



# A Quasi-Score Marginal Approach in Generalized Linear Mixed Models

Catherine Trottier

## ► To cite this version:

Catherine Trottier. A Quasi-Score Marginal Approach in Generalized Linear Mixed Models. RR-3522, INRIA. 1998. inria-00073162

**HAL Id: inria-00073162**

**<https://inria.hal.science/inria-00073162>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***A Quasi-Score Marginal Approach  
in Generalized Linear Mixed Models***

Catherine Trottier

**No 3522**

Octobre 1998

————— THÈME 4 —————



***apport  
de recherche***



# A Quasi-Score Marginal Approach in Generalized Linear Mixed Models

Catherine Trottier

Thème 4 — Simulation et optimisation  
de systèmes complexes  
Projet IS2

Rapport de recherche n° 3522 — Octobre 1998 — 20 pages

**Abstract:** This report deals with the problem of parameter estimation in generalized linear mixed models. Gilmour, Anderson, et Rae (1985) proposed a method of estimation in a probit link model for binomial data. Foulley et Im (1993) adapted this method to Poisson data in a log link model. We propose a unifying formal description including also the case of exponential data in a log link model. This approach enables us to consider other cases such as logit link model for binomial data. Numerical examples are given to illustrate the method.

**Key-words:** Generalized linear mixed model, Variance components, Conditional and marginal model, Quasi-score, Probit and logit link.

*(Résumé : tsvp)*

# Une approche marginale par quasi-score dans les modèles linéaires généralisés mixtes

**Résumé :** Dans ce rapport nous nous intéressons à la question de l'estimation des paramètres d'un modèle linéaire généralisé mixte. Gilmour, Anderson, et Rae (1985) ont proposé une méthode d'estimation pour des données binomiales dans un modèle avec lien probit. Foulley et Im (1993) ont adapté cette méthode à des données poissonniennes dans un modèle avec lien logarithme. Nous proposons une description formelle permettant d'unifier ces deux cas et d'inclure aussi le cas de données exponentielles dans un modèle avec un lien logarithme. Cette approche nous permet d'envisager d'autres cas tels que celui de données binomiales dans un modèle avec lien logit. Des exemples numériques sont donnés pour illustrer cette méthode.

**Mots-clé :** Modèle linéaire généralisé mixte, Composantes de la variance, Modèle conditionnel et modèle marginal, Quasi-score, Lien probit et lien logit.

## 1 Introduction

Generalized linear models constitute a major tool for the analysis of non Gaussian data. When including random effects in the modeling hypothesis, they have led to the class of generalized linear mixed models. This latter category of models has been given an always greater practical interest during the last two decades. This is the case for instance in the field of animal breeding where random effects represent genetic values of animals and where the measurements made may even not be continuous. In this type of model then, the estimation of fixed effects and of variance components (giving an information about the part of variation due to the random effects), and sometimes the prediction of random effects are needed.

Parameter estimation in generalized linear mixed models is a problem which has been tackled according to several approaches but rarely from a general point of view on this class of models. In these models actually, distribution assumptions are made conditionally on the non observed random effects. Thus in order to describe the likelihood, an integral calculus has to be made which is not analytically feasible except for the case where the random effect distribution is assumed to be conjugate. Here we concentrate on normal distribution assumptions for these random effects. Therefore several kind of approximations have been considered. A first approach consists in a numerical approximation of the integral by a Gaussian quadrature, see for instance Anderson et Aitkin (1985). Even with the great development of computation capacities, this seems to be limited to the cases of overdispersion modeling (with one random effect realization per data) or of nested random effects. A second approach suggested by Schall (1991) lies in an iterative approximation of the model conditionally on the random effects. Breslow et Clayton (1993), McGilchrist (1994) ended up with the same method by means of different justifications. The proposition of Gilmour, Anderson, et Rae (1985) (we will denote GAR method in the following) is an alternative approach to those two, by suggesting the calculation of the first two marginal moments and the maximization of the quasi-score function after having approximated the variance matrix. A similar method has been proposed by Breslow et Clayton (1993) with a somewhat different marginal model. Considering random effects as missing data, the EM algorithm has also been adopted (Hinde 1982, Anderson et Aitkin 1985, Jansen et

Hoekstra 1993, Anderson et Hinde 1988) but this seems to be limited as above when calculating the integral in the E-step.

Compared to the method proposed by Schall (1991), the GAR method adopts clearly a different point of view on parameter estimation in generalized linear mixed models. Whereas the former method works on the conditional model, using first its conditional status as generalized linear model ; the latter tends more to lay hands on the marginal model by calculating marginal moments. Thus the integration out of the random effect is not considered at the same stage by the two methods. As the Schall method does not specify any particular model and can be used in all the cases, it is important to investigate the extension of GAR method to more cases than the two cases developed up to now. Thus the aim of this article is to extend the use of the approach developed by Gilmour, Anderson, et Rae (1985). Foulley et Im (1993) already proposed the extension of GAR method to Poisson data with a log link. We suggest here an other formulation of the approach leading to the same method in these two cases but giving the possibility to consider the estimation of parameters in other models.

