

ПРОГНОЗИРОВАНИЕ В СЛУЧАЕ СПЛАЙНОВОЙ РЕГРЕССИИ ПРИ НЕИЗВЕСТНОМ РАСПОЛОЖЕНИИ УЗЛОВ

В. В. Казаченок

Белорусский государственный университет

Минск, Беларусь

E-mail: Kazachenok@bsu.by

Исследуется плотность распределения вероятностей прогнозируемых значений зависимой переменной при описании экспериментальных данных обобщенным сплайном m -го порядка дефекта k с известной дисперсией и неизвестным расположением узлов.

Ключевые слова: временной ряд, регрессия, обобщенный сплайн, байесовский подход.

Пусть наблюдения y_l в моменты времени t_l , $l = 1, \dots, n$ описываются сплайном m -го порядка дефекта k

$$y_l = \sum_{v=0}^m c_v t_l^v + \sum_{i=1}^{N-1} \sum_{j=0}^{k-1} a_{ij} (t_l - t_i^*)_+^{m-j} + u_l \quad (1)$$

с узлами в точках t_i^* , $i=1, \dots, N-1$.

Здесь c_v , $v = 0, \dots, m$, и a_{ij} , $i = 1, \dots, N-1$, $j = 0, \dots, k-1$, – неизвестные параметры; u_l , $l = 1, \dots, n$, независимые и нормально распределенные случайные величины с нулевым средним и дисперсией σ^2 ($0 < \sigma^2 < +\infty$); функция $(x)_+$ определяется следующим образом:

$$\begin{cases} x, \text{ при } x > 0, \\ (x)_+ = 0, \text{ при } x \leq 0. \end{cases}$$

Дефект k означает, что сплайн (1) имеет непрерывные производные во всех точках до $(m-k)$ -го порядка включительно.

Обобщенным сплайном m -го порядка дефекта k будем называть сплайн (1) при условии, что значения $(k+1)-x$ производных в узлах t_i^* , $i=1, \dots, N$ для соседних полиномов отличаются на величину v , где v , в общем случае, случайная величина с математическим ожиданием r и дисперсией Ψ^2 . Можно показать, что при $r=0$ и $\Psi^2 \rightarrow 0$ обобщенный сплайн дефекта k превращается в обычный сплайн дефекта $(k-1)$, а при $\Psi^2 \rightarrow \infty$ получаем обычный сплайн дефекта k , т. е. можно получить обобщенные сплайны дефекта k , приближающиеся к обычным сплайнам дефекта k или $(k-1)$ в зависимости от задаваемого значения $\frac{\sigma^2}{\Psi^2}$.

Таким образом, обобщенный сплайн дефекта k является промежуточным звеном между обычными сплайнами дефектов k и $(k-1)$, что позволяет строить сплай-

новые модели, адекватно описывающие результаты достаточно широкого класса экспериментов.

Оценивание методом наименьших квадратов неизвестных коэффициентов c_v , $v = 0, \dots, m$, и a_{ij} , $i=1, \dots, N-1$, $j = 0, \dots, k-1$, обобщенного сплайна дефекта k проводится в [1]. При неизвестном расположении узлов t_i^* , $i = 1, \dots, N-1$, их поиск можно осуществить минимизацией остаточной суммы квадратов также и по t_i^* , $i = 1, \dots, N-1$ [2].

Рассмотрим s будущих наблюдений (v_i, w_i) , $i = 1, \dots, s$, где $v_i = t_{n+i} - i$ -ое известное значение независимой переменной; $w_i = y_{n+i} - i$ -ое неизвестное значение зависимой переменной.

Пусть $\tau = \sigma^{-2}$, $w = (w_1, \dots, w_s)^T$, $v = (v_1, \dots, v_s)^T$, $\bar{c}_v = (c_0, c_1, \dots, c_m)$,

$$\bar{a}_{ij} = (a_{10}, \dots, a_{N-1,0}, a_{11}, \dots, a_{N-1,1}, \dots, a_{1,k-1}, \dots, a_{N-1,k-1}), \bar{t}^* = (t_1^*, \dots, t_{N-1}^*).$$

Плотность распределения вероятностей прогнозируемой выборки w при неизвестной дисперсии σ^2 исследовалась в [3].

