

LARGE SAMPLE VARIANCES OF MAXIMUM LIKELIHOOD ESTIMATORS
OF VARIANCE COMPONENTS.

BU-255-M

S. R. Searle

May, 1968

Biometrics Unit, Cornell University, Ithaca, New York

Introduction

Asymptotic properties of maximum likelihood (M.L.) estimators are usually considered in terms of sample size tending to infinity. When data are from a linear model involving one or more classifications, then the concept of sample size tending to infinity must be specified in a manner that takes into account the sample sizes in each subclass of the model. Hartley and Rao (1967) have discussed this problem, and it is within their framework that we here consider asymptotic variances. Iterative procedures for deriving ML estimators are presented in that paper; an expression for asymptotic variances of those estimators is now obtained.

Model

We first note that in the most general terms, all linear models of the customary form $\underline{y} = \mu \underline{1} + \underline{X} \underline{\beta} + \underline{e}$ can be considered as mixed models. For, in the usual fixed model, there is one random term, \underline{e} ; and in the random model there is one fixed effect, μ ; thus, without loss of generality, any linear model can be considered as a mixed model and expressed as

$$\underline{y} = \underline{X} \underline{\beta} + \underline{Z} \underline{u}$$

where \underline{y} is a vector of N observations; $\underline{\beta}$ is a $p \times 1$ vector of fixed effects, that include the general mean μ ; \underline{u} is a vector of random effects, that include the error terms \underline{e} ; and \underline{X} and \underline{Z} are known matrices---often, but not always, design matrices. The random effects are further specified as having zero mean and a

variance-covariance matrix \underline{A} , that involves q variance components $\sigma_1^2, \dots, \sigma_q^2$.

In addition, for purposes of using maximum likelihood we adopt normality assumptions and so have \underline{u} distributed as $N(\underline{0}, \underline{A})$. Thus \underline{y} has variance-covariance matrix $\underline{V} = \underline{Z}\underline{A}\underline{Z}'$ and is distributed as $N(\underline{X}\underline{\beta}, \underline{V})$ where elements of \underline{V} are functions of the q variance components.

Likelihood and variances

For the above model the likelihood of the data is

$$(2\pi)^{-\frac{1}{2}N} |\underline{V}|^{-\frac{1}{2}} \exp -\frac{1}{2}(\underline{y} - \underline{X}\underline{\beta})' \underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta})$$

and, apart from a constant, the logarithm of this is

$$L = -\frac{1}{2} \log |\underline{V}| - \frac{1}{2}(\underline{y} - \underline{X}\underline{\beta})' \underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta}).$$

Now the variance-covariance matrix of the large sample M.L. estimators of the p elements of $\underline{\beta}$ and the q variance components is minus the inverse of the expected value of the Hessian of L with respect to these $p + q$ parameters. The sub-matrices of this Hessian are:

$$\underline{L}_{\underline{\beta}\underline{\beta}}, \text{ a } p \times p \text{ matrix of terms } \frac{\partial^2 L}{\partial \beta_h \partial \beta_k} \text{ for } h, k = 1, \dots, p;$$

$$\underline{L}_{\underline{\beta}\sigma^2}, \text{ a } p \times q \text{ matrix of terms } \frac{\partial^2 L}{\partial \beta_h \partial \sigma_j^2} \text{ for } h = 1, \dots, p \text{ and } j = 1, \dots, q;$$

$$\text{and } \underline{L}_{\sigma^2\sigma^2}, \text{ a } q \times q \text{ matrix of terms } \frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2} \text{ for } i, j = 1, \dots, q.$$

Then the matrix we seek is

$$\underline{V}_{\text{ML}} = \begin{bmatrix} \text{var } (\tilde{\underline{\beta}}) & \text{cov } (\tilde{\underline{\beta}} \tilde{\underline{\sigma}}^2) \\ \text{cov } (\tilde{\underline{\sigma}}^2 \tilde{\underline{\beta}}) & \text{var } (\tilde{\underline{\sigma}}^2) \end{bmatrix}$$

$$= \begin{bmatrix} -E(\underline{L}_{\beta\beta}) & -E(\underline{L}_{\beta\sigma^2}) \\ -E(\underline{L}_{\beta\sigma^2})' & -E(\underline{L}_{\sigma^2\sigma^2}) \end{bmatrix}^{-1}$$

where $\tilde{\underline{\beta}}$ is the M.L. estimator of $\underline{\beta}$ and $\tilde{\underline{\sigma^2}}$ is the vector of M.L. estimators of the q variance components.