In Section 2, after a description of the model hypothesis, we present our reading of the original approach, introducing at the same time some notation. We then propose our extension of the method making use of a different approach. The object of Section 3 is to consider the particular case of a log link exponential model according to both approaches and then the one of a logit link binomial model for which our approach is necessary. In Section 4, we give numerical results, first comparing the probit link and logit link procedures on a small data set and then observing the behaviour of this method on simulations.

## 2 The estimation method

### 2.1 Model hypothesis

Let  $y$  be the  $N \times 1$  vector of observations (realization of the random vector  $Y$ ). Given the unobserved  $q \times 1$  vector of random effects  $\xi$ , the generalized linear mixed model is defined as a generalized linear model with the three usual assumptions (McCullagh et Nelder 1989) :

- the  $Y_i$ 's ( $i = 1, \dots, N$ ) are assumed to be conditionally independent, with distribution from the exponential family with conditional mean

$$E(Y_i|\xi) = \mu_{\xi,i} ,$$

and conditional variance

$$\text{var}(Y_i|\xi) = v(\mu_{\xi,i}) ,$$

where  $v$  is the variance function of the distribution including sometimes an unknown dispersion parameter  $\phi$ .

- explanatory vectors  $x'_i$  and  $u'_i$  are the  $i^{\text{th}}$  lines of  $X$  and  $U$  matrices for respectively fixed and random effects and enter the model through the linear predictor

$$\eta_\xi = X\beta + U\xi ,$$

$\beta$  being the  $p \times 1$  fixed effects parameter vector.  $X$  and  $U$  are respectively  $N \times p$  and  $N \times q$  known matrices.  $X'$  means  $X$  transpose.

- the link between the conditional mean  $\mu_\xi$  and the conditional predictor  $\eta_\xi$  is done through the link function  $g$  ( $h = g^{-1}$ ) :

$$\eta_\xi = g(\mu_\xi) .$$

Concerning the random effects, we assume in the following  $J$  random effects  $\xi = (\xi'_1, \dots, \xi'_J)'$  where the  $q_j \times 1$  ( $q = \sum_{j=1}^J q_j$ ) vector  $\xi_j$  concerns the  $j^{\text{th}}$  random effect with variance component  $\sigma_j^2$  :

$$\xi_j \sim \mathcal{N}(0, \sigma_j^2 A_j) ,$$

with  $A_j$  a known matrix. So  $\xi \sim \mathcal{N}(0, G)$  where the block-diagonal matrix  $G$  is defined by  $G = \text{diag}\{\sigma_j^2 A_j\}_{j=1, \dots, J}$ .

## 2.2 A review of the GAR method

The method developed by Gilmour, Anderson, et Rae (1985) consists in the maximization of the quasi-score function built on the first two marginal moments. In this section we present our interpretation of this approach relaxing



one of the hypothesis. We consider the two models up to now investigated with this methodology : the binomial probit link model, model **M1** and the Poisson log link model, model **M2**; and we introduce some notation to unify the approach in both cases. In these two models, we have :

**M1** :  $v(\mu_{\xi,i}) = \mu_{\xi,i}(1 - \mu_{\xi,i})/n_i$  and  $g = \Phi^{-1}$  where  $\Phi$  is the standard normal cumulative distribution function (we work here with the frequencies  $Y_i$  i.e. when  $O_i \sim \mathcal{Bin}(n_i, \mu_i)$  we define  $Y_i = O_i/n_i$ ),

**M2** :  $v(\mu_{\xi,i}) = \mu_{\xi,i}$  and  $g = \log$ .

In order to set up a method for parameter estimation, Gilmour, Anderson, et Rae (1985) and Foulley et Im (1993) adopted the following four steps reasoning.

- Step 1 : as it is difficult in generalized linear mixed models (except for the Gaussian case) to get the marginal density function of the observations, the authors propose to only calculate the first two marginal moments. This gives for the marginal mean :

**M1** :  $\mu_i = E(Y_i) = \Phi\{x'_i\beta/(1 + \alpha_i^2)^{\frac{1}{2}}\}$ ,

**M2** :  $\mu_i = E(Y_i) = \exp(x'_i\beta + \alpha_i^2/2)$ ,

where  $\alpha_i^2 = u'_iGu_i$ . The parameter  $\alpha_i^2$  often coincides with a variance component parameter insofar as the lines of  $G$  are often composed of a one and zeros. We do not assume, unlike Gilmour et al. and Foulley and Im, the  $\alpha_i^2$ 's to be constant. This prevents us from setting up  $\tilde{\beta} = \beta/(1 + \alpha^2)^{\frac{1}{2}}$  and from analyzing this integration as a change in scale for the fixed effects as suggested by Foulley et Manfredi (1991).

