Aplikace matematiky

Marie Hušková; Jan Ámos Víšek; Dana Vorlíčková

Some remarks on experimental design

Aplikace matematiky, Vol. 31 (1986), No. 1, 19–26

Persistent URL: http://dml.cz/dmlcz/104181

31 (1986) APLIKACE MATEMATIKY No. 1, 19—26

SOME REMARKS ON EXPERIMENTAL DESIGN

Marie Hušková, Jan Ámos Víšek, Dana Vorlíčková

(Received November 21, 1984)

Summary. The paper deals with the experimental design which is optimal in the following sense: it satisfies the cost requirements simultaneously with a satisfactory precision of estimates. The underlying regression model is quadratic. The estimates of unknown parameters of the model are explicitly derived.

Keywords: experimental design, the model of linear and quadratic regression, estimation of parameters.

1. INTRODUCTION

During the cooperation of the Faculty of Mathematics and Physics of Charles University with the Institute of Nuclear Research the authors were to suggest an optimal design of an experiment the results of which could be described with the help of the model of linear or quadratic regression, and to find estimates of unknown parameters for a given regression model and experimental design. Another requirement was to obtain estimates as precise as possible under cost restrictions (i.e. restrictions on the number of observations). The so called D-optimal design (see [1], part IV.2.1 for a definition) satisfies the requirement of precision of estimates because it minimizes the generalized variance of the vector of least square estimates of the unknown parameters. This design is known in the case of linear regression (see e.g. [1]), and it satisfies the cost restrictions. The construction of the D-optimal design in the case of quadratic regression with more than foui regressors is too time-consuming (the time necessary for computing grows very rapidly with the increasing number of regressors), and moreover, the resulting design prescribes so many observations that they usually cannot be performed. Obviously, at the points of the D-optimal design which have the minimal probability it is necessary to carry out at least one measurement. Large differences in probabilities of distinct points of the D-optimal design lead to a great number of measurements required. The total numbers of measurements N for different numbers of regressors m are presented in the following table for approximately D-optimal designs, obtained after certain rounding-off.

m	2	3	4	5	6	7
N	12	43	145	491	1641	5323
m	8	9	10			
N	16673	50531	148905			

Therefore, we have suggested an experimental design which is not D-optimal but satisfies the cost requirement and at the same time exhibits a satisfactory precision of estimates.

The considered regression model as well as the suggested design are described in Section 2, moreover, the procedure of obtaining the design is introduced there. Estimates of parameters of this model were constructed explicitly, and they are the contents of Section 3.

2. OPTIMAL DESIGN FOR THE MODEL OF QUADRATIC REGRESSION

Denote the result of the i-th experiment by $y(\mathbf{x}_i)$, $\mathbf{x}'_i = (1, x_{i1}, \ldots, x_{i,k-1})$, $|x_{ij}| \leqq \leqq 1$, and suppose that

$$\text{(1)} \qquad \mathrm{E}y(\mathbf{x}_i) = \sum_{j=1}^{m} \beta_j f_j(\mathbf{x}_i), \qquad \operatorname{var} y(\mathbf{x}_i) = \sigma^2,$$

where $\beta_1, \ldots, \beta_m$, $\sigma^2 > 0$ are unknown parameters, and $f_j(\mathbf{x}_i)$, $j = 1, \ldots, m$, are components of a vector $\mathbf{f}(\mathbf{x})$, generated by all possible products of elements of the set $\{1, x_1, x_2, \ldots, x_{k-1}\}$, so that the vector $\mathbf{f}(\mathbf{x})$ has $\binom{k}{2}$ components. Denote by

$$\begin{pmatrix} \mathbf{x}_1, \ldots, \mathbf{x}_k \\ N_1, \ldots, N_k \end{pmatrix}$$

an exact design of the experiment. Let ξ^* stand for an optimal design if it satisfies

$$\text{(2)} \qquad \max_{x \in X} \mathrm{d}(\mathbf{x}, \xi^*) = m,$$

where

$$\mathrm{d}(\mathbf{x}, \xi) = \mathbf{f}'(\mathbf{x}) \mathbf{M}^{-1}(\xi) \mathbf{f}(\mathbf{x}),$$

