Prace Naukowe Instytutu Techniki Cieplnej i Mechaniki Płynów
Politechniki Wrocławskiej
Nr 31 Konferencje Nr 5 Nr 31 1988

Ewaryst KAFAJŁOWICZ*
Wojciech MYSZKA**

MULTIVARIATE REGRESSION - AN INTERPLAY BETWEEN COMPUTATIONAL ASPECTS AND EXPERIMENT DESIGN

The classical regression parameter estimation problem is considered. It is shown that for a multivariate regression large computational savings can be gained, if we impose certain, but not severe, restrictions on a regression function and on a experiment design. The proposed approach exploites properties of Kronecker's product of matrices.

1. INTRODUCTION

In this paper the classical regression parameters estimation problem is considered from a view-point of real-life computational abilities in conjunction with experiment design possibilities. An approach is proposed, which allows to reduce computational burden in estimating multivariable regression parameters by exploiting a special structure of the information matrix, resulting when a certain class of experiment designs can be applied.

Our results are motivated by the following consideratins. Suppose that estimated regression is of the form:

$$EY(x)=a^T f(x) \qquad (1.1)$$

where $x \in R^r$ is vector of independent variables, $f : R^r \longrightarrow R^m$ is a vector of known linearily independent functions, while $a \in R^m$ is a vector of unknown parameters to be estimated. The estimation is based on uncorrelated measurements $Y(x_1)$, $Y(x_2),\ldots,Y(x_N)$ performed at points $x_1,x_2,\ldots,x_N$. It is a standard piece of the regression theory (see e.g. [3]) that the least squares estimate (LSE) $\hat{a}$ of the vector $a \in R^m$ is the

* Technicalical University of Wroclaw, Institute of Engineering Cybernetics
** Technical University of Wroclaw, Institute of Material Sciences and Technical Mechanics

solution of the normal equations

$$\bar{M}\hat{a} = \bar{p} \tag{1.2}$$

where m x m information matrix

$$\bar{M} = \sum_{n=1}^{N} f(x_n) f^T(x_n) \tag{1.3}$$

while

$$\bar{p} = \sum_{n=1}^{N} f(x_n) \cdot Y(x_n) \tag{1.4}$$

Let us consider computational aspects of (1.2). In the multivariable case the most comman choice of components in f(x) is

$$h^{(i_1)}(x^{(1)}) h^{(i_2)}(x^{(2)}), \ldots, h^{(i_r)}(x^{(r)}); i_j = 1,2,\ldots,k \tag{1.5}$$

where $h^{(j)}$, $j=1,2,\ldots,k$ are chosen functions of one variable (e.g. polynomials of trigonometric functions), while $x^{(1)}, x^{(2)}, \ldots, x^{(r)}$ are components of x. Usually, all elements (1.5) enter into f(x), since there is no a priori knowledge, which of them are of importance. This leads to the necessity of solving (1.2) with $m = k^r$ equations.

The approach proposed in this paper is applicable to a large (but not arbitrary) class of f(x) and to a large (but also not arbitrary) class of experiment designs, leading to an essential reduction of computational burden. Restrictions imposed on f(x) and an experimental design are of a structural type only, as it will be shown in section 2. In section 3 we present a computational algorithm, while in section 4 its advanteges are discussed.

2. ASSUMPTIONS

We need a slight change of notation, namely by $x^{(i)}$, $i = 1,2,\ldots,r$ we denote r_i- dimensional subvector of the vector of independent variables $x \in R^r$ (not necesarilly one dimensional as in Section 1). We restrict our attention to a large class of the regression functions, wchich can be expressed as:

$$f(x) = \overline{\otimes}_{i=1}^{r} g_i(x^{(i)}) \tag{2.1}$$

where $g_i : R^{r_i} \to R^{m_i}$, $i=1,2,\ldots,r$ are vectors of the linearilly independent functions defined on a compact sets $X_i \subset R^{r_i}$. In (2.1) $\overline{\otimes}$ denotes r - fold Kronecker's product. We recall (see [2]) that for $l_1 \times l_2$ and $k_1 \times k_2$ matrices A, B their Kroneker's product $C = A \otimes B$ is $l_1 \cdot k_1 \times l_2 \cdot k_2$ matrix, which is composed from the blocks $a_{ij} \cdot B$, $i=1,2,\ldots,l_1$, $j=1,2,\ldots,l_2$. Thus, f(x) is $m = m_1 \cdot m_2 \cdot \ldots \cdot m_r$ - dimensional vector and

the typical element of f(x) is of the form

$$\prod_{i=1}^{r} \otimes \prod f_i^{j_i}(x^{(i)}), \quad j_i=1,2,\ldots,m_i \tag{2.2}$$

where $f_i^{(l)}(x^{(i)})$ is l-th component of the vector $f_i(x^{(i)})$.

