International Journal of Research and Review
DOI: https://doi.org/10.52403/ijrr.20210937
Vol.8; Issue: 9; September 2021
Website: www.ijrrjournal.com
E-ISSN: 2349-9788; P-ISSN: 2454-2237

Short Communication

Maximum Likelihood Estimation using the EM Algorithm

Ahsene Lanani

Department of Mathematics, Faculty of Exact Sciences, University Frères Mentouri Constantine 1, 25000 Algeria

ABSTRACT

This paper yields with the Maximum likelihood estimation using the EM algorithm. This algorithm is very used to solve nonlinear equations with missing data. We estimated the linear mixed model parameters and those of the variance-covariance matrix. The considered structure of this matrix is not necessarily linear.

Key Words: Algorithm EM; Maximum likelihood; Mixed linear model

1. INTRODUCTION

In the models using longitudinal data or repeated measurements, we are often confronted with missing data. This loss of data or information is due to several reasons, missing and often death of the corresponding experimental units among others. The pioneers in this field are Dempster et al. [1]. In addition, when we estimate the parameters of a mixed linear model using the maximum likelihood (ML) or the restricted maximum likelihood (REML) method, the normal equations are often nonlinear and consequently do not admit explicit solutions; from where the passage to iterative processes or algorithms. The EM algorithm can be used to estimate such parameters, like those generating the variance-covariance matrix of the model (Dempster et al. [2]; Jennrich and Schluchter [4]); Laird and Ware [5]. Thereafter, others works were developed on this subject [8-10]. An improvement of the convergence of the EM algorithm was carried out by Laird et al. [6]. Lindstrom and Bates [7] compared this algorithm with that of Newton - Raphson and some results were discussed. A recent study concerning the estimate of the parameters generating the variance-covariance matrix of a mixed model by using restricted maximum likelihood is given by Foulley et al. [3]. Section 2 of this paper describes the EM algorithm. An example of application is given in section 3. Section 4 relates to the results, especially those of the estimators of the parameters generating the variance-covariance matrix of the model.

2. EM ALGORITHM

Let X be a random variable of density $f(x/\theta)$ where θ is an unknown parameter. Let us suppose that X is not completely observed; i.e. we observe a part Y of X. Let $Y = Y(X)$, a random variable of density g(y/θ). Let t(x) be a vector of sufficient (exhaustive) statistics for θ. The purpose of the EM algorithm (E for expectation and M for maximization) is to find the value of θ which maximizes the likelihood g(y/θ) being given a value of y. This maximization (normal equations) gives the following equation:

$$E(t(x)/\theta) = E(t(x)/y,\theta) \quad (1)$$

The EM algorithm uses two stages to solve this equation in θ.

First stage. E-Step: We calculate the quantity:

$$t(x) = E(t(x)/y,\theta) \quad (2)$$

Secund stage. M-Step: We solve the equation in θ:

$$E(t(x)/\theta) = t(x) \quad (3)$$

In other words, in E-Step, given an initial value for θ; $\theta^{(p)}$ its value at the stage (p); at the stage (p+1), we calculate the value of $t(x)$ noted by $t^{(p)}$ and which is given by $t^{(p)} = E(t(x)/y, \theta^{(p)})$. In the M-Step, given the value of $t^{(p)}$calculated in the E-Step, we solve the equation in $\theta^{(p+1)}$ which is given by:

$E(t(x)/y, \theta^{(p+1)}) = t^{(p)}$ (4)

3. APPLICATION OF THE EM ALGORITHM

3.1 Model

Let us consider the following linear mixed model

$y_i = X_i\alpha + Z_i b_i + e_i, i = 1, \dots m$ (5)

Where: y_i vector of the responses of dimensions $(n_i \times 1)$. X_i: a known $(n_i \times \text{p})$ design matrix linking α to y_i. α: a $(p \times 1)$ vector of unknown parameters, it is a vector of fixed effects. Z_i: a known $(n_i \times \text{k})$ design matrix linking b_i to y_i. b_i: a $(k \times 1)$ vector of unknown parameters, it is a vector of random effects. b_i is distributed as N(0,D), (normal with mean 0 and covariance matrix D). $D = D(\theta)$), θ: an unknown $(q \times 1)$ vector. e_i: vector of the errors which are supposed to be independent and follows $N(0, \sigma^2 I)$; I the identity matrix. Therefore the variance-covariance matrix of y_i noted by V_i is given by

$V_i = Z_i D Z_i^{\,t} + \sigma^2 I$ (6)

We want to find the maximum likelihood estimator of θ, parameter generating this matrix of variance-covariance; for that, we apply the EM algorithm. We consider that b_i and e_i are observations in addition to y_i. The sufficient statistics noted by t_1 and t_2, used to estimate θ are $\sum e_i^{\,t} e_i$, and $\sum b_i b_i^{\,t}$ respectively.

E-Step: we calculate $t_1^{(p)}$and $t_2^{(p)}$ given by:

$t_1^{(p)} = E(\sum e_i^{\,t} e_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$ (7)

$t_2^{(p)} = E(\sum b_i b_i^{\,t} / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$ (8)

M-Step: we solve the equations in $\theta^{(p+1)}$

$E(\sum e_i^{\,t} e_i, \alpha(\theta^{(p+1)}), \theta^{(p+1)}) = t_1^{(p)}$ (9)

$E(\sum b_i b_i^{\,t} / \alpha(\theta^{(p+1)}), \theta^{(p+1)}) = t_2^{(p)}$ (10)

3.2 Calculus

The parameter θ is composed of σ^2, parameter generating t_1 and of $\left(\frac{1}{2}\right) k(k+1)$ component of t_2. M-Step: in this stage, we will use the expressions (11) and (12) below

$\sigma^2 = \frac{\sum_1^m e_i^{\,t} e_i}{\sum_1^m n_i} = \frac{t_1}{\sum_1^m n_i}$ (11)

$D = m^{-1} \sum_1^m b_i b_i^{\,t} = \frac{t_2}{m}$ (12)

E-Step: having a preliminary value for θ (initial value), we then calculate the estimators of statistics t_1 and t_2.

