J. fo The Korean Operations
Research Society
Vol. 3. No. 1. June, 1978

Research note

THE USE OF MATHEMATICAL PROGRAMMING FOR LINEAR REGRESSION PROBLEMS

Sung Hyun Park*

Abstract

The use of three mathematical programming techniques (quadratic programming, integer quadratic programming and linear programming) is discussed to solve some problems in linear regression analysis. When the criterion is the minimization of the sum of squared deviations and the parameters are linearly constrained, the problem may be formulated as quadratic programming problem. For the selection of variables to find "best" regression equation in statistics, the technique of integer quadratic programming is proposed and found to be a very useful tool. When the criterion of fitting a linear regression is the minimization of the sum of absolute deviations from the regression function, the problem may be reduced to a linear programming problem and can be solved reasonably well.

1. Introduction

The primary purpose of this paper is to demonstrate that mathematical programming techniques in operations research may be effectively used to solve some linear regression preblems in statistics that cannot be solved easily by the conventional statistical tools. Also this paper is hoped to serve as an attempt to narrow the gap between operations research and statistics.

Suppose there are n observations on a p-vector of input variables ($x_1, x_2, \cdots x_p$ where $p<n$), and a scalar response, y, such that the j^{th} response, $j=1, 2, \cdots, n$ is determined by

$$y_j=\beta_o+\sum_{i=1}^{p} \beta_i \, x_{ij}+e_j. \tag{1}$$

The residuals, e_j, are assumed identically and independently distributed, usually normal, with mean zero and unknown variance, σ^2. The inputs x_{ij} are frequently taken to be specified design variables.

The model (1) may be expressed in matrix notation as

$$Y=X\beta+e. \tag{2}$$

Here Y is the n-vector of observed responses, X is the design matrix of dimension $n\times(p+1)$, assumed to have rank $(p+1)$, and β is the $(p+1)$-vector of unknown regression coefficients. To estimate β from the given observations, the method of least squares is perhaps the most widely used technique, undoubtedly due to the mathematical and computational simplicity of the method. Lel $\hat{\beta}$ be the least squared estimator of β. To find the vector $\hat{\beta}$ that minimized

*Colloge of Natural Sciences, Seoul National University

$$e'e=(Y-X\beta)'\ (Y-X\beta), \tag{3}$$

we set the derivative of $e'e$ with respect to β equel to zero and obtain

$$\hat{\beta}=(X'X)^{-1}X'Y \tag{4}$$

which can be shown to have a minimum variance within the class of all unbiased estimators which are a linear function of y.

In practice, it is often the case that prior information on individual or combination(s) of the coefficients is given in the form of an inequality such as

$$R\beta\leq r \tag{5}$$

where R is a $k\times(p+1)$ known coefficient matrix ($k\leq p$) and r is a known q-vector. Note that the unrestricted least squares estimator $\hat{\beta}$ in (4) may not satisfy the inequality restraint (5). Therefore, if the inequality is used in conjunction with the sample information, then the following restricted least squares problem :

$$\min_{\beta}\ e'e=(Y-X\beta)'\ (Y-X\beta) \tag{6}$$

such that $R\beta\leq r$.

This nonlinear programming problem will be discussed in Section 2.

Frequently, the number of independent variables is excessive and the research worker wishes to eliminate those variables which contribute little additional information about the dependent variable, y. Present methods for selecting a subset of the original P variables commonly known as "forward selection" or "backward elimination" procedures which are not necessarily optimum in any sense. In fact, it is not unusual that the set of variables selected by "forward selection" and "backward elimination" methods do not agree. Basically the optimum should be found by investigating all possible regressions, which means that, to select a set of k variables, the $\binom{p}{k}$ possible sets should be checked. For variable-selection techniques, see Draper Smith [1], Mallows [5], Hocking [3] and Park [6]. Clearly, the procedures of selection of variables present a computationnally uonreasonable problem since for large p, the number $\binom{p}{k}$ of regression runs to determine the best fit can be quite large.

In order to achieve the optimum but to avoid the large amount of computation, selection problem is reformulated as an integer quadratic programming problem, which will be presented in Section 3.