Concerning measurements, we assume that they are made in the set $X=X_1 \times X_2 \times \ldots \times X_r$. In order to define an admissible experiment design (AED) we choose r sets of not necessarily distinct points $x_i=\{x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\}$, $x_j^{(i)} \in X_i$, $j=1,2,\ldots,n$, $i=1,2,\ldots,r$ and an admissible experiment design $x \subset X$ is defined as $x = x_1 \times x_2 \times \ldots \times x_r$. Thus a typical measurement point is of the form:

$$\left[x_{j_1}^{(1)}, x_{j_2}^{(2)}, \ldots, x_{j_r}^{(r)}\right], \quad x_{j_i}^{(i)} \in x_i \tag{2.3}$$

while their total number $N=n_1 \cdot n_2 \cdot \ldots \cdot n_r$. The class of all admissible designs is denoted by $\mathcal{X}$. It seems that $\mathcal{X}$ is sufficiently rich for many practical applications. In particular, designs from $\mathcal{X}$ are similar in structure to designs proposed in [1].

For compactness of further formulas it will be convenient to identify each experiment design $x_i \subset X_i$ with a certain discrete measure $\xi_i(x^{(i)})$ on X_i is attached by putting equal masses $p_j^{(i)} = 1/n_i$ to every point $x_j^{(i)}$, $j=1,2,\ldots,n_i$, $i=1,2,\ldots,r$. Then, the product measure $\xi(x)$ on X, defined by

$$\xi(dx)=\prod_{i=1}^{r} \xi_i(dx^{(i)}) \tag{2.4}$$

is attached to the corresponding set $x \in \mathcal{X}$. Measure (2.4) has as the set of support points with the equal masses attached to them. The set of all meadures (2.4) corresponding to $x \in \mathcal{X}$ will be denoted by Ξ.

3. COMPUTATIONAL ALGORITHM

The following lemma allows to recognize a structure of M.

Lemma 1

For every $\xi \in \Xi$ and f(x) of the form (2.1), information matrix M, is of the form:

$$M = \prod_{i=1}^{r} M_i \tag{3.1}$$

where $m_i \times m_i$ matrices M_i, $i=1,2,\ldots,r$ are defined by

$$M_i=\int_{X_i} f_i(x^{(i)}) f_i^T(x^{(i)}) \xi(dx^{(i)}) \tag{3.2}$$

Let as denote $\bar{x}^{(i)} = [x^{(1)}, x^{(2)}, \ldots, x^{(i)}]$, $i=1,2,\ldots,r$ with obvious identifications $\bar{x}^{(1)} = x^{(1)}$ and $\bar{x}^{(r)} = x$. In recursive relationships below, we shall also meet $\bar{x}^{(0)}$, what means - by convention - that a

vector function $Z_i(x^{(0)})$ is in fact a vector, which does not depend on $x^{(i)}$, $i=1,2,\ldots,r$. In such a case we shall write Z_i instead of $Z_i(x^{(0)})$.

Lemma 2

If all matrices M_i $i=1,2,\ldots,r$ are nonsingular, then the solution $\hat{a}$ of (1.2) can be obtained from the following recursive relationships:

$$Z^{(k)}(\bar{x}^{(k-1)}) = \int_{X_k} Z^{(k+1)}(\bar{x}^{(k)}) \, b^{(k)}(x^{(k)}) \, \mu(dx^{(k)}), \quad k=r,r-1,\ldots,1 \tag{3.3}$$

where

$$b^{(k)}(x^{(k)}) = M_k^{-1} \cdot g_k(x^{(k)}), \quad x^{(k)} \in X_k \tag{3.4}$$

$$Z^{(r+1)}(\bar{x}^{(r)}) = Y(x); \quad x = \bar{x}^{(1)} \in X \tag{3.5}$$

Then, iterating (3.6) we get $\hat{a} = Z_1$ for $k=1$.

