Les GL2M : extension de la méthode GAR

Catherine Trottier

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Les GL2M : extension de la méthode GAR

Catherine Trottier

N° 3029

Novembre 1996

THÈME 4

ISSN 0249-6399

INRIA
RHÔNE-ALPES

Les GL2M : extension de la méthode GAR

Catherine Trottier

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n° 3029 — Novembre 1996 — 36 pages

Résumé : Nous nous intéressons ici à l'estimation de paramètres dans des modèles linéaires généralisés mixtes (GL2M). Gilmour, Anderson et Rae en 1985 ont proposé une méthode d'estimation dans un modèle avec lien probit pour des données binomiales. En 1993, Foulley et Im ont adapté la méthode GAR au cas de données poissoniennes. Pour ces deux modélisations, nous proposons une nouvelle lecture de la méthode en levant l'hypothèse d'homogénéité des variances des variables sous-jacentes. Ensuite, nous présentons une adaptation à des données exponentielles et donnons, pour finir, une formalisation qui permet d'envisager le cas de données binomiales dans un modèle avec lien logit.

Mots-clé : Effets aléatoires, composantes de la variance, modèle linéaire généralisé mixte, fonction de lien, modèle marginal, méthode GAR.

(Abstract: pto)

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

GAR's method extended in GL2M

Abstract: In this work we focus on parameter estimation in generalized linear mixed models (GL2M). Gilmour, Anderson and Rae proposed in 1985 an estimation method (GAR) in a probit link model for binomial data. In 1993, Foulley and Im adapted this method to Poisson data. In these two cases we present the GAR's method, relaxing the hypothesis of variance homogeneity of underlying variables. We then propose an adaptation to exponential data and eventually, we give a formal description which enables us to propose a method for binomial data in a model with logit link.

Key-words: Random effects, variance components, generalized linear mixed model, link function, marginal model, GAR's method.

1 Introduction

Développée par Gilmour, Anderson et Rae [10], cette méthode (désignée par GAR dans toute la suite) a été conçue pour permettre l'estimation des effets fixes dans un modèle mixte adapté à des données binomiales. En fait, elle rend aussi possible une prédiction des effets aléatoires ainsi qu'une estimation de leur variance.

Les données binomiales, pour lesquelles la méthode a été initialement mise en place, résultent d'une classification en deux catégories. Cette classification suppose l'existence sous-jacente d'une variable aléatoire de loi normale dont on ne peut observer aucune réalisation mais pour laquelle on sait dire si un seuil a été atteint.
Pour ce type de données discrètes, ou pour leur extension à des données polytomiques résultant d'une classification en plusieurs catégories ordonnées, Gianola [8] constate l'inadaptation des méthodes d'analyse développées pour des données continues distribuées selon une loi normale.

Pour pallier cela, divers auteurs (Gianola & Foulley [9], Harville & Mee [12], Stiratelli et al. [16]) développent des méthodes d'estimation dans des modèles adaptés à ces données. Ces méthodes utilisent des arguments différents (bayésiens notamment) mais sont en finalité équivalentes. La méthode GAR, quant à elle, apparaît comme originale. Elle se base sur l'utilisation de la fonction de quasi-vraisemblance marginale dans le cadre des modèles linéaires généralisés (GLM) (McCullagh & Nelder [13]). L'article fondateur fera l'objet de diverses relectures : Foulley, Gianola & Im [4], Foulley & Manfredi [6]. Une extension aux données multicatégories en sera proposée par les auteurs eux-mêmes : Gilmour, Anderson & Rae [11]. De plus, Foulley & Im [5] en suggèrent une adaptation dans le cas de données poissonniennes.