$$\mathbf{M}(\xi) = \sum_{i=1}^{k} \mathbf{f}(\mathbf{x}_i) \mathbf{f}'(\mathbf{x}_i) \xi(\mathbf{x}_i),$$

$$\xi(\mathbf{x}_i) = \frac{N_i}{N}, \quad N = \sum_{i=1}^{k} N_i.$$

The design satisfying (2) cannot be found analytically, therefore, the programme for its derivation was created, based on the method of steepest descent (see [2], Chapter 2). However, when we started the iteration method from the minimal design ξ_0 under which the matrix $\mathbf{M}(\xi_0)$ is regular, and when we proceeded by the method of steepest descent the calculations for $m \geqq 5$ were too time-consuming. On the other hand, it was shown that the resulting design for $m = 2, 3, 4, 5$ is concentrated on the set

$$(3) \qquad \eta^* = \{\mathbf{x}: x_i = -1 \text{ or } x_i = 0 \text{ or } x_i = 1,\ i = 1, \ldots, k-1\}\,.$$

Probabilities of $\mathbf{x}$ from η^* have only two different values: the points from the set

$$(4) \qquad \eta^{**} = \{\mathbf{x}: x_i = -1 \text{ or } x_i = 1,\ i = 1, \ldots, k-1\}$$

have FCT-times larger probability than the points of $\eta^* - \eta^{**}$. Values of FCT for $m = 2, 3, \ldots, 5$ were derived experimentally. The approximation for FCT for an arbitrary dimension was found by fitting the values by a polynomial of the third degree:

$$(5) \qquad \mathrm{FCT}(n) = \tfrac{1}{6}n^3 - \tfrac{9}{8}n^2 + \tfrac{89}{24}n - \tfrac{5}{2}\,.$$

If the method of steepest descent with the starting design with probabilities calculated with the help of FCT(n) was used the time necessary for the evaluation of the optimal design was reduced substantially.

The resulting procedure for construction of the optimal design for the model (1) may be summarized in the following way:

the design (3) is chosen as starting provided the points of (4) have the probability equal to

$$\mathrm{FCT}/(3^{k-1} + (\mathrm{FCT}\text{-}1)\, 2^{k-1})\,,$$

the points of $\eta^* - \eta^{**}$ have the probability

$$(3^{k-1} + (\mathrm{FCT}\text{-}1)\, 2^{k-1})^{-1}\,,$$

where FCT is approximated by (5) with $n = k - 1$. The iterative procedure prescribed by the method of steepest descent is finished as soon as the design ξ_i is found for which $\max_{x \in \langle 0,1 \rangle} \mathrm{d}(\mathbf{x}, \xi_i) \leqq 1{\cdot}01\, m$. The points of the resulting design ξ_i, which are close to each other from the practical point of view, are associated and their probabilities are summarized.

3. ESTIMATION OF PARAMETERS IN THE MODEL OF QUADRATIC REGRESSION

Consider the regression model (1) which can be written in more detail as

$$(6) \qquad y_i(x_1, \ldots, x_k) = \alpha + \beta_1 x_1 + \ldots + \beta_k x_k + \gamma_{11} x_1^2 +$$
$$+ \gamma_{12} x_1 x_2 + \ldots + \gamma_{1l} x_1 x_l + \ldots + \gamma_{l-1,l} x_{l-1} x_l + \gamma_{ll} x_l^2 +$$

$$+ e_i(x_1, \ldots, x_k)\,, \quad i = 1, \ldots, n(x_1, \ldots, x_k)\,, \quad l \leqq k\,,$$

$$n(x_1, \ldots, x_k) = n_1\,, \quad \text{if} \quad x_i = 0\,, \quad 1 \leqq i \leqq k\,, \quad \text{exists}\,,$$

$$= n_2\,, \quad \text{otherwise}\,,$$

$\alpha, \beta_1, \ldots, \beta_k, \gamma_{11}, \gamma_{12}, \ldots, \gamma_{l-1}, \gamma_{ll}$ are unknown parameters ($m = 1 + k + l(l + 1)/2$), $x_1, \ldots, x_k$ are explanatory variables attaining the values

$$x_i = \begin{matrix} 1 \\ 0 \\ -1 \end{matrix} \quad \text{for} \quad i = 1, \ldots, l\,,$$

$$x_i = \begin{matrix} 1 \\ -1 \end{matrix} \quad \text{for} \quad i = l + 1, \ldots, k\,,$$

