Problems of Selection with Restrictions

By C. Radhakrishna Rao
Indian Statistical Institute, Calcutta

[Received May 1962]

Summary

The paper considers a practically important generalization of the theory of regression. A linear function of a set of variables $x_1, \ldots, x_p$, called predictor variables, is constructed so as to maximize its correlation with a criterion variable y_1, subject to the condition that its correlations with other criterion variables $y_2, \ldots, y_q$ are non-negative. It is suggested that a linear function so determined is useful when selection of individuals is done on the basis of $x_1, \ldots, x_p$ to achieve the maximum possible progress in the mean of y_1, while ensuring that no deterioration takes place in the mean values of $y_2, \ldots, y_q$ in the selected group, compared with the original group of individuals from which selection is made.

1. Introduction

In problems of selection there is a set of predictor variables $(x_1, \ldots, x_p)$, on the basis of which individuals are selected for certain desired characteristics defined in terms of criterion variables $(y_1, \ldots, y_q)$ not observable at the time of selection. It is well known that the selection function which gives the maximum progress in the mean value of y_1 in the selected group is the regression of y_1 on $x_1, \ldots, x_p$. In such a case, it may so happen that with respect to some of the other criteria $y_2, \ldots, y_q$ there is a deterioration in the mean values. Is it then possible to determine the selection function in such a way that maximum possible progress is shown in one characteristic subject to the condition that there is possibly progress but certainly no deterioration in some other characteristics? In the context of linear regression, the mathematical problem is that of determining a linear function of $x_1, \ldots, x_p$ such that its correlation with y_1 is positive and a maximum subject to the condition that its correlations with $y_2, \ldots, y_q$ are all non-negative. We shall show that such a linear function always exists and illustrate its computation.

2. Notation and Preliminary Results

Let $\mathbf{x}$ denote the column vector of $(x_1, \ldots, x_p)$, Λ the dispersion matrix of $\mathbf{x}$ and $\mathbf{c}_i$ the column vector of the covariances $(c_{i1}, \ldots, c_{ip})$ of y_i with $x_1, \ldots, x_p$. The matrix of all column vectors $\mathbf{c}_i$ is denoted by C and the dispersion matrix of $\mathbf{y}$ by $\Sigma = (\sigma_{ij})$. We shall assume that the rank of C is q and that Λ is non-singular and $p \geqslant q$. If $\mathbf{b}$ is an arbitrary vector of p elements then the correlation of y_i and $\mathbf{b}'\mathbf{x}$ is

$$(\mathbf{b}'\mathbf{c}_i)/\sqrt{(\sigma_{ii}\,\mathbf{b}'\Lambda\mathbf{b})}. \tag{2.1}$$

The object is to determine $\mathbf{b}$ such that

$$(\mathbf{b}'\mathbf{c}_1)/\sqrt{(\mathbf{b}'\Lambda\mathbf{b})} \text{ is a maximum}$$

subject to the conditions

$$\mathbf{b}'\mathbf{c}_1 > 0, \quad \mathbf{b}'\mathbf{c}_i \geqslant 0 \quad (i = 2, \ldots, q); \tag{2.2}$$

this is a problem of non-linear programming.

The following lemmas are useful in the solution of the problem.

Lemma 1. For any vector $\mathbf{b}$ of order p for which (2.2) is satisfied, there exists a vector $\mathbf{g}$ of order q such that

(i) $\mathbf{m} = \Lambda^{-1} C\mathbf{g}$ satisfies the conditions (2.2), and

(ii) $\dfrac{\mathbf{m}'\mathbf{c}_1}{\sqrt{(\mathbf{m}'\Lambda\mathbf{m})}} \geqslant \dfrac{\mathbf{b}'\mathbf{c}_1}{\sqrt{(\mathbf{b}'\Lambda\mathbf{b})}}$.