Nous allons ici, dans un premier temps, présenter une relecture de cette méthode pour des données binomiales en relâchant l'hypothèse d'origine selon laquelle les variances des variables sous-jacentes sont homogènes.
La deuxième section montre comment une démarche similaire a pu être adoptée pour traiter le cas de données poissonniennes.
Puis, nous proposons une adaptation pour des données exponentielles dans un modèle avec lien logarithmique (lien non canonique cf [13]).
Ces trois cas ayant été étudiés, nous proposons une formalisation permettant de les unifier. En effet, grâce à une même écriture, il s'avère possible de prendre en compte des arguments d'approximation différents pour chacun des trois cas.
Ceci permet d'envisager d'autres cas et notamment celui, très usité, de données binomiales dans un modèle avec lien (canonique) logistique : c'est ce qui est fait dans la dernière section.

2 GAR - Données binomiales - Lien probit

2.1 Le modèle et les notations

Vu le type de données auquel on s'intéresse (cf introduction) et étant donnée la distribution normale sous-jacente, le modèle adopté est un modèle mixte binomial avec lien probit. Rappelons que dans les GLM, la fonction de lien canonique associée à la loi binomiale est la fonction de lien logit. Ce n'est donc pas ce cas qui est envisagé ici.

On note y le vecteur $(N \times 1)$ des observations, réalisation du vecteur aléatoire Y. Conditionnellement au vecteur d'effets aléatoires : U (vecteur aléatoire non observable de dimension q), on suppose que les composantes Y_i sont indépendantes et que :

$$\forall i \in \{1, ..., N\} \quad Y_i \mid U = u \ \sim \ \mathcal{B}in(n_i, p_{u,i}) \ ,^{1}$$

ou encore de façon équivalente : $Y_i = \sum_{r=1}^{n_i} Y_{ir}$ où $Y_{ir} \mid U = u \ \sim \ \mathcal{B}in(1, p_{u,i})$ indépendantes.
On s'intéresse au vecteur $(N \times 1)$ des fréquences $f_i = \dfrac{y_i}{n_i}$ et on note F_i les variables aléatoires associées, composantes du vecteur $F_{(N \times 1)}$.

Le prédicteur linéaire intervenant dans le modèle comporte une partie effet fixe et une partie effet aléatoire : $\eta_U = X\beta + ZU$ où $X_{(N \times p)}$ et $Z_{(N \times q)}$ sont des matrices connues, $\beta_{(p \times 1)}$ est un vecteur d'effets fixes et $U = (U_1', ..., U_K')'$ un vecteur de K effets aléatoires. La $j^{\text{ème}}$ composante U_j est le vecteur aléatoire de dimension q_j correspondant au $j^{\text{ème}}$ effet aléatoire. On suppose de plus que U de dimension q (avec $q = \sum_{j=1}^{K} q_j$) est distribué de la façon suivante : $U \sim \mathcal{N}(0, G)$ où G prendra la forme d'une matrice diagonale [2] par blocs : $G = \text{diag}\{\alpha_j^2 A_j\}_{j=1,...,K}$. Les matrices $A_{j\,(q_j \times q_j)}$ sont supposées connues et on appelle $\alpha_1^2, ..., \alpha_K^2$ les composantes de la variance.

Le lien entre le prédicteur linéaire et l'espérance conditionnelle des F_i se fait alors par l'intermédiaire de la fonction de répartition de la loi normale centrée réduite :

$$\begin{array}{rcccl} E(F_i|U) & = & p_{U,i} & = & \Phi(\eta_{U,i}) \\ & & & = & \Phi((X\beta + ZU)_i) \\ & & & = & \Phi(x_i'\beta + z_i'U) \end{array}$$

où x_i' et z_i' sont les $i^{\text{èmes}}$ lignes de X et Z respectivement.

1. On indice par U ou u tous les objets dépendants respectivement du vecteur d'effets aléatoires U ou de sa réalisation u.
2. On note : diag$\{m_i\}$ la matrice diagonale M dont les éléments diagonaux sont $M_{ii} = m_i$.

