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

Ahmad Khaldi
Mutah University, Faculty of Science, Jordan
khaldiahmad1221@gmail.com

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 on 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 models studied by many researchers and in various ways in the fields of research, as it was the pioneering work of the scientist Fisher (1918) that led to the development of the method of analysis of variation (ANOVA) in estimating the Variance Component, and Rao was one of the researchers who gave the methods of estimating the Variance Component attention, see (Rao and Kleffe, 1988), (Dietrich and Osborne, 1991), (Ashish, 2003).
Consider the linear model:
$y_{ijk} = \propto + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$
i= 1, 2,…,a j=1,2,…,b k=1,2,….,n
$\propto$: The general center of views
y_{ijk}: views k in row i and column j
α_i: random effect of Grade i
β_j: random effect of column j
γ_{ij}: random effect of the interaction between row i and column j
e_{ijk}: random error for viewing y_ijk
We assume that α_i, β_j, γ_{ij}‘, $e_{ijk}$are independent and have a normal distribution with the expectation of zero and the variance $\sigma_0^2, \sigma_3^2, \sigma_2^2, \sigma_1^2$, respectively. $\sigma_0^2, \sigma_3^2, \sigma_2^2, \sigma_1^2$are called the components of the variance and our goal in this research is to estimate these parameters.

2. Estimation of variance components by nonlinear maximization

Estimating the variance components using the Maximum Likelihood Method is a matter of nonlinear optimization with constraints on the solution, as we will apply the special algorithm to the possibility function in order to estimate the components in an iterative method.

DOI: https://doi.org/10.54216/NIF.030104
Received: May 02, 2023 Accepted: January 13, 2024

Researchers have been very interested in this method, including (Mardia and Marshall,, 1984) as well as (Dietrich and Osborne, 1991).

Variance Component Model:

Form (1) can be rewritten in the form of a linear form and as shown in the following figure: see Rao (1988)

$Y = X\beta + ZU + e$ (2)

Since:

Y: the vector of views having dimension n x1

X: the design matrix has dimension n x p and is defined by

β: a vector with Dimension p x1 of unknown information

Z: a design matrix with Dimension n xqi is defined and has the following form:

$$Z = (Z_1, Z_2, \dots, Z_t)$$

U: a vector with Dimension qi x1 of random effects.

e: a vector with Dimension n x 1 of random errors.

Basic assumptions on the Model (2):

$e \sim N(0, \sigma_e^2 I_n)$

$U \sim N\left(0, \mathrm{diag}(\sigma_1^2 I_{q1}, \sigma_2^2 I_{q2}, \dots, \sigma_r^2 I_{qr})\right)$ (3)

$Cov(U, e) = 0$

According to these assumptions, we have:

$Y \sim N(X\beta, V)$

$Y = (y_1, y_2, \dots, y_n)'$

$V = \sum_{i=1}^{r} \sigma_i^2 Z_i Z_i' + \sigma_e^2 I_e$

$= \sum_{i=0}^{r} \sigma_i^2 Z_i Z_i'$ (4)

If we impose in (4)

$\sigma_0^2 = \sigma_e^2$

$Z_0 Z_0' = I_n$

Parameters $\sigma_0^2, \sigma_1^2, \dots, \sigma_r^2$ are called Variance Component and our goal is to estimate these components from the data we have available.

According to the two entries

$\sigma_i^2 \geq 0 \quad i = 1,2,3$

$\sigma_0^2 > 0$ (5)

It is known that constraint (5) leads to the covariance-covariance matrix V being a positive definite Matrix.

Maximum Likelihood Estimation:

We assume that:

$Y \sim N(X\beta, V)$

We stated that V is a positive definite matrix i.e. a nonsingular Matrix.

