УДК. 519.24

А.А. ПАВЛОВ, д-р техн. наук, проф. каф. АСОИУ НТУУ «КПИ»,
А.В. ЧЕХОВСКИЙ, студент каф. АСОИУ НТУУ «КПИ»

ПОСТРОЕНИЕ МНОГОМЕРНОЙ ПОЛИНОМИАЛЬНОЙ РЕГРЕССИИ. АКТИВНЫЙ ЭКСПЕРИМЕНТ С ОГРАНИЧЕНИЯМИ

Обгрунтовується можливість зведення задачі побудови багатовимірної поліноміальної регресії к послідовності одновимірних регресійних задач в умовах обмеженого активного експерименту і вирішенню відповідних систем алгебраїчних рівнянь з постійними коефіцієнтами.

Обосновывается возможность сведения задачи построения многомерной полиномиальной регрессии к последовательности одномерных регрессионных задач в условиях ограниченного активного эксперимента и решению соответствующих систем алгебраических уравнений с постоянными коэффициентами.

In the article there has been substantiated the possibility of solving a multi-dimensional polynomial regression problem as a sequence of one-dimensional regression problems in a limited experiment and solving the relevant systems of algebraic equations with constant coefficients.

Проблема эффективного восстановления полиномиальной регрессии по настоящее время является актуальной [1-4]. Использование нормированных ортогональных полиномов Форсайта, с учётом рекуррентной формулы их построения [3] позволяет предложить эффективный метод построения многомерной полиномиальной регрессии. Он базируется на следующем анализе построения одномерной полиномиальной регрессии с помощью нормированных ортогональных полиномов Форсайта [3]:

Рассмотрим одномерную модель

$$Y(x) = \theta_0 + \theta_1 x + \ldots + \theta_r x^r + E\,, \qquad (1)$$

где x – детерминированная переменная, значение которой в экспериментах исследователь может задавать произвольно. $\theta_j, j=\overline{0,r}$ - неизвестные коэффициенты, r может быть избыточно большим, E – случайная величина с произвольным распределением, σ_E^2 неизвестна, либо существует её верхняя оценка σ^2. По результатам экспериментов $(x_i, y_i, i=\overline{1,n})$ нужно найти истинное значение r и оценить значение коэффициентов $\theta_j, j=\overline{0,r}$.

Были построены таблицы для оценки дисперсий $\theta_j, j=\overline{0,r}$ для случая, когда значения детерминированного аргумента $x_i, i=\overline{1,n}$ равномерно распределены внутри отрезка с концами $(-a,a), a>0$.

Фрагмент одной из них представлен таблицей 1.

Таблица 1

n	0	1	2	3	4	5
10	$\sigma^2 \cdot 0{,}4005$	$\sigma^2 \cdot 0{,}002$	$\sigma^2 \cdot 4{,}26 \cdot 10^{-06}$	$\sigma^2 \cdot 7{,}55 \cdot 10^{-09}$	$\sigma^2 \cdot 1{,}41 \cdot 10^{-12}$	$\sigma^2 \cdot 1{,}28 \cdot 10^{-15}$
50	$\sigma^2 \cdot 0{,}0706$	$\sigma^2 \cdot 0{,}0005$	$\sigma^2 \cdot 4{,}53 \cdot 10^{-07}$	$\sigma^2 \cdot 1{,}15 \cdot 10^{-09}$	$\sigma^2 \cdot 9{,}28 \cdot 10^{-14}$	$\sigma^2 \cdot 1{,}43 \cdot 10^{-16}$
100	$\sigma^2 \cdot 0{,}0352$	$\sigma^2 \cdot 0{,}0002$	$\sigma^2 \cdot 2{,}22 \cdot 10^{-07}$	$\sigma^2 \cdot 5{,}68 \cdot 10^{-10}$	$\sigma^2 \cdot 4{,}47 \cdot 10^{-14}$	$\sigma^2 \cdot 7{,}02 \cdot 10^{-17}$
200	$\sigma^2 \cdot 0{,}0176$	$\sigma^2 \cdot 0{,}0001$	$\sigma^2 \cdot 1{,}10 \cdot 10^{-07}$	$\sigma^2 \cdot 2{,}84 \cdot 10^{-10}$	$\sigma^2 \cdot 2{,}21 \cdot 10^{-14}$	$\sigma^2 \cdot 3{,}50 \cdot 10^{-17}$
300	$\sigma^2 \cdot 0{,}0117$	$\sigma^2 \cdot 7{,}66 \cdot 10^{-05}$	$\sigma^2 \cdot 7{,}36 \cdot 10^{-08}$	$\sigma^2 \cdot 1{,}89 \cdot 10^{-10}$	$\sigma^2 \cdot 1{,}47 \cdot 10^{-14}$	$\sigma^2 \cdot 2{,}33 \cdot 10^{-17}$
500	$\sigma^2 \cdot 0{,}0070$	$\sigma^2 \cdot 4{,}59 \cdot 10^{-05}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-08}$	$\sigma^2 \cdot 1{,}13 \cdot 10^{-10}$	$\sigma^2 \cdot 8{,}82 \cdot 10^{-15}$	$\sigma^2 \cdot 1{,}40 \cdot 10^{-17}$
1000	$\sigma^2 \cdot 0{,}0035$	$\sigma^2 \cdot 2{,}30 \cdot 10^{-05}$	$\sigma^2 \cdot 2{,}21 \cdot 10^{-08}$	$\sigma^2 \cdot 5{,}67 \cdot 10^{-11}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-15}$	$\sigma^2 \cdot 6{,}99 \cdot 10^{-18}$
5000	$\sigma^2 \cdot 0{,}0007$	$\sigma^2 \cdot 4{,}59 \cdot 10^{-06}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-09}$	$\sigma^2 \cdot 1{,}13 \cdot 10^{-11}$	$\sigma^2 \cdot 8{,}82 \cdot 10^{-16}$	$\sigma^2 \cdot 1{,}40 \cdot 10^{-18}$
10000	$\sigma^2 \cdot 0{,}0004$	$\sigma^2 \cdot 2{,}30 \cdot 10^{-06}$	$\sigma^2 \cdot 2{,}21 \cdot 10^{-09}$	$\sigma^2 \cdot 5{,}67 \cdot 10^{-12}$	$\sigma^2 \cdot 4{,}41 \cdot 10^{-16}$	$\sigma^2 \cdot 6{,}99 \cdot 10^{-19}$