For convenience write

$$d \text{ for } \log |\underline{V}| .$$

Then

$$L = -\frac{1}{2}d - \frac{1}{2}(\underline{y} - \underline{X}\underline{\beta})' \underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta})$$

and

$$\underline{L}_{\beta\beta} = -\underline{X}' \underline{V}^{-1} \underline{X}$$

$$\underline{L}_{\beta\sigma^2} = \left\{ \underline{X}' (\underline{V}^{-1})_{\sigma^2_j} (\underline{y} - \underline{X}\underline{\beta}) \right\} \text{ for } j = 1, 2, \dots, q ;$$

and

$$\underline{L}_{\sigma^2\sigma^2} = \left\{ -\frac{1}{2}d_{\sigma^2_{ij}} - \frac{1}{2}(\underline{y} - \underline{X}\underline{\beta})' (\underline{V}^{-1})_{\sigma^2_{ij}} (\underline{y} - \underline{X}\underline{\beta}) \right\}$$

$$\text{for } i, j = 1, 2, \dots, q.$$

In these expressions

$$d_{\sigma^2_{ij}} = \frac{\partial^2 d}{\partial \sigma^2_i \partial \sigma^2_j}$$

and

$$(\underline{V}^{-1})_{\sigma^2_{ij}} = \frac{\partial^2 \underline{V}^{-1}}{\partial \sigma^2_i \partial \sigma^2_j}$$

$$= \left\{ \frac{\partial v^{r,s}}{\partial \sigma^2_i \partial \sigma^2_j} \right\} \text{ for } r, s = 1, 2, \dots, N,$$

where $v^{r,s}$ is the (r,s) 'th element in \underline{V}^{-1} ; i.e., $(\underline{V}^{-1})_{\sigma^2_{ij}}$ is the matrix \underline{V}^{-1} with every element differentiated with respect to σ^2_i and σ^2_j .

Taking expectations we now have

$$\begin{aligned} E(\underline{L}_{\underline{\beta}\underline{\beta}}) &= -\underline{X}\underline{V}^{-1}\underline{X} \\ E(\underline{L}_{\underline{\beta}\sigma^2}) &= \left\{ \underline{X}'(\underline{V}^{-1})_{\sigma_i^2} E(\underline{y} - \underline{X}\underline{\beta}) \right\} \\ &= 0, \quad \text{for } j = 1, 2, \dots, q; \end{aligned}$$

and

$$\begin{aligned} E(\underline{L}_{\sigma^2\sigma^2}) &= \left\{ -\frac{1}{2}d_{\sigma_i^2\sigma_j^2} - \frac{1}{2}E(\underline{y} - \underline{X}\underline{\beta})'(\underline{V}^{-1})_{\sigma_i^2\sigma_j^2}(\underline{y} - \underline{X}\underline{\beta}) \right\} \\ &\quad \text{for } i, j = 1, 2, \dots, q. \end{aligned}$$

Utilizing the fact that a scalar is its own trace, and that under the trace operation matrix products are cyclically commutable, we have

$$\begin{aligned} E(\underline{L}_{\sigma^2\sigma^2}) &= \left\{ -\frac{1}{2}d_{\sigma_i^2\sigma_j^2} - \frac{1}{2}\text{tr} \left[E(\underline{y} - \underline{X}\underline{\beta})(\underline{y} - \underline{X}\underline{\beta})'(\underline{V}^{-1})_{\sigma_i^2\sigma_j^2} \right] \right\} \\ &= \left\{ -\frac{1}{2}d_{\sigma_i^2\sigma_j^2} - \frac{1}{2}\text{tr} \left[\underline{V}(\underline{V}^{-1})_{\sigma_i^2\sigma_j^2} \right] \right\} \\ &\quad \text{for } i, j = 1, 2, \dots, q. \end{aligned}$$

Hence

$$\underline{V}_{\underline{ML}} \begin{bmatrix} \underline{X}'\underline{V}^{-1}\underline{X} & \underline{0} \\ \underline{0} & \frac{1}{2} \left\{ d_{\sigma_i^2\sigma_j^2} + \text{tr} \left[\underline{V}(\underline{V}^{-1})_{\sigma_i^2\sigma_j^2} \right] \right\} \quad i, j = 1, \dots, q \end{bmatrix}^{-1}$$

