

О НЕПАРАМЕТРИЧЕСКОЙ ОЦЕНКЕ НЕСТАЦИОНАРНОЙ ФУНКЦИИ РЕГРЕССИИ ПО НАБЛЮДЕНИЯМ

Рассматривается проблема восстановления функции регрессии по наблюдениям со случайными ошибками. Предлагаются непараметрические оценки нестационарной функции регрессии. Приводятся результаты статистического моделирования непараметрических алгоритмов. Рассматривается задача моделирования статистических многомерных нестационарных систем с запаздыванием.

Ключевые слова: непараметрические оценки, дискретно-непрерывные процессы, нестационарность, регрессия, идентификация, моделирование.

В многочисленных приложениях часто возникает необходимость в восстановлении функции регрессии $M\{x|u\}$ по наблюдениям случайных величин (u_i, x_i) , $i = \overline{1, s}$, где s – объем выборки наблюдений. В классе непараметрических оценок принята статистика, которая будет представлена далее [1]. Предполагается, что плотность вероятности $p(u)$, $\forall u, u \in \Omega(u)$, $p(u) > 0$ неизвестна. При этом предполагается, что $x(u)$ не меняется с течением времени. Поскольку при моделировании дискретно-непрерывных процессов широко используются регрессионные модели, то существенное значение имеет факт дрейфа во времени характеристик исследуемого объекта. Среди многих факторов, влияющих на последнее, отметим только безусловное старение оборудования, других технических средств, где и протекает интересующий нас процесс. Однако укажем кратко и на другие причины, влияющие на дрейф характеристики процесса, а также отметим некоторые его проявления при моделировании:

– наличие неизвестных переменных либо переменных, не поддающихся измерению (естественно, что для успешного решения задачи идентификации с присутствием неконтролируемых входных переменных процесса их значения должны меняться медленно; в противном случае, и тем более, когда значения таких переменных случайны и приводят к осциллирующему процессу, ожидать хороших моделей не приходится);

– изменение параметров модели объекта во времени (речь в данном случае идет не об изменении параметров модели, как это происходит в ряде рекуррентных алгоритмов идентификации, когда наблюдается сходимость параметров некоторым оценкам, а о дрейфе параметров, связанных именно с изменчивостью объекта идентификации);

– изменение структуры объекта во времени (редко встречающаяся на практике ситуация, тем не менее, интересная с точки зрения теории, побуждающая к построению более универсальных, адаптивных алгоритмов идентификации).

Приведем примеры изменения характеристики нестационарного процесса. Из соображений простоты рассмотрим одномерные статические объекты (рис. 1). На рис. 1, *a* и 1, *б* представлены истинные характеристики стационарного и нестационарного объекта. Характеристика последнего изменилась за некоторый

интервал времени $[t_1, t_2]$. Точки выборки, полученные с обоих объектов, показаны на рис. 1, *б* и 1, *г*. Объемы выборки наблюдений одинаковы, на выход объекта наложена одинаковая аддитивная центрированная помеха.

На рис. 1 приняты следующие обозначения: u – входная переменная процесса; x – выходная переменная. Приведенная иллюстрация, помимо разъяснения, о каких процессах идет речь, подводит к следующей задаче: необходимо определить, имеем ли мы дело со стационарным процессом и присутствуют ли помехи в каналах измерений, или имеющаяся выборка отражает поведение нестационарного процесса, а не просто несет в себе искажения, вносимые помехами измерений.

Далее рассмотрим непараметрические оценки функции регрессии по наблюдениям при неизвестном дрейфе изучаемого процесса.

Непараметрические оценки функции регрессии. Непараметрическая оценка функции регрессии по наблюдениям с ошибками $\{x_i, u_i^{(1)}, u_i^{(2)}, \dots, u_i^{(m)}\}$, $i = \overline{1, s}$, где m – размерность вектора входных переменных, имеет вид [1; 2]

$$x_s(u_1, \dots, u_m) = \sum_{i=1}^s x_i \left[\prod_{j=1}^m \Phi \left(\frac{u^j - u_i^j}{c_s^j} \right) \right] / \sum_{i=1}^s \prod_{j=1}^m \Phi \left(\frac{u^j - u_i^j}{c_s^j} \right), \quad (1)$$