Any vector $\mathbf{b}$ can be expressed as the sum of two vectors, one belonging to the linear manifold generated by the columns of $\Lambda^{-1}C$ and another orthogonal to it, the inner product between two vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being defined as $\boldsymbol{\alpha}'\Lambda\boldsymbol{\beta}$. Hence there exist vectors $\mathbf{g}$ and $\boldsymbol{\epsilon}$ such that

$$\mathbf{b} = \Lambda^{-1} C\mathbf{g} + \Lambda^{-1}\boldsymbol{\epsilon} = \mathbf{m} + \Lambda^{-1}\boldsymbol{\epsilon} = 0, \quad \boldsymbol{\epsilon}'\Lambda^{-1} C = 0.$$

To prove (i), observe that

$$0 \leqslant \mathbf{b}'\mathbf{c}_i = \mathbf{m}'\mathbf{c}_i + \boldsymbol{\epsilon}'\Lambda^{-1}\mathbf{c}_i = \mathbf{m}'\mathbf{c}_i \quad (i = 1, \ldots, q).$$

To prove (ii), we have that

$$\mathbf{b}'\Lambda\mathbf{b} = \mathbf{m}'\Lambda\mathbf{m} + \boldsymbol{\epsilon}'\Lambda^{-1}\boldsymbol{\epsilon} \geqslant \mathbf{m}'\Lambda\mathbf{m}$$

and since $\mathbf{b}'\mathbf{c}_1 = \mathbf{m}'\mathbf{c}_1$, we have

$$\frac{\mathbf{m}'\mathbf{c}_1}{\sqrt{(\mathbf{m}'\Lambda\mathbf{m})}} \geqslant \frac{\mathbf{b}'\mathbf{c}_1}{\sqrt{(\mathbf{b}'\Lambda\mathbf{b})}}.$$

Lemma 1 reduces the problem to that of determining $\mathbf{m}$, which is of the form $\Lambda^{-1} C\mathbf{g}$.

Lemma 2. The problem of determining $\mathbf{g}$ such that, with $\mathbf{m} = \Lambda^{-1} C\mathbf{g}$, conditions (2.2) are satisfied and $\mathbf{m}'\mathbf{c}_1/\sqrt{(\mathbf{m}'\Lambda\mathbf{m})}$ is a maximum is equivalent to that of minimizing a non-negative quadratic form $(\mathbf{u}-\boldsymbol{\xi})' B(\mathbf{u}-\boldsymbol{\xi})$ with $\mathbf{u}$ restricted to non-negative vectors, where B and $\boldsymbol{\xi}$ are computed from the known quantities C and Λ.

Let $\mathbf{v}$ be a vector of order q with non-negative elements only and let $\mathbf{g}$ be a solution of

$$C'\mathbf{m} = C'\Lambda^{-1} C\mathbf{g} = \mathbf{v}. \tag{2.3}$$

Hence

$$\mathbf{g} = (C'\Lambda^{-1} C)^{-1}\mathbf{v} = A\mathbf{v}, \quad \mathbf{m} = \Lambda^{-1} CA\mathbf{v}, \tag{2.4}$$

where

$$A = (C'\Lambda^{-1} C)^{-1},$$

$$\frac{\mathbf{m}'\mathbf{c}_1}{\sqrt{(\mathbf{m}'\Lambda\mathbf{m})}} = \frac{v_1}{\sqrt{(\mathbf{v}'A\mathbf{v})}} = \frac{1}{\sqrt{\{(\mathbf{v}'A\mathbf{v})/v_1^2\}}}. \tag{2.5}$$

Writing $v_i/v_1 = u_i$ $(i = 2, \ldots, q)$ and denoting the elements of A by (a_{ij}), we can write the square of the denominator in the last term of (2.5) as

$$\delta + (\mathbf{u}-\boldsymbol{\xi})' B(\mathbf{u}-\boldsymbol{\xi}) \tag{2.6}$$

where $$B = (a_{ij}) \quad (i, j = 2, \ldots, q),$$
and $$\xi' = (\xi_2, \ldots, \xi_q)$$
is a solution of $$-B\xi = \alpha_1, \alpha_1' = (a_{12}, \ldots, a_{1q}), \tag{2.7}$$
and $$\delta = a_{11} - \sum_{i,j \geq 2}\sum a_{ij}\,\xi_i\,\xi_j.$$

