ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIX, 1 (1987), pp. 109–116

M. MĘCZARSKI (Warszawa)

SEQUENTIAL ESTIMATION METHOD FOR THE MINIMUM OF A QUADRATIC REGRESSION FUNCTION

A quadratic regression model is considered. The existence of the minimum of a regression function is investigated. One constructs the confidence interval with prescribed length and asymptotic confidence level and tries to answer the question about the optimal allocation of experimental points.

1. Introduction. Consider a quadratic regression model

$$\mathrm{E}y(x) = a+bx+cx^2,$$

where $x \in R$, a, b, c are unknown parameters. Assume that we can observe a sequence of independent random variables according to this model

$$y_i = y(x_i) = a+bx_i+cx_i^2+\varepsilon_i, \quad i = 1, 2, \ldots,$$

where (ε_i) is a sequence of independent random variables with identical distribution function, $\mathrm{E}\varepsilon_i = 0$ and $\mathrm{E}\varepsilon_i^2 = \sigma^2$ (unknown), $\sigma^2 \in (0, +\infty)$.

For the existence of the minimum of the considered regression function we need $c > 0$. We do not assume it, so we verify if the value of the least squares estimate (LSE) of c is positive and removed from zero far enough. If not, we decide that the minimum does not exist. If yes, we pass to the second stage, the fixed-width estimation of $x_0 = -b/2c$, at which the function $\mathrm{E}y(x)$ attains its minimum. The confidence interval for x_0 is based on the sequence of observations (y_i); it has a prescribed length $2d$ and a prescribed confidence level, asymptotically as d tends to 0. The construction is sequential. Its nature as the estimation of a mean of the normal distribution and the result of Singh [6] yield that the fixed-sample version of the method is impossible to construct.

The idea of the solution is due to S. K. Perng and Y. L. Tong [5], who gave a sequential solution of the inverse regression problem.

2. Description of the method. Let p be a prescribed confidence level, $p \in (0, 1)$, $\alpha = \Phi^{-1}(p)$, $\beta = \Phi^{-1}((1+p)/2)$ with Φ being the distribution function of the normal $\mathcal{N}(0, 1)$-distribution.

Let $\hat{b}_n$, $\hat{c}_n$, $\hat{a}_n$ be the LSE's of the regression parameters b, c, a respectively, based on the first n observations. Let

$$s_n^2 = (n-3)^{-1} \sum_{i=1}^{n} (y_i - \hat{a}_n - \hat{b}_n x_i - \hat{c}_n x_i^2)^2$$

be the standard estimate of σ^2. Let

$$v_c^2(n) = \operatorname{var}(\hat{c}_n)/(\sigma^2/n), \qquad v_b^2(n) = \operatorname{var}(\hat{b}_n)/(\sigma^2/n).$$

Observe that $v_b^2(n)$ and $v_c^2(n)$ depend only on the sequence $(x_i)_{i=1,\ldots,n}$, i.e. on the plan of the experiment.

Let

$$X_n = \begin{bmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \\ x_1^2 & \ldots & x_n^2 \end{bmatrix}$$

be a $(3 \times n)$-matrix.

Assume

(A) $\qquad \lim_{n\to\infty} (1/n) X_n X_n^T = W, \quad$ a positive definite (3×3)-matrix.

From this

$$\lim_{n\to\infty} v_c^2(n) = v_c^2 > 0, \qquad \lim_{n\to\infty} v_b^2(n) = v_b^2 > 0,$$

where $v_b^2 = v_{22}$, $v_c^2 = v_{33}$ for $[v_{ij}]_{i,j=1,2,3} = W^{-1}$.

The method consists of two stages. In the first stage we check $c > 0$.

Let $d_1 > 0$ be a given number. We continue sampling until the sample size equals

$$N = \inf\{n \geqslant 4 \colon s_n^2 \leqslant (d_1^2 n)/(v_c^2(n)\alpha^2)\}.$$

If $\hat{c}_N < d_1$ we decide that c is not greater than zero and there is no minimum. If $\hat{c}_N \geqslant d_1$ we pass to the second stage, i.e. to the construction of the confidence interval.