где финитная колоколообразная функция $\Phi(\cdot)$ удовлетворяют некоторым условиям сходимости [2]:

$$\lim_{s \rightarrow \infty} \frac{1}{c_s} \Phi((u - u_i) / c_s) = \delta(u - u_i),$$

$$\frac{1}{c_s} \int_{\Omega(u)} \Phi((u - u_i) / c_s) du = 1, \quad (2)$$

где $\delta(u - u_i)$ – дельта-функция Дирака. Параметр размытости c_s удовлетворяет следующим условиям сходимости [1; 2]:

$$c_s > 0, \quad \lim_{s \rightarrow \infty} c_s = 0, \quad \lim_{s \rightarrow \infty} s \cdot c_s^m = \infty. \quad (3)$$

Асимптотические свойства непараметрической оценки (1) достаточно хорошо исследованы, в частности, сходимость оценки x_s доказана в среднеквадратичном и почти наверное [3; 4]. Однако эта непараметрическая оценка (1) непригодна для восстановления регрессионных характеристик нестационарного объекта.

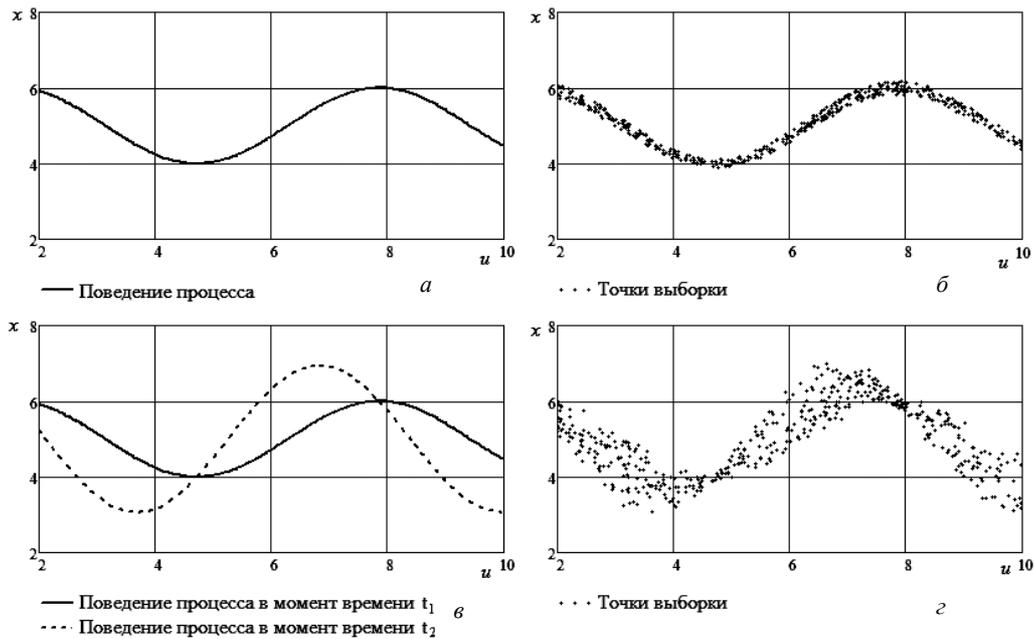


Рис. 1

Нестационарная непараметрическая оценка функции регрессии. Пусть истинная характеристика процесса изменяется во времени в пространстве входных и выходных переменных заранее непредвиденным образом. В этом случае использование всей выборки наблюдений, проводимых в оценке (1), будет искажать оценку функции регрессии, и обусловлено это будет тем, что наблюдения, ранее полученные на объекте, не отражают свойств объекта в настоящий момент времени. Естественно считать, что «старая» информация в виде элементов выборки $\{x_i, u^{(1)}_i, u^{(2)}_i, \dots, u^{(m)}_i, I = 1, 2, \dots\}$ должна использоваться с меньшим весом.

Идея непараметрического алгоритма, учитывающего временной дрейф характеристик исследуемого процесса, состоит в том, что «старая» информация в непараметрической оценке функции регрессии получает меньший вес [2].