$e_i(x_1, \ldots, x_k)$, $i = 1, \ldots, n(x_1, \ldots, x_k)$ are independent random variables with the distribution $N(\mu, \sigma^2)$.

Recall that $n_2 \doteq n_1$ FCT(k), where FCT(k) is given by (5), holds for $x_1, \ldots, x_k$ corresponding to the optimal design found by the method of Section 2.

The model just described covers the case of the linear ($l = 0$) as well as the quadratic regression ($l = k$).

Now, we shall derive the least square estimates of the parameters $\alpha, \beta_1, \ldots, \beta_k, \gamma_{11}, \ldots, \gamma_{ij}, \ldots, \gamma_{ll}$. We shall use the following symbols:

$$n(x_1, \ldots, x_k) = n(x)\,,$$

$$y\,.\,(x_1, \ldots, x_k) = \sum_{i=1}^{n(x)} y_i(x_1, \ldots, x_k)\,,$$

$$y\,.\,(x_1, \ldots, x_{j-1}, .\,, x_{j+1}, \ldots, x_k) = \sum_{x_j} \sum_{i=1}^{n(x)} y_i(x_1, \ldots, x_j, \ldots, x_k)\,,$$

$$y\,.\,(\ldots) = \sum_{x_1} \ldots \sum_{x_k} \sum_{i=1}^{n(x)} y_i(x_1, \ldots, x_k)\,,$$

$$n(1) = \sum_{x_2} \ldots \sum_{x_k} n(1, x_2, \ldots, x_k)\,,$$

$$n(0) = \sum_{x_2} \ldots \sum_{x_k} n(0, x_2, \ldots, x_k)\,,$$

$$n(\ldots, x_j, \ldots) = \sum_{x_1} \ldots \sum_{x_{j-1}} \sum_{x_{j+1}} \ldots \sum_{x_k} n(x_1, \ldots, x_k)\,,$$

$$n = \sum_{x_1} \ldots \sum_{x_k} n(x_1, \ldots, x_k)\,.$$

The following relations obviously hold:

$$n = 2^k n_2 + 2^{k-l}(3^l - 2^l)\, n_1\,,$$

$$n(1, 1) = n(1, 1, \ldots) = n(1, -1, \ldots) = n(-1, 1, \ldots) =$$

$$= n(-1, -1, \ldots) = n(\ldots, \underset{j\text{-té}}{1}, \ldots, \underset{\nu\text{-té}}{1}, \ldots) =$$

$$= 2^{k-2} n_2 + 2^{k-l}(3^{l-2} - 2^{l-2})\, n_1\,, \quad 1 \leqq j < \nu \leqq l\,,$$

$$n(0) = n(\ldots, \underset{j\text{-té}}{0}, \ldots) = 2^{k-l}\, 3^{l-1} n_1\,, \quad 1 \leqq j \leqq l\,,$$

$$n(1) = n(-1, \ldots) = n(\ldots, \underset{j\text{-té}}{1}, \ldots) =$$

$$= 2^{k-1} n_2 + 2^{k-l}(3^{l-1} - 2^{l-1})\, n_1\,, \quad 1 \leqq j \leqq l\,,$$

$$n(\mathrm{k}) = n(\ldots, 1) = n(\ldots, -1) = n(\ldots, \underset{j\text{-té}}{1}, \ldots) =$$

