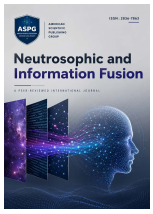
Neutrosophic and Information Fusion

♦ ISSN: 2836-7863 Vol: 03, No. 01, 2024 ♦ PP. 20–23

Article DOI: https://doi.org/10.54216/NIF.030104



FULL LENGTH ARTICLE

On the Problem of the Estimation of Variance Components Based on Non-linear Maximization Approach

Ahmad Khaldi[1,*]

[1] Faculty of Science, Mutah University, Jordan

Email: khaldiahmad1221@gmail.com

Received: May 02, 2023 Accepted: January 13, 2024 ⋆ Corresponding author

ABSTRACT

In this paper, we study the problem of estimating variance components in the two-way classification with interaction in the random-effect linear model by non-linear maximization. We assume the model according to the assumptions and give the theory of derivation of the estimators of these components, then apply these estimators to real data and obtain the estimates. We estimate these components by two other methods: the solution of the expected equation of mean square in the analysis of the variance table, and the minimum-variance quadratic unbiased estimator.

Keywords: Variance ▪ Linear model ▪ Non-linear maximization ▪ Estimation theory

1. INTRODUCTION

The study of the variance component and its analysis is one of the important topics in statistics because of its wide applications. The issue of estimating the variance component is one of the linear-model problems studied by many researchers and in various ways. The pioneering work of Fisher (1918) led to the development of the method of analysis of variance (ANOVA) for estimating variance components. Rao was among the researchers who gave considerable attention to methods of estimating variance components; see Rao and Kleffe (1988), Dietrich and Osborne (1991), and Ashish and Ghosh (2003).

Consider the linear model

$$y_{ijk} = \propto + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \quad \begin{matrix} i = 1,2,\ldots,a, \\ j = 1,2,\ldots,b, \\ k = 1,2,\ldots,n. \end{matrix} \qquad (1)$$

Here, $\propto$ is the general centre of the observations; y_{ijk} is observation k in row i and column j; α_i is the random effect of row i; β_j is the random effect of column j; γ_{ij} is the random interaction effect between row i and column j; and e_{ijk} is the random error associated with y_{ijk}. We assume that α_i, β_j, γ_{ij}, and e_{ijk} are independent and normally distributed with expectation zero and variances σ_0^2, σ_3^2, σ_2^2, and σ_1^2, respectively. These quantities are called the variance components, and the goal of this research is to estimate them.

2. ESTIMATION OF VARIANCE COMPONENTS BY NON-LINEAR MAXIMIZATION

Estimating variance components using the maximum-likelihood method is a non-linear optimization problem with constraints on the solution. A specialized algorithm is therefore applied to the likelihood function to estimate the components iteratively. Researchers who have studied this method include Mardia and Marshall (1984) and Dietrich and Osborne (1991).

2.1 Variance Component Model

Model (1) can be rewritten in linear form as follows (see Rao and Kleffe, 1988):

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{ZU} + \mathbf{e}. \qquad (2)$$

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

Here, $\mathbf{Y}$ is the $n \times 1$ vector of observations; $\mathbf{X}$ is the $n \times p$ design matrix; β is the $p \times 1$ vector of unknown parameters; $\mathbf{Z}$ is an $n \times q_i$ design matrix of the form

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_t);$$

$\mathbf{U}$ is the $q_i \times 1$ vector of random effects; and $\mathbf{e}$ is the $n \times 1$ vector of random errors.

2.2 Basic Assumptions on Model (2)

The model assumptions are

$$\begin{aligned} \mathbf{e} &\sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_n), \\ \mathbf{U} &\sim N\big(\mathbf{0}, \operatorname{diag}(\sigma_1^2 \mathbf{I}_{q_1}, \sigma_2^2 \mathbf{I}_{q_2}, \ldots, \sigma_r^2 \mathbf{I}_{q_r})\big), \\ \operatorname{Cov}(\mathbf{U}, \mathbf{e}) &= \mathbf{0}. \end{aligned} \tag{3}$$

According to these assumptions,

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V}), \qquad \mathbf{Y} = (y_1, y_2, \ldots, y_n)'.$$

The covariance matrix is

$$\begin{aligned} \mathbf{V} &= \sum_{i=1}^{r} \sigma_i^2 \mathbf{Z}_i \mathbf{Z}_i' + \sigma_e^2 \mathbf{I}_e \\ &= \sum_{i=0}^{r} \sigma_i^2 \mathbf{Z}_i \mathbf{Z}_i'. \end{aligned} \tag{4}$$

If, in (4),

$$\sigma_0^2 = \sigma_e^2, \qquad \mathbf{Z}_0 \mathbf{Z}_0' = \mathbf{I}_n,$$

then $\sigma_0^2, \sigma_1^2, \ldots, \sigma_r^2$ are called variance components, and the goal is to estimate them from the available data. The constraints are

$$\sigma_i^2 \geq 0 \quad (i = 1,2,3), \qquad \sigma_0^2 > 0. \tag{5}$$

Constraint (5) makes the covariance matrix $\mathbf{V}$ positive definite.