Для восстановления функции регрессии в нестационарных условиях $x(u, t)$ предлагается ввести в непараметрическую оценку функции регрессии (1) некоторую функцию «памяти», убывающую с увеличением аргумента $\rho = (s-i)$, где i – дискретное время поступления информации, $I = 1, 2, \dots, s$; s – текущее время [2]. Тогда непараметрическая оценка нестационарной функции регрессии по наблюдениям (x, u) будет иметь вид

$$x_s^*(u_1, \dots, u_m) = \frac{\sum_{i=1}^s x_i \prod_{j=1}^m \Phi\left(\frac{u^j - u^j_i}{c_s^j}\right) \rho(s-i)}{\sum_{i=1}^s \prod_{j=1}^m \Phi\left(\frac{u^j - u^j_i}{c_s^j}\right) \rho(s-i)}, \quad (4)$$

где $\rho = (s-i)$ – функция «памяти», удовлетворяющая следующим свойствам:

$$1) i = s \Rightarrow \rho(s-i) = \rho(0) = 1;$$

$$2) \forall i, j \in [1, s], i > j \Rightarrow \rho(s-i) \geq \rho(s-j);$$

$$3) i = 1 \Rightarrow \rho(s-i) = \rho(s-1) \geq 0.$$

Приведем варианты вида функции памяти $\rho(s-i)$:

$$\rho(s-i, \alpha_1, \alpha_2) = \left[\frac{-(s-i) + \alpha_1}{|(s-i) - c_1| + \alpha_2} \cdot 0,5 + 0,5 \right] \alpha_3, \quad (5)$$

$$\alpha_3 = \frac{0,5 \cdot \alpha_1}{\alpha_1 + \alpha_2} + 0,5, \quad (6)$$

где параметр $\alpha_1 \in (0, s)$ определяет положение точки перегиба функции; параметр $\alpha_2 \in (0, \infty)$ определяет «крутизну» функции; параметр α_3 нормирует функцию «памяти» на единицу. Поведение функции (5) (функции «памяти») при различных значениях параметров α_1 и α_2 приведено на рис. 2.

Также функция «памяти» может быть кусочно-постоянной, приводя к скользящему аналогу непараметрической оценки (1).

Введем показатель изменчивости $\delta(t_1, t_2)$ характеристики исследуемого процесса в следующем виде:

$$\delta(t_1, t_2) = \int_{\Omega(u)} |x(u, t_1) - x(u, t_2)| du. \quad (7)$$

Обозначим дискретность измерения переменных (x, u) через Δt . Величина Δt , естественно, будет зависеть от показателя изменчивости $\delta(t_1, t_2)$, где $t_1 < t_2$. Можно считать, что при медленно меняющейся во времени характеристике процесса, т. е. при малых значениях $\delta(t_1, t_2)$, дискретность измерений Δt «входных-выходных» переменных процесса (x, u) больше, чем в случае больших значений $\delta(t_1, t_2)$. В последнем случае параметр Δt – дискретность измерения «входных-выходных» переменных, следует уменьшить. Анализ этого вопроса требует дополнительных исследований имеющейся выборки наблюдений $\{x_i, u^{(1)}_i, u^{(2)}_i, \dots, u^{(m)}_i, I = 1, 2, \dots\}$.

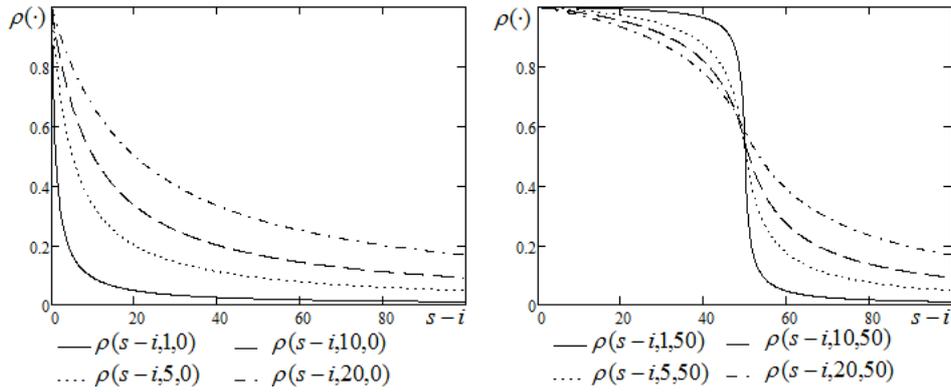


Рис. 2

Другой подход к попытке оценить изменчивость исследуемого процесса состоит в выявлении подобластей в пространстве входных и выходных переменных, которые могут отражать нестационарность процесса. Он основан на применении условной дисперсии $D\{x|u\} = M\{x^2|u\} - [M\{x|u\}]^2$ [5]. Условная дисперсия выходной переменной может быть оценена следующей непараметрической статистикой:

$$D_s(x|u) = \frac{\sum_{i=1}^s x_i^2 \left[\prod_{j=1}^m \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right) \right]}{\sum_{i=1}^s \prod_{j=1}^m \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right)} - \frac{\left[\frac{\sum_{i=1}^s x_i \left[\prod_{j=1}^m \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right) \right]}{\sum_{i=1}^s \prod_{j=1}^m \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right)} \right]^2}{\sum_{i=1}^s \prod_{j=1}^m \Phi\left(\frac{u^j - u_i^j}{c_s^j}\right)} \quad (8)$$

Большие значения оценки (8) (к примеру, выше некоторого порогового значения) могут сигнализировать об изменчивости объекта в рассматриваемых подобластях пространства наблюдений.

Нестационарные адаптивные модели и алгоритмы. Рассмотрим нестационарные адаптивные модели.

Распознавание образов. Оценка (4) с использованием функции «памяти», помимо применения в задаче оценки нестационарной функции регрессии, также может применяться и к непараметрическим алгоритмам классификации, а необходимость в них возникает, когда обучающая выборка накапливалась в течение некоторого времени, за которое расположение классов в признаковом пространстве изменилось, при этом классические алгоритмы классификации, не учитывающее изменение классов во времени, естественно, будут давать плохие результаты.

Рассмотрим случай, когда некоторая ситуация S может быть отнесена к одному из двух классов V_1 или V_2 и характеризуется значениями вектора признаков $v = (v_1, \dots, v_s)$. Задача классификации сводится к построению решающего правила на основании обучающей выборки $\{\bar{v}_s = (v_1, \dots, v_m), \bar{U}_s\}$ объемом s , где \bar{U}_s –

указание учителя о принадлежности S к классу V_1 или V_2 . В этом случае модификация непараметрического алгоритма классификации, учитывающая изменчивость классов, имеет вид

$$U_s(v_1, \dots, v_m) = \frac{\sum_{i=1}^s U_i \left[\prod_{j=1}^m \Phi\left(\frac{v^j - v_i^j}{c_s^j}\right) \right] \rho(s-i)}{\sum_{i=1}^s \prod_{j=1}^m \Phi\left(\frac{v^j - v_i^j}{c_s^j}\right) \rho(s-i)} \quad (9)$$

Управление. Аналогичную предыдущей модификацию можно предложить и для непараметрического алгоритма управления статическим нестационарным объектом:

$$u_s(x) = \frac{\sum_{i=1}^s u_i \Phi\left(\frac{x^* - x_i}{c_s}\right) \rho(s-i)}{\sum_{i=1}^s \Phi\left(\frac{x^* - u_i}{c_s}\right) \rho(s-i)} \quad (10)$$

где x^* – требуемое значение выхода объекта.

Статистическое моделирование нестационарных непараметрических оценок функции регрессии. При исследовании непараметрической оценки нестационарной функции регрессии (4), а также при сравнении результатов с использованием оценки (1) использовался метод статистического моделирования. В качестве уравнения, описывающего моделируемый объект, было выбрано следующее:

$$x(\bar{u}_i) = \sin u_i^{(1)} + \frac{i}{250} \cdot \cos(u_i^{(2)}) - u_i^{(3)} \cdot u_i^{(4)} + \sin(u_i^{(5)} + 0,5), \quad (11)$$

где $i \in [1, s]$, s – объем выборки; $u^{(1)} \in [0, 5]$, $u^{(2)} \in [0, 5]$, $u^{(3)} \in [0, 5]$, $u^{(4)} \in [0, 5]$, $u^{(5)} \in [0, 3]$. Значения входных переменных u генерировались случайно в соответствии с равномерным законом распределения в заданных интервалах.