Значения $x_i, i=\overline{1,n}$ распределены с равным шагом по отрезку $(-50.0;50.0)$

Таблица построена для линии регрессии заданной полиномом пятого порядка. В первой колонке фиксируются различные значения n (количество значений детерминированного аргумента x). В колонках с номером $j(j=\overline{0,5})$ заданы дисперсии коэффициентов $\hat{\theta}_j, j=\overline{0,5}$, как функция σ^2 (σ^2 - это дисперсия E либо её верхняя оценка). Для построения таблицы были найдены все ортогональные полиномы $Q_j(x), j=\overline{0,5}$ (использовались рекуррентные формулы построения нормированных ортогональных полиномов [3], сводя модель (1) к эквивалентной

$$Y(x)=\sum_{j=0}^{r}\omega_j Q_j(x)+E, \tag{2}$$

где $Q_j(x), j=\overline{0,r}$ - j-й нормированный ортогональный полином), а также дисперсии $\hat{\theta}_j$, выраженные через σ_E^2 (либо σ^2).

Приведем качественный анализ таблицы 1. На качественном уровне этот анализ не зависит от величины $a>0$ отрезка разбиения (-*a*, *a*) и величин r – степени полинома [1,2]. Этот результат подтвержден экспериментально.

1. Приведенные значения дисперсий $\hat{\theta}_j, j=\overline{0,5}$ становятся конструктивными, если известна верхняя оценка σ_E^2 дисперсии E. Порядок σ_E^2 можно определить по реализации случайной величины.

$$R^T R=\sum_{i=1}^{n}Y_i^1-\sum_{j=0}^{r}W_j^2,$$

так как

$$M\frac{R^T R}{n-(r+1)}=\sigma_E^2.$$

Далее будет показано, что истинное значение r находят очевидным образом.

2. Чем больше j тем меньше $\hat{\theta}_j$ при фиксированном n. Действительно, при $n=10$

$$D\hat{\theta}_0=\sigma^2\cdot 0.400466, D\hat{\theta}_1=\sigma^2\cdot 0.0024855,$$
$$D\hat{\theta}_2=\sigma^2\cdot 4.26\cdot 10^{-6},\ldots, D\hat{\theta}_5=\sigma^2\cdot 1.28\cdot 10^{-15}$$

т. е. с увеличением j значение $D\hat{\theta}_j$ уменьшается на порядок. Из этого вытекает следующий результат:

3. По минимальному количеству испытаний можно определить истинную степень полинома линии регрессии. В нашем примере при $n=10$ дисперсия коэффициента при $x^2\hat{\theta}_2$ уже равна $\sigma^2\cdot 4{,}26\cdot 10^{-6}$ Т.е. если истинная линия регрессии прямая, то реально оценками $\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_4,\hat{\theta}_5$ будут нули с точностью до соответствующих знаков после запятой (закон 3 сигм).

4. Необходимое количество испытаний n определяется заданной точностью для нахождения $\hat{\theta}_j$ с наименьшим $j(j=0)$ Если эксперименты являются дорогими, то реально эффективно оценивать $\hat{\theta}_j$ нужно начиная с $j=1$ (из анализа таблицы 1 видно, что значения дисперсий $D\hat{\theta}_0$ и $D\hat{\theta}_1$ одного порядка достигаются на числе экспериментов отличающихся на два порядка).

