УДК 519.8

В.И. Рюмкин

ПРОЦЕДУРА НЕПАРАМЕТРИЧЕСКОГО ОЦЕНИВАНИЯ ФУНКЦИОНАЛОВ ОТ РАСПРЕДЕЛЕНИЙ СТАЦИОНАРНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И ЕЕ ПРИМЕНЕНИЕ В ЭКОНОМИЧЕСКИХ И ТЕХНИЧЕСКИХ ПРИЛОЖЕНИЯХ

Построена процедура синтеза сходящихся в среднеквадратичном непараметрических ядерных оценок функционалов от неизвестных распределений по стационарным выборкам слабозависимых случайных величин в условиях равномерно сильного перемешивания. Получены достаточные условия сходимости полученных оценок в среднеквадратичном. Показана асимптотическая нормальность полученных оценок. Рассмотрена возможность применения предложенной процедуры в экономических и технических приложениях.

Ключевые слова: *непараметрическое оценивание, ядерные функции, стационарные выборки, функционал.*

Для решения многих задач прикладной статистики в условиях отсутствия априорных сведений о функции распределения (ф.р.) $F(\boldsymbol{x})$ наблюдаемых многомерных случайных величин требуется оценивание функционалов, представимых в виде [1, 2]

$$H(F(\boldsymbol{x}), F^{(\boldsymbol{\alpha}_1)}(\boldsymbol{x}), \ldots, F^{(\boldsymbol{\alpha}_s)}(\boldsymbol{x})), \quad \boldsymbol{x} \in R^m, \tag{1}$$

где $H : R^{s+1} \to R^1$ – заданная функция, в качестве аргументов которой выступают частные производные $F^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x}) \equiv \dfrac{\partial^{\alpha_q} F(\boldsymbol{x})}{\partial x_1^{\alpha_{q1}} \partial x_2^{\alpha_{q2}} \ldots \partial x_m^{\alpha_{qm}}}$ порядка $\alpha_q = \alpha_{q1} + \ldots + \alpha_{qm}$, $q = \overline{1,s}$, функции распределения $F(\boldsymbol{x})$ в точке $\boldsymbol{x} \in R^m$. Мультииндекс $\boldsymbol{\alpha}_q = (\alpha_{q1}, \ldots, \alpha_{qm})$ здесь и далее представляет собой целочисленный вектор с неотрицательными составляющими.

К функционалам типа (1) относится функция интенсивности

$$\lambda(x) = F^{(1)}(x)/(1 - F(x))$$

(теория надежности и массовое обслуживание [3, 4]), логарифмическая производная плотности

$$\psi(x) = F^{(2)}(x) / F^{(1)}(x)$$

(проверка близких гипотез [5, 6], непараметрическая нелинейная фильтрация марковских процессов [7], оценивание кривой регрессии [8]), функционал потенциала экономической системы [9]. Естественный и наиболее распространенный метод оценивания состоит в подстановке вместо производных $F^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$ их состоятельных оценок $\tilde{F}^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$, $q = \overline{0,s}$. Однако прямое использование этого метода не всегда гарантирует получение оценок нужного качества, если, например, функция

$H(\cdot)$ в (1) имеет некоторые особенности (неограниченность, недифференцируемость и др.). В результате некорректной подстановки может получиться оценка, не имеющая дисперсии или даже математического ожидания.

В данной работе предложена универсальная процедура получения сходящихся в среднеквадратичном оценок функционалов (1) на основе использования непараметрических оценок ф.р. $F(\boldsymbol{x})$ и ее производных $\tilde{F}^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$ ядерного типа и операции усечения больших значений статистик $H(\tilde{F}(\boldsymbol{x}),\tilde{F}^{(\boldsymbol{\alpha}_1)}(\boldsymbol{x}),\ldots,\tilde{F}^{(\boldsymbol{\alpha}_s)}(\boldsymbol{x}))$, получаемых по методу прямой подстановки этих оценок в $H(\cdot)$.

1. Обозначения и определения

Пусть $\{\boldsymbol{x}(j)\in R^m\}$ – стационарная в узком смысле последовательность случайных величин, заданных на пространстве $\{X,B_X,P_{\vec{x}}\}$; $F(\boldsymbol{x})=P_{\vec{x}}(\boldsymbol{x}(j)<\boldsymbol{x})$ – ф.р. элементов последовательности $\{\boldsymbol{x}(j)\}$. Будем говорить, что $\{\boldsymbol{x}(j)\}$ удовлетворяет условию равномерно сильного перемешивания (РСП), если ее коэффициент перемешивания $\varphi_{\boldsymbol{x}}(\tau)$ [10] удовлетворяет условию

$$\varphi_{\boldsymbol{x}}(\tau)=\sup_{A,B}\left\{\frac{P_{\boldsymbol{x}}(AB)-P_{\boldsymbol{x}}(A)P_{\boldsymbol{x}}(B)}{P_{\boldsymbol{x}}(A)}:\ A\in\Im_{<1}^{\boldsymbol{x}},\ B\in\Im_{>\tau}^{\boldsymbol{x}},\ P_{\boldsymbol{x}}(A)\neq 0\right\}\xrightarrow[\tau\to\infty]{}0. \quad (2)$$

Здесь $\Im_{<k}^{\boldsymbol{x}}$, $\Im_{>k}^{\boldsymbol{x}}$ – алгебры, порожденные соответственно семействами $\{\boldsymbol{x}(\tau):\ \tau<k\}$, $\{\boldsymbol{x}(\tau):\ \tau>k\}$.