Lastly, if the residuals in the model (1) are not normally distributed, the regression coefficients generated by the least squares approach are not necessarily the best possible choice. Draper Smith [1] show that, if the residuals are distributed in a double exponential manner, the coefficients estimated by the minimization of the sum of the absolute values of the residuals,

$$\sum_{j=1}^{n}|e_j|=\sum_{j=1}^{n}|y_j-\beta_0-\sum_{i=1}^{p}\beta_i\ x_{ij}| \tag{7}$$

would be the better choice. If this is the case, the estimation of β is a difficult task. For this problem the linear programming technique is proposed in Section 4.

2. Formulation of quadratic programming

It is well known in statistics that the best linear unbiased estimator (BLUE) of β that minimizes $e'e=(Y-X\beta)'(Y-X\beta)$ is $\hat{\beta}=(X'X)^{-1}X'Y$, when the parameter β is unrestricted. Now consider the case when β is restricted in the form of an inequality, $R\beta\leq r$, where R is a $k\times(p+1)$ constant known coefficient matrix ($k\leq p$), and r is a $k\times 1$ known vector.

First, a particular case when an exact linear relation, $R\beta=r$, holds is studied. To find the vector β^* which minimizes $e'e=(Y-X\beta)'(Y-X\beta)$ subject to $R\beta=r$, the following classical Lagrangean problem results.

$$\underset{\beta,\lambda}{\text{Min}}\ f(\beta,\lambda)=(Y-X\beta)'(Y-X\beta)+\lambda'(r-R\beta)$$

where λ is a k-vector of Lagrangean multipliers. A solution to this Lagrangean problem yields a minimizing solution for β.

$$\beta^*=\hat{\beta}+(X'X)^{-1}R'[R(X'X)^{-1-}R']^{-1}(r-R\hat{\beta})$$

which differs from the unrestricted least squares estimator by a linear function of the quantity $(r-R\hat{\beta})$.

Next, consider the case when $R\beta\leq r$. To find the estimator of β (say, $\hat{\beta}$), which minimizes $e'e=(Y-X\beta)'(Y-X\beta)$ subject to $R\beta\leq r$, we can formulate the following quadratic programming problem. Let $\beta=\beta_+-\beta_-$ where $\beta^+\geq 0$ and $\beta_-\geq 0$.

Then, $e'e=(Y-X\beta)'(Y-X\beta)$

$$=\beta'X'X\beta-2\beta'X'Y+Y'Y$$

$$=(\beta_+-\beta_-)'X'X(\beta_+-\beta_-)-2(\beta_+-\beta_-)'X'Y+Y'Y.$$

By rearranging the terms in the above equation for β_+ and β_-, the problem to be solved wil' be

$$\underset{\beta_+,\beta_-}{\text{Min}}\ (\beta_+'X'X\beta_++\beta_-'X'X\beta_--2\beta'_+X'X\beta_--2\beta'_+X'Y+2\beta'_-X'Y+Y'Y)$$

$$\text{subject to } R\beta_+-R\beta_-\leq r \qquad (8)$$

$$\beta_+\geq 0,\ \beta_-\geq 0.$$

Since $X'X$ is a symmetric and positive definite matrix, the problem in (8) is a well defined quadratic programming problem in terms of β_+ and β_-. For the solving-techniques of quadratic programming problems, see Hadley [2].

3. Formulation of integer quadratic programming

As discussed earlier in the introduction for the selection-of-variable problems, the optimum should be found by basically investigating all possible regressions, which means that the $\binom{p}{k}$ possible sets should be checked in order to select a set of k variables out of p variables. Obviously, when p is large, number of regression runs to determine the best fit can be quite large.

In order to achieve this optimum, but to avoid the large number of computations, the selection-of-variable problems reformulated as a constrained optimization. That is to say, the original residual sum of squares in equation (3) will be minimized subject to the condition that at most k of the regression coefficients be nonzero. The constraints can be written as a set of inequalities by introducing a new set of integer variables, z_i, which can only take 0 or 1. For a large value of M, a set of restrictions on the original variables follow:

$$-Mz_i\leq\beta_i\leq Mz_i,\ i=1,2,\cdots,p$$

$$z_1+z_2+\cdots+z_p=k$$

$$z_i=0 \text{ or } 1,\ i=1,2,\cdots,p.$$

Let $z'=(z_1,z_2,\cdots,z_p)$ and $\beta=\beta_+-\beta_-$ as before, where $\beta_i=\beta_{+i}-\beta_{-i}$ The problem can then be written as