Таким образом, точность оценки θ_0 необходимо связывать с полученной числовой оценкой θ_0 (чем больше по модулю это значение, тем достовернее полученный результат). Если оценка θ_0 оказывается недостаточно точной, то полученное выражение для линии регрессии необходимо использовать в тех задачах, для решения которых величина θ_0 не имеет значения (например, сравнение значений линии регрессии для различных значений ее аргумента).

В некоторых задачах массив $x_i, i=\overline{1,n}$ может быть задан заранее и экспериментатор не может его изменить. Тогда до проведения эксперимента можно найти дисперсии $Q_j(x), j=\overline{0,r}$ (r можно задать избыточным) и провести предварительный анализ будущих результатов эксперимента.

При всей эффективности изложенного подхода могут иметь место случаи (например, в теории принятия решений), когда на вход объекта может подаваться последовательность $x_1, x_2, \ldots x_r$, которая может неоднократно повторяться.

Как показывают численные исследования, и в этом случае использование нормированных ортогональных полиномов приводит к эффективному решению задачи восстановления полиномиальной линии регрессии до степени $r-1$ включительно. Так, например, в таблице 2 (аналог таблицы 1) заданы дисперсии коэффициентов для случая полинома 5-й степени и повторяющейся серии, состоящей из 6 значений.

Таблица 2

n	0	1	2	3	4	5
10	$\sigma^2 \cdot 0{,}5$	$\sigma^2 \cdot 0{,}003$	$\sigma^2 \cdot 1{,}88 \cdot 10^{-05}$	$\sigma^2 \cdot 1{,}418 \cdot 10^{-08}$	$\sigma^2 \cdot 1{,}268 \cdot 10^{-11}$	$\sigma^2 \cdot 5{,}837 \cdot 10^{-15}$
50	$\sigma^2 \cdot 0{,}125$	$\sigma^2 \cdot 0{,}0006$	$\sigma^2 \cdot 3{,}977 \cdot 10^{-06}$	$\sigma^2 \cdot 3{,}467 \cdot 10^{-09}$	$\sigma^2 \cdot 2{,}547 \cdot 10^{-12}$	$\sigma^2 \cdot 1{,}308 \cdot 10^{-15}$
100	$\sigma^2 \cdot 0{,}059$	$\sigma^2 \cdot 0{,}0003$	$\sigma^2 \cdot 1{,}893 \cdot 10^{-06}$	$\sigma^2 \cdot 1{,}659 \cdot 10^{-09}$	$\sigma^2 \cdot 1{,}219 \cdot 10^{-12}$	$\sigma^2 \cdot 6{,}265 \cdot 10^{-16}$
200	$\sigma^2 \cdot 0{,}030$	$\sigma^2 \cdot 0{,}00016$	$\sigma^2 \cdot 9{,}641 \cdot 10^{-07}$	$\sigma^2 \cdot 8{,}506 \cdot 10^{-10}$	$\sigma^2 \cdot 6{,}183 \cdot 10^{-13}$	$\sigma^2 \cdot 3{,}197 \cdot 10^{-16}$
300	$\sigma^2 \cdot 0{,}02$	$\sigma^2 \cdot 0{,}0001$	$\sigma^2 \cdot 6{,}363 \cdot 10^{-07}$	$\sigma^2 \cdot 5{,}638 \cdot 10^{-10}$	$\sigma^2 \cdot 4{,}082 \cdot 10^{-13}$	$\sigma^2 \cdot 2{,}116 \cdot 10^{-16}$
500	$\sigma^2 \cdot 0{,}012$	$\sigma^2 \cdot 6{,}187 \cdot 10^{-05}$	$\sigma^2 \cdot 3{,}833 \cdot 10^{-07}$	$\sigma^2 \cdot 3{,}39 \cdot 10^{-10}$	$\sigma^2 \cdot 2{,}459 \cdot 10^{-13}$	$\sigma^2 \cdot 1{,}273 \cdot 10^{-16}$
1000	$\sigma^2 \cdot 0{,}006$	$\sigma^2 \cdot 3{,}08 \cdot 10^{-05}$	$\sigma^2 \cdot 1{,}907 \cdot 10^{-07}$	$\sigma^2 \cdot 1{,}688 \cdot 10^{-10}$	$\sigma^2 \cdot 1{,}224 \cdot 10^{-13}$	$\sigma^2 \cdot 6{,}340 \cdot 10^{-17}$
5000	$\sigma^2 \cdot 0{,}001$	$\sigma^2 \cdot 6{,}168 \cdot 10^{-06}$	$\sigma^2 \cdot 3{,}819 \cdot 10^{-08}$	$\sigma^2 \cdot 3{,}383 \cdot 10^{-11}$	$\sigma^2 \cdot 2{,}45 \cdot 10^{-14}$	$\sigma^2 \cdot 1{,}270 \cdot 10^{-17}$
10000	$\sigma^2 \cdot 0{,}0005$	$\sigma^2 \cdot 3{,}083 \cdot 10^{-06}$	$\sigma^2 \cdot 1{,}909 \cdot 10^{-08}$	$\sigma^2 \cdot 1{,}691 \cdot 10^{-11}$	$\sigma^2 \cdot 1{,}225 \cdot 10^{-14}$	$\sigma^2 \cdot 6{,}348 \cdot 10^{-18}$

