УДК 519.237.5

© А. Ю. Дереза

Дніпропетровський національний університет

УСТОЙЧИВЫЕ ОЦЕНКИ КВАЗИЛИНЕЙНОЙ СПЛАЙН-РЕГРЕССИИ С k УЗЛАМИ

Розглядається числова процедура пошуку стійких оцінок параметрів лінійної та квазілінійної сплайн-регресійної залежності із застосуванням динамічного програмування для пошуку вузлів склейки.

ВВЕДЕНИЕ

В задачах обработки данных испытаний изделий ракетно-космической техники важным является получение адекватных и достоверных моделей. Реализуемая в настоящее время вычислительная технология, основанная на классических методах, не всегда адекватна. Последнее относится к задачам обработки одномерных массивов (восстановлению функции распределения) и многомерных массивов (корреляционный и регрессионный анализ).

Выходом из создавшейся ситуации является реализация новых вычислительных технологий, основанных на сплайн-преобразованиях [1].

Целью работы является рассмотрение вычислительной процедуры нахождения устойчивых оценок параметров линейной и квазилинейной сплайн-регрессионной зависимости с применением динамического программирования для нахождения узлов склейки.

ПРОЦЕДУРА ОЦЕНИВАНИЯ РЕГРЕССИОННЫХ ЗАВИСИМОСТЕЙ

Пусть для оценки регрессионной зависимости заданы массивы данных:

$$\{x_i, y_i, i = 1, ..., n\}$$

для линейной сплайн-зависимости или

$$\{\varphi(x_i), \psi(y_i), \quad i = 1, ..., n\}$$

для квазилинейной сплайн-зависимости, где $y = \psi(y)$, $x = \varphi(x)$ — функции замены координат при преобразовании эмпирических функций, заданных нелинейно, к линейному виду. Например, для регрессионной зависимости вида ae^{bx} они имеют вид $\varphi(x) = x$, $\psi(y) = \ln y$.

Самым простым случаем сплайн-регрессии с $k - 1$ узлом склейки будет линейная регрессия следующего вида:

$$\bar{y}(x) = \begin{cases} a_1 + b_1 x, & x \in [c, x_0), \\ a_i + b_i x, & x \in [x_{i-2}, x_{i-1}),, \\ a_k + b_k x, & x \in [x_{k-2}, d], \end{cases}$$

$$i = 2, ..., k-1.$$

Учитывая условия неразрывности в точках склеивания:

$$a_i = a_{i-1} + (b_{i-1} - b_i)x_{i-2}, \qquad i = 1, ..., k-1,$$

функция регрессии примет вид

$$\bar{y}(x) =$$

$$= \begin{cases} a_1 + b_1 x, & x \in [c, x_0], \\ a_1 + \sum\limits_{j-2}^{i} (b_{j-1} - b_j)x_{j-2} + b_i x, & x \in [x_{i-2}, x_{i-1}], \\ a_1 + \sum\limits_{j-2}^{k} (b_{j-1} - b_j)x_{j-2} + b_k x, & x \in [x_{k-2}, d], \end{cases}$$

$$i = 2, ..., k-1.$$

что эквивалентно

$$\bar{y}(x) = \begin{cases} a_1 + b_1 x, & x \in [c, x_0], \\ \bar{y}(x_{i-2}) + b_i(x - x_{i-2}), & x \in [x_{i-2}, x_{i-1}], \\ \bar{y}(x_{k-2}) + b_k(x - x_{k-2}), & x \in [x_{k-2}, d], \end{cases}$$

$$i = 2, ..., k-1.$$

Таким образом, в классе линейных и квазилинейных функций вводится функция

$$\bar{y}(x) = \sum_{i=1}^{k} (\bar{y}(x_{i-2}) + b_i(x - x_{i-2}))I_i(\omega), \qquad i = 1, ..., k,$$

где

$$I_i(\omega) = \begin{cases} 1, \omega \in [x_{i-2}, x_{i-1}], \\ 0, \omega \neq [x_{i-2}, x_{i-1}], \end{cases}$$

$$x \in [x_{-1}, x_{k-1}], \quad \bar{y}(x_{-1} = a_1 + b_1 x_{-1}.$$

Оценки параметров сплайн-линейной регрессии найдем из условия минимума остаточной дисперсии:

$$\min_\theta S^2_{\text{ост}} = \min_\theta \frac{1}{n-k-1} \sum_{i=1}^{n} [y_i - \bar{y}(x_i, \bar{\theta})]^2 =$$

$$= \min_\theta \frac{1}{n-k-1} \sum_{i=1}^{n_1} [y_i - a_1 - b_1 x_i]^2 +$$

$$+ \sum_{j=2}^{k-1} \sum_{i=n_{j-1}}^{n_j} [y_i - \bar{y}(\tilde{x}_{j-2}) - b_j(x_i - \tilde{x}_{j-2})]^2,$$

