Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires

Christian Lavergne, Catherine Trottier

▶ **To cite this version:**

Christian Lavergne, Catherine Trottier. Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires. [Rapport de recherche] RR-3630, INRIA. 1999. inria-00073045

HAL Id: inria-00073045
https://inria.hal.science/inria-00073045

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.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires

Christian Lavergne, Catherine Trottier

No 3630

Mars 1999

THÈME 4

ISSN 0249-6399

INRIA
RHÔNE-ALPES

Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires

Christian Lavergne, Catherine Trottier

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n 3630 — Mars 1999 — 22 pages

Résumé : Nous nous intéressons à l'estimation des paramètres d'effets fixes et des composantes de la variance dans les modèles linéaires généralisés à effets aléatoires. En prenant un point de vue global sur cette famille de modèles (i.e. sans spécification particulière de loi et pour des effets emboîtés ou non), nous distinguons trois procédures d'estimation. Nous expliquons comment ces trois procédures illustrent trois types de démarche pour l'estimation. L'une est une relecture que nous proposons d'une méthode présentée par Engel & Keen en 1994. Elle est proche de la seconde, mise en place par Schall en 1991, et qui suit un raisonnement conditionnel. Enfin la troisième est une extension d'une méthode présentée en 1985 par Gilmour, Anderson & Rae, qui s'inscrit dans un raisonnement marginal. Nous envisageons donc de comparer ces trois méthodes selon leur comportement vis-à-vis du conditionnement.

Mots-clé : Modèles linéaires généralisés mixtes, Estimation des composantes de la variance, Modèle conditionnel, Modèle marginal.

(Abstract: pto)

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Estimation in Generalized Linear Mixed Models

Abstract: In this paper our interest lies on fixed effect parameters and variance components estimation in generalized linear mixed models, considering these models from a general point of view without specifying a distribution or a particular case for the effects. We distinguish three estimation procedures and we explain how each of them illustrates a kind of reasoning for the estimation. One of them is a rereading we give of a method originally presented by Engel & Keen in 1994. It is close to an other one, set up by Schall in 1991 and which follows a conditional reasoning. Eventually, the third one is an extension of the method proposed by Gilmour, Anderson & Rae in 1985 in a marginal reasoning. We compare these three procedures from the point of view of their analysis of the conditioning.

Key-words: Generalized linear mixed models, Variance components estimation, Conditional model, Marginal model.

1 Introduction

Les modèles linéaires généralisés mixtes (GL2M) sont une combinaison de deux types d'extension des modèles linéaires classiques (LM) : les modèles linéaires généralisés (GLM) pour une extension de la famille de lois considérée, et les modèles linéaires mixtes (L2M) pour l'introduction d'effets aléatoires dans la partie explicative du modèle. Dans un GL2M, ces effets aléatoires sont introduits dans le prédicteur linéaire, de telle sorte que, conditionnellement à ces effets, le modèle est défini comme ayant les propriétés d'un GLM. Ainsi les hypothèses sur la distribution de la variable à expliquer sont énoncées conditionnellement aux effets aléatoires.

Dans une optique de maximisation de la vraisemblance, pour atteindre la distribution marginale de cette variable, il est donc nécessaire de s'affranchir de ce conditionnement. Pour les modèles gaussiens à effets aléatoires, les règles de conditionnement de la loi normale rendent cette opération réalisable. Malheureusement de façon générale, ce déconditionnement n'est pas chose aisée dans la famille exponentielle.

L'élaboration de méthodes d'estimation va donc devoir faire face à ce problème. Pour cela, différentes approches peuvent être suivies qui mènent, en tout état de cause, à des méthodes non exactes, par le biais d'approximations réalisées à différents niveaux selon les raisonnements. Dans des cas particuliers de modèles à effets emboîtés ou avec sur-dispersion (ce qui peut être équivalent à l'introduction d'une réalisation d'un effet aléatoire par donnée), certains auteurs (cf. [1] [2] [11]) ont proposé d'approcher l'intégrale (du calcul de vraisemblance ou de l'algorithme EM) par quadrature gaussienne. D'un point de vue général cependant, ceci n'est plus praticable avec des intégrales multiples.

Nous présentons ici trois types de méthodes. Chacune d'entre elle analyse le modèle de façon différente vis-à-vis de ce déconditionnement aux effets aléatoires. L'une approche la distribution marginale, l'autre mène un raisonnement conditionnel et la dernière, tout en débutant aussi d'un point de vue conditionnel, relâche ensuite le conditionnement.

Dans un premier temps, nous définissons la classe des modèles linéaires généralisés mixtes, en introduisant les notations que nous conserverons dans tout l'article. Nous revenons ensuite succinctement sur quelques techniques d'estimation dans les GLM et les L2M. Cela nous sera utile pour présenter en section 4 les trois méthodes d'estimation considérées. En section 5, nous comparons entre elles ces différentes méthodes relativement à une échelle qualitative de déconditionnement. Nous terminerons par quelques résultats numériques de simulation.

2 Les modèles étudiés

2.1 Définition des GL2M

De même que les effets aléatoires ont été introduits dans les LM, ils le sont tout naturellement au sein des GLM pour donner naissance aux modèles linéaires généralisés mixtes. C'est dans l'expression du prédicteur linéaire qu'une partie aléatoire vient s'ajouter à la

partie fixe. On note y le vecteur des N observations, réalisation du vecteur aléatoire Y, et $\xi = (\xi_1', \ldots, \xi_J')'$ le vecteur de taille q des J effets aléatoires, où ξ_j de taille q_j contient les q_j réalisations du j^{e} effet aléatoire ($\sum_{j=1}^{J} q_j = q$). L'hypothèse gaussienne sur les effets aléatoires qui est introduite classiquement dans les modèles linéaires mixtes, est ici conservée. On associe au j^{e} effet le paramètre de variance σ_j^2 : $\xi_j \sim \mathcal{N}(0, \sigma_j^2 A_j)$, où A_j est une matrice connue. Ainsi $\xi \sim \mathcal{N}(0, G)$ où la matrice diagonale par blocs G est définie par $G = \text{diag}\{\sigma_j^2 A_j\}_{j=1,\ldots,J}$.