Пример 1.

Истинная модель имеет вид $y(x) = 10 + 20x + 30x^2 + 20x^3 + 10x^4 + E.$

Регрессионная модель всегда должна задаваться избыточной. В нашем примере исследователь знает, что регрессионная модель является полиномом не выше пятой степени. Случайная E имеет нулевое математическое ожидание, нормальное распределение $\sigma_E = 50$.

6 чисел в повторяющейся серии входов x_i равномерно распределены по отрезку $(-50,50)$, начиная с -50 с шагом $\frac{100}{6}$.

Для генерации значений случайной величины E использована часть библиотеки расширений для C++ boost.

http://www.boost.org/doc/libs/1_36_0/libs/random/index.html

Результаты эксперимента:

Одномерная регрессия, нормальное распределение ($\sigma_E = 50$)

Исходные коэффициенты:

(10,20,30,20,10,0)

n=10

Ортогональные полиномы:

$$Q_0(x) = 0.316228$$
$$Q_1(x) = 0.180361 + 0.0120241x$$
$$Q_2(x) = -0.271626 + 0.0103967x + 0.000466447x^2$$
$$Q_3(x) = -0.294061 - 0.0239033x + 0.000616654x^2 + 2.15916 \cdot 10^{-5} x^3$$
$$Q_4(x) = 0.165413 - 0.0401629x - 0.00151391x^2 + 4.42556 \cdot 10^{-5} x^3 + 1.16256 \cdot 10^{-6} x^4$$
$$Q_5(x) = 0.424094 + 0.0242158x - 0.00398861x^2 - 0.000108399x^3 + 3.36646 \cdot 10^{-6} x^4 + 7.63981 \cdot 10^{-8} x^5$$

Оценки коэффициентов:

(25.7146,15.5441,30.0167,20.0079,9.99998,-2.59867·10-06)

Дисперсии коэффициентов:

(1250,7.55875,0.0469969,3.54375·10-05,3.17115·10-08,1.45917·10-11)

n=50

Ортогональные полиномы:

$$Q_0(x) = 0.141421$$
$$Q_1(x) = 0.0476467 + 0.00492897x$$
$$Q_2(x) = -0.152967 + 0.00357302x + 0.000204553x^2$$
$$Q_3(x) = -0.108869 - 0.010968x + 0.0002378x^2 + 9.25692 \cdot 10^{-6} x^3$$
$$Q_4(x) = 0.131529 - 0.0149956x - 0.000720587x^2 + 1.68945 \cdot 10^{-5} x^3 + 4.94883 \cdot 10^{-7} x^4$$
$$Q_5(x) = 0.224005 + 0.0160263x - 0.00183274x^2 - 5.56429 \cdot 10^{-5} x^3 + 1.51741 \cdot 10^{-6} x^4 + 3.61598 \cdot 10^{-8} x^5$$

Оценки коэффициентов:

(-2.14096,18.2325,30.0273,20.0024,9.99999,-7.8455·10-07)

Дисперсии коэффициентов:

(312.5,1.59767,0.00994141,8.66813·10-06,6.36862·10-09,3.26884·10-12)

В таблице 3 приведены оценки коэффициентов (точное значение которых равно 10,20,30,20,10,0 соответственно) для числа испытаний 10; 50; 100; 200; 300; 500; 1000; 5000; 10000. В наборе испытаний повторяется серия входных значений длиной 6 значений.

