Applied and Computational Mathematics
2014; 3(5): 268-272
Published online November 10, 2014 (http://www.sciencepublishinggroup.com/j/acm)
doi: 10.11648/j.acm.20140305.22
ISSN: 2328-5605 (Print); ISSN: 2328-5613 (Online)

The equivalence of the Maximum Likelihood and a modified Least Squares for a case of Generalized Linear Model

Ahsene Lanani

MAM Laboratory, Department of Mathematics, University of Constantine 1. Constantine 25000, Algeria

Email address:
alanani1@yahoo.fr

To cite this article:
Ahsene Lanani. The Equivalence of the Maximum Likelihood and a Modified Least Squares for a Case of Generalized Linear Model. *Applied and Computational Mathematics.* Vol. 3, No. 5, 2014, pp. 268-272. doi: 10.11648/j.acm.20140305.22

Abstract: During the analysis of statistical data, one of the most important steps is the estimation of the considered parameters model. The most common estimation methods are the maximum likelihood and the least squares. When the data are considered normal, there is equivalence between the two methods, so there is no privilege for one or the other method. However, if the data are not Gaussian, this equivalence is no longer valid. Also, if the normal equations are not linear, we make use of iterative methods (Newton-Raphson algorithm, Fisher, etc ...). In this work, we consider a particular case where the data are not normal and solving equations are not linear and that it leads to the equivalence of the maximum likelihood method at least squares but modified. At the end of the work, we concluded by referring to the application of this modified method for solving the equations of Liang and Zeger.

Keywords: Maximum Likelihood, Linear Mixed Model, Newton-Raphson Algorithm, Weighted Least Squares

1. Introduction

The frequency of the grouped data in biology, epidemiology and the problems of health in general is at the origin of the increasing interest of the biostatisticians for the methods of statistical analysis which are adapted specifically to the correlated data. The choice of a statistical model among a family of models and an analysis method among so others, is not an easy task. This choice depends on the applicability, the aim, the structure of the sample or the degree of dependence inside the individual groups. For the statistical data analysis, the most frequently used models are those of regression. The linear regression, whose objective is the study of the relation between a variable response (explained variable) and one or more explanatory variables, is based on the linear models (LM). In order to explain variability between the various individuals, random effects were introduced into the explanatory part of the traditional linear models. That gives rise to the linear mixed models (LMM) or random effects models, (sometimes, noted L2M by certain authors).

The adequate model to analyze longitudinal data as well as repeated measurements, balanced and especially unbalanced, is the mixed linear model. This case, in which the outcomes are approximately jointly normal, has been studied by many authors (Harville1977; Laird and Ware1982; Jennrich and Schluchter 1986; Lindstrom and Bates 1988; Chi and Reinsel 1989; Diggle et al. 1994; Foulley et al.2000; Littell et al.2000; Park et al. 2001; Park and Lee 2002). A book treating these models is that of Verbeke and Molenberghs 2000. The problems of calculation are partly solved with the introduction and the implementation of the PROC MIXED procedure from the SAS system, Littell et al.1996 or by BMDP5V (Dixon 1988); Pinheiro and Bates (2000); Galecki and Burzykowski;… In this first approach, the data are supposed to follow a normal distribution and the used method of estimate, is the maximum likelihood. However, when the data are not normally distributed, it is preferable to use other models and other methods of estimate. A recent method introduced by Liang and Zeger 1986, consists of using the marginal distributions of the data at each moment; the data are not supposed to follow a gaussian distribution but rather a law belonging to the exponential families. The method of resolution is based on solving equations known as generalized estimating equations (GEE). The implementation of this method is in the GENMOD procedure of the SAS system, Littell et al.1996. Several works has been done thereafter in this field. One can quote those of Zhao and

Prentice 1990; Prentice and Zhao 1991; Fitzmaurice and Laird 1993; Park 1993; Crowder 1995; Crowder 2001; McCulloch and Searle; Jiang; Stroup... In this paper, we are interested, on the one hand, in the estimate of the mixed linear model parameters; on the other hand, in the resolution of the generalized estimating equations. In section2, we consider a mixed linear model. In section 3, we introduce the generalized estimating equations and their method of resolution. We conclude, in section 4, by showing the relationship between the maximum likelihood method in the generalized linear models framework and the resolution of the generalized estimating equations method named as the iteratively reweighted least squares (IRLS).