$$= 2^{k-1} n_2 + 2^{k-l-1}(3^l - 2^l)\,, \quad l < j \leqq k\,.$$

Minimizing the sum of squares

$$S = \sum_{(x_1, \ldots, x_k)} \ldots \sum \sum_{i=1}^{n(x)} (y_i(x_1, \ldots, x_k) - \alpha - \beta_1 x_1 - \ldots - \beta_k x_k -$$
$$- \gamma_{11} x_1^2 - \gamma_{12} x_1 x_2 - \ldots - \gamma_{1l} x_1 x_l - \ldots - \gamma_{ll} x_l^2)^2$$

we obtain estimates $\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_k, \hat{\gamma}_{11}, \hat{\gamma}_{12}, \ldots, \hat{\gamma}_{ll}$. We successively solve the equations

$$\frac{\partial S}{\partial \gamma_{\mu\nu}} = 0\,, \quad 1 \leqq \mu \mp \nu \leqq l\,,$$

$$\frac{\partial S}{\partial \alpha} = 0\,,$$

$$\frac{\partial S}{\partial \beta_j} = 0\,, \quad 1 \leqq j \leqq k\,,$$

and then we have:

$$\hat{\alpha} = \frac{1}{n}\, y\,.(\ldots) - \frac{2\, n(1)}{n} \sum_{j=1}^{l} \hat{\gamma}_{jj}\,, \tag{7}$$

$$\hat{\gamma}_{\mu\nu} = \frac{1}{4n(1, 1)} (y\,.(\ldots, \underset{\mu\text{-té}}{1}, \ldots, \underset{\nu\text{-té}}{1}, \ldots) - y\,.(\ldots, 1, \ldots, -1, \ldots) - \tag{8}$$
$$- y\,.(\ldots, -1, \ldots, 1, \ldots) + y\,.(\ldots, -1, \ldots, -1, \ldots))\,, \quad 1 \leqq \mu < \nu \leqq l\,,$$

$$\hat{\beta}_j = \frac{1}{2n(1)} (y\,.(\ldots, \underset{j\text{-té}}{1}, \ldots) - y\,.(\ldots, -1, \ldots))\,, \quad 1 \leqq j \leqq k\,. \tag{9}$$

If we put

$$\frac{\partial S}{\partial \gamma_{jj}} = 0\,, \quad 1 \leqq j \leqq l\,,$$

we obtain the system of equations for $\gamma_{11}, \ldots, \gamma_{ll}$ in the form

$$\boldsymbol{A}\hat{\gamma} = \boldsymbol{b},$$

where

$$\boldsymbol{A} = \begin{pmatrix} 1 & a_{01}a_{11} & \cdots & a_{01}a_{11} \\ a_{01}a_{11} & 1 & & a_{01}a_{11} \\ \vdots & & \ddots & \vdots \\ a_{01}a_{11} & a_{01}a_{11} & & 1 \end{pmatrix},$$

$$\hat{\gamma} = (\hat{\gamma}_{11}, \hat{\gamma}_{22}, \ldots, \hat{\gamma}_{ll})',$$

$$\boldsymbol{b} = (b_1, \ldots, b_l)',$$

(10) $$b_j = a_{01}(y\,.\,(\ldots, \underset{j\text{-té}}{1}, \ldots) - y\,.\,(\ldots, \underset{j\text{-té}}{-1}, \ldots)) - a_{00}y\,.\,(\ldots), \quad 1 \leqq j \leqq l,$$

$$a_{01} = \frac{n}{2n(0)\,n(1)},$$

$$a_{11} = 4\left(n(1, 1) - \frac{n^2(1)}{n}\right),$$

$$a_{00} = 1/n(0).$$

The solution can be expressed as

$$\hat{\gamma} = \boldsymbol{A}^{-1}\boldsymbol{b},$$

and the elements a^{ij} of the matrix $\boldsymbol{A}^{-1}$ are

(11) $$a^{ij} = \frac{-a_{01}a_{11}}{(1 - a_{01}a_{11})(1 + (l - 1)\,a_{01}a_{11})}, \quad i \neq j,$$

(12) $$a^{ii} = \frac{1 + (l - 2)\,a_{01}a_{11}}{(1 - a_{01}a_{11})(1 + (l - 1)\,a_{01}a_{11})}.$$

Then,

(13) $$\hat{\gamma}_{jj} = \sum_{i=1}^{l} a^{ji}b_i, \quad j = 1, \ldots, l,$$

with b_i, a^{ji} given by (10), (11), (12).