2.3 Maximum-Likelihood Estimation

Assume that $\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V})$ and that $\mathbf{V}$ is positive definite and hence nonsingular. The likelihood function of $\mathbf{Y}$ is

$$\begin{aligned} L(\beta, \mathbf{V}) = &\left(\frac{1}{\sqrt{2\pi}}\right)^n |\mathbf{V}|^{-1/2} \\ &\times \exp\left[-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)\right]. \end{aligned} \tag{6}$$

Here, $|\mathbf{V}|$ denotes the determinant of $\mathbf{V}$ (Dietrich and Osborne, 1991). Estimating $\mathbf{V}$ and β by maximizing (6) subject to (5) is a constrained non-linear optimization problem. Maximizing (6) is equivalent to maximizing

$$\begin{aligned} \log L(\beta, \mathbf{V}) = C &- \frac{1}{2}\log|\mathbf{V}| \\ &- \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta), \end{aligned} \tag{7}$$

where C is a real constant. Differentiating (7) with respect to β and equating the result to zero gives

$$\frac{\partial \log L(\beta, \mathbf{V})}{\partial \beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\beta. \tag{8}$$

Thus,

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}.$$

Equation (8) gives the generalized least-squares estimator. The inverse sign on $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ denotes the generalized inverse of this matrix; see Searle et al. (1992) and Mardia et al. (1979). If $\mathbf{V}$ is singular, a treatment of $\widehat{\beta}$ is given by Searle (1988). The present goal, however, is to estimate σ_i^2 for $i = 0,1,2,3$ by differentiating (7) with respect to $\mathbf{V}$, which is equivalent to differentiation with respect to σ_i^2.

Let

$$\theta_i = \sigma_i^2, \quad i = 0,1,2,3, \qquad \theta = (\theta_0, \theta_1, \theta_2, \theta_3)'.$$

To differentiate (7) with respect to θ, the following two results are needed:

$$\frac{\partial \log|\mathbf{V}|}{\partial \theta_i} = \operatorname{tr}(\mathbf{V}^{-1}\mathbf{V}_i), \tag{9}$$

$$\frac{\partial \mathbf{V}^{-1}}{\partial \theta_i} = -\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}, \tag{10}$$

where

$$\mathbf{V}_i = \frac{\partial \mathbf{V}}{\partial \theta_i};$$

see Cressie (1993).

2.4 Proof of Equation (9)

Let $\mathbf{V}^{-1} = (\sigma^{ij})$. Then

$$\frac{\partial \log|\mathbf{V}|}{\partial \sigma_{ij}} = \begin{cases} \sigma^{ii}, & i = j, \\ 2\sigma^{ij}, & i \neq j. \end{cases}$$

Applying the chain rule gives

$$\begin{aligned} \frac{\partial \log|\mathbf{V}|}{\partial \theta} &= \sum_i \sum_j \frac{\partial \log|\mathbf{V}|}{\partial \sigma_{ij}} \frac{\partial \sigma_{ij}}{\partial \theta} \\ &= \sum_i \sum_j \sigma^{ij} \frac{\partial \sigma_{ij}}{\partial \theta} = \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta}\right). \end{aligned}$$

Hence,

$$\frac{\partial \log|\mathbf{V}|}{\partial \theta} = \operatorname{tr}\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta}\right),$$

where $\theta = (\theta_0, \theta_1, \ldots, \theta_r)$ and the derivative can be taken with respect to any component θ_i.

Equation (10) follows from $\mathbf{V}^{-1}\mathbf{V} = \mathbf{I}$. Differentiating this identity yields

$$\frac{\partial \mathbf{V}^{-1}}{\partial \theta_i}\mathbf{V} + \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta} = \mathbf{0}. \tag{11}$$