2. Model

Let us consider the mixed linear model:

$$y_i = X_i\alpha + Z_i b_i + e_i \tag{1}$$

Where y_i are observations on the ith individual? α: vector of unknown population parameters, of dimension (p×1). X_i: a design matrix linking α to y_i, known, of dimension (n_i×p) with n_i the individual number of observations. b_i: unknown random vector of the individuals effects, of dimension (k×1). Z_i: a design matrix linking y_i to b_i, known, of dimension (n_i×k). The errors: are independent variables. These models are often called two-stage random-effects models with the stages as follows:

Stage1: For each individual unit $y_i = X_i\alpha+Z_i b_i+e_i$, i=1,…,m. The variables e_i are independent and follows N(0,R_i). Here R_i is an ($n_i \times n_i$) positive-definite covariance matrix; α and b_i are considered fixed. In this stage, we are interested in the intra-individuals variation ('within individual variation') and which is formulated by R_i.

Stage2: The b_i are supposed independent and follow N(0,D). Here D is a (k×k) positive-definite covariance matrix; the α parameters are fixed. In this stage, we are interested in the inter-individuals variation ('between individual variation') and which is formulated by D.

The y_i are independent and follow N($X_i\alpha$, $Z_i D Z_i^T+R_i$). To simplify calculations, we can take R_i=σ²I (homoscedastic models), I is the $n_i \times n_i$ identity matrix. In our case, let us consider the general case by taking R_i and D unspecified (but diagonal), generated by an unknown parameter θ, of dimension (q×1), parameter such as: $R_i = R_i(\theta)$ and D = D(θ). A generalization of this mixed linear model was done by Jones and Boadi-Boateng 1991, where R_i is not necessarily diagonal; e_i is composed of an autoregressive component and a measurement error. In this paper, we are much more interested in the fixed effects estimate, by using the maximum likelihood or the restricted maximum likelihood for the mixed linear models and the resolution of the generalized estimating equations whose solution is an estimate of the fixed effects; consequently we do neither insist on the random effects nor on the parameter generating the variance-covariance matrix of the considered model.

2.1. Estimate of the Model Parameters

2.1.1. Estimation of α and b_i by Assuming that the Variance is known

The estimate of α is defined by the equality

$$\hat{\alpha} = \left(\sum_i X_i^T V_i^{-1} X_i\right)^{-1}\left(\sum_i X_i^T V_i^{-1} X_i Y_i\right) \tag{2}$$

Where

$$Var(y_i) = V_i = Z_i D Z_i^T + R_i, i = 1, \dots, m \tag{3}$$

The estimate of α is that maximizing the likelihood based on the marginal distributions of the data and it is with a minimum variance. The estimate of b_i is given by: $\hat{b_i} = D Z_i^T V_i^{-1}(Y_i - X_i\hat{\alpha})$. It is not the maximum likelihood estimate; but the empirical Bayes one's, given by: $\hat{b_i}$= E ($b_i / y_i, \hat{\alpha}, \theta$). Since $\hat{\alpha}$ and $\hat{b_i}$ are linear functions for Y_i, expressions of their standard errors can be easily calculated and are given by : $Var(\hat{\alpha}) = \left(\sum_i X_i^T V_i^{-1} X_i\right)^{-1}$

$$Var(\hat{b_i}) = D Z_i^T\{V_i^{-1} - V_i^{-1} X_i\left(\sum_i X_i^T V_i^{-1} X_i\right)^{-1} X_i^T V_i^{-1}\} Z_i D$$

2.1.2. Estimation of α and b_i Assuming that the Variance is Unknown

We estimate the variance or the parameter θ who generates it. There are two estimates for θ. The maximum likelihood (ML) estimate, noted by θ_M and the restricted maximum likelihood (REML) estimate, noted by θ_{REML}. To calculate these two estimates, we apply the EM algorithm (Dempster et al.1977; Dempster et al.1981).