Several points about this result are worthy of note. The first is that covariances between large sample M.L. estimators of fixed effects and variance components are zero. Bearing in mind that under conditions of normality the mean of a sample and its sum of squares are independent, this result is not surprising; but its generality is to be observed. An obvious consequence is, of course, that the variance-covariance matrix of the large sample M.L. estimators of the fixed

effects is, from \underline{V}_{ML} , $(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}$; and that for the variance components is the inverse of

$$\frac{1}{2} \left\{ d_{\sigma_i^2 \sigma_j^2} + \text{tr} \left[\underline{V}(\underline{V}^{-1})_{\sigma_i^2 \sigma_j^2} \right] \right\} \quad \text{for } i, j = 1, 2, \dots, q. \quad \text{---(1)}$$

With d being $\log |\underline{V}|$ this matrix is, it will be noted, free of the fixed effects and solely a function of the variance-covariance matrix \underline{V} of the vector of observations \underline{y} . Although $\underline{V} = \underline{Z}\underline{A}\underline{Z}'$ and it is only \underline{A} that involves the variance components, it does not appear to be more useful to write (1) in terms of \underline{Z} and \underline{A} instead of \underline{V} .

Application

The procedure for using (1) is clear: for any model find \underline{V} , $|\underline{V}|$, $d = \log |\underline{V}|$ and \underline{V}^{-1} . Then, for every pair of variance components σ_i^2 and σ_j^2 (including $i = j$), derive $d_{\sigma_i^2 \sigma_j^2}$, $(\underline{V}^{-1})_{\sigma_i^2 \sigma_j^2}$ and $\text{tr} \left[\underline{V}(\underline{V}^{-1})_{\sigma_i^2 \sigma_j^2} \right]$. When there are q components the values $\frac{1}{2} \left\{ d_{\sigma_i^2 \sigma_j^2} + \text{tr} \left[\underline{V}(\underline{V}^{-1})_{\sigma_i^2 \sigma_j^2} \right] \right\}$ will, as in (1), constitute a square matrix of order q , whose inverse yields the variances (and covariances) of the large sample estimators of the σ_i^2 's.

As a simple example consider the model $y = \mu + e_i$ for $i = 1, 2, \dots, N$, where the e_i are $\text{NID}(0, \sigma^2)$. Then

$$\underline{V} = \sigma^2 \underline{I}, \quad |\underline{V}| = \sigma^{2N}, \quad d = N \log \sigma^2, \quad \text{and } \underline{V}^{-1} = (1/\sigma^2)\underline{I}.$$

Hence

$$d_{\sigma^2 \sigma^2} = -N/\sigma^4 \quad \text{and} \quad (\underline{V}^{-1})_{\sigma^2 \sigma^2} = (2/\sigma^6)\underline{I},$$

and so, by substitution in (1), with $\tilde{\sigma}^2$ being the M.L. estimator of σ^2 ,

$$\text{var}(\tilde{\sigma}^2) = \left\{ \frac{1}{2} \left(\frac{-N}{\sigma^4} \right) + \frac{1}{2} \text{tr} \left[(\sigma^2 \underline{I}) \left(\frac{2}{\sigma^6} \underline{I} \right) \right] \right\}^{-1}$$

$$\begin{aligned} &= \left(\frac{-N}{2\sigma^4} + \frac{2N}{2\sigma^4} \right)^{-1} \\ &= \frac{2\sigma^4}{N} \end{aligned}$$

as is to be expected.

Further application of (1), to the random model, 1-way classification, unbalanced data, yields the results given in Searle (1956). In that case $\underline{V} = \sum_i \oplus \underline{V}_i$,

the direct sum of matrices $\underline{V}_i = \sigma_e^2 \mathbf{I}_{n_i} + \sigma_a^2 \mathbf{J}_{n_i}$. Additional application, to the

random model, 2-way classification, unbalanced data is currently being pursued.

The matrix \underline{V} is then $\underline{V} = \sum_i \oplus \underline{V}_i$, with

$$\underline{V}_i = \sum_j \oplus \left(\sigma_e^2 \mathbf{I}_{n_{ij}} + \sigma_b^2 \mathbf{J}_{n_{ij}} \right) + \sigma_a^2 \mathbf{J}_{n_i}.$$

References

Hartley, H. O. and J. N. K. Rao (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54:93-108.

Searle, S. R. (1956). Matrix methods in component of variance and covariance analysis. *Ann. Math. Stat.* 27:737-748.