Для имитации функционирования объекта в условиях помех различной интенсивности на имитируемый выход объекта (11) накладывается центрированная аддитивная помеха по следующему правилу:

$$x_i = x_i^* + \xi_i, \quad (12)$$

где x_i – выход объекта с наложенной на него помехой; x_i^* – выход объекта без помехи; ξ – случайная величина, подчиняющаяся нормальному закону распределения $N(0, \sigma)$, а значение среднеквадратичного отклонения σ определяется в зависимости от заданного уровня помехи e (%) согласно выражению

$$\sigma = \frac{e \left(\max(x(u | i = 0), x(u | i = s)) - \min(x(u | i = 0), x(u | i = s)) \right)}{6 \cdot 100 \%}, \quad (13)$$

в этом случае интервал 6σ будет соответствовать заданному в процентах интервалу от изменения выхода объекта.

Оценка качества моделирования проводилась с использованием следующего критерия:

$$W = \sqrt{\frac{1}{s} \cdot \sum_{l=1}^s (x(u_l) - x_s(u_l, c_s))^2}, \quad (14)$$

где $x(u_l)$ – значение выходной переменной в l -й точке выборки в момент времени s ; $x_s(u_l, c_s)$ – значение выхода модели (1), построенной по имеющейся выборке в той же точке. Аналогично будем проводить оценку качества моделирования при использовании непараметрической модели (4):

$$W' = \sqrt{\frac{1}{s} \cdot \sum_{l=1}^s (x(u_l) - x_s^*(u_l, c_s))^2}, \quad (15)$$

где $x_s^*(u_l, c_s)$ – значение выхода оценки (4).

Оценку ошибок моделирования по нескольким экспериментам, проводимым в одних и тех же условиях, будем проводить усреднением по серии экспериментов.

Результаты исследования качества моделирования в зависимости от уровня шума e (%) приведены в табл. 1, усреднение проводилось по 50 экспериментам при каждом уровне шума, объем выборки $s = 3\,000$.

Сравнение зависимости среднеквадратичных оценок ошибок моделирования (14) и (15) для стандартной (1) и модифицированной (4) непараметрических оценок функции регрессии от уровня шума e (%) приведено на рис. 3.

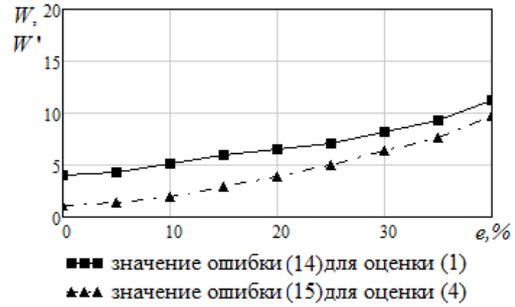


Рис. 3

Можно сделать вывод, что исследуемый алгоритм чувствителен к помехам, а также к точкам выбросов, ввиду того, что больший вес придается последним наблюдениям, даже если они оказываются ошибочными.

Приведем результаты моделирования при увеличении объема выборки за счет уменьшения шага дискретизации снятия наблюдений. То есть увеличим число наблюдений при условии, что последнее наблюдение будет происходить при одинаковых параметрах объекта (табл. 2). Усреднение проводилось по 50 экспериментам при каждом объеме выборки, на выход объекта наложена 5%-я помеха.

Графически сравнение зависимости среднеквадратичных ошибок моделирования (14) и (15), указанных в табл. 2, представлено на рис. 4.

Как видно из рис. 4, при уменьшении шага дискретизации снятия наблюдений по времени (а в данном случае именно за счет этого увеличивался объем выборки) заметна некоторая сходимость модифицированного алгоритма (4).

Увеличим объем наблюдений, оставив шаг дискретизации снятия наблюдений постоянным, сохранив исходный характер нестационарности (11). Объем выборки наблюдений s будет равен 3 000, 5 000, 7 000, 9 000, 12 000. Результаты приведены в табл. 3.

Сравнение зависимостей относительных ошибок моделирования (14) и (15) для стандартной и модифицированной непараметрических оценок функции регрессии по проведенной серии экспериментов представлено на рис. 5.