Étant données les matrices connues et définies par l'expérience, X $(N \times p)$ pour les effets fixes et U $(N \times q)$ pour les effets aléatoires, un GL2M est alors défini par les hypothèses suivantes :

• conditionnellement à ξ, les composantes Y_i $(i = 1, \ldots, N)$ sont indépendantes et distribuées selon une loi de la famille exponentielle, dont la densité s'écrit sous la forme :

$$f_{Y_i|\xi}(y_i, \theta_i) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right\} ,$$

où θ_i est un paramètre *canonique* et ϕ un paramètre de *dispersion*. Les fonctions b et c sont spécifiques à chaque distribution, et la fonction a_i s'écrit $a_i(\phi) = \phi/\omega_i$ avec ω_i un poids connu associé à l'observation i.

On a alors pour l'espérance et la variance conditionnelle de Y_i :

$$\begin{aligned} E(Y_i|\xi) &= b'(\theta_i) = \mu_{\xi,i} , \\ \text{var}(Y_i|\xi) &= a_i(\phi) b''(\theta_i) = a_i(\phi) v(\mu_{\xi,i}) , \end{aligned}$$

où v est appelée fonction de variance.

• le prédicteur additionne linéairement effets fixes et aléatoires. C'est alors un vecteur aléatoire à N composantes,

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

où β $(p \times 1)$ est le vecteur de paramètres d'effets fixes inconnus.

• le lien entre l'espérance conditionnelle μ_ξ et le prédicteur linéaire η_ξ est réalisé au travers de la fonction de lien g $(h = g^{-1})$: $\eta_\xi = g(\mu_\xi)$.
Pour chaque distribution, il existe une fonction de lien particulière $g = b'^{-1}$ appelée fonction de lien *canonique*.

Remarques :

Pour indiquer le conditionnement, nous indiçons par ξ tous les objets qui en dépendent. Par souci de simplification, nous confondons dans la même notation le vecteur aléatoire ξ et sa réalisation. Ainsi le prédicteur linéaire η_ξ pourra être de nature aléatoire ou non le cas échéant.

On additionne effets fixes et effets aléatoires sur une même échelle. Par conséquent conditionnellement à ξ le GL2M conserve toutes les propriétés du GLM (cf. [13] avec l'importance des fonctions de lien g et de variance v). Le modèle se trouve ainsi principalement défini par un raisonnement conditionnel à ξ. Nous le nommerons *modèle conditionnel*.

2.2 Un exemple de modélisation en génétique animale

De nombreux problèmes, dans divers domaines d'application, peuvent être modélisés par des modèles linéaires généralisés mixtes. Citons par exemple la biométrie avec des données présentées dans [18], puis réutilisées dans [8] ou bien le domaine de la fiabilité du logiciel où certains modèles (cf. [7]) peuvent être vu comme des GL2M. L'exemple que nous présentons ici est issu du domaine de la génétique quantitative.

Dans deux troupeaux, des génisses et des vaches donnent naissance à vingt veaux mâles ou femelles, issus de quatre taureaux. Pour chacun, on enregistre la difficulté de vêlage selon deux catégories : facile / difficile. On note aussi l'âge de la mère, le numéro de troupeau et le sexe du veau.

La modélisation retenue considère les effets fixes suivants :

- A : l'âge de la mère à 2 niveaux : A_1 (génisse), A_2 (vache),
- T : le troupeau d'origine à 2 niveaux : T_1, T_2,
- S : le sexe du veau à 2 niveaux : S_1 (mâle), S_2 (femelle),

et comme effet aléatoire P, le père dont on observe 4 réalisations : P_1, P_2, P_3, P_4. On suppose que $P \sim \mathcal{N}(0, \sigma^2)$ et on ne cherche pas à connaître l'effet de chacun des 4 niveaux observés mais plutôt à estimer σ^2, la variabilité dûe à cet effet.
En considérant :

$$\begin{aligned} \beta &= (A_1 + T_1 + S_1, A_2 - A_1, T_2 - T_1, S_2 - S_1)' \\ \xi &= (P_1, P_2, P_3, P_4)' , \end{aligned}$$

les matrices d'incidence X et U s'obtiennent simplement. On peut alors écrire le prédicteur linéaire sous la forme :

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

Conditionnellement à ξ, les composantes de Y sont considérées indépendantes et distribuées selon une loi de Bernoulli :

$$\forall i \in \{1, ..., 20\} , \qquad Y_i|\xi \sim \mathcal{B}(p_{\xi,i}) ,$$

pour laquelle on a alors : $E(Y_i|\xi) = p_{\xi,i}$.
Plusieurs fonctions de lien sont envisageables dont :
- le lien logit (lien canonique) : $\eta_{\xi,i} = log(\frac{p_{\xi,i}}{1 - p_{\xi,i}})$,
- le lien probit : $\eta_{\xi,i} = \Phi^{-1}(p_{\xi,i})$ (Φ désignant la fonction de répartition de la loi normale centrée réduite).

Le modèle adopté est donc un modèle linéaire généralisé à un effet aléatoire.
Dans ce genre de situation, on peut très facilement imaginer un prolongement de modélisation avec deux effets aléatoires. On considère par exemple des taureaux issus de deux races

différentes. L'effet aléatoire associé au taureau peut alors se découper en un effet P_1 dû à la race 1 et un effet P_2 dû à la race 2. Chacun serait distribué avec son propre paramètre de variance : $P_1 \sim \mathcal{N}(0, \sigma_1^2)$, $P_2 \sim \mathcal{N}(0, \sigma_2^2)$. Ainsi, à l'aide des estimations des deux composantes de la variance σ_1^2 et σ_2^2, on pourra comparer les races 1 et 2.

3 Retour sur l'estimation dans les GLM et les L2M

Comme nous l'avons noté, les GL2M constituent une extension des GLM d'une part, et des L2M d'autre part. Les méthodes d'estimation considérées en section 4 s'inspirent donc des techniques propres à chacune de ces deux catégories de modèles. Nous revenons succinctement sur ces procédures d'estimation.