При построении состоятельных оценок $\tilde{F}^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$ неизвестных $F^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$, определяющих значение функционала (1), будем использовать оценки ядерного типа. Для каждой оцениваемой величины $F^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$ рассмотрим m-мерное ядро $\boldsymbol{K}_q(\boldsymbol{u})\equiv\prod_{i=1}^{m}K_{qi}(u_i)$, представляющее собой произведение заданных одномерных ядерных вещественных функций $K_{qi}(u_i):R^1\to R^1$.

В качестве состоятельных оценок $\tilde{F}^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$ величин $F^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})$ рассмотрим оценки ядерного типа

$$\tilde{F}^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})=F_n^{(\boldsymbol{\alpha}_q)}(\boldsymbol{x})\equiv\frac{1}{nh_q^{\alpha_q}}\sum_{j=1}^{n}\boldsymbol{K}_q^{(\boldsymbol{\alpha}_q)}\left(\frac{\boldsymbol{x}-\boldsymbol{x}(j)}{h_q}\right), \quad (3)$$

где $\boldsymbol{K}_q^{(\boldsymbol{\alpha}_q)}\left(\frac{\boldsymbol{x}-\boldsymbol{x}(j)}{h_q}\right)=\prod_{i=1}^{m}\frac{\partial^{\alpha_{qi}}}{\partial x_i^{\alpha_{qi}}}K_{qi}\left(\frac{x_i-x_i(j)}{h_q}\right)$ есть произведение производных одномерных ядерных функций; h_q – параметр размытости. Оценки строятся на основе выборки $X_n\equiv(\boldsymbol{x}(1),\ldots,\boldsymbol{x}(n))$ объема n из генеральной совокупности всех последовательностей $\{\boldsymbol{x}(j)\}$. В (3) и везде далее: $q=\overline{0,s}$; $\boldsymbol{\alpha}\equiv(\alpha_1,\alpha_2,\ldots,\alpha_m)$; $\alpha\equiv\alpha_1+\alpha_2+\ldots+\alpha_m$; если $\boldsymbol{x}\in R^m$ и $\boldsymbol{\alpha}\in R^m$, то тогда $\boldsymbol{x}^{\boldsymbol{\alpha}}\equiv x_1^{\alpha_1}x_2^{\alpha_2}\ldots x_m^{\alpha_m}$; если

$x \in R^1$ и $\boldsymbol{\alpha} \in R^m$, то тогда $x^{\boldsymbol{\alpha}} \equiv x^{\alpha_1+\alpha_2+\ldots+\alpha_m}$; $\mathbf{0} \equiv (0,0,\ldots,0)$; $\mathbf{1} \equiv (1,1,\ldots,1)$; $\boldsymbol{x} > \boldsymbol{y} \Leftrightarrow x_i > y_i$, $i=\overline{1,m}$; $\varphi(n) \prec \psi(n) \Leftrightarrow \varphi(n) < \psi(n)$ $\forall n > N$;

$$\operatorname{sgn}(\boldsymbol{x}) \equiv (\operatorname{sgn}(x_1), \operatorname{sgn}(x_2), \ldots, \operatorname{sgn}(x_m)), \text{ где } \operatorname{sgn}(x_j) = \begin{cases} 1, & x_j > 0, \\ 0, & x_j = 0, \\ -1, & x_j < 0. \end{cases}$$

Определение 1. Последовательность $h \equiv h(n)$ положительных вещественных чисел $h \in H_k$, $h \in H_{k,\varepsilon}$, если она удовлетворяет соответственно условиям

$$\lim_{n\to\infty}\left(h + \frac{1}{nh^{2k-\operatorname{sgn}(k)}}\right) = 0, \quad \lim_{n\to\infty}\left(h + \frac{1}{n^{1-\varepsilon}h^{2k-\operatorname{sgn}(k)}}\right) = 0.$$

Определение 2. Функция $f \in N_{\boldsymbol{\alpha}}(\boldsymbol{x})$, если $f(\boldsymbol{x})$ абсолютно непрерывна на R^m и имеет непрерывную производную

$$f^{(\boldsymbol{\alpha})}(\vec{x}) = \frac{\partial^{\alpha} f(\boldsymbol{x})}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_m^{\alpha_m}}$$

порядка $\boldsymbol{\alpha}$ в точке $\boldsymbol{x}$; $f \in N_{\boldsymbol{\alpha}}^{+}(\boldsymbol{x})$, если $f \in N_{\boldsymbol{\alpha}}(\boldsymbol{x})$ и $\sup |f^{(\boldsymbol{\nu})}(\boldsymbol{x})| < \infty$, $0 < \boldsymbol{\nu} < \boldsymbol{\alpha}$.