Likelihood function: The log-likelihood noted by λ of the data $y_1, \dots, y_m$ is:

$$\lambda = \mathrm{Cst} - \left(\frac{1}{2}\right)\sum_i Log\,|V_i| - \left(\frac{1}{2}\right)\sum_i (y_i - X_i\alpha - Z_i b_i)^T V_i^{-1} (y_i - X_i\alpha - Z_i b_i)$$

Where |Vi| is the Vi determinant; cst represents a constant.

Restricted likelihood function: It is well-known that the maximum likelihood estimates of the variance components are biased. Indeed, in the estimate of the maximum likelihood we do not take into account that the fixed effect α is also estimated. The method introduced by Patterson and Thompson 1971, consists of modifying the likelihood in a restricted likelihood which takes into account the estimate of the fixed effect and which gives unbiased estimates of the variance components. The logarithm noted by λ_R of the restricted likelihood of the data $y_1, \dots, y_m$ is:

$$\lambda_R = \mathrm{Cst} - \left(\tfrac{1}{2}\right)\sum_i Log\,|V_i| - \left(\tfrac{1}{2}\right)\sum_i (y_i - X_i\alpha - Z_i b_i)^T V_i^{-1} (y_i - X_i\alpha - Z_i b_i) - \left(\tfrac{1}{2}\right)\sum_i Log\,\left|X_i^T V_i^{-1} X_i\right|$$

3. Marginal Models. Generalized Estimating Equations

3.1. Introduction

The first and the most traditional approach, is based on the maximum likelihood function but it proved to be insufficient because it leads to asymptotically biased estimators when the variance-covariance matrix is not completely specified. An improvement can be considered by using the restricted likelihood function. However, these two methods are based on knowledge of the distributions which are frequently normal (or belonging to the exponential families). Therefore, this maximum likelihood approach is applied when the data are approximately normal. The second approach that we will consider is applied for normal or not normal data, but it is often applied for the second case. This more recent methodology based on the generalized linear models (GLM), Mc Cullagh and Nelder1989, and on the estimate of quasi-likelihood (QL) developed by Wedderburn 1974, was a great alternative and which leads thereafter to the generalized estimating equations (GEE) developed by Liang and Zeger 1986. Many works were published thereafter in this context, which is that of the marginal model. We can quote, for example, Liang and Zeger 1986; Zeger, Liang and Albert 1988; Zhao and Prentice 1990; Prentice and Zhao 1991; Fitzmaurice and Laird 1993; Park 1993; Crowder 1995; Crowder 2001; among others.

3.2. Longitudinal Data Analysis Using the GLM

3.2.1. Definition of a Generalized Linear Model

Let us consider $y_1, \dots, y_N$ independent random variables, such as: y_i=μ_i+ε_i. A generalized linear model is defined by: a) a distribution law of the ε_i. b) A matrix X of dimension (N×t), of explanatory variables; this defines a linear predictor η=Xβ. c) a link function g, invertible such as g(μ)=Xβ.

3.2.2. Model

In this part, we propose another extension of the GLM to analyze longitudinal data; these data are not supposed to follow a gaussian probability distribution but we suppose at each moment only one form of the marginal distribution (Liang and Zeger 1986). Let us consider y_{it}: outcomes vector of the ith individual at the moment t, of dimension (p×1); t=1...,n_i; i=1...,K. x_{it} : covariates vector, of dimension (p×1). $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$: outcomes vector, of dimension $(n_i \times 1)$ and $X_i = (X_i \ , X_i \ , \dots, X_{in_i})^T$: matrix of the covariates values, of dimension $(n_i \times p)$ for the ith subject. Let us suppose that the marginal density of y_{it} is given by:

$$f(y_{it};\theta_{it};\emptyset) = exp\{[[y_{it}\theta_{it} - b(\theta_{it})]/a(\emptyset)] + c(y_i;\emptyset)\} \quad (4)$$

Where $\theta_{it} = h(\eta_{it})$, $\eta_{it} = x_{it}\beta$ and a; b; c known functions. The h function is monotonic and differentiable. With this formulation, we have (Mc Cullagh et Nelder 1989):

$E(y_{it}) = b'(\theta_{it})$; Var$(y_{it}) = b''(\theta_{it})/\phi$

Where primes in b′ and b′′ denotes first and second differentiation with respect to θ from the function b, which is supposed to be known. ϕ is a scale parameter.