3.1 Estimation dans les GLM

Dans les GLM, lorsque l'on s'intéresse à l'estimation du vecteur de paramètres des effets fixes β, la maximisation de la vraisemblance (cf. [13]) conduit au système d'équations suivant :

$$X' W_\beta^{-1} \frac{d\eta}{d\mu}(y - \mu) = 0 \,, \tag{1}$$

où
$$W_\beta = \text{diag}\{ \text{var}(Y_i) g'(\mu_i)^2 \}_{i=1,\dots,N} = \text{diag}\{ \frac{\phi}{\omega_i} v(\mu_i) g'(\mu_i)^2 \}_{i=1,\dots,N} \,,$$
$$\frac{d\eta}{d\mu} = \text{diag}\{ \frac{d\eta_i}{d\mu_i} \}_{i=1,\dots,N} = \text{diag}\{ g'(\mu_i) \}_{i=1,\dots,N} \,.$$

Ce système n'est bien entendu pas linéaire en β et est donc résolu par une procédure itérative. L'algorithme usuel mis en place est l'algorithme des scores de Fisher. Cet algorithme peut aussi être décrit comme la résolution itérative des équations normales :

$$X' W_\beta^{-1} (z_\beta - X\beta) = 0 \tag{2}$$

où $z_\beta = \eta + \frac{d\eta}{d\mu}(y - \mu) = X\beta + \frac{d\eta}{d\mu}(y - \mu)$. À β fixé, z_β est considéré comme un nouveau vecteur de données,

W_β comme une nouvelle matrice de poids, et on reconnaît alors dans le système (2) les équations classiques des moindres carrés généralisés associées au modèle : $Z_\beta = X\beta + e$ où $E(e) = 0$ et W_β correspond à la matrice de variance de Z_β. Ainsi, à l'itération $[t]$, on utilise la méthode des moindres carrés pondérés dans le modèle que nous notons $\mathcal{M}^{[t]}$: $Z_{\beta^{[t]}} = X\beta + e^{[t]}$ où $E(e^{[t]}) = 0$ et $\text{Var}(e^{[t]}) = W_{\beta^{[t]}}$. Par la suite, nous nommerons ce modèle $\mathcal{M}^{[t]}$: *modèle linéarisé*.

3.2 Estimation par maximum de quasi-vraisemblance

Dans certains problèmes, les hypothèses de modélisation ne permettent pas d'expliciter la fonction de vraisemblance. Wedderburn [21] propose alors une inférence statistique qui utilise uniquement l'information contenue dans les deux premiers moments du vecteur Y. On définit alors la *log-quasi-vraisemblance*, pour un vecteur Y à composantes indépendantes :

$$Q(\mu; y) = \sum_{i=1}^{N} \int_{y_i}^{\mu_i} \frac{y_i - t}{a_i(\phi) v(t)} \, dt \ ,$$

en notant $E(Y) = \mu$ et $V = \mathrm{Var}(Y) = \mathrm{diag}\{a_i(\phi)v(\mu_i)\}_{i=1,\ldots,N}$. Le problème de maximisation de cette log-quasi-vraisemblance revient à annuler la fonction quasi-score en β :

$$\mathcal{U}(\beta) = G' V^{-1} (y - \mu) \ , \qquad \text{où} \quad G = \frac{\partial \mu}{\partial \beta'} \ .$$

Ici $G = KX$ avec $K = \mathrm{diag}\{h(x_i'\beta)\}_{i=1,\ldots,N}$. Pour résoudre les équations qui en découlent, on utilise le schéma itératif suivant :

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

avec $\begin{array}{lcl} \zeta^{[t]} & = & X\beta^{[t]} + K^{[t]\,-1}(y - \mu^{[t]}) \\ W^{[t]} & = & K^{[t]\,-1} V K^{[t]\,-1} \ . \end{array}$

On retrouve le même type de procédure itérative que pour le maximum de vraisemblance dans les GLM. D'ailleurs, si la fonction de variance v choisie est la fonction de variance naturelle associée à une loi de la famille exponentielle, la fonction quasi-score est la fonction score du GLM défini par cette loi. La solution $\hat{\beta}$ du maximum de quasi-vraisemblance est alors l'estimation du maximum de vraisemblance du modèle. Cependant, la quasi-vraisemblance permet d'envisager d'autres fonctions de variance que celles précisément associées aux lois classiques. C'est là un intérêt primordial de cette notion. On poursuit alors une inférence sans s'appuyer sur une hypothèse de loi de probabilité. Tout ceci est aussi transposable au cas de composantes non indépendantes (cf. [13]).

3.3 Estimation dans les L2M

Parmi les nombreux articles traitant des problèmes de l'estimation dans les L2M, les ouvrages de Searle, Casella et McCulloch [17] et de Rao et Kleffe [15] font référence. De nombreuses techniques ont été proposées pour traiter de cette question et nous retenons ici l'estimation par maximisation de la vraisemblance (ML) ou de la vraisemblance restreinte (REML), en revenant sur deux procédures : une résolution directe et une résolution par l'intermédiaire des équations du modèle mixte (équations dites de Henderson).

3.3.1 Procédure directe

L'hypothèse fondamentale du modèle linéaire mixte est l'hypothèse de loi normale sur les erreurs ainsi que sur les effets aléatoires. Ceci implique la normalité de la distribution marginale de Y avec $E(Y) = X\beta$, $\mathrm{Var}(Y) = R + UGU' = \sum_{j=0}^{J} \sigma_j^2 V_j = \Gamma$ où $R = \sigma_0^2 V_0$ et $V_j = U_j A_j U_j'$, $\forall j \in \{1, ..., J\}$. La maximisation de la vraisemblance pour obtenir les estimations ML conduit alors au système suivant :

$$\left\{ \begin{array}{lcll} X'\,\Gamma^{-1}\,X\,\beta & = & X'\,\Gamma^{-1}\,y & \\ \mathrm{tr}(\Gamma^{-1} V_j) & = & y'\,P\,V_j\,P\,y & j = 0, ..., J, \end{array} \right. \tag{3}$$

où $P = \Gamma^{-1}(I - X(X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1})$.

En s'intéressant à la maximisation de la partie de la vraisemblance indépendante de l'effet fixe β, on aboutit à l'estimation REML proposée par Thompson en 1962. Cette technique permet de prendre en compte la perte de degrés de liberté lors de l'estimation de β et correspond aussi à une méthode ML sur une certaine transformation des données. L'équation en β est alors supprimée. La procédure d'estimation peut être obtenue à partir des équations pour les composantes du système (3) en remplaçant Γ^{-1} par P.