The solution of (2.7) is
$$\xi_i = \frac{c_i'\Lambda^{-1}c_1}{c_1'\Lambda^{-1}c_1} \quad (i = 2, \ldots, q) \tag{2.8}$$
and
$$\delta = \frac{1}{c_1'\Lambda^{-1}c_1},$$
which are simple functions of known quantities c_i and Λ^{-1}. Now
$$\sup_{\mathbf{m}} \frac{\mathbf{m}'c_1}{\sqrt{(\mathbf{m}'\Lambda\mathbf{m})}} = \sup_{\mathbf{u}\geq 0}\{\delta + (\mathbf{u}-\xi)'B(\mathbf{u}-\xi)\}^{-\frac{1}{2}}$$
$$= \{\delta + \inf_{\mathbf{u}\geq 0}(\mathbf{u}-\xi)'B(\mathbf{u}-\xi)\}^{-\frac{1}{2}}. \tag{2.9}$$

The problem is thus reduced to that of minimizing the non-negative quadratic form
$$Q(\mathbf{u}) = (\mathbf{u}-\xi)'B(\mathbf{u}-\xi)$$
with the restriction $\mathbf{u} \geq 0$. If $\mathbf{u}_0' = (u_{20}, \ldots, u_{q0})$ is the minimizing vector, then the optimum vector $\mathbf{m}$ is found from (2.4)
$$\mathbf{m} = \Lambda^{-1}CA\mathbf{v}_0$$
and the selection function is
$$\mathbf{v}_0'AC'\Lambda^{-1}\mathbf{x}, \tag{2.10}$$
where $$\mathbf{v}_0' = (1, u_{20}, \ldots, u_{q0}).$$

The correlation between y_1 and the best selection function (multiple regression) when there are no restrictions on other criterion variables is
$$R_1 = 1/\sqrt{(\delta\sigma_{11})}. \tag{2.11}$$
With the restriction that the changes in mean values of other criterion variables are to be in specified directions if possible, or otherwise zero, the correlation between y_1 and the selection function (2.10) reduces to
$$R_2 = 1/\sqrt{[\sigma_{11}\{\delta + \min_{\mathbf{u}\geq 0}(\mathbf{u}-\xi)'B(\mathbf{u}-\xi)\}]}. \tag{2.12}$$
If the restriction is such that no change is desired in the mean values of other criterion variables, a problem considered by Kempthorne and Nordskog (1959) and Tallis (1962), the correlation is obtained by putting $\mathbf{u} = 0$ in (2.12) and is
$$R_3 = 1/\{\sigma_{11}(\delta + \xi'B\xi)\}^{\frac{1}{2}}.$$
It may be seen that $$R_1 \geq R_2 \geq R_3, \tag{2.13}$$

which implies that selective efficiency possibly increases by generalizing the restriction that no change should occur in other criterion variables to changes in specified directions if possible.

The correlation coefficient between the index and the criterion variable y_k ($k \neq 1$) is

$$\frac{u_k \sqrt{\sigma_{11}}}{\sqrt{\sigma_{kk}}} R_q.$$

This means that changes in the y_k's can easily be estimated and also illustrates that when $u_k = 0$ no change in the expected value of y_k will result from the selection.

3. Discussion of Special Cases

The case $q = 2$ is quite simple. The quadratic form $(\mathbf{u}-\boldsymbol{\xi})' B(\mathbf{u}-\boldsymbol{\xi})$ reduces to

$$a_{22}\left(u_2 - \frac{\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_2}{\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_1}\right)^2. \tag{3.1}$$

If $\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_2 \geqslant 0$, then the minimum value of (3.1) for non-negative u_2 is zero and the multiple regression of y_1 on $x_1, \ldots, x_p$, $\mathbf{c}_1' \Lambda^{-1} \mathbf{x}$, without the constant term, automatically satisfies the desired restriction on y_2. If $\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_2 < 0$, then the minimum is attained when $u_2 = 0$ and by the procedure indicated in (2.10) the selection function is found to be

$$\mathbf{c}_1' \Lambda^{-1} \mathbf{x} - \frac{\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_2}{\mathbf{c}_2' \Lambda^{-1} \mathbf{c}_2} \mathbf{c}_2' \Lambda^{-1} \mathbf{x}, \tag{3.2}$$

which is a linear combination of the multiple regression equations of y_1 and y_2 on $x_1, \ldots, x_p$. The square of the correlation between y_1 and the selection function (3.2) is

$$\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_1 - \frac{(\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_2)^2}{\mathbf{c}_2' \Lambda^{-1} \mathbf{c}_2} \tag{3.3}$$

divided by σ_{11}. The square of the multiple correlation between y_1 and $x_1, \ldots, x_p$ is $\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_1 / \sigma_{11}$ and the reduction due to the restriction on y_2, when $\mathbf{c}_1' \Lambda^{-1} \mathbf{c}_2 < 0$, is given by the second term in (3.3).

