n(y)$ является состоятельной оценкой для двумерной совместной плотности $g(y)$.

Если, кроме того, выполняются условия леммы при $r_1 \geqslant 4$, $r_2 \geqslant 5$ (или $r_1 \geqslant 5$, $r_2 \geqslant 4$) и $q_1(n) \sim \alpha_1 n^{1/r_1}$, $q_2(n) \sim \alpha_2 n^{1/r_2}$, то

$$M(g_n(y) - g(y))^2 = O\left(\frac{1}{n^{1 - \frac{2}{r_1} - \frac{2}{r_2}}}\right).$$

Следствие. Оценка $t_n(x, y_1, y_2) = \frac{g_n(y_1, y_2)}{p_n(x)}$ является состоятельной оценкой для плотности вероятности перехода $t(x, y_1, y_2) = \frac{g(y_1, y_2)}{p(x)}$.

3. Теперь рассмотрим интегральную среднеквадратичную сходимость. Имеем

$$I_n = M \int_{-\infty}^{\infty} (p_n(x) - p(x))^2 dx = M \int_{-\infty}^{\infty} \sum_{j=0}^{q(n)} (a_{jn} - a_j) \varphi_j(x) -$$

$$- \sum_{j=1+q(n)}^{\infty} a_j \cdot \varphi_j(x))^2 dx.$$

Используя ортонормированность функций Эрмита, получим

$$M\int_{-\infty}^{\infty}(p_n(x)-p(x))^2\,dx=\sum_{j=0}^{q(n)}M(a_{jn}-a_j)^2+\sum_{j=q(n)+1}^{\infty}a_j^2.$$

Так как $Ma_{jn}=a_j$, то

$$I_n=\sum_{j=0}^{q(n)}Da_{jn}+\sum_{j=q(n)+1}^{\infty}a_j^2.$$

При оценке Da_{jn} воспользуемся теми же рассуждениями, что и при оценке A

$$Da_{jn}=\frac{1}{n^2}\sum_{k=1}^{n}D\varphi_j(x_k)+\frac{2}{n^2}\sum_{1\leqslant i<k\leqslant n}\operatorname{cov}(\varphi_i(x_i),\ \varphi_j(x_k)),$$

$$\frac{1}{n^2}\sum_{k=1}^{n}D\varphi_j(x_k)=\frac{1}{n}D\varphi_j(x_1)<c_1^2\frac{1}{n}.$$

Применяя лемму 7.1 [1], получим

$$\frac{2}{n^2}\sum_{i<k}|\operatorname{cov}(\varphi_j(x_i),\ \varphi_j(x_k))|<N_2\frac{1}{n},$$

где $N_2=4\gamma^{1/2}\rho^{1/2}\ (1-\rho^{1/2})^{-1}\ c_1^2$.

Таким образом,

$$I_n<\frac{1+q(n)}{n}\cdot[c_1^2+N_2]+\sum_{j=q(n)+1}^{\infty}a_j^2. \tag{6}$$

Теорема 3. Пусть $q(n)=o(n)$. Тогда, при сделанных выше предположениях, оценка $p_n(x)$ является состоятельной оценкой для плотности вероятности $p(x)$ в смысле интегральной среднеквадратичной сходимости. Если, кроме того, выполняются условия леммы для $p(x)$ при $r\geqslant 2$ и $q(n)\sim\beta n^{1/r}$, когда $n\to\infty$, то

$$I_n=O\left(\frac{1}{n^{\frac{r-1}{r}}}\right).$$

Действительно, из (6), используя лемму для начальной (одномерной) плотности $p(x)$, имеем

$$I_n<\frac{1+q(n)}{n}[c_1^2+N_2]+\sum_{j=q(n)+1}^{\infty}\frac{c^2(r)}{(2j)^r}<$$

$$<\frac{1+q(n)}{n}[c_1^2+N_2]+\frac{c^2(r)}{2^r(r-1)(1+q(n))^{r-1}}.$$

Аналогично доказывается следующая теорема.

Теорема 4. Пусть $q_1(n)\ q_2(n)=o(n)$. Тогда, при сделанных выше предположениях, оценка $q_n(y)$ для двумерной совместной плотности $g(y)$ является состоятельной оценкой в смысле интегральной среднеквадратичной сходимости.