Now we proceed sampling until the sample size equals

$$M = \inf\left\{m \geqslant N \colon \frac{s_m^2}{m}\left(v_b^2(m) + v_c^2(m)\,\hat{b}_m^2/\underline{\hat{c}}_m^2\right) \leqslant (4d_2^2\,\underline{\hat{c}}_m^2)/\beta^2\right\},$$

where $\underline{\hat{c}}_m = \max(d_1, \hat{c}_m)$. Then the interval

$$I_M = [-\hat{b}_M/(2\underline{\hat{c}}_M) - d_2,\ -\hat{b}_M/(2\underline{\hat{c}}_M) + d_2]$$

is accepted as the confidence interval for x_0. The confidence level is characterized by the probability of correct decision (CD) as follows:

$$P(\mathrm{CD}) = \begin{cases} P_c(\hat{c}_N \geqslant d_1,\ x_0 \in I_M) & \text{if} \quad c > 0, \\ P_c(\hat{c}_N < d_1) & \text{if} \quad c \leqslant 0. \end{cases}$$

Hence for $c > 0$ is $P(x_0 \in I_M) \geqslant P(\mathrm{CD})$.

3. The properties of the method. For the quadratic regression model, with LSE's of the regression coefficients, under condition (A), with the stopping rules N and M defined above, the following three theorems are true:

THEOREM 1.

(a) (i) $\lim_{d_1 \to 0} N = \infty$ *w.p.*1,

(ii) $\lim_{d_1 \to 0} (d_1^2 N)/(v_c^2 \sigma^2 \alpha^2) = 1$ *w.p.*1,

(b) (i) $\lim_{d_2 \to 0} M = \infty$ *w.p.*1,

(ii) $\lim_{d_2 \to 0} (4d_2^2 M)/(\beta^2 \sigma^2 (v_b^2/\underline{c}^2 + v_c^2 b^2/\underline{c}^4)) = 1$ *w.p.*1.

THEOREM 2.

(a) (i) $(\forall d_1 > 0)\ \mathrm{E}N < \infty,\ P(N < \infty) = 1,$

(ii) $\lim_{d_1 \to 0} (d_1^2 \mathrm{E}N)/(v_c^2 \sigma^2 \alpha^2) = 1,$

(b) (i) $(\forall d_2 > 0)\, \mathrm{E}M < \infty,\ P(M < \infty) = 1,$

(ii) $\lim_{d_2 \to 0} (4d_2^2 \mathrm{E}M)/(\beta^2 \sigma^2 (v_b^2/\underline{c}^2 + v_c^2 b^2/\underline{c}^4)) = 1.$

THEOREM 3. (i) *If* $c < 0$ *then* $\lim_{d_1 \to 0} P(\mathrm{CD}) = 1$.

(ii) *If* $c = 0$ *then* $\lim_{d_1 \to 0} P(\mathrm{CD}) = p$.

(iii) *If* $c > 0$ *then* $\lim_{d_1 \to 0} \lim_{d_2 \to 0} P(\mathrm{CD}) = p$.

4. Proofs. We need the following four lemmas:

LEMMA 1 (Chow and Robbins [1]). *Let* (z_n) *be any sequence of random variables such that* $z_n > 0$ *w.p.*1, $\lim_{n \to \infty} z_n = 1$ *w.p.*1. *Let* $f(n)$ *be any sequence of constants such that*

$$f(n) > 0, \qquad \lim_{n \to \infty} f(n) = \infty, \qquad \lim_{n \to \infty} f(n)/f(n-1) = 1$$

and for each $t > 0$ *define*

$$N = N(t) = \inf\{k \geqslant 1\colon\ z_k \leqslant f(k)/t\}.$$

Then N is well-defined and non-decreasing as a function of t, $\lim_{t\to\infty} N = \infty$, $\lim_{t\to\infty} \mathrm{E}N = \infty$ and $\lim_{t\to\infty} f(N)/t = 1$ w.p.1.

LEMMA 2 (Chow and Robbins [1]). *If the conditions of Lemma 1 hold, if $\lim_{n\to\infty} f(n)/n = 1$, if for N defined in Lemma 1 there is $\mathrm{E}N < \infty$, $\limsup_{t\to\infty} \mathrm{E}(Nz_N)/\mathrm{E}N \leqslant 1$ and if there exists a sequence of constants $g(n)$ such that $g(n) > 0$, $\lim_{n\to\infty} g(n) = 1$, $z_n \geqslant g(n) z_{n-1}$, then $\lim_{t\to\infty} \mathrm{E}N/t = 1$.*

Remark. Lemma 2 holds with

(i) $g(n)$ being a sequence of random variables such that $g(n) > 0$ w.p.1, $z_n \geqslant g(n) z_{n-1}$ w.p.1,

(ii) $\limsup_{t\to\infty} \mathrm{E}[(N-m_0) z_N]/\mathrm{E}N \leqslant 1$ for any fixed integer m_0.