Записи $f \in N_{\boldsymbol{\alpha}}(R^m)$ и $f \in N_{\boldsymbol{\alpha}}^{+}(R^m)$ означают, что $f(\boldsymbol{x})$ удовлетворяет соответственно условиям $f \in N_{\boldsymbol{\alpha}}(\boldsymbol{x})$ и $f \in N_{\boldsymbol{\alpha}}^{+}(\boldsymbol{x})$ при любых $\boldsymbol{x} \in R^m$.

Определение 3. Ядро $K(u) \in L_1$, если оно измеримо по Борелю, причем $\int |K(u)|\, du < \infty$; $K(u) \in L_1^{+}$, если $K(u) \in L_1$ и, кроме того, $\sup |K(u)| < \infty$.

$\boldsymbol{K}(\boldsymbol{u}) \in \boldsymbol{L}_1$, если $K_i(u_i) \in L_1$, $i = \overline{0,m}$; $\boldsymbol{K}(\boldsymbol{u}) \in \boldsymbol{L}_1^{+}$, если $K_i(u_i) \in L_1^{+}$, $i = \overline{0,m}$.

Определение 4. $K(u) \in Z_{s,v}$, если

$$\int K(u) u^j du = \begin{cases} \begin{cases} 0, & j = \overline{0, s-1},\ j \neq v, \\ (-1)^v v!, & j = v, \end{cases} \\ \omega_{s,v} \neq 0, \quad j = s. \end{cases}$$

$\boldsymbol{K}(\boldsymbol{u}) \in \boldsymbol{Z}_{\boldsymbol{s},\boldsymbol{v}}$, если $K_i(u_i) \in Z_{s_i, v_i}$, $i = \overline{0,m}$.

Определение 5. Ядро $K(u) \in B$, если $K(u)$ абсолютно непрерывно на R^1, $K(u) = K(-u)$, $\int K(u) du = 1$, $K(u) \in L_1$; $\boldsymbol{K}_{\boldsymbol{q}}(\boldsymbol{u}) \in \boldsymbol{B}$, если $K_{q_i}(u_i) \in B$, $i = \overline{1,m}$; $B^{+} \equiv K \cap L_1$; $\boldsymbol{B}^{+} \equiv \boldsymbol{B} \cap \boldsymbol{L}_1$.

При написании данной работы везде далее рассматриваются стационарные последовательности с.в. из класса Seq_p, коэффициент РСП $\varphi(\tau)$ которых с ростом τ убывает не медленнее степенной функции (см. определение 6 и Замечание 1).

Определение 6. Последовательность РСП с.в. $\{\boldsymbol{x}(j)\} \in Seq_p$, $p > 0$, если

$$\sum_{\tau=1}^{n} (1 - \tau/n)(\varphi_{\boldsymbol{x}}(\tau))^{1/p} = O(1).$$

Замечание 1 . Если $\varphi_{\boldsymbol{x}}(\tau) = c_1\tau^{-(1+\varepsilon)}$, $\varepsilon > 0$, то $\{\boldsymbol{x}(j)\} \in Seq_p$, где $p \in (0,\ 1+\varepsilon)$. Действительно, в этом случае $\sum_{\tau=1}^{n}(1-\tau/n)(\tau)^{-(1+\varepsilon)/p} = O\left((n)^{1-(1+\varepsilon)/p}\right) + O(1) = O(1)$, если $0 < p \le\ 1+\varepsilon$.

Подтверждением того, насколько широкое распространение находят последовательности РСП с.в. класса Seq_p в практических приложениях, служит следующий пример. Рассмотрим последовательность с.в. $\{x(j)\}$, где $x(j)$ представляет собой номер исхода в j-м испытании для однородной цепи Маркова с конечным числом состояний. Как показано ([10, с. 466–467]), «при достаточно широких условиях» последовательность $\{x(j)\}$ стационарна и удовлетворяет условию РСП с коэффициентом перемешивания $\varphi_x(\tau) \le K\rho^{\tau}$, $\rho < 1$. Следовательно, $\{x(j)\} \in Seq_p$. Отметим, что марковские процессы широко используются в различных практических приложениях: радиофизике, автоматике, телемеханике, ядерной физике [11,12], экономике [13] и др.

Рассмотрим последовательность с.в.

$$\xi_j^{(\boldsymbol{\alpha})} \equiv \boldsymbol{K}^{(\boldsymbol{\alpha})}\left(\frac{\boldsymbol{x}-\boldsymbol{x}(j)}{h}\right) - \boldsymbol{E}\boldsymbol{K}^{(\boldsymbol{\alpha})}\left(\frac{\boldsymbol{x}-\boldsymbol{x}(j)}{h}\right), \tag{4}$$

порождаемую выборкой X_n.