3.2.3. Generalized Estimating Equations

Let us consider $V = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} a(\phi)$. The $R(\alpha)$ matrix which is of dimension (n×n), is called working correlation matrix. We suppose that the data are balanced $(n_i = n)$. The generalized estimating equations are given by:

$$\textstyle\sum_i D_i^T V_i^{-1} S_i = 0 \quad (5)$$

Where, $D_i = d\{b'(\theta_{it})\}/d\beta = A_i \Delta_i X_i$, where $A_i = diag(b''(\theta_{it}))$, matrix of dimension (nxn); $S_i = y_i$-a$'(\theta_{it})$. The β estimate, noted by $\hat{\beta}_G$ is the solution of equation (5).

3.3. Maximum Likelihood for GLM

Let us suppose that the marginal density of y_{it} is given by:

$$f(y_i;\theta_i;\emptyset) = exp\{[[y_i\theta_i - b(\theta_i)]/a(\emptyset)] + c(y_i;\emptyset)\}$$

The log-likelihood l_i is given by: $l_i(\theta_i, \emptyset; y_i) = \{[[y_i\theta_i - b(\theta_i)]/a(\emptyset)] + c(y_i;\emptyset)\}$

For N observations, we have: L(β)=$\sum_i l_i$.

$$\frac{\partial l_i}{\partial \beta_j} = \frac{y_i - \mu_i}{Var(y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}$$

The likelihood equations are then given by:

$$\sum_i \frac{y_i - \mu_i}{Var(y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}; j = 1, \dots, t$$

These likelihood equations, which are equivalent to the GEE equations of Liang and Zeger 1986, are nonlinear according to β. Their resolution to find the β estimator, noted by $\hat{\beta}$, requires iterative method, which will be discussed below. The algorithm that we will use is the Fisher Scoring algorithm and thus, the rate of convergence of β to $\widehat{\beta}$, depends on the information matrix. We know that for a generalized linear model, we have:

$$E\left(\frac{\partial^2 l_i}{\partial \beta_h \partial \beta_j}\right) = -E\left(\frac{\partial l_i}{\partial \beta_h}\right)\left(\frac{\partial l_i}{\partial \beta_j}\right) =$$

$$-E\left[\frac{y_i-\mu_i}{Var(y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ih} \frac{y_i-\mu_i}{Var(y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right] = -\frac{x_{ij}x_{ih}}{Var(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

Thus: $E\left(\frac{\partial^2 L(\beta)}{\partial \beta_h \partial \beta_j}\right) = -\sum_i \frac{x_{ij}x_{ih}}{Var(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2, j = 1, \dots, t$

The information matrix, whose elements are: $\left(\frac{\partial^2 L(\beta)}{\partial \beta_h \partial \beta_j}\right)$, is noted and given by : Inf=X′WX, where W is the diagonal matrix, of which elements according to the main diagonal, are :$W_i = \frac{1}{Var(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$

3.3.1. Algorithm of Resolution

We recall that the Newton-Raphson algorithm is given by:

$$\beta^{k+1} = \beta^k - (H^k)^{-1} . q^k$$

The H elements are: $\frac{\partial^2 L(\beta)}{\partial \beta_h \partial \beta_j}$. Those of q are: $\frac{\partial L(\beta)}{\partial \beta_j}$