((i) requires the Egorov theorem for a proof, (ii) is a slight modification.)

LEMMA 3 (Srivastava [7]). *If condition (A) holds for the linear q-dimensional regression model, i.e. W is a positive definite $(q+1)\times(q+1)$-matrix, if N satisfies the conclusion of Lemma 1 with $f(n) = n$, then*

$$\sigma^{-1}(X_N X_N^T)^{1/2}(\hat{\mu}_N - u) \xrightarrow[t\to\infty]{\mathscr{D}} \mathscr{N}(0, I_{q+1}),$$

where u is a $q+1$-dimensional regression parameter vector, $\hat{\mu}_n$ is the LSE of u, I_{q+1} is a unit $(q+1)\times(q+1)$-matrix.

LEMMA 4 (Drygas [2]). *Under the conditions of Lemma 3*

$$\lim_{n\to\infty} \hat{\mu}_n = u \text{ w.p.1.}$$

Proof of Theorem 1. From Lemma 4 we have strong consistency of $\hat{b}_k$ and $\hat{c}_k$, from Gleser [4] $\lim_{k\to\infty} s_k^2 = \sigma^2$ w.p.1. We use Lemma 1 with:

(a) $z_k = (s_k^2 v_c^2(k))/(\sigma^2 v_c^2)$, $f(k) = k$, $t = d_1^{-2} v_c^2 \alpha^2 \sigma^2$;

(b) $z_k = \dfrac{s_k^2 (v_b^2(k)/\underline{\hat{c}}_k^2 + v_c^2(k)\hat{b}_k^2/\underline{\hat{c}}_k^4)}{\sigma^2 (v_b^2/\underline{c}^2 + v_c^2 b^2/\underline{c}^4)}$, $f(k) = k$, $t = \beta^2\sigma^2(v_b^2/\underline{c}^2 + v_c^2 b^2/\underline{c}^4)/(4d_2^2)$.

Proof of Theorem 2. We use Lemma 2 with

(a) $g(n) = [(n-4) v_c^2(k)]/[(n-3) v_c^2(k-1)]$, see Remark (ii),

(b) $g(n) = \dfrac{n-4}{n-3} \cdot \dfrac{v_b^2(k)/\underline{\hat{c}}_k^2 + v_c^2(k)\hat{b}_k^2/\underline{\hat{c}}_k^4}{v_b^2(k-1)/\underline{\hat{c}}_{k-1}^2 + v_c^2(k-1)\hat{b}_{k-1}^2/\underline{\hat{c}}_{k-1}^4}$, see Remark (i), (ii).

The proof follows the lines of Chow and Robbins' proof of the asymptotic efficiency of their procedure [1] and is omitted.

Proof of Theorem 3. Lemma 3 implies that $(\hat{c}_N - c)/((\sigma/\sqrt{N}) v_c(N))$ is asymptotically $\mathscr{N}(0, 1)$-distributed as $d_1 \to 0$. Since $(d_1\sqrt{N})/(v_c(N)\sigma)$ tends to α w.p.1 as $d_1 \to 0$, we have for $c < 0$ the following:

$$\lim_{d_1 \to 0} P(\mathrm{CD}) = \lim_{d_1 \to 0} P\big([\sqrt{N}(\hat{c}_N - c)]/(\sigma v_c(N)) < \sqrt{N}(d_1 - c)/(\sigma v_c(N))\big)$$
$$= \lim_{d_1 \to 0} P\big(\mathcal{N}(0, 1) \leqslant \alpha - \alpha c/d_1\big) = 1.$$

If $c = 0$ then

$$\lim_{d_1 \to 0} P(\mathrm{CD}) = \lim_{d_1 \to 0} P\big(\hat{c}_N/((\sigma/\sqrt{N})\, v_c(N)) \leqslant d_1/((\sigma/\sqrt{N})\, v_c(N))\big) = p.$$