Если, кроме того, выполняются условия леммы при $r_1 \geqslant 2, r_2 \geqslant 3$, (или $r_1 \geqslant 3, r_2 \geqslant 2$) и $q_1(n) \sim \alpha_3 n^{1/r_1}$, $q_2(n \sim \alpha_4) n^{1/r_2}$, когда $n \to \infty$, то

$$M \int_{R_2} (g_n(y) - g(y))^2 dy = O\left(\frac{1}{n^{1-\frac{1}{r_1}-\frac{1}{r_2}}}\right).$$

4. Рассмотрим асимптотическое распределение оценок $p_n(x)$ и $q_n(y)$. $p_n(x)$ можно представить в виде

$$p_n(x) = \sum_{k=1}^{n} \eta_k,$$

где

$$\eta_k = \frac{1}{n} \sum_{j=0}^{q(n)} \varphi_j(x_k)\varphi_j(x).$$

Непосредственно проверяется, что существует положительный предел

$$\lim_{n\to\infty} M\left\{\frac{\sqrt{n}}{1+q(n)} \sum_{k=1}^{n} [\eta_k - M\eta_k]\right\}^2 = \sigma^2.$$

Следовательно, оценка $p_n(x)$ асимптотически нормальна ([1], стр. 207, теорема 7.5), т. е. при $n \to \infty$

$$P\left\{\frac{\sqrt{n}}{1+q(n)} [p_n(x) - Mp_n(x)] \leqslant \lambda\right\} \to \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\lambda} e^{-\frac{x^2}{2\sigma^2}} dx.$$

Оценку $g_n(y)$ представим в виде

$$g_n(y) = \sum_{k=1}^{n} \eta_k,$$

где

$$\eta_k = \frac{1}{n} \sum_{i=0}^{q_1(n)} \sum_{j=0}^{q_2(n)} \varphi_{ij}(y)\psi_{ij}(y_k).$$

Непосредственно проверяется, что существует положительный предел

$$\lim_{n\to\infty} M\left\{\frac{\sqrt{n}}{(1+q_1(n))(1+q_2(n))} \sum_{k=1}^{n} [\eta_k - M\eta_k]\right\}^2 = \sigma_0^2.$$

Таким образом, оценка $q_n(y)$ асимптотически нормальна с параметрами $(0, \sigma_0^2)$.

ЛИТЕРАТУРА

1. Дуб Дж. Л. Вероятностные процессы, ИЛ, М., 1956.
2. Сираждинов С. Х. К теории многомерных полиномов Эрмита.— Труды Института математики и механики АН УзССР, 1949, вып. 5, 70—95.
3. Надарая Э. А. О непараметрических оценках плотности вероятности и регрессии.— Теория вероят. и ее прим., 1965, **10**, № 1, 199—203.

4. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности.— Теория вероят. и ее прим. 1969, **14**, № 1, 156—161.
5. Ченцов Н. Н. Оценка неизвестной плотности распределения по наблюдениям,— ДАН СССР, 1961, **147**, № 1, 45—48.
6. Parzen E. On estimation of probability function and mode.— Ann. Math. Statist., 1962, 33, 1065—1076.
7. Murthy V. K. Estimation of probability density.— Ann. Math. Statist., 1965, 36, 3, 1027—1031.
8. Schwartz S. C. Estimation of probability density by an orthogonal series.— Ann. Math. Statist., 1967, **38**, 4, 1261—1265.
9. Roussas G. G. Nonparametric estimation in Markov processes.— Ann. Inst. Statist. Math., 1969, **21**, 1, 73—87.
10. Roussas G. G. Nonparametric estimation of the trasition distribution function of a markov process.— Ann. Math. Statist, 1969, **40**, 34, 1187—1195.
11. Bhattacharya P. K. Estimation of a probability density function and its derivatives.— Sankhya, 1967, Ser. A, 29, part 4, 373—382.
12. Vaduva I. Contributii la teoria estimatiilar statistice ale densitatilar de repartitic si aplicatii.— Stuilii si cercetarimat. Acad. R. S. R., 1968, 20, 8, 1207—1276.

M. A. Mirzahmedov, Sh. A. Khasimov

ON ESTIMATION OF THE PROBABILITY DENSITY

Summary

Let $x_1, \ldots, x_n, \ldots$ is a stationary Markov process, $p(x) \equiv p$ and $g(y_1, y_2) \equiv g$ are initial and two-dimensional jointly probability density of process, respectively.

In this paper estimations p_n, g_n for p, g by orthonormal functions of Hermite are constructed. Estimations p_n, g_n are asymptotically normal.

Поступила в редколлегию 1.IX 1971.