The Canadian Journal of Statistics
Vol. 8, No. 1, 1980, Pages 41–45
La Revue Canadienne de Statistique

Categorical information and the singular linear model

A. FEUERVERGER and D. A. S. FRASER*

University of Toronto

Key words and phrases: Linear model, singular variance matrix, categorical information, necessary reduction, inference base.
AMS 1980 *subject classifications:* Primary 62J05; secondary 62F10.

ABSTRACT

C. R. Rao (1978) discusses estimation for the common linear model in the case that the variance matrix $\sigma^2\mathbf{Q}$ has known singular form $\mathbf{Q}$. In the more general context of inference, this model exhibits certain special features and illustrates how information concerning unknowns can separate into a categorical component and a statistical component. The categorical component establishes that certain parameters are known in value and thus are not part of the statistical inference.

1. INTRODUCTION

In a recent article, C. R. Rao (1978) examines least squares and linear unbiased estimation for the linear model with possibly singular variance matrix: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathscr{E}(\mathbf{e}) = \mathbf{0}$, $\mathscr{V}\!ar(\mathbf{e}) = \sigma^2\mathbf{Q}$ with $\mathbf{Q}$ known and possibly singular. In the cases when $\mathbf{X}$ has less than full column rank, or $\mathbf{Q}$ is singular, certain special features arise; in particular the usual least squares algebra based on $\mathbf{X}$ and $\mathbf{Q}$ is not immediately available. With these special cases in mind, Rao (1971, 1972, 1973, 1976, 1978) has developed the least squares and linear unbiased analysis in terms of generalized inverse matrices.

For the model with rank deficiency for the columns of $\mathbf{X}$ or for the matrix $\mathbf{Q}$, Rao notes that the model could be transformed to the simpler case where standard formulas are available and remarks "thus, in a sense, there is no new problem to be solved". Somewhat later, however, he comments that this does not "preclude us from attempting a unified theory (a single method to cover all situations) of linear estimation applicable to the general model" with possible rank deficiencies, and then develops the general notation that covers these cases. However this emphasis on general notation deflects from certain notions of general importance in inference, and tends to mask certain basic inference results that are available for this model with data.

Point estimation—the production of a value for the parameters and nothing further—occupies a fairly restricted role in inference overall. With measures of reliability added however, and the results extended, say as confidence intervals, then only can estimation address the more substantive needs of inference.

In this larger context of inference the linear model in the singular case provides a striking example of how information concerning unknowns can separate into a

* Research completed in part while on leave at the Université de Genève during 1978–1979.

categorical component and a statistical component. This separation is currently under investigation in a general context, cf. Brenner and Fraser (1979). Details of this separation are examined here for the linear model with singularities. In particular we find that in certain cases some parameters will be known in value and thus are not part of the *statistical* inference. We point out that the consequences of singularity in **X**—nonidentifiability of certain parameters—have been noted in the literature. The direct consequences of the singularity of **Q**, however, have not been noted.

For background, we require first a brief survey concerning the definition of the model, the inference base, and the necessary reduction of an inference base.

2. MODEL, DATA, AND CATEGORICAL INFORMATION

A formal study of inference argues for a formal starting point. Ideas for such a starting point are currently being developed and in this section we present a brief survey adequate for the example to be discussed in Section 3. For the more general inference framework much additional discussion would be needed: see, for example, Fraser (1979).

The purpose of an applied investigation is to obtain in whole or in part a satisfactory model for a system of interest. For this the background information concerning the system and related systems leads to the statistical *model for the system*; this records the set of possible descriptions for the system together with a parameter that indexes these descriptions and corresponds to the unknowns of the system.

A planned investigation of a system will determine what variables are to be examined, what performances are to be made, what randomization is to be applied, and what unknowns are the object of the investigation. The background information then leads to the *model for the investigation*; this records the set of mathematical-probabilistic descriptions needed for the unknowns of the investigation. The model for the investigation is not an arbitrary construct. Rather it is a formal and specific presentation of the background information as needed for the unknowns of the investigation. In detail there are four requirements and in context these would be interpreted to a reasonable approximation: (1) the model is *descriptive*—the components and variables of the model correspond to objective components and variables of the investigation; (2) the model is *exhaustive*—objective components and variables of the investigation have a corresponding component and variable in the model; (3) the model is *probabilistic*—emphasizing certain descriptive aspects of the use of probability: marginal probabilities for what is observed and conditional probabilities for what is unobserved in relation to an observed value on a probability space or an observed value of an objective function; (4) the model is *consistent*—the model is the set of possible descriptions for unknowns of the system and such additional descriptions as are needed for an internally consistent model.

The model for an investigation together with the data from the investigation constitute the formal *inference base*. In turn, *inference* is the theory and analysis used to determine and present the available information concerning the unknowns. As initially presented, an inference base may contain arbitrary elements of notation and structure beyond that required for the unknowns. Certain basic reduction methods, *necessary methods*, have been developed to accomplish the elimination of the arbitrary elements from an initial presentation of an inference base, cf. Fraser (1979).

