MAXIMUM LIKELIHOOD FOR GENERALIZED LINEAR MODEL AND GENERALIZED ESTIMATING EQUATIONS

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

ABSTRACT
The most frequent methods to analyze statistical data are the regression methods, whereby the maximum likelihood method or that of least squares. On the one hand, in the context of the longitudinal data or repeated measurements, these data are often unbalanced; on the other hand, their law of probability are not normal rending the passage to other models with other methods of estimate possible. In this paper, we initially recalled the maximum likelihood method applied to a mixed linear model; thereafter, we introduced the generalized estimating equations and their method of resolution. Finally, we showed the relationship between the maximum likelihood method and the iteratively reweighted least squares ones.

Key words: IRLS, GEE equations, GLM, Maximum Likelihood, Mixed Linear Model.

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; 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). 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 Liang and Albert 1988; Zhao and Prentice 1990; Prentice and Zhao 1991; Liang et al.1992; Fitzmaurice and Laird 1993; Park 1993; Crowder 1995; Crowder 2001... 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:
\[ y_i = X_i \alpha + Z_i b_i + e_i \]  
(1)
y_i: observations on the ith individual. \( \alpha \): vector of unknown population parameters, of dimension (p×1). X_i: a design matrix linking \( \alpha \) to \( y_i \), known, of dimension (n_i×p) with n_i the 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×p). The errors: are independent variables. These models are often called two-stage random-effects models

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 follow N(0,R_i). Here \( R_i \) is an \((n_i \times n_i)\) positive-definite covariance matrix; \( \alpha \) 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 \((k \times k)\) positive-definite covariance matrix; the \( \alpha \) 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 \( R_i = \sigma^2 I \), I is the \( n_i \times n_i \) identity matrix; the errors are homogeneous. In our case, let us consider the general case by taking \( R_i \) and D unspecified (but diagonal), generated by an unknown parameter \( \theta \), of dimension (q×1), parameter such as: \( R_i = R_i(\theta) \) and D = D(\( \theta \)). 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 not 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 \( \alpha \) and b_i by supposing that the variance is known
\[ Var(y_i) = V_i = Z_i D Z_i^T + R_i, i = 1, ..., m \]  
(2)
\[ \hat{\alpha} = (\sum_i X_i^T V_i^{-1} X_i)^{-1} (\sum_i X_i^T V_i^{-1} Y_i) \]  
(3)
The estimate of \( \alpha \) 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: \( \hat{\alpha} = E (b_i/\hat{y}_i, \hat{\alpha}, \theta) \). 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: \[ Var(\hat{\alpha}) = (\sum_i X_i^T V_i^{-1} X_i)^{-1} \]  
\[ Var(\hat{b}_i) = D Z_i^T \{ V_i^{-1} - V_i^{-1}X_i(\sum_i X_i^T V_i^{-1} X_i)^{-1}X_i^T V_i^{-1}\}Z_i D \]

2.1.2 Estimation of \( \alpha \) and b_i by supposing that the variance is unknown
We estimate the variance or the parameter \( \theta \) who generates it. There are two estimates for \( \theta \). The maximum likelihood (ML) estimate, noted by \( \theta_M \) and the restricted maximum likelihood (REML) estimate, noted by \( \theta_{REML} \). 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 \( \lambda \) of the data \( y_1, ..., y_m \) is:
\[ \lambda = -\left(\frac{1}{2}\right) \sum_i Log |V_i| - \left(\frac{1}{2}\right) \sum_i (Y_i - X_i \alpha - Z_i b_i)^T V_i^{-1} (Y_i - X_i \alpha - Z_i b_i) \]  
\[ |V_i| \] 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 \( \alpha \) 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 \( \lambda_R \) of the restricted likelihood of the data \( y_1, ..., y_m \) is:

\[
\lambda_R = -\left( \frac{1}{2} \right) \sum_i \log |V_i| - \left( \frac{1}{2} \right) \sum_i (y_i - X_i\alpha - Z_i b_i)^T V_i^{-1} (y_i - X_i\alpha - Z_i b_i) - \left( \frac{1}{2} \right) \sum_i \log |X_i^T V_i^{-1} X_i|
\]

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), McCullagh 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 and Liang 1986; Zeger, Liang and Albert 1988; Zhao and Prentice 1990; Prentice and Zhao 1991; Liang et al.1992; 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 \( y_1, ..., y_N \) independent random variables, such as: \( y_i=\mu_i+\epsilon_i \). A generalized linear model is defined by: a) a distribution law of the \( \epsilon_i \), b) A matrix \( X \) of dimension \((N\times t)\), of explanatory variables; this defines a linear predictor \( \eta=X\beta \). c) a link function \( g \), invertible such as \( g(\mu)=X\beta \).