3.3.2 Les équations de Henderson

L'estimation des paramètres d'effets fixes et des composantes de la variance d'un L2M peut aussi être obtenue à partir des solutions des équations de Henderson. Ces équations, proposées par Henderson, Kempthorne, Searle et VonKrosig [10], sont particulièrement intéressantes lorsqu'une prédiction de l'effet aléatoire est aussi nécessaire. Elles permettent d'obtenir simultanément la prédiction BLUP (Best Linear Unbiased Predictor) de ξ et l'estimation BLUE (Best Linear Unbiased Estimator) de β. Elles résultent de la maximisation en β et ξ de la loi jointe (Y, ξ), construite comme produit de la loi conditionnelle de $Y|\xi$ et de la loi de ξ. Ce système d'équations est le suivant :

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}U \\ U'R^{-1}X & U'R^{-1}U + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \xi \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ U'R^{-1}y \end{pmatrix}. \tag{4}$$

À partir des solutions de ce système, les estimations ML et REML des composantes de la variance peuvent être obtenues par le schéma itératif suivant :

ML

$$\left[\begin{array}{lcll} \sigma_j^{2(m+1)} & = & \dfrac{\xi_j^{\prime(m)}\, A_j^{-1}\, \xi_j^{(m)}}{q_j - \mathrm{tr}(P_{jj}^{(m)})} & , \quad j = 1, \ldots, K \\ \sigma_0^{2(m+1)} & = & \dfrac{y'\, V_0^{-1}\, (y - X\beta^{(m)} - U\xi^{(m)})}{N} & \end{array} \right. \tag{5}$$

REML

$$\left[\begin{array}{lcll} \sigma_j^{2(m+1)} & = & \dfrac{\xi_j^{\prime(m)} A_j^{-1} \xi_j^{(m)}}{q_j - \mathrm{tr}(Q_{jj}^{(m)})} & , \quad j = 1, \ldots, K \\ \sigma_0^{2(m+1)} & = & \dfrac{y' V_0^{-1} (y - X\beta^{(m)} - U\xi^{(m)})}{N - \mathrm{rg}(X)} & \end{array}\right. \tag{6}$$

où $P = (I + U'R^{-1}UG)^{-1}$, $Q = (I + U'SUG)^{-1}$ avec $S = R^{-1}(I - X(X'R^{-1}X)^{-1}X'R^{-1})$, et où P_{jj} (resp. Q_{jj}) désigne la $j^{ème}$ sous matrice de dimension $(q_j \times q_j)$ de la matrice P (resp. Q).

4 Estimation dans les GL2M

Nous décrivons dans leurs grandes lignes trois types de procédure pour l'estimation dans un GL2M.

4.1 Une relecture de la méthode EK

La méthode que nous décrivons dans ce paragraphe peut être vue comme une relecture de la méthode (que nous nommons méthode EK) proposée par Engel et Keen en 1994 [4]. Elle se décompose en deux étapes, tirant profit de la double nature du GL2M et utilisant successivement les procédures rappelées aux sections 3.1 et 3.3. Ainsi, la première étape consiste en une linéarisation conditionnelle à ξ et analogue à celle utilisée dans les GLM. La deuxième étape procède ensuite à l'estimation dans ce modèle linéarisé, vu comme un L2M. Cette méthode découle donc tout naturellement de ce qui a été présenté dans la section précédente.

4.1.1 Étape de linéarisation

En se plaçant conditionnellement à ξ, le modèle considéré est un GLM que l'on peut écrire :

$$Y = g^{-1}(\eta_\xi) + \varepsilon \, ,$$

avec les hypothèses de lois conditionnelles adéquates. Sachant ξ, on peut donc reprendre la démarche de 3.1 pour les GLM, en introduisant la variable dépendante $z_{\beta,\xi} = X\beta + U\xi + (y - \mu_\xi)g'(\mu_\xi)$. On est ensuite amené à considérer les équations normales dans le *modèle linéarisé* noté $\mathcal{M}_\xi$:

$$Z_{\beta,\xi} = X\beta + U\xi + e \, ,$$

où $\begin{array}{lcl} E(Z_{\beta,\xi}|\xi) & = & X\beta + U\xi \, , \\ \mathrm{Var}(Z_{\beta,\xi}|\xi) & = & \mathrm{Var}(e|\xi) = \mathrm{Var}(\varepsilon g'(\mu_\xi)|\xi) = \mathrm{Var}((Y - \mu_\xi)g'(\mu_\xi)|\xi) = W_{\beta,\xi} \, . \end{array}$

En se plaçant conditionnellement à ξ, l'effet aléatoire peut être considéré comme un paramètre. On pourrait alors l'estimer par moindres carrés généralisés itérés selon la technique des

GLM. La procédure de moindres carrés itérés utiliserait alors la matrice des poids ou matrice de variance du vecteur des erreurs conditionnel à ξ : $W_{\beta,\xi} = \text{diag}\{\ g'(\mu_{\xi,i})^2 \text{var}(\varepsilon_i|\xi)\ \}_{i=1,\dots,N}$. Ce qui dans le cas d'un lien canonique n'est autre que $W_{\beta,\xi} = \text{diag}\{\ a_i(\phi) g'(\mu_{\xi,i})\ \}_{i=1,\dots,N}$. Ainsi, on nommera plus précisément $\mathcal{M}_\xi$ *modèle linéarisé conditionnel.* Le tableau I donne l'expression de cette matrice de variance dans le cas d'un lien canonique pour les lois classiques de la famille exponentielle.

	lien canonique	matrice variance *conditionnelle*
$\frac{\mathcal{B}(n,\pi)}{n}$	$g(x) = \ln(\frac{x}{1-x})$	$W_{\beta,\xi} = \{_d\ \frac{1}{n_i} \frac{(1+e^{(X\beta+U\xi)_i})^2}{e^{(X\beta+U\xi)_i}}\ \}$
$\mathcal{P}(\lambda)$	$g(x) = \ln(x)$	$W_{\beta,\xi} = \{_d\ \frac{1}{e^{(X\beta+U\xi)_i}}\ \}$
$\mathcal{E}xp(\lambda)$	$g(x) = \frac{1}{x}$	$W_{\beta,\xi} = \{_d\ (X\beta+U\xi)_i^2\ \}$
$\mathcal{G}(a,\lambda)$	$g(x) = \frac{1}{x}$	$W_{\beta,\xi} = a\{_d\ (X\beta+U\xi)_i^2\ \}$
$\mathcal{N}(\mu,\sigma_0^2)$	$g(x) = x$	$W_{\beta,\xi} = \{_d\ \sigma_0^2\ \}$

TAB. I – *Matrice de variance du modèle linéarisé* **conditionnel** *associé aux lois usuelles de la famille exponentielle.*

Remarquons que dans le cas de la loi normale, lien identité, cette étape de linéarisation n'a bien entendu aucun effet. Le modèle $\mathcal{M}_\xi$ est le modèle initial.