The Fisher scoring algorithm is:

$$\beta^{k+1} = \beta^k + (\mathrm{Inf}^k)^{-1} . q^k \quad (6)$$

From where:

$$\mathrm{Inf}^k \beta^{k+1} = \mathrm{Inf}^k \beta^k + q^k \quad (7)$$

Where the Inf^k elements, are given by: $E(-\frac{\partial^2 L(\beta)}{\partial\beta_h \partial\beta_j})$, evaluated at β^k. The right part of the equation (7) is the vector whose elements are given by:

$$\sum_j \left[\sum_i \frac{x_{ij}x_{ih}}{Var(y_i)} (\frac{\partial\mu_i}{\partial\eta_i})^2 \beta_j^k\right] + \sum_i \frac{(y_i - \mu_i{}^k)}{Var(y_i)} x_{ih} \frac{\partial\mu_i}{\partial\eta_i}.$$

Thus $\beta^k + q^k = X'W^k z^k$, where z^k is the vector whose elements are: $z^k = \sum_j x_{ij}\beta_j^k + (y_i - \mu_i{}^k)\frac{\partial\eta_i{}^k}{\partial\mu_i{}^k} = \eta_i{}^k + (y_i - \mu_i{}^k)\frac{\partial\eta_i{}^k}{\partial\mu_i{}^k}$. The equation (7) becomes $(X'W^k X)\beta^{k+1} = X'W^k z^k$. These equations are the normal equations in the weighted least squares method to fit a linear model, with z^k like dependent variable; X, the model matrix and the matrices W^k are the weights. The solutions of these equations are given by:

$\beta^{k+1} = (X'W^k X)^{-1}(X'W^k z^k)$. The z vector is a linearized form of the link function at μ, evaluated at y and is given by: $g(y_i) \cong g(\mu_i) + (y_i - \mu_i)g'(\mu_i) = \eta_i + (y_i - \mu_i)\frac{\partial\eta_i}{\partial\mu_i} = z_i$. The variable z has ith element approximated by $z_i^k$for the kth cycle of the iterative scheme. At this cycle, we make a regression of z^k on X with weights W^k to obtain a new estimate β^{k+1}. This estimate yields a new linear predictor value and which is given by $\eta^{k+1} = X\beta^{k+1}$and a new value for the dependent variable, z^{k+1} for the next cycle. The maximum likelihood estimate is the limit of β^kas k→∞. we remark that the estimator of the generalized linear model is equivalent at that of weighted least square but modified since the weight change at each cycle of the process, this last method is called iterative reweighted least squares (IRLS). To begin the iterative process, we take y as initial value of the μ estimate. So we have the initial estimate of the weight matrix W and then, the initial estimator of β.

4. Conclusion

We outlined the maximum likelihood method applied to a mixed linear model. Thereafter, we introduced the generalized linear models (GLM) as well as the generalized estimating equations (GEE). We showed, on the one hand, that the maximum likelihood equations for the GLM are the same as that of Liang and Zeger (the GEE equations). On the other hand, we showed the existing relationship between the maximum likelihood for the GLM and the GEE resolution method which is the iteratively reweighted least squares ones (IRLS).

Acknowledgments

The author is very grateful to the anonymous referee for his (or her) useful comments, corrections and suggestions, which greatly improved the presentation of the paper.

References

[1] E. M. Chi, and G. C. Reinsel, "Models for longitudinal data with random effects and AR(1) errors," JASA, Theory and Methods, Vol.84, No.406. pp. 452-459, 1989.

[2] M. Crowder, "On the use of a working matrix correlation in using GLM for repeated measures," Biometrika, 82, vol. 2, pp. 407-410, 1995.

[3] M. Crowder, "On repeated measures analysis with misspecified covariance structure," JRSS, B 63, part1, pp. 55-62, 2001.

[4] A. P. Dempster, N. M. Laird, and R. B. Rubin, "Maximum likelihood with incomplete data via the E-M algorithm," JRSS, B39, pp. 1-38, 1977.

[5] A. P. Dempster, D. B. Rubin, and R. K. Tsutakawa, "Estimation covariance component models," JASA, vol. 76, pp. 341-353, 1981.

[6] P. J. Diggle, K. Y. Liang, and S. L. Zeger, "Analysis of longitudinal data," Oxford science publications, 1994.