If the maximum Likelihood function of the variable Y is written in the following form:

$L(\beta, V) = (\frac{1}{\sqrt{2T}})^n |V|^{-\frac{1}{2}} exp[-\frac{1}{2}(Y - X\beta)V^{-1}(Y - X\beta)]$ (6)

Where $|V|$ means the determinant of the Matrix V

See: (Dietrich and Osborene, 1991)

For the purpose of estimating the two parameters V, β in such a way as to maximize the maximum Likelihood function of the function $L(\beta, V)$, the optimization must be in accordance with constraint (5) and this is a question in the subject of nonlinear constrained optimization.

The maximization of the function (6) for β and V is equivalent to the maximization of $log(\beta, V)$that is:

$log L(\beta, V) = C. -\frac{1}{2} log|V| - \frac{1}{2}(Y - X\beta)' V^{-1}(Y - X\beta)$ (7)

Where C is a real constant.

Taking the derivative of the function (7) for β and equating it to zero, we obtain an estimate of $\hat{\beta}$

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

DOI: https://doi.org/10.54216/NIF.030104
Received: May 02, 2023 Accepted: January 13, 2024

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

(8) is called the general least squares estimator GLs and the-sign of the Matrix $(X\, V^{-1}X)$ means the general inverse of this matrix see (Searle, 1992), (Mardia et al, 1979)

If the Matrix V is a Singular, then there is a treatment of the estimator $\hat{\beta}$ see (Searle, 1988), but our goal in this study is to estimate σ_i^2 (i=0,1,2,3)by returning to the derivation of (7) for V and this is equivalent to the derivation for σ_i^2 ,i=0,1,2,3.

We assume that:

$\theta_i = \sigma_i^2$

i=0,1,2,3

We consider the vector:

$\theta = (\theta_0, \theta_1, \theta_2, \theta_3)'$

So in order to derive (7) for θ We must prove the following two results:

$$\frac{\partial log/V/}{\partial \theta_i} = trV^{-1}V_i \qquad (9)$$

$$\frac{\partial V^{-1}}{\partial \theta_i} = -V^{-1}V_iV^{-1} \qquad (10)$$

Since:

$$V_i = \frac{\partial V}{\partial \theta_i}$$

See: (Cressie, 1993)

Proof of the equation (9):

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

Since:

$V^{-1} = (\sigma^{ij})$

Then applying the chain rule:

$$\frac{\partial log/V/}{\partial \theta} = \sum_i \sum_j \frac{\partial log/V/}{\partial \sigma_{ij}} \frac{\partial \sigma_{ij}}{\partial \theta} = \sum_i \sum_j \sigma^{ij} \frac{\partial \sigma_{ij}}{\partial \theta} = trV^{-1}\frac{\partial V}{\partial \theta}$$

hence:

$$\frac{\partial log/V/}{\partial \theta} = trV^{-1}\frac{\partial V}{\partial \theta}$$

It is natural to derive:

$\theta = (\theta_0, \theta_1, \ldots., \theta_r)$

Is the derivative of one of the components θ_i

Equation (10) can also be proved in the following form:

From the equation: $V^{-1}V = I$

We get the following:

$$\frac{\partial V^{-1}}{\partial \theta_i}V + V^{-1}\frac{\partial V}{\partial \theta} = 0 \qquad (11)$$

So using equations (7) and (9) we have (r) of the equations yields from:

DOI: https://doi.org/10.54216/NIF.030104
Received: May 02, 2023 Accepted: January 13, 2024

$$\frac{\partial logL(\beta,V)}{\partial \theta_i} = -\frac{1}{2} trV^{-1}V_i + \frac{1}{2} WV^{-1}V_iV^{-1}W \quad (12)$$

i=1,2,…r

$W = Y - F\beta$

These equations can be solved numerically, and for Computational purposes it is possible to write:

$$\frac{\partial logL(\beta,V)}{\partial \theta_i} = -\frac{1}{2} tr\, \hat{G}\, \hat{V}_i = 0 \qquad i=1,2,\ldots,r \quad (13)$$

$G = V^{-1} - SS'$

$S = V^{-1}W$

These equations can be solved only iteratively, and one of the solution methods includes the information matrix we need other results from the derivation so we get:

$$\frac{\partial^2 V^{-1}}{\partial \theta_i \partial \theta_j} = \mathrm{B_{ij}} V^{-1} + \mathrm{B_{ij}} V^{-1} - R_{ij} \quad (14)$$

then:

$\mathrm{V_{ij}} = \frac{\partial^2 V}{\partial \theta_i \partial \theta_j}$

$R_{ij} = \mathrm{V_{ij}} V^{-1}$

$\mathrm{B_{ij}} = V^{-1}\mathrm{V_i} V^{-1}\mathrm{V_j}$

And we get equation (14) from deriving (15) and simplifying it.