4.1.2 Étape d'estimation

C'est maintenant davantage l'aspect extension d'un L2M qui prédomine. Dans le modèle linéarisé conditionnel $\mathcal{M}_\xi$:

$$Z_{\beta,\xi} = X\beta + U\xi + e\ ,$$

on redonne à ξ sa nature aléatoire et l'on plonge alors ce modèle dans la structure d'un L2M où :

$$\begin{array}{lcl} E(Z_{\beta,\xi}) & = & X\beta \\ \text{Var}(Z_{\beta,\xi}) & = & UGU' + E(W_{\beta,\xi}) \\ & = & UGU' + W_\beta \\ & = & \Gamma\ . \end{array}$$

La matrice de variance des erreurs de ce modèle linéaire mixte est donc $W_\beta = E(W_{\beta,\xi}) = E(\text{diag}\{g'(\mu_{\xi,i})^2 \text{var}(\varepsilon_i|\xi)\}_{i=1,\dots,N})$. Cette fois-ci c'est la matrice de variance marginale qui intervient. Et on fera référence à ce modèle $\mathcal{M}$ par *modèle linéarisé marginal.*

Le modèle $\mathcal{M}$ obtenu est un peu particulier dans la mesure où la matrice de variance W_β des erreurs e dépend des paramètres modélisant l'espérance. On procède donc à l'estimation de façon itérative. À chaque itération, les valeurs de $z_{\beta,\xi}$ et W_β sont calculées et on estime les paramètres du modèle ainsi défini. Pour une adaptation de type ML, on utilise les systèmes (3), ou (4) et (5) décrits précédemment, en remplaçant R par W_β et y par $z_{\beta,\xi}$; l'adaptation au cas REML étant tout aussi aisée.

Toute la difficulté réside dans le calcul de W_β. Dans le cas d'un lien canonique et sous l'hypothèse $\xi \sim \mathcal{N}(0, G)$ avec G diagonale par blocs : $G = \text{diag}\{\ \sigma_j^2 A_j\ \}_{j=1,\ldots,J}$, ceci conduit au calcul de $E(\text{diag}\{\ \phi/\omega_i\ g'(\mu_{\xi,i})\ \}_{i=1,\ldots,N})$. Ce qui donne les résultats exhibés dans le tableau II pour les lois classiques de la famille exponentielle.

loi	matrice variance *marginale* [a, b]
$\dfrac{\mathcal{B}(n,\pi)}{n}$	$W_\beta = \{_d\ \dfrac{2}{n_i}[1 + \text{ch}((X\beta)_i)\exp[\sum_{j=1}^{K}\dfrac{\sigma_j^2 U_{ji} A_j U'_{ji}}{2}]\ \}$
$\mathcal{P}(\lambda)$	$W_\beta = \{_d \exp[-(X\beta)_i + \sum_{j=1}^{K}\dfrac{\sigma_j^2 U_{ji} A_j U'_{ji}}{2}]\ \}$
$\mathcal{E}xp(\lambda)$	$W_\beta = \{_d\ (X\beta)_i^2 + \sum_{j=1}^{K}\sigma_j^2 U_{ji} A_j U'_{ji}\ \}$
$\mathcal{G}(a,\lambda)$	$W_\beta = a\{_d\ (X\beta)_i^2 + \sum_{j=1}^{K}\sigma_j^2 U_{ji} A_j U'_{ji}\ \}$
$\mathcal{N}(\mu,\sigma_0^2)$	$W_\beta = \{_d\ \sigma_0^2\ \}$

TAB. II – *Matrice de variance du modèle linéarisé* **marginal** *associé aux lois usuelles de la famille exponentielle.*

[a] On note U_{ji} la i^{e} ligne de la matrice U_j, matrice concernant le j^{e} effet aléatoire.

[b] On note ch la fonction cosinus hyperbolique : $\text{ch}(x) = \dfrac{e^x + e^{-x}}{2}$.

Nous voyons dans le tableau II que le calcul de $E(W_{\beta,\xi})$ ne pose pas de problème particulier dans le cas de lien canonique. Il est aussi réalisable dans certains cas de liens non canoniques. Remarquons, par exemple, que pour la loi exponentielle – lien log, ces deux matrices sont égales : $W_\beta = W_{\beta,\xi} = I_N$. Notons aussi que dans le cas de la loi normale (lien identité), on a $W_{\beta,\xi} = W_\beta = \sigma_0^2\, I_N$. Pour autant, ce calcul n'est pas toujours réalisable analytiquement, et c'est notamment le cas pour la loi binomiale muni du lien probit.

4.2 La méthode Schall

La méthode décrite par Schall [16] consiste aussi en une linéarisation du modèle conditionnellement à ξ, puis en l'estimation des paramètres dans le modèle linéarisé par utilisation des équations de Henderson pour les modèles linéaires mixtes. Finalement, on retrouve dans [16] les deux étapes, décrites précédemment, de linéarisation et d'estimation et la mise en place d'une procédure itérative.

Lors de la linéarisation, la variable dépendante est introduite de la même façon. Cependant, le modèle linéarisé adopté par la suite, diffère. En effet, le point de vue de Schall est de se placer dans le modèle $\mathcal{M}_\xi$ que nous avons appelé *modèle linéarisé conditionnel*:

$$\mathcal{M}_\xi \ : \quad Z = X\beta + U\xi + \varepsilon \ ,$$

où la matrice de variance des résidus est $W_{\beta,\xi} = \mathrm{Var}(\varepsilon|\xi)$ et la structure de variance considérée est $\Gamma_\xi = UGU' + W_{\beta,\xi}$. À aucun moment il ne considère le calcul de $W_\beta = E(W_{\beta,\xi})$, la matrice de variance marginale des résidus. L'analyse de ce modèle comme un L2M implique alors que ξ retrouve sa nature aléatoire mais seulement partiellement puisque l'on maintient, à l'intérieur de la structure de variance, la matrice de variance conditionnelle des résidus. Pourtant, ce raisonnement peut être justifié par le fait que Schall utilise les équations de Henderson pour obtenir les estimations dans le L2M associé. Or, nous l'avons évoqué dans la section 3.3.2, la construction de ces équations se base sur la loi du couple (Y, ξ) comme produit de la loi conditionnelle de Y à ξ et de la loi de ξ. Ainsi dans l'approximation normale de cette loi conditionnelle, c'est bien la matrice $W_{\beta,\xi}$ qui intervient et c'est donc bien celle que l'on retrouve lors de la construction des équations. Ceci peut donc justifier le fait de conserver le conditionnement et donc $W_{\beta,\xi}$. Par contre, si, pour l'estimation, on cherche à utiliser les équations directes ML (3) ou REML, ces équations étant construites à partir de la distribution marginale de Y, c'est la structure de variance $\Gamma = UDU' + W_\beta$ qui interviendrait plus naturellement. Cette ambiguïté naît de l'approximation du modèle linéarisé par un modèle gaussien alors que $W_{\beta,\xi} \neq W_\beta$. Dans un L2M, on aurait en effet pour les erreurs $\mathrm{Var}(e|\xi) = \mathrm{Var}(e) = R$. Le choix de $\mathcal{M}_\xi$ offre l'avantage de ne pas être limité dans l'utilisation des fonctions de lien, la matrice $W_{\beta,\xi}$ étant toujours explicite.