Таблица 1

Значение усредненных оценок ошибок моделирования в зависимости от уровня помех

$e, \%$	0	5	10	15	20	25	30	35	40
W	4,08683	4,35962	5,1858	5,97146	6,50943	7,11131	8,16638	9,34669	11,26575
W'	1,10454	1,42067	1,91831	2,9245	3,90674	4,93265	6,40269	7,70785	9,69883

Таблица 2

Значение усредненных оценок ошибок моделирования в зависимости от объема выборки

s	3 000	5 000	7 000	9 000	11 000	20 000
W	4,35962	4,35114	4,28764	4,25306	4,20386	4,18571
W'	1,42067	0,71623	0,30293	0,15271	0,14397	0,11291

Значение усредненных оценок ошибок моделирования в зависимости от объема выборки

S	3 000	5 000	7 000	9 000	12 000
$E_{\text{отн}}$	4,35962	5,65906	8,00766	11,24024	17,08449
$E^*_{\text{отн}}$	1,42067	1,56532	1,62685	1,7869	2,05778

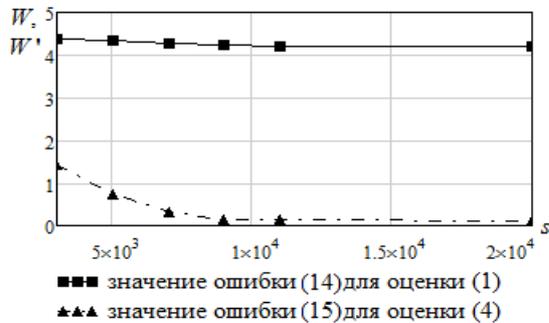


Рис. 4

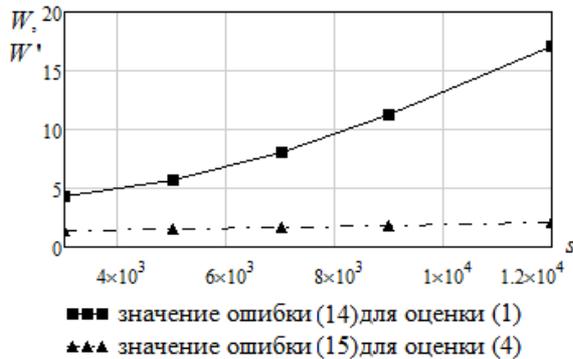


Рис. 5

Последний рисунок, отражает наиболее существенный момент, отмеченный ранее: непараметрическая оценка (1) непригодна для восстановления регрессионных характеристик нестационарного объекта.

В ходе численных исследований определено, что использование предлагаемой непараметрической оценки нестационарной функции регрессии дает более точные результаты оценивания, чем известная непараметрическая оценка, кроме случаев с большим

значением помехи. Рассматриваемый алгоритм более чувствителен к помехам.

Однако в любом случае адекватность получаемой модели зависит от реального объекта, скорости изменения характеристик объекта, частоты дискретизации снятия наблюдений, наличия априорной информации и качества получаемых данных.

При исчерпывающей выборке наблюдений позволительно использовать лишь ее часть для получения более точных оценок. Предлагаемый вид функции «памяти» рекомендуется использовать в тех ситуациях, когда невозможно определить вес каждого наблюдения конкретно. В противном случае, необходимо видоизменять функцию «памяти» в соответствии с имеющейся информацией.

Библиографические ссылки

1. Медведев А. В. Элементы теории непараметрических систем управления // Актуальные проблемы информатики, прикладной математики и механики. Ч. 3. Информатика. Новосибирск ; Красноярск : Изд-во СО РАН. 1996. С. 87–112.
2. Медведев А. В. Адаптация в условиях непараметрической неопределенности // Адаптивные системы и их приложения. Новосибирск : Наука, 1978. С. 4–34.
3. Кошкин Г. М., Пивен И. Г. Непараметрическая идентификация стохастических объектов : науч. пособие // Хабаровск : ДВО РАН, 2009. С. 336.
4. Васильев В. А., Добровидов А. В., Кошкин Г. М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей / отв. ред. Н. А. Кузнецов. М. : Наука, 2004. С. 508.
5. Дисперсионная идентификация / Н. С. Райбман [и др.] ; под ред. проф. Н. С. Райбмана. М. : Наука, 1981. С. 336.

A. V. Medvedev, A. V. Faustov

ABOUT NONPARAMETRIC ESTIMATION OF NONSTATIONARY REGRESSION FUNCTION ACCORDING OBSERVATIONS

A problem of backup of mathematical models of nonstationary processes in a non-parametric uncertainty is considered in the article. Estimations of nonstationary regression function are offered. Results of numerical simulation are given. A problem of static multi-dimensional nonstationary systems modeling with delay is considered as well.

Keywords: nonparametric estimations, discrete-continuous processes, nonstationarity, regression, identification, modeling.

© Медведев А. В., Фаустов А. В., 2010