In the case $c > 0$, standardized linear combinations of LSE's of regression parameters are asymptotically normal as well. Hence

$$\lim_{d_1 \to 0} \lim_{d_2 \to 0} P(\mathrm{CD}) = \lim_{d_1 \to 0} \lim_{d_2 \to 0} P\left(\hat{c}_N \geqslant d_1, |-\hat{b}_M/(2\hat{c}_M) - x_0| \leqslant d_2\right)$$
$$= \lim_{d_1 \to 0} \lim_{d_2 \to 0} P\left(\hat{c}_N \geqslant d_1, \frac{|\hat{b}_M + 2x_0 \underline{\hat{c}}_M|}{[(\sigma^2/M)(v_b^2(M) + v_c^2(M) b^2/c^2)]^{1/2}} \leqslant \frac{2d_2 \underline{\hat{c}}_M}{[(\sigma^2/M)(v_b^2(M) + v_c^2(M) b^2/c^2)]^{1/2}}\right)$$
$$= \lim_{d_1 \to 0} P\big(\hat{c}_N \geqslant d_1, |\mathcal{N}(0, 1)| \leqslant \beta\big) = p,$$

because $\underline{c} = c$ for d_1 small enough and $\lim_{d_1 \to 0} P(\hat{c}_N \geqslant d_1) = 1$.

5. On practical realization of the method. Planning of experiments. The quantities $v_b^2(n)$ and $v_c^2(n)$ depend only on the plan of the experiment, i.e. on the sequence $(x_i)_{i=1,\ldots,n}$. This choice is essential because of the necessity of computing the regression coefficients for each n. Moreover, using a suitable plan of experiment one can reduce the expected number of observations in the sense of Theorem 2. Concerning this latter problem one is led to choosing such a plan that $v_c^2(n)$ and $v_b^2(n) + v_c^2(n) b^2/\underline{c}^2$ are minimized. Theorems on the optimal planning of experiments (Fedorov [3]) show that one cannot minimize $v_c^2(n)$ and $v_b^2(n)$ simultaneously. We decide to minimize $v_c^2(n)$ in order to get the first stage shorter and a decreased influence of $b^2/\underline{c}^2$ on the length of the second stage in the sense of Theorem 2.

The minimization of $v_c^2(n)$ is a problem of D-optimal truncated planning. We interpret this problem as L-optimal planning with nonsingular information matrix (see the definition below), where L is a linear functional on the set of quadratic matrices. This functional ought to be positive for positive definite matrices. Accordingly, let L be the value of the last element of the last line of a matrix

$$L\big((n/\sigma^2)\operatorname{cov}(\hat{a}_n, \hat{b}_n, \hat{c}_n)\big) = v_c^2(n).$$

Definition. The matrix $M = n^{-1}(X_n X_n^T)$ is called an *information matrix of a plan of experiment in a regression model with n observations.*

LEMMA 5 (Theorem 2.9.2* of Fedorov [3]). *The following conditions are equivalent:*

(1) *the plan* ε^* *minimizes* $L(D)$, *where* $D = (n/\sigma^2)\operatorname{cov}(\hat{a}_n, \hat{b}_n, \hat{c}_n)$;

(2) $\max_{x \in [-1,1]} \varphi(x, \varepsilon^*) = L(D)$, *where*

$$\varphi(x, \varepsilon) = L(D[1, x, x^2]^T [1, x, x^2] D).$$

The information matrices of plans satisfying (1) *and* (2) *are equal.*

In Lemma 5 one means by a plan the following assignment: $p_j \mapsto z_j$, $j = 1, \ldots, k$, $k \leqslant n$, where $p_j \in [0, 1]$ is a fraction of the number of observations, $p_1 + p_2 + \ldots + p_k = 1$ and z_j is a number from $[-1, 1]$, $z_j \neq z_l$ for $j \neq l$, the z_j's are the values of x_i, $i = 1, \ldots, n$. We denote this assignment by $\varepsilon = \begin{Bmatrix} p_1, \ldots, p_k \\ z_1, \ldots, z_k \end{Bmatrix}$.

THEOREM 4. *For each* $n \geqslant 4$ *and for* $x_i \in [-1, 1]$, $i = 1, \ldots, n$, *the least value of* $v_c^2(n)$ *in the quadratic regression model is* 4. *The optimal plan has the information matrix*

$$M = \begin{bmatrix} 1 & \overline{x}_n & \overline{x_n^2} \\ \overline{x}_n & \overline{x_n^2} & \overline{x_n^3} \\ \overline{x_n^2} & \overline{x_n^3} & \overline{x_n^4} \end{bmatrix} \quad (\textit{where } \overline{\cdot}_n \textit{ denotes averaging of } n \textit{ numbers})$$

with the optimal value

$$M^* = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

This plan has the unique form $\varepsilon^* = \begin{Bmatrix} -1, & 0, & 1 \\ 1/4, & 1/2, & 1/4 \end{Bmatrix}$ *for every* $n \geqslant 4$.