Depuis sa mise en place, cette méthode Schall a reçu divers éclairages. En effet, d'autres démarches suivies par certains auteurs ont abouti aux mêmes équations et se sont donc avérés être autant de façon de justifier cette méthode.
Il y a d'une part la méthode de quasi-vraisemblance pénalisée de [3], où l'on aboutit au même système (résolu itérativement) de Henderson dans le modèle linéarisé $\mathcal{M}_\xi$ de Schall. D'autre part, la même idée est reprise dans le cadre plus général des HGLM (Hierarchical Generalized Linear Models qui englobent les GL2M) par [12]. Dans ces modèles, les auteurs définissent la *h-vraisemblance* comme la vraisemblance jointe définie à partir de la loi jointe de (Y, ξ). Dans le cas d'une hypothèse normale sur la loi de $Y|\xi$ ainsi que sur celle de ξ, cette *h-vraisemblance* n'est autre que la vraisemblance jointe de Henderson. Dans le cas des

GL2M, la maximisation de cette vraisemblance jointe est identique à la maximisation de la quasi-vraisemblance pénalisée. Or, ils montrent que la solution en β de la maximisation jointe de la h-vraisemblance maximise la vraisemblance marginale approchée à l'aide de l'approximation de Laplace. La solution en ξ, quant à elle, est ensuite utilisée pour l'estimation des composantes par une procédure de nouveau équivalente dans le cas des GL2M à celle de [16], [3] ou [14]. Enfin, cette méthode trouve aussi une justification du point de vue bayésien adopté dans [19]. En s'intéressant à la recherche du mode a posteriori de β et ξ, lorsque l'on suppose un a priori non informatif uniforme sur β, le raisonnement bayésien aboutit aux même équations.

4.3 La méthode GAR étendue

Gilmour, Anderson et Rae ont proposé en 1985 [9] une méthode d'estimation (méthode GAR) dans un modèle à effets aléatoires pour des données binomiales avec un lien probit. Foulley et Im [6] l'ont étendue au cas d'une distribution de Poisson et son extension à un cadre plus large de modèles est présentée dans [20]. Nous revenons ici sur l'idée qui fonde la démarche adoptée (notamment vis-à-vis du conditionnement).

Abandonnant l'idée d'atteindre la distribution marginale de Y, cette méthode se base, dans un premier temps, sur le calcul des deux premiers moments marginaux. Ce calcul permet ensuite la définition d'une fonction de quasi-vraisemblance, que l'on cherchera à maximiser pour procéder à l'estimation (cf. 3.2). Cependant, cette fonction de quasi-vraisemblance est construite, non pas à partir de l'expression exacte de la variance, mais à l'aide d'une approximation de celle-ci. Or on reconnaît dans la variance approchée, la forme de la structure de variance d'un L2M. Il est alors possible d'utiliser les méthodes d'estimation propres aux L2M. Ainsi, pour l'estimation de β, le maximum de quasi-vraisemblance (construite sur l'espérance et la variance approchée) sera aussi solution du système de Henderson du L2M associé. La résolution de ce système permet en outre l'estimation des composantes de la variance à partir de la prédiction de ξ obtenue. L'utilisation des équations de Henderson implique une réintroduction artificielle d'effets aléatoires, mais ayant les mêmes propriétés de variance que les effets aléatoires d'origine.

Dans cette méthode, la linéarisation est réalisée par approximation de la variance marginale. Le raisonnement suivi par ces auteurs débute donc par un déconditionnement. Ce n'est ensuite qu'au niveau du *modèle marginal* (ou *modèle déconditionné*) qu'intervient la linéarisation, contrairement à la méthode Schall où la première étape consiste à effectuer cette linéarisation.

5 Comparaison des trois procédures

5.1 Échelle qualitative de déconditionnement

Comme nous l'avons signalé au 4.3, l'idée première de la méthode GAR est de se libérer du conditionnement pour procéder ensuite à l'estimation dans le modèle déconditionné associé. Ce modèle marginal est linéarisé par approximation de la variance. Ainsi, cette méthode s'oppose à celle de Schall, pour laquelle le raisonnement est mené conditionnellement au vecteur aléatoire ξ. Même après la linéarisation, c'est toujours la matrice conditionnelle $W_{\beta,\xi}$ qui intervient dans le modèle. D'où le nom du L2M associé $\mathcal{M}_\xi$: *modèle linéarisé conditionnel.* La méthode présentée au 4.1 apparaît alors comme intermédiaire entre ces deux méthodes puisqu'on relâche ce conditionnement en travaillant avec la matrice de variance déconditionné W_β. Lorsqu'on plonge le modèle linéarisé dans la structure d'un L2M, l'approximation normale peut se réaliser de deux façons :

- soit sur la loi conditionnelle de Z à ξ. Ceci permet alors d'expliquer l'utilisation de la matrice $W_{\beta,\xi}$ dans les équations de Henderson. Et il est possible de construire, utilisant $W_{\beta,\xi}$, un système direct équivalent (avec Γ_ξ).
- soit sur la loi marginale de Z. Ceci permet alors d'expliquer l'utilisation de la matrice W_β dans le système direct. Et il est possible de construire, utilisant W_β, des équations de Henderson équivalentes.

Cependant, ces deux approches, équivalentes dans le cas normal, ne le sont plus dans le cas général du fait que $W_\beta \neq W_{\beta,\xi}$.

Ainsi, alors que le positionnement du problème initial et la définition même du GL2M, se font conditionnellement à ξ, les trois méthodes peuvent être différenciées au regard de leur degré de déconditionnement. Si l'on mesurait cet indice pour chacune des méthodes, la méthode GAR se situerait à une extrémité de l'échelle, Schall à l'autre et la méthode entre les deux. On aurait cependant tendance à rapprocher ces deux dernières méthodes puisqu'elles suivent une même première étape de linéarisation, alors qu'à l'inverse c'est d'abord le déconditionnement qui prime dans la méthode GAR, la linéarisation n'intervenant qu'ensuite.