Таблица 3

	$\theta_0=10$	$\theta_1=20$	$\theta_2=30$	$\theta_3=20$	$\theta_4=10$	$\theta_5=0$
10	15.0406	24.1986	29.9586	19.9937	10.0001	$2.95043 \cdot 10^{-06}$
30	23.386	20.2097	29.9182	19.9962	10.0001	$2.37303 \cdot 10^{-06}$
50	30.578	19.6912	29.8692	19.9975	10.0001	$2.33215 \cdot 10^{-06}$

Продолжение табл. 3

60	24.1369	20.9136	29.9141	19.997	10.0001	$1.66418 \cdot 10^{-06}$
70	14.9404	20.2851	30.0798	20.0011	9.99993	$-1.2811 \cdot 10^{-06}$
80	2.98935	19.1196	29.977	20.0007	10	$-5.87023 \cdot 10^{-09}$
90	5.00373	18.4626	29.9589	20.0003	10.0001	$8.49835 \cdot 10^{-07}$
100	10.8082	22.1	29.9784	19.9969	10	$1.25764 \cdot 10^{-06}$
110	10.6553	21.8771	29.9152	19.9952	10.0001	$2.43644 \cdot 10^{-06}$
120	10.6901	19.5917	30.0119	20.0011	9.99999	$-5.49008 \cdot 10^{-07}$
150	10.588	20.1336	30.0241	20.0004	9.99998	$-3.7002 \cdot 10^{-07}$
200	-4.46675	19.2351	30.0359	20.0016	9.99998	$-5.67059 \cdot 10^{-07}$
300	4.80789	19.4225	30.0078	20.0006	10	$-1.31689 \cdot 10^{-07}$
500	9.80126	20.1821	29.9964	19.9997	10	$1.43024 \cdot 10^{-07}$

Рассмотрим многомерный случай.

Возможность для одномерного случая практически гарантировано находить степень полинома линии регрессии, вычислять с допустимой вероятностью с заданной погрешностью коэффициенты этого полинома позволяют предложить достаточно эффективную процедуру восстановления многомерной полиномиальной линии регрессии (при условии реализации активного эксперимента).

Пусть многомерная модель задаётся в виде

$$y(\bar{x}) = \sum_{\forall(i_1 \ldots i_t) \in K} \sum_{\forall(j_1 \ldots j_t) \in K(i_1 \ldots i_t)} b_{i_1 \ldots i_t}^{j_1 \ldots j_t} (x_{i_1})^{j_1} \cdot (x_{i_2})^{j_2} \ldots (x_{i_t})^{j_t} + E, \tag{3}$$

где $\bar{x} = (x_1 \ldots x_n)^T$ - детерминированный вектор входных переменных, x_i - i-я компонента вектора $\bar{x}$, $b_{i_1 \ldots i_t}^{j_1 \ldots j_t}$ - неизвестные коэффициенты, j_l, i_l - натуральные числа; E - случайная величина с нулевым математическим ожиданием и ограниченной неизвестной дисперсией σ_E^2 (как и в одномерном случае возможно известна верхняя оценка σ^2).

Модель (3) является избыточной – возможно некоторые из коэффициентов $b_{i_1 \ldots i_t}^{j_1 \ldots j_t}$ равны нулю. Для удобства дальнейшего изложения линию регрессии модели (3) представим иначе.

$$\sum_{l=1}^{n} \sum_{\forall(i_1 \ldots i_t) \in K_l} \sum_{\forall(j_1 \ldots j_t) \in K_l(i_1 \ldots i_t)} b_{i_1 \ldots i_t}^{j_1 \ldots j_t} (x_{i_1})^{j_1} \cdot (x_{i_2})^{j_2} \ldots (x_{i_t})^{j_t}. \tag{4}$$

Составляющие

$$\sum_{\forall(i_1 \ldots i_t) \in K_1} \sum_{\forall(j_1 \ldots j_t) \in K_1(i_1 \ldots i_t)} b_{i_1 \ldots i_t}^{j_1 \ldots j_t} (x_{i_1})^{j_1} \cdot (x_{i_2})^{j_2} \ldots (x_{i_t})^{j_t} \tag{5}$$

содержит все слагаемые из (3), в каждую из которых входит компонента x_1, а составляющие

$$\sum_{\forall(i_1 \ldots i_t) \in K_l} \sum_{\forall(j_1 \ldots j_t) \in K_l(i_1 \ldots i_t)} b_{i_1 \ldots i_t}^{j_1 \ldots j_t} (x_{i_1})^{j_1} \cdot (x_{i_2})^{j_2} \ldots (x_{i_t})^{j_t}, l = \overline{2, n} \tag{6}$$

содержат все слагаемые из (3), в каждую из которых входит компонента x_l, за исключением тех составляющих, которые вошли в (4) и (5) для $\forall(i_1, \ldots i_t) \in K_m \forall(j_1, \ldots j_t) \in K_m(i_1, \ldots i_t), m = \overline{1, l-1}.$

Рассмотрим составляющую (5).

Обозначим через $M_j^1, j = \overline{1, n_1}$ количество слагаемых, каждая из которых содержит x_1 в j-й степени.

$$M^1 = \max_j M_j^1, j = \overline{1, n_1},$$

n_1 – максимальная степень полинома от переменной x_1.