It remains to estimate the parameter σ^2. The residual sum of squares S is evaluated according to the formula

(14) $$S = \sum_{(x_1,\ldots,x_k)} \!\!\cdots \sum \sum_{i=1}^{n(x)} (y_i(x_1, \ldots, x_k) - \frac{1}{n}\,y\,.\,(\ldots))^2 -$$

$$- 2n(1)\sum_{j=1}^{l} \hat{\beta}_j^2 - 2n(k)\sum_{j=l+1}^{k} \hat{\beta}_j^2 -$$

$$- 4n(1, 1)\sum_{1 \leqq \mu < \nu \leqq l}\sum \hat{\gamma}_{\mu\nu}^2 - 2\,\frac{n(0)\,n(1)}{n}\sum_{j=1}^{l} \hat{\gamma}_{jj}^2 -$$

$$- 8\left(n(1,1) - \frac{n^2(1)}{n} \right) \sum_{1 \leq \mu < \nu \leq l} \sum \hat{\gamma}_{\mu\mu} \hat{\gamma}_{\nu\nu} ,$$

where $\hat{\beta}_j$, $\hat{\gamma}_{\mu\nu}$, $\hat{\gamma}_{\nu\nu}$ are replaced by the expressions (9), (8), (13), respectively. Then we put

$$\hat{\sigma}^2 = \frac{1}{n - m} S_e , \tag{15}$$

with $m = 1 + k + l(l + 1)/2$.

4. Remark. The original aim of the authors was to publish here programmes for evaluating the design and the estimates of the unknown parameters of the model. However, they appeared too long for publication in a journal which does not specialize in computational methods. Readers who are interested in the programmes are invited to contact the authors directly.

References

[1] *A. Pázman:* Fundaments of Optimization of Experiment (in Slovak). Bratislava 1980.

[2] *V. V. Fedorov:* Theory of Optimal Experiment (Design of Regression Experiments) (in Russian). Moskva 1971.

[3] *P. F. Andrukovič, T. I. Golikova, S. G. Kostina:* Designs of second order on a hypercube whose properties are close to D-optimal (in Russian). In: New Ideas of Experimental Design. Nauka, Moskva 1969.

Souhrn

NĚKOLIK POZNÁMEK K PLÁNOVÁNÍ EXPERIMENTŮ

MARIE HUŠKOVÁ, JAN ÁMOS VÍŠEK, DANA VORLÍČKOVÁ

V článku je konstruován plán experimentu, který je optimální v tom smyslu, že splňuje podmínky na náklady experimentu a zároveň skýtá odhady s uspokojující přesností. Předpokládá se, že pozorování vyhovují modelu kvadratické regrese. Odhady neznámých parametrů modelu jsou podrobně odvozeny.

Резюме

НЕСКОЛЬКО ЗАМЕЧАНИЙ К ПЛАНИРОВАНИЮ ЭКСПЕРИМЕНТОВ

MARIE HUŠKOVÁ, JAN ÁMOS VÍŠEK, DANA VORLÍČKOVÁ

В работе построен план регрессионного эксперимента, который удовлетворяет требованию ограниченного количества измерений и одновременно дает довольно точные оценки неизвест-

ных параметров модели. Предполагается, что регрессия квадратическая. Добавление содержит вычисления оценок неизвестных параметров.

Authors' addresses: Dr. *Marie Hušková*, CSc., MFF UK Sokolovská 83, 186 00, Praha 8, Dr. *Jan Ámos Víšek*, CSc., ÚTIA ČSAV, Pod vodárenskou věží 4, 182 08 Praha 8, Dr. *Dana Vorlíčková*, CSc., MFF UK Sokolovská 83, 186 00 Praha 8.