We introduce the notation  $\eta = X\beta$  which we call the marginal linear predictor, and

$\eta^* = M\eta + C$  the h-marginalized linear predictor.  $M$  is the diagonal matrix containing the multiplicative effect due to the integration of the random effects, and  $C$  is an off-set vector. The previous expressions for the marginal mean can then be written

$$\mu = E(Y) = h(\eta^*) ,$$

with :

**M1** :  $M = \text{diag}\{1/(1 + \alpha_i^2)^{\frac{1}{2}}\}_{i=1,\dots,N}$  and  $C = 0_N$ ,

**M2** :  $M = I_N$  and  $C = (\alpha_0^2/2, \dots, \alpha_N^2/2)'$ .

For the marginal variance matrix  $V$  of  $Y$ , we have :

$$V = V_0 + V_1 ,$$

where

$$\begin{aligned} \mathbf{M1} : \quad V_0 &= \text{diag}\{[\mu_i(1 - \mu_i) - \{\Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi(\eta_i^*)^2\}]/n_i\}_{i=1, \dots, N} \\ V_1 &= \{\Phi_2(\eta_i^*, \eta_j^*, t_{ij}) - \Phi(\eta_i^*)\Phi(\eta_j^*)\}_{i,j=1, \dots, N} \end{aligned}$$

with  $t_{ij} = u_i'Gu_j / \{(1 + \alpha_i^2)^{\frac{1}{2}}(1 + \alpha_j^2)^{\frac{1}{2}}\}$  and  $\Phi_2$  is the cumulative distribution function of the bivariate standard normal distribution :  $\Phi_2(y_1, y_2, \rho) = P[(Y_1 \leq y_1) \cap (Y_2 \leq y_2)]$ , where  $Y_1 \sim \mathcal{N}(0, 1)$ ,  $Y_2 \sim \mathcal{N}(0, 1)$  and  $\text{corr}(Y_1, Y_2) = \rho$ .

$$\begin{aligned} \mathbf{M2} : \quad V_0 &= \text{diag}\{\mu_i\}_{i=1, \dots, N} \\ V_1 &= \{\mu_i\mu_j\{\exp(u_i'Gu_j) - 1\}\}_{i,j=1, \dots, N}. \end{aligned}$$

• Step 2 : let us define the following diagonal matrices  $K = \text{diag}\{h'(\eta_i^*)\}_{i=1, \dots, N}$  and  $L = KM$ . When variance components are known and with the two moments  $\mu$  and  $V$  derived from step 1, fixed effect parameters can be estimated by maximizing the quasi-likelihood, using the following iterative algorithm, see McCullagh et Nelder (1989) ( $t$  denotes the iteration step)

$$(X'W^{[t]-1}X)\beta^{[t+1]} = X'W^{[t]-1}\zeta^{[t]} \quad (1)$$

where  $\zeta^{[t]} = X\beta^{[t]} + L^{[t]-1}(y - \mu^{[t]})$  and  $W^{[t]} = L^{[t]-1}V^{[t]}L^{[t]-1}$ .

• Step 3 : in practice variance components are unknown. The previous procedure cannot therefore be directly used. To face this problem, the authors propose to approximate the variance matrix  $V$  and to recognize in this approximation a structure which enables the estimation of fixed effects and variance components parameters. For small values of variance components,  $V$  is approximated by

$$\tilde{V} = \tilde{V}_0 + \tilde{V}_1$$

$$\begin{aligned} \mathbf{M1} : \quad \tilde{V}_0 &= \text{diag}\{[\mu_i(1 - \mu_i) - t_{ii}\varphi(\eta_i^*)^2]/n_i\}_{i=1, \dots, N} \\ \tilde{V}_1 &= \left\{ \varphi(\eta_i^*)(1 + \alpha_i^2)^{-\frac{1}{2}} u_i'Gu_j \varphi(\eta_j^*)(1 + \alpha_j^2)^{-\frac{1}{2}} \right\}_{i,j=1, \dots, N} , \end{aligned}$$

where  $\varphi$  is the standard normal density function.

$$\begin{aligned} \mathbf{M2} : \quad & \tilde{V}_0 = \text{diag}\{\mu_i\}_{i=1,\dots,N} \\ & \tilde{V}_1 = \{\mu_i\mu_j \quad u_i'Gu_j\}_{i,j=1,\dots,N} . \end{aligned}$$

This approximation of  $V$  can thus be written in both cases :

$$\tilde{V} = \tilde{V}_0 + LUGU'L .$$

• Step 4 :  $V$  being approximated suchlike, this leads to the following approximation for  $W$  :

$$\tilde{W} = R + UGU' ,$$

where  $R = L^{-1}\tilde{V}_0L^{-1}$ . In  $\tilde{W}$  we recognize the variance structure of a linear mixed model. The maximum quasi-score estimation of  $\beta$  from (1) with  $\tilde{W}$  can be therefore obtained by solving iteratively the Henderson's mixed model equations associated with this variance structure

$$\begin{pmatrix} X'R^{[t]-1}X & X'R^{[t]-1}U \\ U'R^{[t]-1}X & U'R^{[t]-1}U + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \xi \end{pmatrix} = \begin{pmatrix} X'R^{[t]-1}\zeta^{[t]} \\ U'R^{[t]-1}\zeta^{[t]} \end{pmatrix} . \quad (2)$$

This is an iterative procedure setting up and solving at each step these equations with the current values of the parameters and for the working vector  $\zeta^{[t]}$ . As a by-product, we then get an estimation of the variance components (Patterson et Thompson 1971) and a prediction of the random effects.