Фиксируем M^1 наборов значений компонент $x_2^s \ldots x_n^s$, $s = \overline{1, M^1}$ На числа $x_i^s, i = \overline{2, n}, s = \overline{1, M^1}$ накладывается единственное условие – определенные ниже квадратные матрицы должны быть невырожденными.

Реализуем M^1 экспериментов в каждом из которых (s-м, $s = \overline{1, M^1}$) переменные $x_2 \ldots x_n$ принимают фиксированные значения $x_i^s, i = \overline{2, n}$ а переменная x_1 изменяется повторяющимися сериями, как это было описано для случая построения одномерной регрессии.

При фиксированных значениях переменных $x_2 \ldots x_n$ в s-м эксперименте ($s = \overline{1, M^1}$) многомерная линия регрессии превращается в полином от переменной x_1 степени n_1.

Погрешность оценок точных значений коэффициентов линии регрессии находят либо с использованием закона трёх сигм (при этом предполагается, что верхняя оценка DE известна), либо эмпирическим способом, который заключается в следующем:

При последовательном вычислении коэффициентов линии регрессии, как функции числа экспериментов, добиваются монотонного увеличения числа постоянных разрядов после запятой в значениях оценок коэффициентов. Эти разряды и задают эмпирическую оценку точности вычислений коэффициентов линии регрессии.

По всем экспериментам ($s=\overline{1,M^1}$) оценки $\hat{\theta}_j^s, j=\overline{1,n_1}$ ранжируем по возрастанию их погрешностей при фиксированном j. Получим n_1 проранжированных последовательностей оценок коэффициентов $\theta_j^{s_1},...\theta_j^{s_{M^1}}$ ($j=\overline{1,n_1}$).

Эти результаты позволяют сформировать n_1 систем линейных уравнений, решениями которых являются значения всех коэффициентов $b_{i_1...i_t}^{j_1...j_t}$ в выражении (5), кроме коэффициента – константы.

Действительно в каждом из s экспериментов, неизвестные коэффициенты $\theta_j^s(j=\overline{1,n_1})$ одномерной полиномиальной регрессии степени n_1 от переменной x_1 определяются следующим образом: необходимо из всех членов выражения (содержащих переменную x_1 в степени j) вынести x_1^j. Полученное выражение для θ_j^s содержит только M_1^1 неизвестных коэффициентов вида $b_{i_1...i_t}^{j_1...j_t}$, т. к. в каждом s-м эксперименте при изменении значений переменной x_1 переменные $x_i, i=\overline{2,n}$ принимают одно и тоже фиксированное значение $x_i^s, i=\overline{2,n}$. Таким образом для построения системы линейных уравнений для нахождения M_1^1 коэффициентов вида $b_{i_1...i_t}^{j_1...j_t}$ надо использовать первые M_1^1 чисел $\theta_j^{s_1},...\theta_j^{s_{M^1}}$ (они имеют наименьшую дисперсию).

Для определения верхних оценок точности нахождения M_1^1 коэффициента вида $b_{i_1...i_t}^{j_1...j_t}$, полученную систему линейных уравнений условно запишем так:

$$A\begin{pmatrix} x_1 \\ \vdots \\ x_{M_1^1} \end{pmatrix}=\begin{pmatrix} \hat{\theta}_1^{s_1} \\ \vdots \\ \hat{\theta}_1^{s_{M_1^1}} \end{pmatrix}, \qquad (7)$$

где $x_l, l=\overline{1,M_1^1}$ - переменные (соответствующие M_1^1 переменным вида $b_{i_1...i_t}^{j_1...j_t}$).

Пусть погрешность оценок $\hat{\theta}_1^{s_l}, l=\overline{1,M_1^1}$ по модулю не превышает чисел $\Delta_1^{s_l}, l=\overline{1,M_1^1}$. Тогда максимальная (точная либо эмпирическая) величина

погрешности нахождения точных значений M_1^1 соответствующих коэффициентов вида $b_{i_1...i_t}^{j_1...j_t}$ имеет вид:

$$\max_{j=1,M_1^1}\{\max(\sum{}^{(+)}a_{jl}^{-1}\Delta_1^{s_l}, \sum{}^{(-)}\left|a_{jl}^{-1}\right|\Delta_1^{s_l})\}, \qquad (8)$$

где $\sum_l{}^{(+)}a_{jl}^{-1}\Delta_1^{s_l}$ берется по всем $l=\overline{1,M_1^1}$, для которых $a_{jl}^{-1}\geq 0$; $\sum_l{}^{(-)}a_{jl}^{-1}\Delta_1^{s_l}$ берется по всем $l=\overline{1,M_1^1}$ для которых $a_{jl}^{-1}<0$; a_{jl}^{-1} - jl-й элемент матрицы A^{-1}.

Как указывалось выше, предполагается, что $x_i^s, i=\overline{2,n}, s=\overline{1,M_1^1}$ выбраны так, что матрица A^{-1} существует.