где $\tilde{x}_0$ ... $\tilde{x}_0$ $\tilde{x}_{k-2}$ — фиксированные узлы склеивания, а n_j, $j = 1, \ldots, k-1$, определяются из условий

$$n_1 : x_{n1} \in [c, \tilde{x}_0] \,\&\, x_{n1+1} \in (\tilde{x}_0, \infty),$$

$$n_j : x_{n1} \in [\tilde{x}_{j-2}, \tilde{x}_{j-1}] \,\&\, x_{nj+1} \in (\tilde{x}_{j-1}, \infty),$$

$$j = 2, \ldots, k-2, \qquad n_{k-1} = n.$$

Пусть

$$\min_{\bar{\theta}} S^2_{\text{ост}} = \min_{\bar{\theta}} \frac{1}{n-k-1}\left(\Sigma_1 + \sum_{j=2}^{k-1} \Sigma_j\right),$$

тогда оценки параметров $\hat{a}_1$, $\hat{b}_1$ находят из решения системы $\partial\Sigma_1/\partial a_1 = 0$, $\partial\Sigma_1/\partial b_1 = 0$. Оценив $\hat{\bar{y}}(\tilde{x}_0) = \hat{a}_1 + \hat{b}_1\tilde{x}_0$, найдем $\hat{b}_2$ из равенства $\partial\Sigma_2/\partial b_2 = 0$. Далее последовательно найдем $\hat{b}_i$ из $\partial\Sigma_i/\partial b_i = 0$, $i = 1, \ldots k-1$.

Для нахождения оптимальных узлов склеивания предлагается процедура на основании динамического программирования. Критерием оптимальности выбора узлов является достижение минимума остаточной дисперсии после проведения оценки параметров регрессии.

Определим рекурентную функцию $F(k, n)$ так, чтобы величина $F(k, n)/(n-k-2)$ являлась минимальной остаточной дисперсией для сплайн-регрессии с k узлами склеивания, восстановленной по первым n точкам:

$$F(0, n) = \min_{a_1, b_1} \sum_{i=1}^{n} [y_i - a_1 - b_1 x_i]^2,$$

$$F(k, n) =$$

$$= \min_{t=1+2k,\, n-3} \left(F(k-1, t) + \min_{b_{k+1}} \sum_{i=t+1}^{n} [y_i - \bar{y}(x_i, \bar{\theta})]^2\right),$$

$$k > 0, \tag{1}$$

где $\bar{\theta}$ — оценки параметров, соответствующие вычислению $F(k-1, t)$.

Местоположение узлов определяется переменными t_i, при которых достигается минимум в правой части соотношения (1). А именно, $\tilde{x}_i = x_{ti}$.

Узел может быть уточнен путем изменения функции $F(k, n)$:

$$F(k, n) = \min_{t=1+2k,\, n-3} \Bigg(F(k-1, t) +$$

$$+ \min_{b_{k+1},\, t \le x_{k-1} < t+1} \sum_{i=t+1}^{n} [y_i - \bar{y}(x_i, \bar{\theta})]^2\Bigg), \quad k > 0, \tag{2}$$

В этом случае местоположение узлов определяется непосредственно переменными $\tilde{x}_i$, которые соответствуют минимуму правой части соотношения (2).

Заметим, что для линейной сплайн-регрессии нахождение минимума по b_{k+1} и $\tilde{x}_{k-1}$ может быть реализовано при помощи численных методов. Поскольку точность вычисления узла оказывает незначительное влияние на остаточную дисперсию, то допустимым является применение небольшого числа итераций метода координатного спуска.

Устойчивые оценки для сплайн-квазилинейной регрессии получают на основании робастной вычислительной схемы [1], описанной ниже.

1. Вычисление начальных значений оценок параметров $\bar{\theta}(0)$ из условия минимума функционала

$$\min_\theta S_0^2 = \min_\theta \frac{1}{n-k-1} \sum_{i=1}^{n} [\psi(y_i) - \bar{y}(\varphi(x_i), \bar{\theta}^{(0)})]^2 \psi_i,$$