Введем неограниченно возрастающую числовую последовательность d_n с помощью равенства

$$\frac{1}{d_n} \equiv \sum_{i=0}^{s}\frac{1}{nh_i^{2\boldsymbol{\alpha}_i-\operatorname{sgn}(\boldsymbol{\alpha}_i)}}\ . \tag{5}$$

В соответствии с (4) и (5) положим

$$\beta_{\boldsymbol{\alpha}} = \lim_{n\to\infty}\frac{1}{n}\frac{1}{h^{\operatorname{sgn}(\boldsymbol{\alpha})}}\boldsymbol{D}\left(\sum_{j=1}^{n}\xi_j^{(\boldsymbol{\alpha})}\right) =$$

$$= \lim_{n\to\infty}\frac{1}{h^{\operatorname{sgn}(\boldsymbol{\alpha})}}\left\{\boldsymbol{E}\left(\xi_1^{(\boldsymbol{\alpha})}\right)^2 + 2\sum_{\tau=1}^{n-1}(1-\tau/n)\boldsymbol{E}\left(\xi_1^{(\boldsymbol{\alpha})}\xi_{1+\tau}^{(\boldsymbol{\alpha})}\right)\right\},$$

$$\sigma_{\boldsymbol{\alpha}} \equiv \lim_{n\to\infty}\beta_{\boldsymbol{\alpha}}\frac{d_n}{nh_i^{2\boldsymbol{\alpha}-\operatorname{sgn}(\boldsymbol{\alpha})}}\ .$$

Обозначим $\omega_{\boldsymbol{\mu},\boldsymbol{\alpha}} \equiv \int(\boldsymbol{u})^{\boldsymbol{\mu}-\boldsymbol{\alpha}}\boldsymbol{K}^{(1)}(\boldsymbol{u})d\boldsymbol{u}\frac{F^{(\boldsymbol{\mu})}(\boldsymbol{x})}{(\boldsymbol{\mu}-\mathbf{1})!}$.

2. Усеченные оценки функционалов

Рассмотрим функционал $H(\boldsymbol{\tau}(\boldsymbol{x}))$, где $\boldsymbol{\tau}(\boldsymbol{x}) \equiv (F^{(\mathbf{0})}(\boldsymbol{x}), F^{(\boldsymbol{\alpha}_1)}(\boldsymbol{x}),\ldots,F^{(\boldsymbol{\alpha}_s)}(\boldsymbol{x}))$. Положим

$$\boldsymbol{t}_n(\boldsymbol{x}) \equiv \left(t_{0n}(\boldsymbol{x}),\ldots,t_{sn}(\boldsymbol{x})\right),\ t_{in}(\boldsymbol{x}) \equiv F_n^{(\boldsymbol{\alpha}_i)}(\boldsymbol{x}),$$

$$\boldsymbol{\tau}_n(\boldsymbol{x}) \equiv \boldsymbol{E}\boldsymbol{t}_n(\boldsymbol{x}),\ \ \tau_{in}(\boldsymbol{x}) \equiv \boldsymbol{E}F_n^{(\boldsymbol{\alpha}_i)}(\vec{x})\ .$$

Без ограничения общности везде далее

$$\sum_{j=1}^{m}\alpha_{sj} > \sum_{j=1}^{m}\alpha_{ij}\text{ , если } i \neq s\ .$$

В некоторых случаях корректное использование прямого метода подстановки при оценивании $H(\boldsymbol{\tau}(\boldsymbol{x}))$ дает неудовлетворительный результат, так как наличие особенностей функции $H(\cdot)$ приводит к «выбросам» и несостоятельности получаемых таким способом оценок $H(\boldsymbol{t}_n)$.

Для избежания нежелательных эффектов такого сорта введем операцию усечения больших значений $H(\boldsymbol{t}_n)$:

$$G(\boldsymbol{t}_n) = \begin{cases} H(\boldsymbol{t}_n), & |H(\boldsymbol{t}_n)| < Cd_n^{\gamma}, \\ C\,\mathrm{sign}(H(\boldsymbol{t}_n))d_n^{\gamma}, & |H(\boldsymbol{t}_n)| \geq Cd_n^{\gamma}. \end{cases} \tag{6}$$

Здесь C,γ – положительные константы.

Определенная таким образом процедура усечения (6) задает класс $P_{C,\gamma}(H)$ зависящих от n функционалов, которые сходятся в среднеквадратичном к $H(\boldsymbol{\tau}(\boldsymbol{x}))$ при $n \to \infty$. Следующая теорема дает условия, при которых усеченная оценка $G(\boldsymbol{t}_n) \in P_{C,\theta}(H)$ сходится в среднеквадратичном к $H(\boldsymbol{\tau}(\boldsymbol{x}))$ при $n \to \infty$.

Теорема 1. Пусть элементы выборки X_n взяты из последовательности слабозависимых случайных величин $\{\boldsymbol{x}(j)\} \in Seq_p$, $p > 1$, причем $F \in N_{\boldsymbol{\mu}}^{+}(\boldsymbol{x})$, $\mu_k \geq \alpha_{ik} + 1$, $i = \overline{0,s}$; $k = \overline{1,m}$.