[7] W. J. Dixon, "BMDP statistical software," manual volume 2, University of California press: Berkley, CA, 1988.

[8] G. M. Fitzmaurice, and N.M. Laird, "A likelihood based method for analysing longitudinal binary responses," Biometrika, vol. 80, 1, pp. 141-151, 1993.

[9] J. L. Foulley, F. Jaffrézic, and C. R. Granié, "EM-REML estimation of covariance parameters in gaussian mixed models for longitudinal data analysis," Genet. Sel. Evol. Vol. 32, 129-141, 2000.

[10] A. Galecki and T. Burzykowski, ''Linear Mixed-effects Models using R," Springer, 2012.

[11] D. A. Harville, "Maximum likelihood approaches to variance components estimation and to related problem," JASA, Vol.72, No.358, pp. 320-329, 1977.

[12] R. I. Jennrich, and M. D. Scluchter, "Unbalanced Repeated-Measure Models with Structured Covariance Matrices," Biometrics, vol. 42, pp. 805-820, 1986.

[13] J. Jiang, "Linear and Generalized Linear Mixed models and their applications," Springer, 2007.

[14] H. R. Jones, and Boadi-Boateng, "Unequally spaced longitudinal data with AR(1) serial correlation," Biometrics, vol. 47, pp. 161-175, 1991.

[15] N. M. Laird, and J. H. Ware, "Random effect models for longitudinal data," Biometrics, 38, pp. 963-974, 1982.

[16] K. Y. Liang, and S. Zeger, "Longitudinal data analysis using GLM," Biometrika, vol. 73, 1, pp. 13-22, 1986.

[17] M. J. Lindstrom and D. M. Bates, "Newton-Raphson and E-M algorithm for linear mixed effect models for repeated - measures data," JASA, V83, N 404, pp. 1014-1022, 1988.

[18] R. C. Littell, G. A. Milliken, W. W. Stroup, and R. D. Wolfinger, "System for mixed models," SAS institute inc: Cary, N.C, 1996.

[19] R. C. Littell, J. Pendergast, and R. Natarajan, “Modelling covariance Structure in the analysis of repeated measures data,” Statist. Med. V19, pp. 1793-1819, 2000.

[20] P. McCullagh, and J. A. Nelder, “Generalized linear model,” 2nd edition. Chapman and Hall London, 1989.

[21] C. E. McCulloch, and S. R. Searle, “Generalized Linear Mixed Models”, Wiley 2001.

[22] T. A. Park, “A comparison of the GEE approach with the maximum likelihood approach for repeated measurements,” Stat. Med. V12, pp. 1723-32, 1993.

[23] T. Park, and Y. J. Lee, “Covariance models for nested measures data: analysis of ovarian steroid secretion data,” Statist. Med. V21, pp. 143-164, 2002.

[24] H. D. Patterson, and R. Thompson, “Recovery of inter-block information when block sizes are unequal,” Biometrika, vol. 58, pp. 545-554, 1971.

[25] J. C. Pinheiro, and D. M. Bates, “Mixed Effect Models in S and S-Plus,” Springer 2000.

[26] R. L. Prentice, and L. P. Zhao, “Estimating equations for parameters in means and covariance of multivariate discrete and continuous responses,” Biometrics 47, pp. 825-839, 1991.

[27] G. Verbeke, and G. Molenberghs, “Linear Mixed Models for longitudinal data. New York: Springer, 2000.

[28] R. W. M. Wedderburn, “Quasi-likelihood function; GLM and the Gauss-Newton method,” Biometrika 61, 3, pp. 439-447, 1974.

[29] V. Witkovsky, “Matlab algorithm mixed.m for solving Henderson's mixed model equations,” Institute of Measurement Science Slovak Academy of Sciences, 2002.

[30] L. P. Zhao, and R. L. Prentice, “Correlated binary regression using a quadratic exponential model,” Biometrika 77, pp. 642-648, 1990.

[31] S. L. Zeger, K. Y. Liang, and P. S. Albert, “Models for longitudinal data: a generalized equation approach,” Biometrics 44, pp. 1049-1060, 1988.