### 2.3 A new approach for the GAR method

The approach of Section 2.2 is described for the two particular models **M1** and **M2**. When considering other types of models, we face the problem of the marginal moments calculation and above all the one of the marginal variance matrix  $V$  which is far from being easy. We thus propose to first approximate the conditional model and calculate the marginal moments only afterwards. This approximation leads to an easier marginal calculus and, as we will see, works out the same approximated marginal variance matrix  $\tilde{V}$ . As  $\tilde{V}$  is actually the matrix used in the procedure, we will eventually end up with the same estimates.

Let us assume that the observations  $y$  are realizations of a random vector  $\tilde{Y}$  which components are conditionally independent and with conditional mean

$$\tilde{\mu}_{\xi,i} = h(\eta_i^*) + h'(\eta_i^*) \tilde{u}_i' \xi ,$$

where  $\tilde{u}_i'$  is the  $i^{\text{th}}$  line of  $\tilde{U} = MU$ , and conditional variance

$$\text{var}(\tilde{Y}_i|\xi) = v(\tilde{\mu}_{\xi,i}) ,$$

with the same variance function  $v$  as in the original model.

We introduce the notation  $\tilde{\eta}_\xi = \eta^* + MU\xi$ , which we call the approximated conditional linear predictor. As we can see, it is obtained by adding to the h-marginalized linear predictor  $\eta^*$  the random part pre-multiplied by the same multiplicative effect matrix  $M$ . This can be interpreted for the conditional predictor  $\eta_\xi$  as a change in scale of the regressors :  $\tilde{X} = MX, \tilde{U} = MU$ . The conditional mean  $\tilde{\mu}_\xi$  so defined can be seen as a first order approximation of  $h(\tilde{\eta}_{\xi,i})$  at  $\eta_i^*$  for  $\xi$  close to zero.

Having approximated the conditional model, we then proceed to the marginal moments calculation in this new model on  $\tilde{Y}$ . We get for the marginal mean

$$E(\tilde{Y}_i) = h(\eta_i^*)$$

and for the marginal variance matrix

$$\begin{aligned} \tilde{V} &= \text{diag}\{E(v(\tilde{\mu}_{\xi,i}))\}_{i=1,\dots,N} + K\tilde{U}G\tilde{U}'K \\ &= \text{diag}\{E(v(\tilde{\mu}_{\xi,i}))\}_{i=1,\dots,N} + LUGU'L \end{aligned}$$

The first term of this sum gives

$$\begin{aligned} \mathbf{M1} : \quad E\{v(\tilde{\mu}_{\xi,i})\} &= E[\tilde{\mu}_{\xi,i}(1 - \tilde{\mu}_{\xi,i})/n_i] \\ &= [\Phi(\eta_i^*)(1 - \Phi(\eta_i^*)) - \varphi(\eta_i^*)^2(MU)_i'G(MU)_i]/n_i \\ \mathbf{M2} : \quad E\{v(\tilde{\mu}_{\xi,i})\} &= E[\tilde{\mu}_{\xi,i}] \\ &= \mu_i . \end{aligned}$$

We thus end up in this new approach with exactly the same marginal moments as the approximated marginal moments in the original approach. Constructing the quasi-scores on the basis of these moments, we obtain the same estimates as

described in Section 2.2. The approximated variance structure from the original model and the variance structure of the approximated model are the same. The procedure is defined similarly solving iteratively Henderson's equations.

In this new approach therefore, instead of calculating first the marginal moments and then approximating them, we first approximate the model and calculate the marginal moments afterwards. This is a way to avoid the direct calculation of  $V$  which is actually not useful. This is particularly interesting, as we will see below, when the marginal variance matrix cannot be calculated. The approximation is so done one step before. And by this, it encompasses for model **M1** and **M2** two different types of approximations when done further on. Nevertheless to define the approximated conditional model on  $\tilde{Y}$  the matrices  $K$ ,  $M$  and  $C$  have to be known. They are identified once the marginal expectation is calculated. This is for sure a limitation of this approach insofar as we are doing the underlying assumption that we are able to calculate the marginal mean and to write it  $E(Y_i) = h(\eta_i^*)$ . But as we will see in the model considered below, this hypothesis is verified.

### 3 Other models

#### 3.1 Exponential log link model

We here consider the model :

$$\begin{aligned} Y_i|\xi &\sim \mathcal{Exp}(\mu_{\xi,i}) \\ \text{and } \mu_{\xi,i} &= \exp(\eta_{\xi,i}) . \end{aligned}$$

In this case, as in the two previous cases, both approaches can be used. Indeed we do not encounter any special difficulty to calculate the marginal moments. For this model, we have :

$$\begin{aligned} E(Y_i|\xi) &= \mu_{\xi,i} = \exp(x'_i\beta + u'_i\xi) , \\ \text{and } \text{var}(Y_i|\xi) &= v(\mu_{\xi,i}) = \mu_{\xi,i}^2 . \end{aligned}$$

Thus following the original approach, we compute the marginal moments :

$$\begin{aligned} E(Y_i) &= \mu_i = \exp(x'_i\beta + u'_iGu_i/2) , \\ \text{var}(Y_i) &= \mu_i^2 \{2 \exp(u'_iGu_i) - 1\} , \\ \text{cov}(Y_i, Y_j) &= \mu_i\mu_j \{\exp(u'_iGu_j) - 1\} . \end{aligned}$$