Пусть, кроме того,

1. $h_0 \in H_0$; $h_j \in H_{\alpha_j,\delta_j}$, $\delta_j = \dfrac{2\mu(\mu-1) - w_j(2w_j - 1) + w_s}{(2\mu-1)(\mu + w_j - 1))}$, $j \geq 1$;

2. $\lim\limits_{u\to-\infty} K_0(u) = 0$, $\lim\limits_{|u|\to\infty}\left(u^{\mu_i - \alpha_{ki} + j}K_{\alpha_{ki}}^{(j)}(u)\right) = 0$, $j = \overline{1,\alpha_{ki}-1}$, $k = \overline{1,s}$; $i = \overline{1,m}$; $\boldsymbol{K}_q^{(\mathbf{1})}(\boldsymbol{u}) \in \boldsymbol{B}$; $\boldsymbol{K}_q^{(\boldsymbol{1})}(\boldsymbol{u}) \in \boldsymbol{Z}_{\boldsymbol{\mu}-\boldsymbol{\alpha}_q,\mathbf{0}}$;

3. $H(\boldsymbol{z}) = H(z_0, z_1, \ldots, z_s)$ – вещественная функция, которая в некоторой окрестности точки $\boldsymbol{\tau}$ непрерывна и имеет непрерывные первые и вторые производные по аргументам $z_j \in R^1$, причем $\infty > H_s(\boldsymbol{\tau}) > 0$;

4. $G(\boldsymbol{t}_n) \in P_{C,\theta}(H)$, где

$$C = \max_{j=\overline{0,s}}\left\{\sup_{\boldsymbol{u}}\left|\boldsymbol{K}_j^{(\boldsymbol{\alpha}_j)}(\boldsymbol{u})\right|\right\}; \tag{7}$$

$$\theta = \frac{2\mu - 1}{2(\mu - w_s)}\ . \tag{8}$$

Тогда усеченная оценка $G(\boldsymbol{t}_n) \in P_{C,\theta}(H)$ сходится к $H_s(\boldsymbol{\tau})$ при $n \to \infty$ в среднеквадратичном. При этом оптимальные последовательности параметров размытости $\{h_j\}$ имеют вид

$$\begin{cases} h_0^* = \operatorname*{Argmin}_{h_0} u^2(G(\boldsymbol{t}_n)) \approx o\left(\dfrac{1}{n}\right)^{\frac{\mu-w_s}{\mu(2\mu-1)}}, \\ h_j^* = \operatorname*{Argmin}_{h_j} u^2(G(\boldsymbol{t}_n)) = \left(\dfrac{(2w_s-1)\mu!\sigma_s^2(\boldsymbol{x})}{2(\mu-w_s)\kappa_s^2 F^{(\mu)}(\boldsymbol{x})}\dfrac{1}{n}\right)^{\frac{\mu+w_s-1}{2\mu-1}} \times \\ \times\left(\dfrac{(2w_j-1)(\mu-w_s-1)!H_s(\boldsymbol{\tau})\kappa_s}{(2w_s-1)(\mu-w_j-1)!H_j(\boldsymbol{\tau})\kappa_j}\right)^{\frac{1}{\mu+w_j-1}}, \qquad j=\overline{1,s}. \end{cases} \tag{9}$$

а оптимальная СКО $u^2(G(\boldsymbol{t}_n))$ равна

$$u^2(G(\boldsymbol{t}_n)) \equiv E\left(G(\boldsymbol{t}_n)-H(\boldsymbol{\tau})\right)^2 \approx$$

$$\approx \left(\frac{1}{n}f(\boldsymbol{x})\right)^{\frac{2(\mu-w_s)}{2\mu-1}} \cdot R_{\mu,w_s} \cdot \left(\sigma_{\vec{r}_s}\right)^{\frac{(\mu-w_s)}{2\mu-1}} \cdot \left(\kappa_s\right)^{\frac{2(2w_s-1)}{2\mu-1}} \cdot H_s^2(\boldsymbol{\tau}) \cdot \left(F^{(\boldsymbol{\mu})}(\vec{x})\right)^{\frac{2(2w_s-1)}{2\mu-1}}, \tag{10}$$

где
$$R_{\mu,w_s} = \left(\frac{2\mu-1}{2(\mu-w_s)}\right) \cdot \left(((\mu-1)!)^2\,\frac{2w_s-1}{2(\mu-w_s)}\right)^{-\frac{2w_s-1}{2\mu-1}} -$$

не зависящая от ядерных функций константа,

$$\kappa_s = \int(\boldsymbol{u})^{\boldsymbol{\mu}-1} K_s^{(\boldsymbol{\alpha}_s)}(\boldsymbol{u})d\boldsymbol{u}\,; \quad \sigma_{\boldsymbol{\alpha}_s} = \int\left(K_s^{(\boldsymbol{\alpha}_s)}(\boldsymbol{u})\right)^2 d\boldsymbol{u} + \frac{\gamma_{\boldsymbol{\alpha}_s}}{f(\boldsymbol{x})}.$$

В (9) и (10) $\mu = \sum_{i=1}^{m}\mu_i$, $w_s = \sum_{i=1}^{m}\alpha_{si}$.

Следующая теорема устанавливает асимптотическую нормальность оценок усеченного типа.

Теорема 2. В условиях теоремы 1

$$\sqrt{d_n}\left(G\left(F_n^{(\boldsymbol{0})}(\boldsymbol{x}),\dots,F_n^{(\boldsymbol{\alpha}_s)}(\boldsymbol{x})\right) - G\left(F^{(\boldsymbol{0})}(\boldsymbol{x}),\dots,F^{(\boldsymbol{\alpha}_s)}(\boldsymbol{x})\right)\right) (\Rightarrow N_1(\lambda_n;\quad \rho),$$

где $\lambda_n = H_s(\boldsymbol{\tau})\sqrt{d_n} \cdot \int(\boldsymbol{u})^{\boldsymbol{\mu}-\boldsymbol{\alpha}} K^{(1)}(\boldsymbol{u})d\boldsymbol{u}\dfrac{F^{(\boldsymbol{\mu})}(\boldsymbol{u})}{(\boldsymbol{\mu}-\boldsymbol{1})!} \cdot \left(h_s^*\right)^{\mu-w_s}$; $\quad \rho = H_s^2(\boldsymbol{\tau})\sigma_{\boldsymbol{\alpha}_s}(\boldsymbol{x})$.