где

$$\psi_i = \left(\frac{\partial\psi(y_i)}{y_i}\right)^{-2}.$$

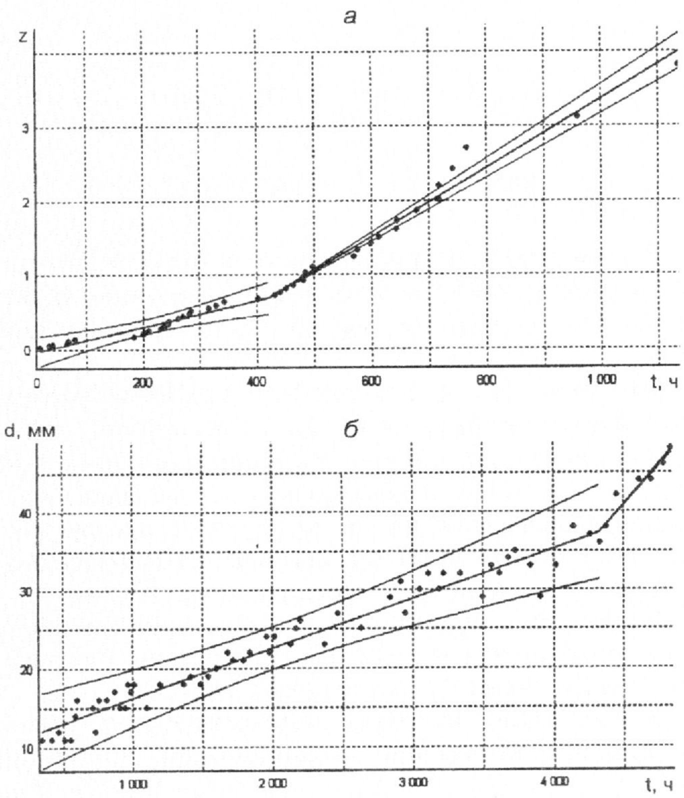
2. На последующих итерациях ($j = 1, 2, \ldots$) уточняется значение оценок параметров $\bar{\theta}^{(j)}$ поиском минимума функционала

$$\min_\theta S_j^2 =$$

$$= \min_\theta \frac{1}{n-k-1} \sum_{i=1}^{n} [\psi(y_i) - \bar{y}(\varphi(x_i), \bar{\theta}^{(j)})]^2 \psi_i^{(j)}.$$

3. Проверяется условие $S^2_{j-1} \le S^2_j$. Если последнее выполняется, то берут $S_0^2 = S^2_{j-1}$, и робастная процедура заканчивает свою работу, а искомые оценки параметров $\hat{\theta} = \theta^{(j-1)}$.

Приведенная вычислительная процедура получения устойчивых оценок сплайн-регрессионной зависимости нашла свою реализацию в системе автоматизированной обработки статистических данных ViStA 2.0. С ее помощью был обработан массив данных о времени отказа из условий эксплуатации космического аппарата 11ф619. На рисунке *а* график восстановленной регрессии $Z(t)$ наложен на корреляционное поле. Узел в точке 420 соответствует переходу космического аппарата в новое состояние, в котором вероятность отказа повышается.



Графики сплайн-регрессии для распределения времени на отказ космического аппарата 11ф619 (*а*) и для состояния соединения «втулка блока — плунжер» гидропривода трансмиссии ГСТ-90 (*б*)

Был обработан также массив данных динамики изменения зазора в сопряжении «втулка блока — плунжер» от наработки качающего узла гидронасоса гидропривода трансмиссии ГСТ-90. Восстановленная линейная сплайн-регрессия приведена на рисунке *б*. Узел в точке 4329 соответствует переходу системы в новое состояние. На основании проведенного анализа возможно дать рекомендацию о целесообразности проведения ремонтно-восстановительных работ через 4329 часов эксплуатации агрегата.

Анализ приведенных реализаций подтверждает работоспособность описываемой вычислительной процедуры. Полученные оценки являются более адекватными. Таким образом, робастная параметризация квазилинейной сплайн-регрессии позволяет получить статистические оценки, устойчивые относительно возможных отклонений реальной регрессии от теоретической. Достоверность результатов достигается как за счет «взвешивания» наблюдений, так и путем учета характера нелинейного преобразования модели.

ВЫВОДЫ

1. Предложенный метод предлагается использовать для обработки статистических данных о сроках активного существования как космических аппаратов в целом, так и их подсистем.
2. Вычислительная процедура была реализована в системе автоматизированной обработки статистических данных ViStA 2.0

1. Бабак В. П., Білецький А. Я., Приставка О. П., Приставка П. О. Статистична обробка даних. — Київ: МІВВЦ, 2001.— 388 с.
2. Минков Д. П., Митков А. Л., Тодоров Т. И. Робастный алгоритм построения одного класса нелинейных регрессионных моделей // Заводская лаборатория.—1985.—№ 5— С. 64—67.

ROBUST ESTIMATIONS OF QUASI-LINEAR SPLINE-REGRESSION WITH *k* NODES

A. Yu. Dereza

We consider a computational procedure for deriving of robust estimations of linear and quasi-linear spline-regression with the application of dynamic programming for finding nodes.