Since for any matrix of constants C:

$$\frac{\partial}{\partial \theta_i} trCV = trC\mathrm{V_i}$$

So from (9) we get:

$$\frac{\partial^2 log\, V}{\partial \theta_i\, \partial \theta_j} = tr(\mathrm{B_{ij}} - R_{ij})$$

And by deriving (14) for θ_j

Using (9) and (10) we find that

$$2\frac{\partial^2 log\, (\beta,V)}{\partial \theta_i\, \partial \theta_j} = -tr(\mathrm{R_{ij}} - \mathrm{B_{ij}}) - W'(\mathrm{B_{ij}} + \mathrm{B_{ji}} - \mathrm{R_{ij}})\, V^{-1}W \quad (15)$$

Since $\mathrm{R_{ij}}, \mathrm{B_{ij}}$are defined as before.

Since:

E (*W*) =0

E (*W CW*) = *tr VC*

By taking the expectation to (10)

$$E\left[\frac{-\partial^2 logL(\beta,V)}{\partial \theta_i\, \partial \theta_j}\right] = \frac{1}{2} tr\mathrm{B_{ij}} \quad (16)$$

Where $\mathrm{B_{ij}}$ is defined in (14)

Suppose that:

$$A = \frac{1}{2} tr\mathrm{B_{ij}} \quad (17)$$

we have:

DOI: https://doi.org/10.54216/NIF.030104
Received: May 02, 2023 Accepted: January 13, 2024

$$\frac{\partial^2 logL(\beta,V)}{\partial\beta\,\partial\theta_j} = -F'\,V^{-1}V_jW \quad (18)$$

$$E\left[\frac{\partial^2 logL(\beta,V)}{\partial\beta\,\partial\theta_i}\right] = 0 \quad (19)$$

Because:

$$E(W) = E[Y - F\beta] = 0$$

And also:

$$\frac{\partial^2 logL(\beta,V)}{\partial\beta^2} = F\,V^{-1}F$$

If the information matrix of the pair (β, θ) is:

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

And the covariance matrix of the estimator $\hat{\beta}$ is

$$COV(\hat{\beta}) = (FV^{-1}F)^{-1}$$

And the covariance matrix of the estimator $\hat{\theta}$ is

$$COV(\hat{\theta}) = A^{-1}$$

A is defined as in (17).

Equations (8) and (13) can be solved by iteration if we assume:

θ_0: the initial or initial value of the parameter θ from (8) the value β can be obtained using the Newton-Raphson method to update θ at each step using:

$$\theta_{k+1} = \theta_k + A_k^{-1}(\delta_k) \quad (21)$$

Where A_kThe Matrix A is calculated at θ_k:

$$\delta_k = \frac{-1}{2}trGV_i$$

Is the vector of derivatives by equation (13) calculated at $\theta_k$and β_k

To update β, we use (8) in each iteration, knowing that our goal is to estimate θ_i only.

Scientific aspect:

The scientific application of the maximum Likelihood Method to real data, which represents the achievement scores in testing workers on four training methods at three age levels, two factors were selected for each training method and age level, and then two methods of analysis of variance (ANOVA) and the method of the smallest unbiased quadratic variance (MIVQUE) were applied to the same data.