Using (7) and (9), the resulting equations are

$$\begin{aligned} \frac{\partial \log L(\beta, \mathbf{V})}{\partial \theta_i} = &-\frac{1}{2}\operatorname{tr}(\mathbf{V}^{-1}\mathbf{V}_i) \\ &+ \frac{1}{2}\mathbf{W}\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{W}, \quad i = 1,2,\ldots,r. \end{aligned} \tag{12}$$

where $\mathbf{W} = \mathbf{Y} - \mathbf{F}\beta$. For computational purposes, (12) may

be written as

$$\frac{\partial \log L(\beta,\mathbf{V})}{\partial \theta_i} = -\frac{1}{2}\operatorname{tr}(\widehat{\mathbf{G}}\widehat{\mathbf{V}}_i) = 0, \quad i = 1,2,\ldots,r, \tag{13}$$

with

$$\mathbf{G} = \mathbf{V}^{-1} - \mathbf{S}\mathbf{S}', \qquad \mathbf{S} = \mathbf{V}^{-1}\mathbf{W}.$$

These equations can be solved only iteratively. One solution method uses the information matrix, for which further derivative results are needed:

$$\frac{\partial^2 \mathbf{V}^{-1}}{\partial \theta_i \partial \theta_j} = \mathbf{B}_{ij}\mathbf{V}^{-1} + \mathbf{B}_{ij}\mathbf{V}^{-1} - \mathbf{R}_{ij}, \tag{14}$$

where

$$\mathbf{V}_{ij} = \frac{\partial^2 \mathbf{V}}{\partial \theta_i \partial \theta_j}, \qquad \mathbf{R}_{ij} = \mathbf{V}_{ij}\mathbf{V}^{-1}, \qquad \mathbf{B}_{ij} = \mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j.$$

Equation (14) is obtained by differentiation and simplification. For any constant matrix $\mathbf{C}$,

$$\frac{\partial}{\partial \theta_i}\operatorname{tr}(\mathbf{C}\mathbf{V}) = \operatorname{tr}(\mathbf{C}\mathbf{V}_i).$$

Therefore, from (9),

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \theta_i \partial \theta_j} = \operatorname{tr}(\mathbf{B}_{ij} - \mathbf{R}_{ij}).$$

Using (9) and (10) gives

$$\begin{aligned} 2\frac{\partial^2 \log L(\beta,\mathbf{V})}{\partial \theta_i \partial \theta_j} = & -\operatorname{tr}(\mathbf{R}_{ij} - \mathbf{B}_{ij}) \\ & - \mathbf{W}'(\mathbf{B}_{ij} + \mathbf{B}_{ji} - \mathbf{R}_{ij})\mathbf{V}^{-1}\mathbf{W}. \end{aligned} \tag{15}$$

Since $E(\mathbf{W}) = 0$ and

$$E(\mathbf{W}'\mathbf{C}\mathbf{W}) = \operatorname{tr}(\mathbf{V}\mathbf{C}),$$

taking expectations gives

$$E\left[-\frac{\partial^2 \log L(\beta,\mathbf{V})}{\partial \theta_i \partial \theta_j}\right] = \frac{1}{2}\operatorname{tr}(\mathbf{B}_{ij}). \tag{16}$$

Suppose that

$$\mathbf{A} = \frac{1}{2}\operatorname{tr}(\mathbf{B}_{ij}). \tag{17}$$

Then

$$\frac{\partial^2 \log L(\beta,\mathbf{V})}{\partial \beta\, \partial \theta_j} = -\mathbf{F}'\mathbf{V}^{-1}\mathbf{V}_j\mathbf{W}, \tag{18}$$

and

$$E\left[\frac{\partial^2 \log L(\beta,\mathbf{V})}{\partial \beta\, \partial \theta_i}\right] = 0, \tag{19}$$

because $E(\mathbf{W}) = E[\mathbf{Y} - \mathbf{F}\beta] = 0$. Also,

$$\frac{\partial^2 \log L(\beta,\mathbf{V})}{\partial \beta^2} = \mathbf{F}\mathbf{V}^{-1}\mathbf{F}.$$

The information matrix of the pair (β,θ) is

$$\mathbf{M}(\beta,\theta) = \begin{pmatrix} \mathbf{F}\mathbf{V}^{-1}\mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix}.$$

The covariance matrices of the estimators are

$$\operatorname{Cov}(\widehat{\beta}) = (\mathbf{F}\mathbf{V}^{-1}\mathbf{F})^{-1}, \qquad \operatorname{Cov}(\widehat{\theta}) = \mathbf{A}^{-1}.$$

Equations (8) and (13) can be solved iteratively. If θ_0 is the initial value, then β is obtained from (8), while the Newton–Raphson method updates θ at each step according to

$$\theta_{k+1} = \theta_k + \mathbf{A}_k^{-1}(\delta_k). \tag{21}$$

Here, $\mathbf{A}_k$ is evaluated at θ_k, and

$$\delta_{k,i} = -\frac{1}{2}\operatorname{tr}(\mathbf{G}\mathbf{V}_i)$$

is the derivative vector from (13), evaluated at θ_k and β_k. Equation (8) is used to update β in each iteration, although the main goal is to estimate θ_i.