Proof. First we restrict ourselves to plans for which $\overline{x_n} = \overline{x_n^3} = 0$. Then $v_c^2(n) = (\overline{x_n^4} - (\overline{x_n^2})^2)^{-1}$. So we have to maximize $n^{-1} \sum_{i=1}^{n} (x_i^4 - n^{-1} \sum_{i=1}^{n} x_i^2)^2$. This expression has the largest value iff the x_i^2's are allocated in $[0, 1]$ equally on both sides of $\overline{x_n^2}$ and attain the values 0 and 1. Thus the plan has the form $\begin{Bmatrix} -1, & 0, & 1 \\ 1/2-\tau, & 1/2, & \tau \end{Bmatrix}$, $\tau \in (0, 1/2)$. Since $\overline{x_n} = \overline{x_n^3} = 2\tau - 1/2 = 0$, $\tau = 1/4$. Denote such a plan by ε^*. This is a unique optimal plan among the symmetric plans. Its information matrix has the form M^* and

$$(n/\sigma^2)\operatorname{cov}(\hat{a}_n, \hat{b}_n, \hat{c}_n) = D^* = \begin{bmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{bmatrix}.$$

By Lemma 5 this plan is also optimal among all the plans

$$\varphi(x, \varepsilon^*) = 16x^4 - 16x^2 + 4 \quad \text{and} \quad L(D^*) = 4 = \max_{x \in [-1,1]} \varphi(x, \varepsilon^*).$$

Every optimal plan has the same information matrix M^*, so always $\overline{x_n} = \overline{x_n^3} = 0$. Hence the plan ε^* is uniquely determined.

For n not divisible by 4 the following procedure satisfies the necessary condition of optimality (Theorem 3.3.1 of Fedorov [3]) and is intuitively justifiable since it retains the symmetry of the optimal plan:

for $n = 4k+1$,

$$\begin{Bmatrix} -1, & 0, & 1 \\ k/n, & (2k+1)/n, & k/n \end{Bmatrix},$$

for $n = 4k+2$,

$$\begin{Bmatrix} -1, & 0, & 1 \\ (k+1)/n, & 2k/n, & (k+1)/n \end{Bmatrix},$$

for $n = 4k+3$,

$$\begin{Bmatrix} -1, & 0, & 1 \\ (k+1)/n, & (2k+1)/n, & (k+1)/n \end{Bmatrix}.$$

Then $v_c^2(n)$ tends to 4 and $v_b^2(n)$ tends to 2 as $n \to \infty$.

In practice the number of observations proves to be most reduced when we extend the sample by four in every step and when we allocate the observations according to the optimal plan. Then

$$\operatorname{cov}(\hat{a}_n, \hat{b}_n, \hat{c}_n) = (\sigma^2/n) \begin{bmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{bmatrix}$$

for each n divisible by 4. This enables a recursive computation of the regression parameters.

Acknowledgements. I wish to thank Prof. R. Zieliński, Dr. J. Koronacki and Dr. B. Gołdys from the Institute of Mathematics, Polish Academy of Sciences, for helpful discussions and criticism.

References

[1] Y. S. Chow and H. Robbins, *On the asymptotic theory of fixed-width sequential confidence intervals for the mean*, Ann. Math. Statist. 36 (1965), p. 457-462.

[2] H. Drygas, *Weak and strong consistency of the least squares estimators in regression models*, Z. Wahrsch. verw. Geb. 34 (1976), p. 119-127.

[3] V. V. Fedorov, *Theory of optimum experiment* (in Russian), Nauka, Moskva 1971.

[4a] L. J. Gleser, *On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters*, Ann. Math. Statist. 36 (1965), p. 463-467.

[4b] —, *Correction to "On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters"*, ibid. 37 (1966), p. 1053-1055.

[5] S. K. Perng and Y. L. Tong, *A sequential solution to the inverse linear regression problem*, Ann. Statist. 2 (1974), p. 535-539.

[6] R. Singh, *Existence of bounded-length confidence intervals*, Ann. Math. Statist. 34 (1963), p. 1474-1485.

[7] M. S. Srivastava, *On fixed-width confidence bounds for regression parameters and mean vector*, J. Roy. Statist. Soc. Ser. B 29 (1967), p. 132-140.

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
00-656 WARSZAWA

CENTRAL SCHOOL OF PLANNING AND STATISTICS
DEPARTMENT OF MATHEMATICAL STATISTICS
02-554 WARSZAWA

Received on 1984.08.08