De façon imagée, on peut ensuite ajouter une notion d'*élasticité* à cette échelle. En effet, ces différentes méthodes, appliquées au cas de la loi normale - lien identité, sont en fait toutes identiques. Il n'y a effectivement plus d'approximation ou de linéarisation. Mais elles se différencient de plus en plus lorsque l'on s'écarte du modèle gaussien, donc lorsque les approximations effectuées pour la linéarisation sont de moins en moins valides. On tirera donc plus ou moins sur l'"élastique" selon les lois et les fonctions de lien choisies.

Il est important de constater que la méthode EK, dans certains cas (lorsque $W_{\beta,\xi} = W_\beta$), s'identifie à la méthode Schall, alors que pour d'autres cas, elle s'identifie à la méthode GAR.

Ce qui la positionne bien dans son rôle intermédiaire. Nous avons, pour souligner cela, résumé dans le tableau III les différents cas en indiquant :
- les équivalences,
- le cas où, faute de savoir calculer W_β, la méthode EK n'est pas définie,
- le cas où la méthode GAR n'a pu être étendue (la colonne GAR de ce tableau faisant référence soit à la méthode d'origine, soit à ses extensions).

		Schall		EK		GAR
binomiale	probit	×		?		×
	logit	×		×		×
Poisson	logarithme	×		×		×
	identité	×		×	=	×
exponentielle	logarithme	×	=	×		×
	inverse	×		×		?
	identité	×		×	=	×
normale	identité	×	=	×	=	×

TAB. III – *Existence et équivalences des trois méthodes.*

Ces trois méthodes permettent donc de couvrir un éventail allant du raisonnement marginal pour GAR au raisonnement conditionnel pour Schall.

5.2 Des résultats numériques

5.2.1 Des données réelles

Nous considérons dans un premier temps les données du tableau IV contenant 18 observations et tout à fait analogue à l'exemple de modélisation discuté en section 2.2. C'est un extrait d'un jeu de données présenté dans [5]. Il concerne la difficulté de vêlage et utilise un modèle binomial. Dans ce modèle, deux facteurs environnementaux ont été retenus comme effets fixes avec respectivement 2 et 3 niveaux et sans interactions ; d'où un vecteur de paramètres d'effets fixes de taille 4. D'autre part, un seul effet aléatoire est considéré : le taureau père avec 4 réalisations.

No	Facteurs environnementaux		Père	Catégories	
				facile	difficile
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

TAB. IV – *Un exemple test concernant les difficultés de vêlage.*

Dans le tableau V, nous donnons les résultats d'estimation de β et σ^2 des différentes procédures dans une approche REML et ML pour un modèle avec lien logit. Le tableau VI fournit ces même résultats pour un modèle avec lien probit. Dans ce cas, la procédure EK n'est pas définie.

La question n'est pas ici de comparer les approches ML et REML (on trouvera dans [17] une discussion à ce sujet dans le cadre des L2M). Au vu de ces résultats, nous ne remarquons pas de différences notables entre ces trois méthodes, qui restent cohérentes.

5.2.2 Des données simulées

Nous comparons maintenant le comportement de ces méthodes d'estimation sur des jeux de données simulées. Nous considérons pour cela le cas d'un modèle binomial avec lien logit. Dans ce cas, nous utilisons pour GAR l'extension à la méthode d'origine. Dans le tableau VII, nous présentons les résultats obtenus. Nous avons pris un plan d'expérience avec un

Lien logit		Schall	EK	GAR étendue
REML	$\hat{\sigma}^2$	0.2318	0.2499	0.2254
	$\hat{\beta}$	−0.9672	−0.9705	−0.9942
		−0.9742	−1.0752	−0.9641
		−0.1763	−0.1281	−0.1780
		−0.3730	−0.2638	−0.3792
ML	$\hat{\sigma}^2$	0.0777	0.0926	0.0766
	$\hat{\beta}$	−1.1153	−1.1006	−1.1206
		−0.7595	−0.8683	−0.7539
		−0.1301	−0.1090	−0.1327
		−0.2866	−0.2073	−0.2896

TAB. V – *Résultats estimation des trois méthodes : modèle binomial - lien logit.*

Lien probit		Schall	GAR
REML	$\hat{\sigma}^2$	0.0659	0.0654
	$\hat{\beta}$	−0.6083	−0.6219
		−0.5210	−0.5189
		−0.1294	−0.1158
		−0.1931	−0.2169
ML	$\hat{\sigma}^2$	0.0202	0.0240
	$\hat{\beta}$	−0.6952	−0.6854
		−0.4102	−0.4195
		−0.0890	−0.0863
		−0.1348	−0.1677

TAB. VI – *Résultats estimation des trois méthodes : modèle binomial - lien probit.*

seul effet aléatoire ayant 4 réalisations selon un plan équilibré. Pour chaque valeur de la variance σ^2 de cet effet (allant de 0.05 à 2), nous avons simulé 200 vecteurs de données de longueur 40. Le tableau contient alors le résumé (moyenne, écart-type) des 200 estimations par chacune des trois méthodes.

À l'aide de ce tableau et d'autres simulations réalisées, nous pouvons faire les remarques suivantes.

- Pour $\sigma^2 = 2$, la méthode GAR donne des résultats délirants. En effet, plus les valeurs de σ^2 en simulation sont élevées, plus la méthode se détériore. Ceci s'explique par le fait

		Valeurs estimées					
		Schall		**EK**		**GAR étendue**	
Valeurs simulées		$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\beta}$
$\sigma^2 = 0.05 \quad \beta = 1$	moy.	0.0558	1.0016	0.0558	1.0031	0.0569	1.0013
	e.t.	0.0484	0.1055	0.0484	0.1057	0.0508	0.1057
$\sigma^2 = 0.5 \quad \beta = 1$	moy.	0.4924	0.9637	0.4931	0.9656	0.5831	0.9642
	e.t.	0.4122	0.3439	0.4124	0.3448	0.8470	0.3510
$\sigma^2 = 1 \quad \beta = 1$	moy.	0.9141	0.9679	0.9151	0.9701	1.1351	0.9734
	e.t.	0.6960	0.4591	0.6964	0.4604	1.1373	0.4713
$\sigma^2 = 2 \quad \beta = 1$	moy.	1.8247	0.8822	1.8229	0.8849	242.898	1.5926
	e.t.	1.3365	0.6127	1.3306	0.6145	3288.39	7.6353

TAB. VII – *Résultats simulation comparée des trois méthodes : modèle binomial - lien logit.*

que la méthode GAR utilise une approximation de la variance. Or, cette approximation est valide pour les faibles valeurs de σ^2. Donc ces résultats confortent l'idée que cette méthode est à utiliser avec précautions dans un domaine de validité de l'approximation. Les méthodes Schall et EK continuent de bien se comporter pour des grandes valeurs de σ^2.