Therefore: at the (p) stage or in $\theta^{(p)}$ (preliminary value), we have the expressions $t_1^{(p)}$ and $t_2^{(p)}$ which are given by the formulas (13) and (14) below

$t_1^{(p)} = E(\sum e_i^{\,t} e_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$

$= \sum_1^m e_i(\theta^{(p)})^t e_i(\theta^{(p)}) + trvar(e_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$

(13)

Where tr and var means trace and variance respectively.

$t_2^{(p)} = E(\sum b_i b_i^{\,t} / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$

$\sum_1^m b_i(\theta^{(p)}) b_i(\theta^{(p)})^t + var(b_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$

(14)

Where:

$e_i(\theta^{(p)}) = E(e_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)}) = y_i - X_i\, \alpha(\theta^{(p)}) - Z_i\, b_i(\theta^{(p)})$

With $\alpha(\theta^{(p)})$ and $b_i(\theta^{(p)})$ given by

$$\alpha(\theta^{(p)}) = (\sum_1^m X_i^{\,t} W_i(\theta^{(p)}) X_i)^{-1} (\sum_1^m X_i^{\,t} W_i(\theta^{(p)}) y_i)$$

$$b_i(\theta^{(p)}) = D(\theta^{(p)}) Z_i^{\,t} W_i(\theta^{(p)}) (y_i - X_i\, \alpha(\theta^{(p)}))$$

Where $W_i = V_i^{-1}$

Recall that the estimators of α and b_i are, the maximum likelihood estimator for α and the estimator of generalized least squares or the empirical Bayes estimator for b_i which is given by:

$$b_i(\theta^{(p)}) = E(b_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$$

To calculate $E(e_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$ and $trvar(e_i / y_i, \alpha(\theta^{(p)}), \theta^{(p)})$, it is necessary to calculate the distribution of e_i conditionally at $(y_i, \alpha(\theta^{(p)}), \theta^{(p)})$. The same method is done to calculate $b_i(\theta^{(p)})$ and $var(b\_i/y\_i, \alpha(\theta^\wedge((p))), \theta^\wedge((p)))$.

To have the maximum likelihood estimator of θ, that we will note by θ_M, we start with a suitable initial value of θ, we make then iterations between (13) and (14) stage defining the E-Step, and (11) and (12) stage defining the M-Step. At convergence, we do not have only θ_M, but also $\alpha(\theta_M)$ and $b_i(\theta_M)$ of the calculation of the last E-Step.

Remark: for initial values for the estimators of α and $b_i(\theta_M)$ we can take for example the ordinary least squares estimators, which are given by

$$\alpha_0 = \left(\sum_{i=1}^{m} X_i{}^t X_i\right)^{-1} \left(\sum_{i=1}^{m} X_i{}^t y_i\right)$$

$$b_i = (Z_i{}^t Z_i)^{-1}(y_i - X_i\,\alpha_0), \quad i = 1, \dots, m$$

4. CONCLUSION

We are interested on the EM algorithm although it admits certain disadvantages, because it converges slowly towards a local solution rather than a global one (Lindstrom & Bates [7]). The advantage is that it applies not only to estimate the fixed parameters and those of the variance-covariance matrix of the model, but also to estimate the random parameters of the considered model. Another advantage is that it gives estimators of the parameters generating the variance-covariance matrix, whatever its form (not necessarily linear, as in the majority of the old methods, those of Henderson among others).

Acknowledgement: None

Conflict of Interest: None

Source of Funding: None

REFERENCES

1. Dempster AP, Laird NM, Rubin DB. Maximum likelihood with incomplete data via the E-M algorithm. Jour. Roy. Stat. Soc. 1977; B39: 1-38
2. Dempster AP, Rubin DB, Tsutakawa RK. Estimation covariance component models. Jour. Amer. Stat. Assoc. 1981 ; 76 : 341-353
3. Foulley JL, Jaffrézic F, Granié CR. EM-REML estimation of covariance parameters in Gaussian mixed models for longitudinal data analysis. Genet. Sel. Evol., 2000; 32: 129 - 141
4. Jennrich RI, Scluchter MD. Unbalanced Repeated-Measures Models with Structured Covariance Matrices. Biometrics. 1986; 42: 805 - 820
5. Laird NM, Ware JH. Random effects models for longitudinal data. Biometrics. 1982 ; 38: 963-974
6. Laird NM, Lange N, Stram D. Maximum likelihood computations with repeated measures: application of the E-M algorithm. Jour. Amer. Stat. Assoc.1987 ; 82(397) : 97-105
7. Lindstrom MJ, Bates DM. Newton-Raphson and E-M algorithm for linear mixed effect models for repeated - measures data. Jour. Amer. Stat. Assoc. 1988 ; 83(404): 1014-1022
8. Verbeke G, Molenberghs G, Linear mixed models for longitudinal data. New York: Springer, 2000.
9. Littell RC, Pendergast J, Natarajan R. Modelling covariance Structure in the analysis of repeated measures data. Stat. in Med. 2000 ; 19 :1793-1819
10. Crowder M. On repeated measures analysis with misspecified covariance structure. Jour. Roy. Stat. Soc. 2001 ; B63: 55-62

How to cite this article: Lanani A. Maximum likelihood estimation using the EM algorithm. *International Journal of Research and Review*. 2021; 8(9): 275-277. DOI: *https://doi.org/10.52403/ijrr.20210937*

******