Using the two-classification random model with the presence of interaction as in (1).

$$Y = F \propto +Z_1U_1 + Z_2U_2 + Z_3U_3 + e \quad (22)$$

Where:

F= (1,1,….,1)

$$E(Y) = F \propto$$

$$var(Y) = \sigma_1^2Z_1Z'_1 + \sigma_2^2Z_2Z'_2 + \sigma_3^2Z_3Z'_3 + \sigma_0^2Z_0Z'_0$$

Where:

$$Z_0Z'_0 = I$$

DOI: https://doi.org/10.54216/NIF.030104
Received: May 02, 2023 Accepted: January 13, 2024

Assuming that the model is dual-classified with the presence of interaction and data as in the following table:

Table 1: Data for two-Category model with the presence of 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

Note that the number of rows a=3 and the number of columns b=4, since there are two views in each cell n=2

Table 2: The results of the estimate in the Maximum Likelihood Method

Parameters (σ_i^2)	**the greatest possible (MLE)**
σ_0^2	19.0
σ_1^2	25.2
σ_2^2	0
σ_3^2	0

Table 3: Comparison of results

parameters σ_i^2	**the way of the possible (MLE)**	**method of contrast analysis (ANOVA)**	**Method (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

Note that the negative estimates of σ_2^2 are zero, see (Marshall and Mardia, 1985)

As shown in the table of comparison of the results of the three methods for estimating the variance components we noted that the results of the method of the Maximum Likelihood Method are close to the results of the method of variance analysis and the method of estimating the smallest unbiased quadratic variance, which indicates the success of the method of estimating the variance components by the method of non-linear maximization, the method of the Maximum Likelihood Method, and this is the goal of the research.

DOI: https://doi.org/10.54216/NIF.030104
Received: May 02, 2023 Accepted: January 13, 2024

3. Conclusion

In non-linear maximization it is necessary to use the Maximum Likelihood Method.

1. The y_{ijk}observations in the Maximum Likelihood Method should be independent as well as in the variance analysis method. However, in the method of the smallest unbiased squared variance the data can be uncorrelated. When using the latter method, we assumed the independence of the views.
2. The Maximum Likelihood Method and the estimator of the smallest unbiased quadratic variance does not require the design to be balanced, as in the method of variance analysis.
3. Negative estimates of components are interpreted to be equal to zero because the variance may not be negative.
4. In the maximum likelihood method, when finding estimates of the variance components using iteration, we notice the loss of the bias-free property, that is, the maximum likelihood method is sensitive to the bias-free property.
5. The Maximum Likelihood Method can be modeled on other designs that are broader than a two-partition design with interaction, for example, the universal design.

References

[1] Cressie, N. (1993), Statistics for Spatial Data, Wiley, New York.
[2] Dietrich, C. R. and Osborne, M. R. (1991), Estimation of Covariance Parameter in Cringing Vai Restricted Maximum Likelihood, J. Math. Geo. Vol. 23. No. 1. PP. 119-135.
[3] Mardia, Kent .J.T. and Bibby, J.A. (1979), Multivariate Analysis, Academic press, London.
[4] Marshall, R. J.and Mardia, K.V.(1985) , Minimum Norm Quadratic Estimation of Components of spatial Covariance Math. Geol. Vol. 17, No. 5,p.517-525
[5] Mardia, K. V. and Marshall, R. J. (1984), Maximum likelihood Estimation of Models for Residual Covariance In Spatial Regression Biometrika, 73, 135-146.
[6] Rao, C. R. and kleffe, J. (1988) : Estimation of variance component and application north Holland, Amsterdam
[7] Searle, S. R. (1988). Lecture notes on Variance Component. Biometrics unit, Cornell University, Ithaca, New York.
[8] Searle, S. R. Casella, G. and Mc culloch, C. E. (1992): Variance Component. Wiley, New York.
[9] ASHISH, DAS, Himadri Ghosn (2003): Estimation of Ratio of Variance Component and Optimal Dialed Cross Designs. http: // www.isid.ac.in / ~ statmath / eprints Indian statistical institute, Delhi centre 7.

DOI: https://doi.org/10.54216/NIF.030104
Received: May 02, 2023 Accepted: January 13, 2024