$$\underset{\beta_+,\beta_-}{\text{Min}}\ (\beta_+-\beta_-)'X'X(\beta_+-\beta_-)-2(\beta_+-\beta_-)'X'Y+Y'Y$$

subiect to $\beta_{+i}-\beta_{-i}\leq Mz_i$

$$\beta_{-i}-\beta_{+i}\leq -Mz_i$$
$$z_1+z_2+\cdots+z_p=k$$
$$z_i=0 \text{ or } 1$$
$$\beta_{+i}\geq 0,\ \beta_{-i}\geq 0 \text{ for } i=1,2,\cdots,p.$$

If desired, the equality $\sum_{i=1}^{p} z_i=k$ can be expressed as two inequalities, $\sum_{i=1}^{p} z_i\leq k$ and $-\sum_{i=1}^{p} z_i\leq -k$.

For the solution produres of integer quadratic programming problems, the reader may consult Kunzi Oettli [4], Wolf [7] or Hadley [2]. The author used the Kunzi-Oettli algorithm to solve the quadratic integer programming problem formulated above for a few moderate size regression porblems, and the computational results were found to encouraging. The author is currently working on some computational aspects of the Kunzi-Oettli algorithm and hope to report it soon.

4. Formulation of linear programming

The problem of fitting the regression model, $Y=X\beta+e$ in equation (2), by minimizing the sum of absolute deviations may be put in the form of a linear programming problem as follows. Decompose the vector β into its positive and negative parts β_+ and β_- so that $\beta=\beta_+-\beta_-$ where $\beta_+\geq 0$ and $\beta_-\geq 0$ as before. Similarly decompose the residuls $e=Y-X\beta$ into $e=e_+-e_-$ where $e_+\geq 0$ and $e_-\geq 0$. Then the problem

$$\text{Min } S=\sum_{j=1}^{p}|e_j|$$
$$\text{subject to } X\beta+e-Y=0$$

is equivalent to the linear program

$$\text{Min } S=1'e_+ +1'e_-$$

suject to

$$X\beta_+ -K\beta_- +e_+ -e_- -Y=0$$
$$\beta_+\geq 0,\ \beta_-\geq 0,\ e_+\geq 0,\ e_-\geq 0$$

where $1'$ is a $(1\times n)$ row vector whose elements are ones. The most covenient form of this linear program for our purposes is

$$\text{Max} -S=1'X\beta_+ -1'X\beta_- -21'e_- -1'Y$$

suject to

$$X\beta_+ -X\beta_- -e_-\leq Y$$
$$\beta_+\geq 0,\ \beta_-\geq 0,\ e_-\geq 0.$$

If we set $A=(X, -X, -I)$, $b=Y, c'=(1'X, -1'X, -21')$ aud $r'=(\beta'_+, \beta'_-, e'_-)$, then the matrix formulation of the linear program is

$$\text{Max} -S=c'r-1'Y$$

subject to

$$Ar\leq b$$
$$r\geq 0,$$

which is the general form of usual linear programming problems. When the program has been solved to obtain r^* and the associated minimum (that is, $-\max(-S)$), the estimate of the parameter β is recovered from $r^*=(\overset{*}{\beta}_{+}, \overset{*}{\beta}_{-}, \overset{*}{e}_{-})$ by setting $\beta^*=\overset{*}{\beta}_{+}-\overset{*}{\beta}_{-}$.

REFRENCES

1. Draper, N.R. and Smith, H. Applied Regression Analysis, Wiley, New York, 1966.
2. Hadley, G. Nonlinear and dynamic programming. Addison-Wesley, 1964.
3. Hocking, R.R. The analysis and selection of variables in linear regression. Biometrics 32, page 1-49, 1976.
4. Kunzi, H.P. and Oettli, W. Integer quadratic programming, Recent advances in mathematical programming. Mc Graw Hill Book Co., 1963.
5. Mallows, C.L. L. Some commments on Cp. Technometrics 15, page 661∼675, 1973.
6. Park, S.H. Selection of polynomial terms for response surface experiments, Biometrics 33, page 225∼229, 1977.
7. Wolfe, P. The simplex method for quaratic programming, Econometrica 27, page 382∼398, 1959.