The next practically important case is that of $q = 3$. The quadratic form to be minimized is

$$Q(u_2, u_3) = a_{22}(u_2 - \xi_2)^2 + 2a_{23}(u_2 - \xi_2)(u_3 - \xi_3) + a_{33}(u_3 - \xi_3)^2,$$

where a_{ij} and ξ_i are as defined in (2.5) and (2.8). A number of cases arise depending on the signs of $\xi_2, \xi_3, \ldots$.

Case (i). Suppose that $\xi_2 \geqslant 0$, $\xi_3 \geqslant 0$. The minimum of Q is zero and the multiple regression of y_1 on $x_1, \ldots, x_p$ is the selection function.

Case (ii). Suppose that $\xi_2 < 0$, $\xi_3 \geqslant 0$. The minimum of Q is attained on the boundary $u_2 = 0$. To determine the value of u_3, we solve the equation

$$\frac{1}{2} \frac{dQ(0, u_3)}{du_3} = a_{33}(u_3 - \xi_3) - a_{23}\xi_2 = 0,$$

obtaining

$$u_3 = \frac{a_{23}}{a_{33}} \xi_2 + \xi_3. \tag{3.4}$$

If $a_{23}\xi_2+a_{33}\xi_3 \geqslant 0$, then the minimum value of Q is attained when $u_{20}=0$ and u_{30} has the right-hand side value in (3.4). If $a_{23}\xi_2+a_{33}\xi_3<0$, then the minimum is attained at $u_{20}=0$, $u_{30}=0$. The selection function is determined as indicated in (2.10). The case of $\xi_2 \geqslant 0$, $\xi_3<0$ is treated in a similar way.

Case (iii). Suppose that $\xi_2<0$, $\xi_3<0$. There are three possible pairs of values at which the minimum might be attained:

$$u_{20}=0, \quad u_{30}=\frac{a_{23}}{a_{33}}\xi_2+\xi_3;$$

$$u_{20}=\frac{a_{23}}{a_{22}}\xi_3+\xi_2, \quad u_{30}=0;$$

$$u_{20}=0, \quad u_{30}=0.$$

Out of these we need consider only the pairs where both the coordinates are non-negative and then choose that pair for which Q is a minimum.

When $q>3$, the number of different cases to be considered is large. When each $\xi_i \geqslant 0$, the minimum of Q is zero. But in the other cases the algorithms developed for general quadratic programming (Charnes and Cooper, 1961, pp. 682–687) may have to be adopted. It may, however, be observed that by replacing $\mathbf{u}'=(u_2,\ldots,u_q)$ by $\mathbf{w}'=(w_2^2,\ldots,w_q^2)$ in Q, the problem reduces to that of minimizing a quartic in $w_2,\ldots,w_q$ without any restrictions. No great simplification seems to result by transforming the problem in this way. As mentioned earlier, the practically important cases correspond to $q=2$ and 3 for which the solution is simple, as already indicated. The selective efficiency may go down rapidly with increase in the value of q.

It is well known that a large number of statistical techniques were first developed for applications to biological problems. The present paper shows that the entire new field of programming problems, i.e. of optimization subject to linear inequalities, could have been developed for application to biological problems several years earlier, soon after the concepts of correlation and regression were introduced.

REFERENCES

CHARNES, A. and COOPER, W. W. (1961), *Management Models and Industrial Applications of Linear Programming*, 2. New York: Wiley.

KEMPTHORNE, O. and NORDSKOG, A. W. (1959), "Restricted selection indices", *Biometrics*, **15**, 10–19.

TALLIS, G. M. (1962), "A selection index for optimum genotype", *Biometrics*, **18**, 120–122.