Lemma 2 contains a prototype of our algorithm. It remains to express the above formulas in the terms of available data. To this end, let us note that at measurement points $(x_{j_1}^{(1)}, x_{j_2}^{(2)}, \ldots, x_{j_r}^{(r)})$, $j_i=1,2,\ldots,n_i$, $i=1,2,\ldots,r$ formula (3.3) reads as:

$$Z_{j_1 j_2 \ldots j_{k-1}}^{(k)} = \sum_{j_k=1}^{n_k} Z_{j_1 j_2 \ldots j_k}^{(k+1)} \cdot b_{j_k}^{(k)} \tag{3.7}$$

for $j_i=1,2,\ldots,n_i$ and $i=1,2,\ldots,k-1$, while $k=r,r-1,\ldots,1$. In (3.7) we introduced the following notations:

$$Z_{j_1 j_2 \ldots j_k}^{(k+1)} \hat{=} Z^{(k+1)}(x_{j_1}^{(1)}, x_{j_2}^{(2)}, \ldots, x_{j_k}^{(k)})$$

$$Z_{j_1 j_2 \ldots j_r}^{(r+1)} \hat{=} Y(x_{j_1}^{(1)}, \ldots, x_{j_r}^{(r)})$$

$$b_{j_k}^{(k)} \hat{=} b^{(k)}(x_{j_k}^{(k)}) = M_k^{-1} \cdot g_k(x_{j_k}^{(k)})$$

4. COMMENTS

Before discussing computational savings of the above algorithm in comparison with the ordinary least squares, some comments are in place.

1. The orthogonal regression and ordinary LSE can be imbedied into the proposed scheme, as the extreme special case.

2. Having a typical libray subroutine for multiresponse least squares estimation it is easy to implement our approach by nesting this subroutine into r-step loop of a programme.

3. One can notice that at k-th iteration of our algorithm we have a set of "partial" regression in our disposal. In this respect, our approach posseses a common feature with an identification procedure proposed in [1]. Detailed analysis of similarities and differences of these two approaches is outside the scope of this paper.

Comparing BPLSA and the classical least squares algorithm (CLSA),

given by (1.2)-(1.4), we can distinguish two sources of computational savings. The first of them comes from reduction in size of linear equations, which have to be solved in BPLSA and CLSA.

The second source of computational savings is in forming matrix M, and p in (1.2). Let us assume that the vectors $g_i(x_j^{(i)})$ $j_i = 1,3,\ldots,n_i$, $i=1,2,\ldots,r$ are given and $m_1=m_2=\ldots=\mu$ Then it can be shown that the ratio (number of operations in CLSA)/(number of operations in BPLSA)$\to\mu^r$ for $N\to\infty$. Thus, for a large sample size, reduction of computational burden is of the same order as the number of estimated parameters. This reductin increases when the number of estimated parameters increases. For exemple, if we have 4 independent variables $x^{(i)}\in R$ and third order polynomials are fitted ($\mu=4$), then the reduction is of order 256. We shall not prove the above facts about complexity, since proofs are easy but cumbersone.

5. REFERENCES

[1] Kasprzak W., Lysik B.; Dimensional Analysis in Experiment Design (in Polish), Ossolineum, Wrocław 1974
[2] Lankaster P.; Theory of matrices, Academic Press, New York, 1969
[3] Rao C. R.; Linear Statistical Inference and its Applications, Wiley, New York, 1965

МНОГОМЕРНАЯ РЕГРЕССИЯ – СВЯЗЬ МЕЖДУ ВЫЧИСЛИТЕЛЬНЫМИ АСПЕКТАМИ И ПЛАНИРОВАНИЕМ ЭКСПЕРИМЕНТОВ

В работе рассматриваются классические вопросы регрессионного оценивания параметров. Показывается, что для многомерной регресси достигается большая экономия вычислений. В предлагаемом подходе используются свойства произведения матриц Кронекера.