So,  $V = V_0 + V_1$  with

$$\begin{aligned} V_0 &= \text{diag}\{\mu_i^2 \exp(u'_iGu_i)\}_{i=1,\dots,N} \\ V_1 &= \{\mu_i\mu_j (\exp(u'_iGu_j) - 1)\}_{i,j=1,\dots,N} . \end{aligned}$$

The approximation of this matrix leads to  $\tilde{V} = \tilde{V}_0 + \tilde{V}_1$  with

$$\begin{aligned} \tilde{V}_0 &= \text{diag}\{\mu_i^2(1 + u'_iGu_i)\}_{i=1,\dots,N} \\ \tilde{V}_1 &= \{\mu_i\mu_j u'_iGu_j\}_{i,j=1,\dots,N} . \end{aligned}$$

Thus, the approximated variance structure is  $\tilde{W} = R + UGU'$  where  $R = \text{diag}\{1 + u'_iGu_i\}_{i=1,\dots,N}$ .

Now, following our approach and since  $E(Y_i) = \exp(x'_i\beta + u'_iGu_i/2) = h(\eta_i^*)$ , we define  $M = I_N$ ,  $C = (u'_1Gu_1/2, \dots, u'_NGu_N/2)'$  and  $\eta^* = MX\beta + C$ . Therefore, we get  $K = \text{diag}\{\exp(x'_i\beta + u'_iGu_i/2)\}_{i=1,\dots,N}$ ,  $L = KM = K$  and the conditional approximated model is defined through

$$\begin{aligned} E(\tilde{Y}_i|\xi) &= \tilde{\mu}_{\xi,i} = (1 + u'_i\xi) \exp(x'_i\beta + u'_iGu_i/2) \\ \text{var}(\tilde{Y}_i|\xi) &= v(\tilde{\mu}_{\xi,i}) = (1 + u'_i\xi)^2 \exp\{2(x'_i\beta + u'_iGu_i/2)\} . \end{aligned}$$

We then obtain for the two marginal moments in the approximated model:

$$\begin{aligned} E(\tilde{Y}_i) &= \exp(x'_i\beta + u'_iGu_i/2) \\ \tilde{V} &= \tilde{V}_0 + LUGU'L \end{aligned}$$

$$\begin{aligned} \text{where } \tilde{V}_0 &= E[\text{diag}\{(1 + u'_i\xi)^2 \exp\{2(x'_i\beta + u'_iGu_i/2)\}\}_{i=1,\dots,N}] \\ &= \text{diag}\{(1 + u'_iGu_i) \exp(2x'_i\beta + u'_iGu_i)\}_{i=1,\dots,N} . \end{aligned}$$

In this approach, we end up with the same variance structure as in the first approach above

$$\begin{aligned} \tilde{W} &= L^{-1}\tilde{V}L^{-1} + UGU' \\ &= \text{diag}\{1 + u'_iGu_i\}_{i=1,\dots,N} + UGU' . \end{aligned}$$

In this case again, as for models **M1** and **M2** in Section 2.3, we verify that the two approaches lead to the same approximated marginal moments. And we may use the system (2) to get the expected estimates.

### 3.2 Binomial link model

We here consider the model :

$$\begin{aligned} O_i|\xi &\sim \mathcal{Bin}(n_i, \mu_{\xi,i}) \\ \text{and } \mu_{\xi,i} &= \text{logit}^{-1}(\eta_{\xi,i}) , \end{aligned}$$

where  $\text{logit}^{-1}(x) = \exp(x)/\{1 + \exp(x)\}$ . As in Section 2.2, we work here with the frequencies  $Y_i = O_i/n_i$ .

This is a case where the original approach cannot be used since the marginal variance matrix cannot be calculated explicitly and even the marginal expectation, as we will see thereafter, is approximated. Our approach is thus needed to propose a marginal procedure for parameter estimation in this model.

The conditional model hypothesis give

$$\begin{aligned} E(Y_i|\xi) &= \mu_{\xi,i} = \exp(\eta_{\xi,i})/\{1 + \exp(\eta_{\xi,i})\} \\ \text{var}(Y_i|\xi) &= v(\mu_{\xi,i}) = \mu_{\xi,i}(1 - \mu_{\xi,i})/n_i . \end{aligned}$$

Let us first consider the marginal expectation  $E(Y_i) = E[\exp(\eta_{\xi,i})/\{1 + \exp(\eta_{\xi,i})\}]$ , which cannot be directly calculated. This is the reason why we use the usual approximation of the logistic function (Zeger, Liang, et Albert 1988) in two ways. Indeed, we use it first by writing  $E(Y_i|\xi) \approx \Phi(c\eta_{\xi,i})$  where  $c$  can be set up to  $c = (3^{\frac{1}{2}}16)/(15\pi)$ . We then get

$$E(Y_i) \approx E[\Phi(c\eta_{\xi,i})] ,$$

which is equal to

$$\begin{aligned} E[\Phi(c\eta_{\xi,i})] &= E[\Phi(cx'_i\beta + cu'_i\xi)] \\ &= \Phi\{cx'_i\beta/(1 + c^2\alpha_i^2)^{\frac{1}{2}}\} \end{aligned}$$

where  $\alpha_i^2 = u_i' G u_i$  as in Section 2.2.