Мы же далее будем предполагать, что дисперсия σ^2 задана, т. е. τ – известно. При этом расположение узлов t_i^* , $i = 1, \dots, N-1$ остается неизвестным.

Рассмотрим вначале несобственные априорные предположения для неизвестных параметров [4]:

1. Задана априорная плотность распределения вероятностей $p_0(\bar{t}^*)$.
2. Параметры \bar{c}_v и \bar{a}_{ij} равномерно распределены в R^{m+1} и $R^{k(N-1)}$ соответственно.
3. Параметры \bar{t}^* и $(\bar{c}_v, \bar{a}_{ij})$ являются независимыми.

Используя теорему Байеса и заданную выше априорную информацию, можно определить совместное априорное распределение неизвестных параметров $P(\bar{t}^*, \bar{c}_v, \bar{a}_{ij})$. Исключение из полученного выражения \bar{t}^* путем интегрирования дает апостериорную плотность вероятностей $P(\bar{c}_v, \bar{a}_{ij})$. Используя (1), можем выписать плотность распределения вероятностей $w = (w_1, \dots, w_s)^T$ при заданных \bar{c}_v и \bar{a}_{ij} , которую обозначим через $P(w|\bar{c}_v, \bar{a}_{ij})$. Тогда $P(w, \bar{c}_v, \bar{a}_{ij}) = P(w|\bar{c}_v, \bar{a}_{ij})P(\bar{c}_v, \bar{a}_{ij})$.

Интегрирование $P(w, \bar{c}_v, \bar{a}_{ij})$ по параметрам \bar{c}_v и \bar{a}_{ij} приводит к следующей теореме.

Теорема. Плотность распределения вероятностей прогнозируемой выборки представима взвешенной суммой нормальных распределений с вектором средних θ и положительно-определенной матрицей R

$$P(w) = \sum_{\bar{t}^*} K_{\bar{t}^*} N(w; \theta, R), w \in R^s,$$

где значения $K_{\bar{t}^*}$ представляют собой апостериорное дискретное распределение вероятностей \bar{t}^* , которое вычисляется подобно [5].

Явные выражения для $K_{\bar{t}^*}$ в некоторых частных случаях можно найти в [6].

Сформулированная теорема сохраняет силу и при собственных априорных предположениях для неизвестных параметров, наиболее часто используемых в подобных задачах [4]:

1. Задана априорная плотность распределения вероятностей $p_0(\vec{t}^*)$.
2. Условное распределение \bar{c}_v и \bar{a}_y при заданных τ и \vec{t}^* является нормальным с известным вектором средних и корреляционной матрицей $(\tau P)^{-1}$, где P – положительно-определенная матрица.
3. Параметры \vec{t}^* и (\bar{c}_v, \bar{a}_y) являются независимыми.

ЛИТЕРАТУРА

1. Казаченок В. В. Построение сплайновой регрессии по экспериментальным данным // Вестн. БГУ. Сер. 1. Физ. Мат. Информ. 1997. № 1. С. 70–71.
2. Medvedev G. A., Kazachenok V. V. Estimation of discontinuous regression function // Detection of Changes in Random Processes. New York: Optimization Software, 1987. Р. 126–130.
3. Казаченок В. В. Построение сплайновой регрессии в задачах прогнозирования // Теория вероятностей, математическая статистика и их применение: Материалы науч. конф. Минск: БГУ, 2004. С. 32–34.
4. Land M., Broemeling L. Bayesian forecasting with changing linear models // Commun. Statist., 1983. V. 12. N. 12. P. 1421–1430.
5. Казаченок В. В. Построение разрывных регрессионных моделей при стохастических ограничениях. Байесовский подход // Статистические проблемы управления. Вильнюс, 1988. Вып. 83. С. 181–186.
6. Казаченок В. В. Оценивание разрывной функции регрессии в задачах прогнозирования // Применение случайного поиска при решении прикладных задач: Тез. докл. 4-го рабочего совещания. Таштагол, 26–28 ноября 1986 г. Кемерово: КГУ, 1986. С. 138–140.