Аналогично строятся все остальные системы линейных уравнений (правыми частями которых являются столбцы $\begin{pmatrix}\hat{\theta}_l^{s_1} & ... & \hat{\theta}_l^{s_{M_1^1}}\end{pmatrix}^T$, $l=\overline{2,n_1}$) для нахождения всех остальных коэффициентов $b_{i_1...i_t}^{j_1...j_t}$ из выражения (5). Аналогично строятся все оценки вида (8).

Процедуры нахождения всех неизвестных коэффициентов $b_{i_1...i_t}^{j_1...j_t}$ из выражений (6) для $l=\overline{2,n_1}$ полностью повторяют процедуру изложенную для выражения (5).

Оценка константы в выражении (3) может быть получена как среднее арифметическое по всем проведенным испытаниям разностей $y_i-(\hat{y}(\overline{x_i})-\theta_0)$, где y_i - значение выходной переменной модели, когда на вход подается векторное значение $\overline{x_i}$, а выражение $\hat{y}(\overline{x_i})-\theta_0$ - это значение выражения (4) для $\overline{x_i}$, из которой исключен коэффициент θ_0 и вместо коэффициентов $b_{i_1...i_t}^{j_1...j_t}$ подставлены полученные их оценки.

Если верхняя оценка σ_E^2 не известна, то её можно эффективно оценить как среднее арифметическое оценок σ_E^2 по всем одномерным регрессиям.

Обобщения.

Очевидно, что полученные результаты применимы для случая, когда в выражении (3) вместо переменных $x_1,...x_n$ поставить переменные $z_1,...z_m$ $m<n$, где $z_j=f_j(\overline{x_j})$, $j=\overline{1,m}$ где компонентами векторов $\overline{x_j}$ являются компоненты вектора $\overline{x}$, и множества компонент векторов $\overline{x_j}$, $j=\overline{1,m}$ не пересекаются. f_j - непрерывные функции, ограниченные при ограниченных значениях своих аргументов.

Задача построения многомерной регрессии очевидным образом обобщается на случай, когда при построении одномерных регрессий на модель действуют разные случайные величины E_l (l – номер одномерной регрессии). $ME_l = 0, DE_l = \delta^2_{E_l} < \infty$. В общем виде распределения случайных величин E_l (при фиксированном l) могут не совпадать между собой. Анализ формул показывает, что при построении одномерных регрессий в экспериментах на регрессионную модель аддитивно могут воздействовать независимые случайные величины E_l с различными распределениями, имеющие нулевые математические ожидания и одинаковые дисперсии, для фиксированного l. Для разных l дисперсии $\sigma^2_{E_l}$ могут быть различными.

Пример 2. (многомерная регрессия)

Исходная модель линии регрессии задаётся в виде следующего избыточного полинома:

$$y = 30 + 20x_1 + 25x_2 + 5x_3 + 10x_1x_2 + 7x_2^2 + 8x_1^2x_2 + 22x_1x_3 + 0x_2^2x_3^2 + 0x_1x_2x_3 + E. \quad (9)$$

Обозначим через $a_0, a_1, a_2, ..., a_9$ коэффициенты линии регрессии, которые считаются неизвестными.

E – случайная величина, имеющая нормальное распределение ME=0, σ^2_R=50.

В этом примере $K_1 = \{1;1,2;1,3;1,2,3\}$; $K_1(1) = 1$; $K_1(1,2) = \{1,1;2,1\}$; $K_1(1,3) = \{1,1\}$; $K_1(1,2,3) = \{1,1,1\}$. Аналогично определяются все $K_l, K_l(i_1,...i_t), l = \overline{2,3}$.

Для переменной x_1 последовательно фиксируются 4 пары значений x_2, x_3, и для каждой из них восстанавливается одномерная регрессия от переменной x_1, коэффициенты которой позволяют составить:

- систему из четырёх равенств для нахождения коэффициентов a_1, a_4, a_7, a_9 (коэффициенты в (9) при x_1 в первой степени);
- одно равенство для нахождения а6 (коэффициент в (9) при x1 во второй степени);

Для переменной x_2 последовательно фиксируются 2 пары значений x_1, x_3, и для каждой из них восстанавливается одномерная регрессия от переменной x_2. Составляется равенство для нахождения коэффициента a_2, а также система из двух уравнений для нахождения a_5, a_8. Затем для переменной x_3 фиксируется пара значений x_1, x_2, и для неё восстанавливается одномерная регрессия от переменной x_3, что позволяет найти оставшийся коэффициент a_3.

В таблице 4 приведены оценки точных коэффициентов многомерной регрессии, полученные для различного количества числа экспериментов n для

каждой одномерной регрессии, при этом на вход подавались повторяющиеся серии из 5 элементов.