Now using once again this approximation of the logistic function, we finally have

$$E(Y_i) \approx \exp\{x_i' \beta / (1 + c^2 \alpha_i^2)^{\frac{1}{2}}\} / \left[ 1 + \exp\{x_i' \beta / (1 + c^2 \alpha_i^2)^{\frac{1}{2}}\} \right] = h(\eta_i^*) = \tilde{\mu}_i .$$

Thus, we define :

- $M = \text{diag}\{(1 + c^2 \alpha_i^2)^{-\frac{1}{2}}\}_{i=1, \dots, N}$ ,
- $K = \text{diag}\{h'(\eta_i^*)\}_{i=1, \dots, N}$   
 $= \text{diag}\{\exp(\eta_i^*) \{1 + \exp(\eta_i^*)\}^{-2}\}_{i=1, \dots, N}$   
 $= \text{diag}\{\tilde{\mu}_i (1 - \tilde{\mu}_i)\}_{i=1, \dots, N}$ ,
- $L = \text{diag}\{\tilde{\mu}_i (1 - \tilde{\mu}_i) / (1 + c^2 \alpha_i^2)^{\frac{1}{2}}\}_{i=1, \dots, N}$ ,
- $C = 0_N$ .

And the conditional approximated model verifies

$$\begin{aligned} E(\tilde{Y}_i | \xi) &= \tilde{\mu}_{\xi, i} = \tilde{\mu}_i + \tilde{\mu}_i (1 - \tilde{\mu}_i) \tilde{u}_i' \xi \\ \text{var}(\tilde{Y}_i | \xi) &= v(\tilde{\mu}_{\xi, i}) = \tilde{\mu}_{\xi, i} (1 - \tilde{\mu}_{\xi, i}) / n_i . \end{aligned}$$

We now calculate the variance matrix of  $\tilde{Y}$ , which is the marginal second moment in the approximated model.

$$\tilde{V} = \tilde{V}_0 + LUGU'L$$

where  $\tilde{V}_0$  is the diagonal matrix with  $i^{\text{th}}$  diagonal element

$$\begin{aligned} \tilde{V}_{0_{ii}} &= E[\tilde{\mu}_{\xi, i} (1 - \tilde{\mu}_{\xi, i}) / n_i] \\ &= \tilde{\mu}_i (1 - \tilde{\mu}_i) \{1 - \tilde{\mu}_i (1 - \tilde{\mu}_i) \tilde{u}_i' G \tilde{u}_i / n_i\} . \end{aligned}$$

Eventually we have in this case the following variance structure  $\tilde{W} = R + UGU'$  with

$$R = \text{diag} \left\{ \left[ -u_i' G u_i + (1 + c^2 \alpha_i^2) / \{\tilde{\mu}_i (1 - \tilde{\mu}_i)\} \right] / n_i \right\}_{i=1, \dots, N} .$$

And this variance structure of a linear mixed model is used to work out and iterate the Henderson's mixed model equations as in Section 2.2.



### 3.3 Identity link models

The identity link receives a particular place in this methodology since the marginal mean is then defined by  $\mu_i = x_i' \beta = \eta_i^*$ , which leads to  $M = K = L = I_n$  and  $C = 0_N$ . So the original model and the approximated model turn out to be the same. And the two approaches end up in a unique one. Moreover we do not need any additional approximation in that case.

This link may be used for instance in a Poisson or an exponential model:

- Poisson model:

$$v(\mu) = \mu \text{ so } \tilde{V}_0 = R = \text{diag}\{E(\tilde{\mu}_{\xi,i})\}_{i=1,\dots,N} = \text{diag}\{x_i' \beta\}_{i=1,\dots,N}$$

- exponential model:

$$v(\mu) = \mu^2 \text{ so } \tilde{V}_0 = R = \text{diag}\{E(\tilde{\mu}_{\xi,i}^2)\}_{i=1,\dots,N} = \text{diag}\{(x_i' \beta)^2 + u_i' G u_i\}_{i=1,\dots,N}.$$

## 4 Numerical examples

### 4.1 A small data set

We first consider a small data set of 18 observations concerning calving difficulties. This is a simplified extract of a data set presented in Foulley (1998). In this test example, for each observation, the number of calvings in each of two categories (easy/difficult) is reported. A binomial model is used and the explanatory variables considered consist of two environmental factors of fixed effects (with respectively 2 and 3 levels), and one random effect : the sire, with 4 levels and variance component  $\sigma^2$ . This data set is given in Table 1.

Table 2 gives the GAR parameter estimates and their precision for a probit link model and for a logit link model, making use in the latter case of the procedure presented in Section 3.2.

These results illustrate once again the link between logit and probit functions. It is well known that parameter estimates in fixed effect binomial model with probit link and logit link are expected to be proportional with a coefficient  $c$  related to the approximation of the logit function by the probit function ( $c = (3^{\frac{1}{2}}16)/(15\pi)$ ). In the case of random effect models, we here get the same

No	Environmental factors		Sire	Categories	
				easy	difficult
1	1	1	1	50	13
2	1	2	1	35	22
3	1	1	1	36	6
4	1	2	1	18	3
5	1	2	1	14	4
6	1	1	2	24	12
7	1	1	2	18	3
8	1	2	2	49	5
9	2	1	2	77	4
10	2	2	3	16	5
11	2	1	3	40	17
12	2	2	3	27	3
13	2	3	3	97	14
14	2	2	3	35	4
15	2	3	4	36	3
16	2	3	4	14	4
17	2	2	4	49	8
18	2	2	4	34	2

Table 1: A test data set concerning calving difficulties.

type of relationship.