2.5 Scientific Application

The maximum-likelihood method is applied to real data representing workers' achievement scores under four training methods and three age levels. Two observations were selected for each combination of training method and age level. Analysis of variance (ANOVA) and the minimum-variance quadratic unbiased estimator (MIVQUE) were then applied to the same data.

Using the two-way random-effects model with interaction as in (1),

$$\mathbf{Y} = \mathbf{F}\propto + \mathbf{Z}_1\mathbf{U}_1 + \mathbf{Z}_2\mathbf{U}_2 + \mathbf{Z}_3\mathbf{U}_3 + \mathbf{e}. \tag{22}$$

Here,

$$\mathbf{F} = (1,1,\ldots,1)', \qquad E(\mathbf{Y}) = \mathbf{F}\propto,$$

and

$$\begin{aligned} \operatorname{Var}(\mathbf{Y}) = & \sigma_1^2\mathbf{Z}_1\mathbf{Z}_1' + \sigma_2^2\mathbf{Z}_2\mathbf{Z}_2' \\ & + \sigma_3^2\mathbf{Z}_3\mathbf{Z}_3' + \sigma_0^2\mathbf{Z}_0\mathbf{Z}_0', \\ \mathbf{Z}_0\mathbf{Z}_0' = & \mathbf{I}. \end{aligned}$$

Assuming a two-way classification with interaction, the data are shown in Table 1.

Table 1. Data for the two-category model with interaction

	1	2	3	4	Total
1	10 14	16 22	12 18	9 19	120
2	23 25	17 21	24 32	18 24	184
3	13 17	8 12	16 12	7 19	104
					408

There are $a = 3$ rows, $b = 4$ columns, and $n = 2$ observations in each cell. The maximum-likelihood estimates are reported in Table 2.

Table 2. Maximum-likelihood estimates

Parameter	MLE
σ_0^2	19.0
σ_1^2	25.2
σ_2^2	0
σ_3^2	0

Table 3 compares the estimates produced by maximum likelihood, ANOVA, and MIVQUE.

Table 3. Comparison of estimation results

Parameter	MLE	ANOVA	MIVQUE
σ_0^2	19.0	20.83	20.83
σ_1^2	25.2	25.17	25.17
σ_2^2	0	−1.78	−1.78
σ_3^2	0	0.92	0.92

The negative estimates of σ_2^2 are interpreted as zero; see Marshall and Mardia (1985). The comparison shows that the maximum-likelihood estimates are close to the ANOVA and MIVQUE estimates. This indicates the success of estimating variance components through non-linear maximization by the maximum-likelihood method, which is the objective of the research.

3. CONCLUSION

In non-linear maximization, it is necessary to use the maximum-likelihood method. The principal conclusions are as follows:

1. The y_{ijk} observations in the maximum-likelihood method should be independent, as in the analysis-of-variance method. In the minimum-variance quadratic unbiased method, however, the data can be uncorrelated. When using the latter method here, independence of the observations was assumed.

2. The maximum-likelihood method and the minimum-variance quadratic unbiased estimator do not require a balanced design, unlike the analysis-of-variance method.

3. Negative estimates of variance components are interpreted as zero because a variance cannot be negative.

4. When maximum-likelihood estimates of the variance components are found iteratively, the unbiasedness property is lost; that is, the maximum-likelihood method is sensitive to the unbiasedness property.

5. The maximum-likelihood method can be applied to designs broader than a two-way classification with interaction, such as the universal design.

REFERENCES

[1] N. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.

[2] C. R. Dietrich and M. R. Osborne, "Estimation of covariance parameters in kriging via restricted maximum likelihood," *Mathematical Geology*, vol. 23, no. 1, pp. 119–135, 1991.

[3] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. London: Academic Press, 1979.

[4] R. J. Marshall and K. V. Mardia, "Minimum norm quadratic estimation of components of spatial covariance," *Mathematical Geology*, vol. 17, no. 5, pp. 517–525, 1985.

[5] K. V. Mardia and R. J. Marshall, "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, vol. 71, no. 1, pp. 135–146, 1984.

[6] C. R. Rao and J. Kleffe, *Estimation of Variance Components and Applications*. Amsterdam: North-Holland, 1988.

[7] S. R. Searle, "Lecture notes on variance components," Biometrics Unit, Cornell University, Ithaca, New York, Tech. Rep., 1988.

[8] S. R. Searle, G. Casella, and C. E. McCulloch, *Variance Components*. New York: Wiley, 1992.

[9] A. Das and H. Ghosh, "Estimation of ratio of variance component and optimal diallel cross designs," Indian Statistical Institute, Delhi Centre, New Delhi, India, Tech. Rep., 2003.