As part of the investigation of necessary methods, some major attention has focused on a classification of the types of information that can be available concerning unknowns: (1 *categorical*—a yes-no concerning possibilities for an unknown; (2) *frequency*—a nondegenerate probability or frequency description for an unknown; (3) *diffuse*—embracing information not of the preceding types. An important and interesting special case arises if the information segregates into types (1) and (2) with (3) absent. For a general investigation of this classification see Brenner and Fraser (1979).

For the singular linear model we will find a clear separation of the categorical and frequency information. As a consequence of this certain parameters become known in value and the statistical model applies just to the remaining frequency type information.

3. SINGULAR LINEAR MODEL

Consider the linear model $\mathscr{M}$ given by $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathscr{E}(\mathbf{e}) = \mathbf{0}$, $\mathscr{V}ar\,(\mathbf{e}) = \sigma^2\mathbf{Q}$ with $\mathbf{Q}$ known and singular. In addition the distribution form for the error may be known, or known up to a parameter λ, or in some extensive class of distributions possibly with regularity restrictions. Also let $\mathbf{y} = \mathbf{y}^o$ be the observed response from the system under investigation. Then we will be concerned with the inference base $\mathscr{I} = (\mathscr{M}, \mathbf{y}^o)$.

The determination of categorical information can arise in various ways. For example, if the distribution is known, or known up to a shape parameter λ, then certain conditioning occurs necessarily. Although fundamentally important, this conditioning, for example, Fraser (1976, Chapter 11) is essentially distinct from our concerns centering on the singular variance matrix $\mathbf{Q}$ and we do not examine it here. Rather we shall focus on the categorical consequences of the singularity of $\mathbf{Q}$. We shall also avoid some minor complications by assuming that $\mathbf{e}$ has nonzero density wherever permitted by the singularity of $\mathbf{Q}$: thus $\mathbf{e}$ is assumed to have nonzero density throughout $\mathscr{L}(\mathbf{Q})$, the linear space spanned by the column vectors of $\mathbf{Q}$.

We now turn to the linear model $\mathscr{M}$ with data $\mathbf{y}^o$ and focus on the categorical information consequent to the singularity of the variance matrix $\mathbf{Q}$.

The location vector of the response distribution lies in $\mathscr{L}(\mathbf{X})$, the error vector lies in $\mathscr{L}(\mathbf{Q})$. Our purpose is to examine the categorical information and we try to keep specialized notation to a minimum: orthonormal bases for subspaces are easily derived by routine computational procedures. Accordingly, let $\mathbf{N}$ be a matrix of basis vectors giving $\mathscr{L}(\mathbf{X}) \cap \mathscr{L}(\mathbf{Q}) = \mathscr{L}(\mathbf{N})$; $\mathbf{W}$ be a matrix of basis vectors giving $\mathscr{L}(\mathbf{X} \perp \mathbf{Q}) = \mathscr{L}(\mathbf{W})$, the orthogonal complement of $\mathscr{L}(\mathbf{N})$ in $\mathscr{L}(\mathbf{X})$; $\mathbf{M}$ be a matrix of basis vectors giving $\mathscr{L}(\mathbf{Q} \perp \mathbf{X}) = \mathscr{L}(\mathbf{M})$, the orthogonal complement of $\mathscr{L}(\mathbf{N})$ in $\mathscr{L}(\mathbf{Q})$; and $\mathbf{R}$ be a matrix of basis vectors giving $\mathscr{L}(\mathbb{E}^n \perp \mathbf{X}, \mathbf{Q}) = \mathscr{L}(\mathbf{R})$, the orthogonal complement of $\mathscr{L}(\mathbf{X}, \mathbf{Q})$ in $\mathbb{E}_n$.

The model $\mathscr{M}$ then has the following components:

(1) $\mathbf{R}'\mathbf{y} = \mathbf{0}$
(2) $\mathbf{W}'\mathbf{y} = \mathbf{W}'\mathbf{X}\boldsymbol{\beta}$
(3) $\mathbf{N}'\mathbf{y} = \mathbf{N}'\mathbf{X}\boldsymbol{\beta} + \mathbf{e}_1$; $\mathscr{E}(\mathbf{e}_1) = \mathbf{0}$, $\mathscr{V}ar\,(\mathbf{e}_1) = \sigma^2\mathbf{N}'\mathbf{Q}\mathbf{N}$
(4) $\mathbf{M}'\mathbf{y} = \mathbf{e}_2$; $\mathscr{E}(\mathbf{e}_2) = \mathbf{0}$, $\mathscr{V}ar\,(\mathbf{e}_2) = \sigma^2\mathbf{M}'\mathbf{Q}\mathbf{M}$.

(In fact without loss of generality we may presume $\mathbf{N}'\mathbf{Q}\mathbf{N}$ and $\mathbf{M}'\mathbf{Q}\mathbf{M}$ to be identity matrices.) For the inference base with model $\mathscr{M}$ and data $\mathbf{y}^o$ we now have the following segregation of the information:

(i) $\mathbf{R}'\mathbf{y}^o = \mathbf{0}$. If in fact $\mathbf{R}'\mathbf{y}^o \neq \mathbf{0}$ then the inference base is self-contradictory denying the validity of the model or data.