Let us denote  $\beta_P$  (resp.  $\beta_L$ ) the fixed effect parameters in the probit link model (resp. logit link model), and  $\alpha_{i_P}^2$  (resp.  $\alpha_{i_L}^2$ ) the parameter related to the variance component  $\sigma^2$ . From Table 2, we have for the fixed effects :  $\hat{\beta}_P/\hat{\beta}_L = (0 \cdot 6255, 0 \cdot 5382, 0 \cdot 6507, 0 \cdot 5720)'$ . This estimation is based on the marginal models for the expectations :

$E(Y_i) = \Phi(x_i' \beta_P / (1 + \alpha_{i_P}^2)^{\frac{1}{2}})$  in the probit case,

$E(Y_i) = \text{logit}^{-1}(x_i' \beta_L / (1 + c^2 \alpha_{i_L}^2)^{\frac{1}{2}}) \approx \Phi(cx_i' \beta_L / (1 + c^2 \alpha_{i_L}^2)^{\frac{1}{2}})$  in the logit case.

Now, estimating also  $\sigma^2$ , we have here  $c(1 + \hat{\alpha}_{i_P}^2)^{\frac{1}{2}} / (1 + c^2 \hat{\alpha}_{i_L}^2)^{\frac{1}{2}} = 0 \cdot 5846$ ,

	probit link		logit link	
	estimate	standard error	estimate	standard error
$\hat{\beta}_1$	-0.6219	0.1958	-0.9942	0.3526
$\hat{\beta}_2$	-0.5189	0.1859	-0.9641	0.3382
$\hat{\beta}_3$	-0.1158	0.1282	-0.1780	0.2236
$\hat{\beta}_4$	-0.2169	0.1908	-0.3792	0.3637
$\hat{\sigma}^2$	0.0654		0.2254	

Table 2: Estimates and standard errors of fixed effect and variance component for the calving data using a binomial model

which is very close to the above given ratios.

As regards the variance component estimation, in our approach the approximated conditional models considered are the followings :

$$\begin{aligned}
 \text{probit case : } E(\tilde{Y}_i|\xi) &= \Phi(x'_i\beta_P/(1+\alpha_{i_P}^2)^{\frac{1}{2}}) \\
 &\quad + \varphi(x'_i\beta_P/(1+\alpha_{i_P}^2)^{\frac{1}{2}}) (1+\alpha_{i_P}^2)^{-\frac{1}{2}} u'_i \xi \quad , \\
 \text{logit case : } E(\tilde{Y}_i|\xi) &= \text{logit}^{-1}(x'_i\beta_L/(1+c^2\alpha_{i_L}^2)^{\frac{1}{2}}) \\
 &\quad + \text{logit}^{-1'}(x'_i\beta_L/(1+c^2\alpha_{i_L}^2)^{\frac{1}{2}}) (1+c^2\alpha_{i_L}^2)^{-\frac{1}{2}} u'_i \xi \quad ,
 \end{aligned}$$

logit<sup>-1'</sup> being the derivative of logit<sup>-1</sup>.

Now,

$$\text{logit}^{-1'}(x'_i\beta_L/(1+c^2\alpha_{i_L}^2)^{\frac{1}{2}}) \approx c (1+c^2\alpha_{i_L}^2)^{-\frac{1}{2}} \varphi(cx'_i\beta_L/(1+c^2\alpha_{i_L}^2)^{\frac{1}{2}}) .$$

We thus expect to have

$$\alpha_{i_P}^2/\alpha_{i_L}^2 \sim c^2(1+\alpha_{i_P}^2)/(1+c^2\alpha_{i_L}^2) ,$$

which is here estimated to  $0.5846^2 = 0.3418$ , whereas we get in fact the relatively close ratio  $\hat{\alpha}_{i_P}^2/\hat{\alpha}_{i_L}^2 = 0.29$ .

These results therefore mean that our procedure for the logit link model stay in agreement with the original GAR procedure for the probit link model.

It also shows that the proportionality of probit and logit estimates for fixed effect models is also present for random effect models.

## 4.2 Simulation results

We now observe the performance of this extended GAR procedure on simulated data. We simulated 200 data sets and we give in the first part of Table 3 the mean and standard error of the 200 estimated values obtained for each parameter. We consider here a log link exponential model and we are so in position to use the extension described in Section 3.1. The model involves one fixed parameter and one random vector with 8 realizations in an unbalanced experimental design. The size of the data vector is 72. The simulated value of  $\beta$  is set up to 3. And results are given for different values of  $\sigma^2$ .