Теорема 1 является конструктивной и дает способ построения оценок $G(\boldsymbol{t}_n) \in P_{C,\theta}(H)$ функционала $H(\boldsymbol{\tau})$, сходящихся в среднеквадратичном к его истинному значению. Как следует из (10), асимптотическая СКО $u^2(G(\boldsymbol{t}_n))$ таких оценок мажорируется степенной функцией, которая при увеличении параметра гладкости μ ф.р. $F(\boldsymbol{x})$ неограниченно сближается с функцией $1/n$:

$$u^2(G(\boldsymbol{t}_n)) \prec C \cdot R_{\mu,w_s}(n)^{-\frac{2(\mu-w_s)}{2\mu-1}}.$$

Здесь C не зависит от μ и s.

Известно, что скорость сходимости СКО $\approx 1/n$ имеет место в параметрической постановке задачи оценивания [1, 5]. Тот факт, что скорость сходимости

СКО предлагаемых усеченных непараметрических оценок при соблюдении некоторых условий регулярности (условия 2 и 3 теоремы 1) неограниченно сближается с нижней границей сходимости СКО в параметрической постановке, говорит о перспективности процедуры усечения (6).

3. Применение «усеченной» процедуры непараметрического оценивания функционалов от распределений стационарных последовательностей в экономических и технических приложениях

Предложенная в разделе 2 процедура усеченного непараметрического оценивания терминальных функционалов может быть реализована при решении различных задач экономики и техники. Одной из таких задач является задача распознавания пространственно распределенных объектов. Задачи распознавания пространственно распределенных объектов возникают в многих приложениях экономики, техники, метеорологии (лесных пожаров, посевных площадей, объектов загрязнения, затопления, тайфунов, ледниковых полей и т.д.). Описанная выше процедура оценивания в сочетании с процедурой функционального шкалирования [14] может быть использована для решения подобных задач.

Пусть каждому объекту $s \in S$ из множества S однозначно соответствует вектор количественных дискрипторов $\boldsymbol{x}_s \in \Im \subseteq R^n$; причем для любого $\boldsymbol{x} \in \Im$ существует вероятность $P_A(\boldsymbol{x})$ его принадлежности к классу A и вероятность $P_B(\boldsymbol{x}) = 1 - P_A(\boldsymbol{x})$ принадлежности к классу B.

Пусть $E = \{e_i \in S,\ i = \overline{1,n}\} \subseteq S$ – выборочное множество объектов, совокупность дескрипторов $\boldsymbol{x}(i) = \{\boldsymbol{x}_j(i), j = \overline{1,m}\}$ которых образует обучающую выборку $\{\boldsymbol{x}(n) \in R^m\}$ слабозависимых случайных величин с неизвестной функцией распределения $F(\boldsymbol{x})$; о каждом из объектов $e_i \in E$ сообщается, к какому классу он принадлежит. Требуется восстановить функцию достоверности $P_A(\boldsymbol{x})$.

Рассмотрим функцию

$$I(\vec{x}) = \begin{cases} 1, & \boldsymbol{x} \in A, \\ 0, & \boldsymbol{x} \in B, \end{cases}$$

которая в точке $\boldsymbol{x}$ равна 1 с вероятностью $P_A(\boldsymbol{x})$ и 0 – с вероятностью $P_B(\boldsymbol{x})$. Ее математическое ожидание равно $P_A(\boldsymbol{x})$, т.е. той функции, которую надо восстановить. Пусть $G(y,z) = y - z$ – «опорная» функция двух вещественных переменных, которая вместе с $I(\boldsymbol{x})$ определяет на исходном множестве S матричную структуру $Q = \| q_{ik} \|_{m \times m}$, элементы которой определяются следующим образом:

$$q_{ik} = G(I(\boldsymbol{x}(i)), I(\boldsymbol{x}(k))) = I(\boldsymbol{x}(i)) - I(\boldsymbol{x}(k)).$$

Предположим, что $I(\boldsymbol{x})$ есть информация о значениях $P_A(\boldsymbol{x})$, сообщаемая с помехой $\xi(\boldsymbol{x})$, математическое ожидание которой равно нулю:

$$I(\boldsymbol{x}) = P_A(\boldsymbol{x}) + \xi(\boldsymbol{x}), \quad E[I(\boldsymbol{x})] = P_A(\boldsymbol{x}).$$

Аппроксимируем искомую функцию $P_A(\boldsymbol{x})$ некоторой неизвестной параметризованной функцией-шкалой $\varphi(\boldsymbol{x},\boldsymbol{b})$ из заданного класса вещественных функций $\Phi=\{\varphi(\boldsymbol{x},\boldsymbol{b})\}$:

$$P_A(\boldsymbol{x})=\begin{cases} z=\varphi(\boldsymbol{x},\boldsymbol{b}), & 0<z<1, \\ 0, & z\le 0, \\ 1, & z\ge 1. \end{cases}$$

Здесь $\boldsymbol{b}\in R^p$ – неизвестный векторный параметр.