(ii) $\mathbf{W}'\mathbf{y}^o = \mathbf{W}'\mathbf{X}\boldsymbol{\beta}$. This is categorical information that asserts that certain linear functions of $\boldsymbol{\beta}$ are known in value; this is a nonstatistical result and occurs whenever $\mathscr{L}(\mathbf{X}) \not\subset \mathscr{L}(\mathbf{Q})$.

(iii) $\mathbf{N}'\mathbf{y}^o$ is an observed value for $\mathbf{N}'\mathbf{X}\boldsymbol{\beta} + \mathbf{e}_1$. This is statistical information in which the elements of $\mathbf{N}'\mathbf{y}^o$ are measurements with error $\mathbf{e}_1$ on the linear functions $\mathbf{N}'\mathbf{X}\boldsymbol{\beta}$ of $\boldsymbol{\beta}$.

(iv) $\mathbf{M}'\mathbf{y}^o$ is an observed value for $\mathbf{e}_2$. This is statistical information in which $\mathbf{M}'\mathbf{y}^o$ is an error observation.

(v) Any particular $\boldsymbol{\beta}$-parameter difference in $\mathscr{L}^{\perp}(\mathbf{X}'\mathbf{X})$ is unidentifiable by the model with data.

The statistical analysis of (iii) and (iv) is well established for the normal case, is well developed subject to computer integrations for a λ-family for variation, and opens to large-sample nonparametric possibilities more generally. Our primary focus here, however, is on the explicit and clear initial separation of the categorical component (ii) as part of the overall inference procedure: certain parameter functions are known in value and are not part of the statistical investigation.

4. CONCLUDING REMARKS AND SUMMARY

Our primary contention is that categorical information should be clearly and explicitly separated from statistical information in the formal examination of an inference base. We have done this for the linear model and found that categorical information will sometimes specify the exact values for certain parameters and that the statistical information takes the form of a nonsingular linear model.

The linear model with singularities has had substantial attention in the literature, almost all of which has been concerned with the specialized activity of point estimation. In this paper our purpose has been to point out that something very fundamental must precede this—the separation and interpretation of categorical information which gives the exact values for certain parameters.

Directly, in terms of the model and data we can summarize as follows. If $\mathbf{y}^o \notin \mathscr{L}(\mathbf{X}, \mathbf{Q})$ then the inference base is self-contradictory. Accepting the inference base, we obtain the categorical information that any location parameters of $\mathscr{L}(\mathbf{X} \perp \mathbf{Q})$ are known exactly. We then obtain the statistical information that the location parameters of $\mathscr{L}(\mathbf{X}) \cap \mathscr{L}(\mathbf{Q})$ have separate direct measurements and that the observations for $\mathscr{L}(\mathbf{Q} \perp \mathbf{X})$ provide error measurements marginal to the joint error distribution in $\mathscr{L}(\mathbf{Q})$.

The singular linear model provides an important example of the presence of categorical information in a problem apparently just statistical. The recognition and initial separation of this component is fundamental. An applied statistician should certainly be interested in the opportunity to calculate certain parameters exactly.

RÉSUMÉ

C. R. Rao (1978) étudie l'estimation avec le modèle linéaire classique dans le cas où la matrice de covariance est $\sigma^2\mathbf{Q}$, $\mathbf{Q}$ étant une matrice singulière connue. Pour la théorie de l'inférence, ce modèle présente un intérêt bien particulier car il illustre comment l'information concernant des paramètres inconnus peut être séparée en une composante catégorique et une

composante statistique. La composante catégorique implique qu'en fait certains paramètres ont une valeur connue et ne font pas partie de l'inférence statistique.

REFERENCES

Brenner, D., and Fraser, D. A. S. (1979). On foundations for conditional probability with statistical models—when is a class of functions a function? *Statist. Hefte* (*N.F.*), 20, 148–159.

Fraser, D. A. S. (1976). *Probability and Statistics: Theory and Applications.* Duxbury Press, North Scituate, Mass.

Fraser, D. A. S. (1979). *Inference and Linear Models.* McGraw-Hill, New York.

Rao, C. R. (1971). Unified theory of linear estimation. *Sankhyā Ser. A*, 33, 371–394. (Correction: *Sankhyā Ser. A*, 34 (1972), 477.)

Rao, C. R. (1972). Some recent results in linear estimation. *Sankhyā Ser. B*, 34, 369–378.

Rao, C. R. (1973). Unified theory of least squares. *Comm. Statist.*, 1, 1–8.

Rao, C. R. (1976). Estimation of parameters in a linear model. *Ann. Statist.* 4, 1023–1037.

Rao, C. R. (1978). Least squares theory for possibly singular models. *Canad. J. Statist.*, 6, 19–23.

Received 25 January 1979
Revised 14 February 1980

Department of Statistics
University of Toronto
Toronto, Ontario
M5S 1A1