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 \( i \)th individual at the moment \( t \), of dimension \((p\times 1)\); \( i=1,...,n_i \); \( t=1,...,K \). \( x_{it} \) : covariates vector, of dimension \((p\times 1)\). \( y_{it} = (y_{i1}, y_{i2}, ..., y_{in_i})^T \): outcomes vector, of dimension \((n_i \times 1)\) and \( X_i = (x_{i1}, x_{i2}, ..., x_{in_i})^T \): matrix of the covariates values, of dimension \((n_i \times p)\) for the \( i \)th subject. Let us suppose that the marginal density of \( y_{it} \) is given by:

\[
f(y_{it}; \theta_{it}; \phi) = \exp\left\{ [y_{it} - b(\theta_{it})]/a(\phi) \right\} + c(y_{it}; \phi)
\]

Where \( \theta_{it} = h(\eta_{it}) \), \( \eta_{it} = x_{it} \beta \), the \( h \) function is monotonic and differentiable. With this formulation, we have (Mc Cullagh et Nelder 1983):

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

Where primes in \( b' \) and \( b'' \) denotes first and second differentiation with respect to \( \theta \) from the function \( b \), which is supposed to be known. \( \phi \) a scale parameter.
3.2.3 Generalized estimating equations

Let us consider $V = A^2 R(a) A^2$, where $R(a)$ is the working correlation matrix. We suppose that the data are balanced ($n_i = n$). The generalized estimating equations are given by:

$$
\sum_i D_i^T V_i^{-1} S_i = 0
$$

(5)

Where, $D_i = d(W(\theta_i))/d\beta = A_i \Delta_i X_i$, where $A_i = diag(\beta^\top \phi_i)$, matrix of dimension (nxn). The $\beta$ 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:

$$
f(y_{it}; \theta; \phi) = \exp\{[y_{it} - b(\theta_i)]/a(\phi) + c(y_{it}; \phi)\}
$$

The log-likelihood $l_i$ is given by:

$$
l_i(\theta_i; \phi; y_i) = \{[y_{it} - b(\theta_i)]/a(\phi) + c(y_{it}; \phi)\}
$$

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

$$
\frac{\partial l_i}{\partial \beta} = \frac{y_{it} - \mu_i}{\text{Var}(y_{it})} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}
$$

The likelihood equations are then given by:

$$
\sum_{i} \frac{y_{it} - \mu_i}{\text{Var}(y_{it})} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = 0, \quad j = 1, ..., t
$$

These likelihood equations, which are equivalent to the GEE equations of Liang and Zeger, 1986, are nonlinear according to $\beta$. Their resolution to find the $\beta$ 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 $\beta$ to $\hat{\beta}$, depends on the information matrix. We know that for a generalized linear model, we have:

$$
E\left(\frac{\partial^2 L(\beta)}{\partial \theta_h \partial \beta_j}\right) = -E\left(\frac{\partial l_i}{\partial \beta_j}\right) = -E\left[\frac{y_{it} - \mu_i}{\text{Var}(y_{it})} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} - \frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right] = -\frac{x_{ij} x_{ik}}{\text{Var}(y_{it})} \left(\frac{\partial \mu_i}{\partial \eta_i} x_{ij}\right)^2
$$

Thus: $E\left(\frac{\partial^2 L(\beta)}{\partial \theta_h \partial \beta_j}\right) = -\sum_{i} \frac{x_{ij} x_{ik}}{\text{Var}(y_{it})} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2, \quad j = 1, ..., t$

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

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: $\frac{\partial^2 L(\beta)}{\partial \theta_h \partial \beta_j}$. Those of $q$ are: $\frac{\partial L(\beta)}{\partial \beta_j}$

The Fisher scoring algorithm is:

$$
\beta^{k+1} = \beta^k + (\text{Inf}^k)^{-1} q^k
$$

(6)

From where:

$$
\text{Inf}^k \beta^{k+1} = \text{Inf}^k \beta^k + q^k
$$

(7)
Where the Inr\(k\) elements, are given by: 
\[ E\left(-\frac{\partial^2 L(\hat{\beta})}{\partial \hat{\beta}_k \partial \hat{\beta}_j}\right) \]
the vector whose elements are given by:
\[
\sum_j \left[ \sum_i \frac{x_i x_{ik}}{\text{Var}(y)} \frac{\partial \mu_{ik}}{\partial \eta_{ij}} z_j^k \right] + \sum_i \left( \frac{y_i - \mu_i^k}{\text{Var}(y)} \right) x_i \frac{\partial \mu_i^k}{\partial \eta_{ij}}. 
\]
Thus \( \beta^k + q^k = X'W^k z^k \), where \( z^k \) is the vector whose elements are:
\[
z^k = \sum_j x_j y_{i,j}^k + (y_i - \mu_i^k) \frac{\partial \eta_i^k}{\partial \mu_i} = \eta_i^k + (y_i - \mu_i^k) \frac{\partial \eta_i^k}{\partial \mu_i}. 
\]
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 \( \mu \), evaluated at \( y \) and is given by:
\[
g(y_i) \equiv g(\mu_i) + (y_i - \mu_i) g'(\mu_i) = \eta_i + (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} = z_i. 
\]
The variable \( z \) has \( i \)th element approximated by \( z_i^k \) for the \( k \)th cycle of the iterative scheme. At this cycle, we make a regression of \( z_i^k \) on \( X \) with weights \( W^k \) to obtain a new estimate \( \beta^{k+1} \). This estimate yields a new linear predictor value and which is given by \( \eta_i^{k+1} = X \beta^{k+1} \) and a new value for the dependent variable, \( z_i^{k+1} \) for the next cycle. The maximum likelihood estimate is the limit of \( \beta^k \) as \( k \to \infty \), 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 \( \mu \) estimate. So we have the initial estimate of the weight matrix \( W \) and then, the initial estimator of \( \beta \).

4 CONCLUSIONS
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).

REFERENCES