Simulated values			$\sigma^2 = 0 \cdot 05$	$\sigma^2 = 0 \cdot 5$	$\sigma^2 = 1$	$\sigma^2 = 2$
<b>GAR</b>	$\hat{\sigma}^2$	mean	0 · 0714	0 · 4893	0 · 9657	1 · 6172
		s.e.	0 · 1014	0 · 3984	0 · 8119	1 · 2624
	$\hat{\beta}$	mean	2 · 9743	2 · 9840	2 · 9432	2 · 9653
		s.e.	0 · 1519	0 · 2999	0 · 4470	0 · 5712
<b>SCHALL</b>	$\hat{\sigma}^2$	mean	0 · 0733	0 · 5177	1 · 0280	2 · 0345
		s.e.	0 · 1023	0 · 3266	0 · 5940	1 · 2648
	$\hat{\beta}$	mean	2 · 9647	2 · 9737	2 · 9546	2 · 9482
		s.e.	0 · 1533	0 · 2772	0 · 3847	0 · 5223

Table 3: Means and standard errors of estimations on 200 simulated data sets from a log link exponential model using GAR and SCHALL procedure

Concerning the fixed effect, whatever the simulated values of variance component  $\sigma^2$  are, means of estimations give back the simulated  $\beta$ , and non surprisingly standard error increases as  $\sigma^2$  increases. Concerning the variance component, even if means of estimations are close to the simulated value, the procedure seems to have more difficulties to reach higher values of  $\sigma^2$ . Standard errors stay around the mean value.

Let us note that for  $\sigma^2 = 0.05$ , in around one third of the cases, the estimation of  $\sigma^2$  tends to zero. This is the case when simulated data are marginally close to exponential distribution and so the random effect does not introduce a much high perturbation.

In the second part of Table 3, we give the estimation results of the SCHALL procedure on the same simulated data sets. As regards fixed effects, this procedure gives the same results as the GAR procedure. However, concerning variance component estimation, it performs better for high values of  $\sigma^2$  : We do not encounter the same problem as for GAR procedure in this range of values. Moreover, all the estimation dispersions are somewhat smaller with the SCHALL procedure.

## 5 Discussion

A conditional approach to parameter estimation in generalized linear mixed models has been proposed by Schall (1991). It is conditional in the sense that its first step is a conditional linearization of the model. This procedure can be used for any model in this category. The procedure proposed by Gilmour, Anderson, et Rae (1985) in the case of a probit link mixed model lies more on a marginal approach since it first tends to get information on the marginal model. It was thus important to study the extension of this procedure to more models in the category. Even if we were able to propose such an extension for cases described in Section 2, it is however not valid for any generalized linear mixed model due to the limitation of the marginal expectation calculation mentioned in Section 2.3. Let us note however that besides the models considered in this paper, the approach may also be used for the inverse-link exponential model. In that case, the marginal expectation is hard to be calculated but after an

approximation of the conditional expectation the estimation method can then be set up without much difficulty.

Eventually it would be interesting to investigate more in deep this comparison between marginal and conditional procedures. With the results given in Table 3, in the case of an exponential log link model, we are led to prefer the use of the SCHALL procedure insofar as it seems to be less limited in the range of variance component values. Nevertheless a more general and extensive study should be done. And the different types of approximations used in each approach should also be more enlightened.

## References

- Anderson, D. A. et M. Aitkin (1985). Variance Components Models with Binary Response : Interviewer Variability. *Journal of the Royal Statistical Society, B* 47(2), 203–210.
- Anderson, D. A. et J. P. Hinde (1988). Random effects in generalized linear models and the EM algorithm. *Communications in Statistics - Theory and Methods* 17(11), 3847–3856.
- Breslow, N. E. et D. G. Clayton (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88(421), 9–25.
- Foulley, J. L. (1998). Heteroskedastic threshold models with applications to the analysis of calving difficulties. *International Bull Evaluation Service* 18, 3–11.
- Foulley, J. L. et S. Im (1993). A marginal quasi-likelihood approach to the analysis of poisson variables with generalized linear mixed models. *Genetics, Selection, Evolution* 25, 101–107.
- Foulley, J. L. et E. Manfredi (1991). Approches statistiques de l'évaluation génétique des reproducteurs pour des caractères binaires à seuils. *Genetics, Selection, Evolution* 23, 309–338.

- Gilmour, A. R., R. D. Anderson, et A. L. Rae (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72(3), 593–599.
- Hinde, J. (1982). Compound poisson regression models. In R. Gilchrist (Ed.), *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, Number 14 in Lecture Notes in Statistics, pp. 109–121. Springer-Verlag.
- Jansen, J. et J. A. Hoekstra (1993). The analysis of proportions in agricultural experiments by a generalized linear mixed model. *Statistica Neerlandica* 47(3), 161–174.
- McCullagh, P. et J. Nelder (1989). *Generalized Linear Models*. (second ed.). Chapman and Hall.
- McGilchrist, C. A. (1994). Estimation in Generalized Mixed Models. *Journal of the Royal Statistical Society, B* 56(1), 61–69.
- Patterson, H. D. et R. Thompson (1971). Recovery of interblock information when the block sizes are unequal. *Biometrika* 58, 545–554.
- Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* 78(4), 719–727.
- Zeger, S. L., K. Y. Liang, et P. S. Albert (1988). Models for longitudinal data : a generalized estimating equation approach. *Biometrics* 44, 1049–1060.



---

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,  
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY  
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex  
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN  
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex  
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

---

Éditeur  
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399