Выводы.

В статье приведен конструктивный метод восстановления многомерной полиномиальной регрессии, представленной избыточным описанием, с использованием ограниченного активного эксперимента. Показано, что при использовании нормированных ортогональных полиномов Форсайта эту задачу можно свести к последовательности задач восстановления одномерных регрессий и решению систем линейных уравнений с постоянными коэффициентами. На основе анализа проведенных вычислительных экспериментов приведены конкретные практические рекомендации по использованию предложенного метода.

Таблица 4

Кол-во испы-таний	Исходные коэффициенты				
	30 (a_0)	20 (a_1)	25 (a_2)	5 (a_3)	10 (a_4)
	Оценки коэффициентов				
10	10.5772	12.4504	22.8376	3.16269	8.87862
50	13.2718	20.2274	25.3698	5.64317	9.66775
60	33.9547	20.7736	20.8223	4.46369	9.8577
70	-40.7656	18.5612	17.1159	0.55941	10.3737
80	20.0037	19.4024	24.9283	4.04724	9.96117
90	6.20488	19.3809	9.38889	5.13412	10.488
100	48.7494	19.2368	21.5281	3.38334	9.89669
110	37.6077	20.007	24.0356	5.50484	10.0555
120	30.6137	20.0776	25.4948	4.85337	10.0752
130	29.3955	19.8699	24.9731	3.72106	10.1493
140	31.2402	20.5324	24.725	4.83088	10.1734
150	32.9463	19.9379	25.1489	4.87948	10.0259
160	15.6351	19.5216	24.8536	4.52153	10.0786
170	72.9976	14.7554	25.2271	-1.5704	8.93467
180	32.0477	19.9483	25.0244	5.24743	10.0251
190	31.9689	20.3207	25.1918	5.2927	10.0511
200	34.8878	19.9577	25.1329	5.49405	9.97445
210	27.4535	19.8989	24.6718	4.60196	9.95048
220	26.0901	20.4691	22.2607	7.24466	10.1058
230	33.8137	19.5863	24.1966	4.25518	9.96735

Продолжение табл. 4

240	20.1789	18.7338	27.139	7.25653	10.142
250	29.0536	20.1413	25.0356	5.17554	9.97691
10	7.02319	7.9899	22.8982	-0.0014	0.1062
50	7.00446	7.96222	22.2583	0.0001	-0.4694
60	6.9918	7.99977	21.8531	$-1.6\cdot10^{-05}$	0.0395
70	5.47399	7.9944	22.038	0.0164	-0.0751
80	6.98585	7.99806	22.2087	0.0012	0.0015
90	7.01084	8.00375	21.7812	-0.0004	0.1810
100	7.01277	8.01916	22.1694	-0.0002	0.0476
110	7.00452	8.0045	22.0726	$4.3\cdot10^{-05}$	-0.0053
120	7.00057	8.001	22.0546	$-7.4\cdot10^{-05}$	0.0222
130	7.01671	8.00023	22.1026	-0.0004	-0.0032
140	6.99941	7.98936	21.9122	$2.8\cdot10^{-05}$	-0.0012
150	7.00474	8.00166	22.0428	$-2.9\cdot10^{-05}$	0.0123
160	6.99686	8.003	22.0597	0.0024	-0.0022
170	7.00643	7.99947	22.654	$-8.8\cdot10^{-05}$	0.1230
180	7.00282	7.99986	22.0078	-0.0027	0.0005
190	6.99492	7.99932	22.0145	0.0004	0.0005
200	7.03261	8.00014	21.9848	-0.0010	0.0175
210	6.99115	8.00185	21.9572	0.0006	-0.0203
220	6.99466	7.99958	21.9537	$5.7\cdot10^{-05}$	-0.0450
230	7.01846	8.00317	22.1897	-0.0005	0.0325
240	6.99426	7.99965	21.8917	0.0003	0.0179
250	7.00901	7.99995	22.0079	$4.7\cdot10^{-06}$	-0.0016

Список литературы: ***1***. *Павлов А.А., Чеховский А.В.* Сведение задачи построения многомерной регрессии к последовательности одномерных задач // Вісник НТУУ «КПІ» Інформатика, управління та обчислювальна техніка. – 2008 р. - №48. ***2***. *Павлов А.А., Чеховский А.В.* Построение многомерной полиномиальной регрессии (активный эксперимент). // Системні дослідження та інформаційні технології, Інститут прикладного системного аналізу НАН України та Міносвіти і науки України, №1 2009 р. (в печати). ***3***. *Д. Худсон.* Статистика для физиков. Москва, Мир, 1970. ***4***. *Радченко С.Г.* Устойчивые методы оценивания статистических моделей: Монография – К.: ПП «Санспарель», 2005. – 504 с.

Поступила в редколлегию 28.01.09