Точность шкалы $\varphi(\boldsymbol{x},\boldsymbol{b})$ определяется некоторым терминальным функционалом [14]

$$J(\varphi(\boldsymbol{x},\boldsymbol{b}))=J(\varphi(\boldsymbol{x},\boldsymbol{b}),F(\boldsymbol{x}),Q,D(P_A(\boldsymbol{x})),$$

ставящим в соответствие каждой шкале число, отражающее степень близости заданной структуры Q и матричной структуры $D(P_A(\boldsymbol{x}))=\| d_{ik}\|_{m\times m}$, порожденной шкалой $\varphi(\boldsymbol{x},\boldsymbol{b})$. Элементы матрицы $D(P_A(\boldsymbol{x}))$ равны $d_{ik}=P_A(\boldsymbol{x}(i))-P_A(\boldsymbol{x}(k))$.

Искомая функция достоверности при этом будет определяться оптимальной шкалой $\varphi^*(\boldsymbol{x})=\varphi(\boldsymbol{x},\boldsymbol{b}^*)$, для которой значения функционала $J(\varphi(\boldsymbol{x},\boldsymbol{b}))$ минимально, а искомые значения параметров шкалы при этом будут являться функционалами, зависящими от неизвестного распределения $F(\vec{x})$:

$$b_i^*=G_i(F(\boldsymbol{x}))=\underset{b_i}{\operatorname{Argmin}}\{J(\varphi(\boldsymbol{x},\boldsymbol{b}^*)\,\|\,b_i),F(\boldsymbol{x}),Q,D(P_A(\boldsymbol{x}))\},\quad i=\overline{1,p}\,.$$

Здесь $(\boldsymbol{b}^*\,\|\,b_i)=(b_1^*,b_2^*,\dots,b_{i-1}^*,b_i,b_{i+1}^*,\dots,b_p^*)$.

В качестве базовой оценки $F_n(\boldsymbol{x})$ плотности распределения $F(\boldsymbol{x})$ выборки X_n рассмотрим непараметрическую оценку ядерного типа (3)

Теорема 3. Пусть элементы выборки X_n взяты из последовательности слабозависимых случайных величин $\{\boldsymbol{x}(j)\}\in Seq_p$, $p>1$, причем $F\in N_{\boldsymbol{\mu}}^+(\boldsymbol{x})$, $\mu_k\ge\alpha_{ik}+1$, $i=\overline{0,s}$; $k=\overline{1,m}$. Пусть, кроме того, справедливы условия 1–3 теоремы 1. Тогда оценки $b_{in}=G_i(F_n(\boldsymbol{x}))\in P_{C,\gamma}(F)$ функционала $G_i(F(\boldsymbol{x}))$ асимптотически нормальны и сходятся в среднеквадратичном при $n\to\infty$, причем для среднеквадратичной ошибки справедливо

$$\boldsymbol{E}(b_i^*-G_i(F_n(\boldsymbol{x})))^2\cong O\left(\frac{1}{n}\right)^{2(\mu-1)/2\mu},\quad n\to\infty,\ i=\overline{1,p}\,.$$

Утверждение теоремы означает, что при определенных условиях регулярности, накладываемых на $F(\boldsymbol{x})$, скорость сходимости среднеквадратичного отклонения искомых параметров шкалы может быть как угодно близкой к $1/n$, где n – объем выборки X_n, что в итоге гарантирует хорошее качество распознавания.

Заключение

Предложена процедура, реализующая двухшаговую схему непараметрического оценивания функционалов от распределений стационарных последовательностей. На первом шаге с помощью операции усечения (6) исходному функционалу

(1) ставится в соответствие функционал $G(F(\boldsymbol{x}),F^{(\alpha_1)}(\boldsymbol{x}),\ldots,F^{(\alpha_s)}(\boldsymbol{x}))$ из класса «усеченных» функционалов $G(\boldsymbol{t}_n)\in P_{C,\gamma}(H)$; на втором – производится подстановка многомерной статистики $\boldsymbol{t}_n \equiv (F_n(\boldsymbol{x}),F_n^{(\alpha_1)}(\boldsymbol{x}),\ldots,F_n^{(\alpha_s)}(\boldsymbol{x}))$ на место $\boldsymbol{\tau}$ в функционал $G(\boldsymbol{\tau})$. Получены достаточные условия сходимости полученных оценок в среднеквадратичном. Показана их асимптотическая нормальность.

Наиболее важной характеристикой предложенной процедуры оценивания является ее универсальность, поскольку она не зависит от конкретного вида отображения $H: R^{s+1}\to R^1$ и распределения исходной выборки.

Рассмотрена возможность применения предложенной процедуры в экономических и технических приложениях. Предложена процедура распознавания пространственно распределенных стохастических объектов на основе функционального шкалирования [14] и непараметрического оценивания плотности распределения выборки дескрипторов наблюдаемых объектов [15]. Предполагается, что на исходном допустимом множестве объектов существует парное качественное отношение, определяющее его структуру как множества, состоящего из двух подмножеств-классов. В предлагаемой процедуре распознавания заданное внешнее качественное отношение аппроксимируется количественным, на основе которого производится построение оптимальной шкалы – интегрального агрегирующего показателя распознаваемых объектов, определяющей в итоге искомую процедуру распознавания.