- On ne constate pas de différence significative entre les méthodes EK et Schall. Ceci n'est peut être pas surprenant dans la mesure où la différence entre ces deux méthodes réside dans l'éloignement des matrices $W_{\beta,\xi}$ et $W_\beta = E(W_{\beta,\xi})$. Or, prendre la moyenne sur 200 simulations a peut-être aussi pour effet de moyenniser la différence entre $W_{\beta,\xi}$ et W_β, qui logiquement devrait être faible. Cependant, en regardant de plus près les trajectoires sur ces 200 simulations, même différentes, elles restent très proches.

- Lorsqu'on augmente le nombre de réalisations q de l'effet aléatoire (de 4 à 8 par exemple), on améliore la réponse de la méthode GAR pour les grandes valeurs de σ^2. De plus, on observe qu'à N/q fixé, les résultats se détériorent avec N diminuant.

Ces différentes tendances se retrouvent de façon générale pour d'autres modèles simulés. Nous avons aussi essayé de changer les matrices U de plan des effets aléatoires mais sans noter de résultats réellement différents.

Notons pour finir que ces trois méthodes, pour σ^2 petit, donnent des résultats très proches en moyenne.

6 Conclusion

Nous avons donc tenté de comparer ces trois procédures en exhibant les liens qu'il existait entre elles ainsi que leurs différences de démarche vis-à-vis du conditionnement du modèle initial. Sur les jeux de données simulés, en moyenne, le comportement de ces procédures est semblable même si des différences peuvent apparaître sur des jeux de données isolés. Notons cependant que, pour des raisons pratiques, ces simulations ont été réalisées pour un faible nombre de réalisations de l'effet aléatoire (4 ou 8).

Il reste que l'on ne dispose pas de résultats théoriques sur les propriétés de ces estimateurs, permettant d'aller plus loin dans leur comparaison. Des avancées ont pu être réalisées, notamment dans [12] en ce qui concerne l'estimateur des effets fixes $\hat{\beta}$. Cela reste plus délicat pour les composantes de la variance.

D'autre part, il serait intéressant de pouvoir mettre en place des critères de choix entre ces différentes démarches, suggérant de privilégier telle ou telle méthode selon le modèle considéré.

Références

[1] D. A. Anderson and M. Aitkin. Variance Components Models with Binary Response : Interviewer Variability. *Journal of the Royal Statistical Society, B*, 47(2): 203–210, 1985.

[2] D. A. Anderson and J. P. Hinde. Random effects in generalized linear models and the EM algorithm. *Communications in Statistics - Theory and Methods*, 17(11): 3847–3856, 1988.

[3] N. E. Breslow and D. G. Clayton. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421): 9–25, 1993.

[4] B. Engel and A. Keen. A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, 48(1): 1–22, 1994.

[5] J.-L. Foulley. Heteroskedastic threshold models with applications to the analysis of calving difficulties. *International Bull Evaluation Service*, 18: 3–11, 1998.

[6] J.-L. Foulley and S. Im. A marginal quasi-likelihood approach to the analysis of poisson variables with generalized linear mixed models. *Genetics, Selection, Evolution*, 25: 101–107, 1993.

[7] O. Gaudoin, C. Lavergne, and J.-L. Soler. A generalized geometric de-eutrophication software reliability model. *IEEE Trans. on Reliability*, 43(4): 536–541, 1994.

[8] D. Gianola and J.-L. Foulley. Sire evaluation for ordered categorical data with a threshold model. *Genetic Selection Evolution*, 15(2): 201–224, 1983.

[9] A. R. Gilmour, R. D. Anderson, and A. L. Rae. The analysis of binomial data by a generalized linear mixed model. *Biometrika*, 72(3): 593–599, 1985.

[10] C. R. Henderson, O. Kempthorne, S. R. Searle, and C.N. VonKrosig. Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15: 192–218, 1959.

[11] J. Hinde. Compound poisson regression models. In R. Gilchrist, editor, *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, number 14 in Lecture Notes in Statistics, pages 109–121. Springer-Verlag, 1982.

[12] Y. Lee and J. A. Nelder. Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society, B*, 58(4): 619–678, 1996.

[13] P. McCullagh and J. Nelder. *Generalized Linear Models.* Chapman and Hall, second edition, 1989.

[14] C. A. McGilchrist. Estimation in Generalized Mixed Models. *Journal of the Royal Statistical Society, B*, 56(1): 61–69, 1994.

[15] C. R. RAO and J. KLEFFE. *Estimation of variance components and applications.* Number 3 in Series in Statistics and Probability. North Holland, 1988.

[16] R. SCHALL. Estimation in generalized linear models with random effects. *Biometrika*, 78(4): 719–727, 1991.

[17] S. R. SEARLE, G. CASELLA, and C. E. MCCULLOCH. *Variance components.* John Wiley & Sons, 1992.

[18] L. R. SHAEFFER and J. W. WILTON. Methods of sire evaluation for calving ease. *Journal of Dairy Science*, 59: 544–551, 1976.

[19] R. STIRATELLI, N. LAIRD, and J. H. WARE. Random-Effects Models for Serial Observations with Binary Response. *Biometrics*, 40: 961–971, 1984.

[20] C. TROTTIER. A quasi-score marginal approach in generalized linear mixed models. Rapport de recherche Inria no 3522 (soumis), 1998.

[21] R. W. M. WEDDERBURN. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, 61: 439–447, 1974.

Table des matières

1 Introduction **3**

2 Les modèles étudiés **3**
2.1 Définition des GL2M 3
2.2 Un exemple de modélisation en génétique animale 5

3 Retour sur l'estimation dans les GLM et les L2M **6**
3.1 Estimation dans les GLM 6
3.2 Estimation par maximum de quasi-vraisemblance 7
3.3 Estimation dans les L2M 7
3.3.1 Procédure directe 8
3.3.2 Les équations de Henderson 8

4 Estimation dans les GL2M **9**
4.1 Une relecture de la méthode EK 9
4.1.1 Étape de linéarisation 9
4.1.2 Étape d'estimation 10
4.2 La méthode Schall 12
4.3 La méthode GAR étendue 13

5 Comparaison des trois procédures **14**
5.1 Échelle qualitative de déconditionnement 14
5.2 Des résultats numériques 15
5.2.1 Des données réelles 15
5.2.2 Des données simulées 16

6 Conclusion **19**

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e 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