ЛИТЕРАТУРА

1. *Васильев В.А., Добровидов А.В., Кошкин Г.М.* Непараметрическое оценивание функционалов от распределений случайных последовательностей. М.: Наука, 2004. 508 с.
2. *Рюмкин В.И.* Непараметрическое оценивание одного класса функционалов // Обозрение прикл. и промышл. матем. 2003. Т.10. Вып. 3. С. 735.
3. *Вопросы* математической теории надежности / под ред. Б.В. Гнеденко. М.: Радио и связь, 1983. 376 с.
4. *Гнеденко Б.В., Коваленко И.Н.* Введение в теорию массового обслуживания. М.: Наука, 1987. 336 с.
5. *Боровков А.А.* Математическая статистика. М.: Наука, 1984. 472 с.
6. *Кушнир А.Ф.* Асимптотически оптимальные критерии для регрессионной задачи проверки гипотез // Теория вероятностей и ее применения. 1968. Т. 13. Вып. 4. С. 682–700.
7. *Добровидов А.В.* Асимптотически ε-оптимальная непараметрическая процедура нелинейной фильтрации стационарных последовательностей с неизвестными статистическими характеристиками // Автоматика и телемеханика. 1984. № 12. С. 40–49.
8. *Надарая Э.А.* Непараметрическое оценивание плотности вероятностей и кривой регрессии. Тбилиси: Изд-во Тбилисского университета, 1983. 192 с.
9. *Беленький В.З.* О понятии потенциала экономической системы // ЭНСР. 2006. № 1(32). С. 17–28.
10. *Ибрагимов И.А., Линник Ю.В.* Независимые и стационарно связанные величины. М.: Наука, 1965. 524 с.
11. *Кошкин Г.М., Рюмкин В.И.* Оптимальные непараметрические алгоритмы выделения сигналов на фоне стационарной слабозависимой помехи // Математическое и программное обеспечение анализа данных. Минск, 1990. С. 114.
12. *Тихонов В.И., Миронов М.А.* Марковские процессы. М.: Советское радио, 1977. 488 с.
13. *Дынкин Е.Б., Юшкевич А.А.* Управляемые марковские процессы и их приложения. М.: Наука, 1977. 172 с.

14. *Авен П.О., Мучник И.Б., Ослон А.А.* Функциональное шкалирование. М.: Наука, 1988. 182 с.
15. *Рюмкин В.И.* Процедура функционального шкалирования пространственно распределенных объектов // Обозрение прикл. и промышл. матем. 2004. Т. 11. Вып. 4. С. 916–917.

Рюмкин Валерий Иванович
Томский государственный университет
E-mail: vir@mail.tomsknet.ru
Поступила в редакцию 12 мая 2012 г.

Rjumkin Valeriy I. (Tomsk state university). **Procedure of a nonparametric estimation of terminal type functionals and its application in economic and engineering applications.**

Keywords: nonparametric estimation, kernel functions, stationary samplings, functional.

The problem of an estimation of functionals of terminal type $H(\vec{\tau}(\vec{x})) \equiv H(F^{(\vec{0})}(\vec{x}), F^{(\vec{r}_1)}(\vec{x}),\ldots,F^{(\vec{r}_s)}(\vec{x}))$ on samples of stationary random quantities is considered. Here $H : R^{s+1} \to R^1$ is a known function, and $\vec{x} \in R^m -$ is fixed, on stationary sampling weekly dependent random quantities with an unknown distribution function $F(\vec{x})$.

The procedure implementing the two-step plan of a nonparametric estimation of the given functional is proposed. On the first step by means of truncation operation to an initial functional $G(F^{(\vec{0})}(\vec{x}), F^{(\vec{r}_1)}(\vec{x}),\ldots,F^{(\vec{r}_s)}(\vec{x}))$ puts in correspondence a functional from a class of "sectional" functionals $G(\vec{t}_n) \in P_{C,\gamma}(H)$; on second – statistics substitution $\vec{t}_n \equiv (F_n^{(\vec{0})}(\vec{x}), F_n^{(\vec{r}_1)}(\vec{x}),\ldots$ $\ldots, F_n^{(\vec{r}_s)}(\vec{x}))$ into place $\vec{\tau}$ in a functional $G(\vec{\tau})$ is produced. Convergence in mean square of estimates is proved. Velocity of convergence СКО of proposed sectional nonparametric estimates at observance of some regularity conditions superimposed on $H : R^{s+1} \to R^1$ and $F(\vec{x})$, beyond all bounds approaches with lower bound of convergence СКО in parametric setting.

The most important performance of the proposed procedure of an estimation is its scalability as it does not depend on a specific view of mapping and allocation of initial sampling. The given procedure can be used for the solution of various practical problems of economy and technics.

As an example the recognition problem spatially distributed stochastic objects on the basis of functional scaling and a nonparametric estimation of a density function of sampling of descriptors of observable objects is considered.