Comme nous l'avons déjà mentionné, ce lien est inhérent à l'hypothèse d'existence de variables normales sous-jacentes. En effet, pour chacune des variables Y_{ir}, notons L_{ir} la variable latente associée : $Y_{ir} = 1 \Leftrightarrow L_{ir} > 0$. Supposons ensuite que $L_{ir} = x_i'\beta + z_i'U + \varepsilon_{ir}$ où $\varepsilon_{ir} \sim \mathcal{N}(0,1)$ indépendantes, on obtient alors :

$$\begin{aligned} p_{U,i} &= E(F_i|U) = E(Y_{ir}|U) = P(Y_{ir} = 1|U) = P(L_{ir} > 0|U) \\ &= 1 - \Phi(-x_i'\beta - z_i'U) \\ &= \Phi(x_i'\beta + z_i'U) \end{aligned}$$

De plus, on peut calculer les covariances et corrélations entre ces variables sous-jacentes. On a :

$$\forall i \in \{1, ..., N\} \forall r, r' \in \{1, ..., n_i\} \ \mathrm{cov}(L_{ir}, L_{ir'}) = z_i' G z_i \ ,$$

et

$$\forall i, j \in \{1, ..., N\} \forall r \in \{1, ..., n_i\}, \forall r' \in \{1, ..., n_j\} \ \mathrm{cov}(L_{ir}, L_{jr'}) = z_i' G z_j \ .$$

Ce qui, en notant $\sigma_i^2 = z_i' G z_i$, nous permet d'obtenir :

$$\begin{aligned} \mathrm{corr}(L_{ir}, L_{ir'}) &= \frac{\sigma_i^2}{1 + \sigma_i^2} &= t_{ii} \\ \mathrm{corr}(L_{ir}, L_{jr'}) &= \frac{z_i' G z_j}{\sqrt{1 + \sigma_i^2}\sqrt{1 + \sigma_j^2}} &= t_{ij} \ . \end{aligned}$$

On désignera par $T_{(N \times N)}$ la matrice de corrélations des variables sous-jacentes.

Remarquons ici que le paramètre introduit $\sigma_i^2 = z_i' G z_i$ coïncide bien souvent avec la composante de la variance α_i^2. En effet, dans de nombreux cas comme par exemple dans des modèles liés à la sélection animale (cf. Gianola et Foulley [9]), le vecteur ligne z_i' n'est composé que d'un seul 1 et de 0.

2.2 La méthode

La méthode se découpe principalement en deux étapes. Les effets aléatoires n'étant pas directement observés, on se libère dans un premier temps du conditionnement pour pouvoir travailler dans un modèle marginal où Y vecteur à expliquer est observé. Ensuite, après approximation de la matrice de variance de Y, on pourra par l'intermédiaire des équations de Henderson procéder à l'estimation des composantes de la variance.

2.2.1 Etape de "marginalisation" et estimation de β

Cette étape repose sur le calcul des espérances et matrice de variance marginales. On "intègre" donc les effets aléatoires dont la présence introduit une dépendance entre les données. Ensuite, les expressions obtenues permettent de construire la fonction de quasi-vraisemblance marginale (à défaut de pouvoir obtenir la vraisemblance marginale, la densité n'ayant pas d'expression explicite) ; on procède alors à l'estimation du vecteur des paramètres inconnus β.

Calcul de l'espérance marginale: Π

$$\begin{aligned} E(F_i) \quad = \quad \Pi_i \quad &= \quad E(E(F_i|U)) \\ &= \quad E(p_{U,i}) \\ &= \quad E(\Phi(x_i'\beta + z_i'U)) \end{aligned}$$

on obtient (cf. annexe A) : $\Pi_i = \Phi(\frac{x_i'\beta}{\sqrt{1+\sigma_i^2}})$.

On notera par la suite $\eta_i^* = \frac{x_i'\beta}{\sqrt{1+\sigma_i^2}}$, que nous désignerons par prédicteur linéaire dans le modèle marginalisé.

Ainsi, dans ce modèle marginalisé, le lien entre l'espérance et le prédicteur linéaire se fait par l'intermédiaire de la fonction de lien probit. Cette conservation du lien lors du passage de l'espérance conditionnelle à l'espérance marginale est une propriété inhérente à la fonction de lien inverse Φ. Nous verrons cependant qu'elle s'applique à d'autres cas. Elle s'avère essentielle pour la mise en place de cette méthode.

Nous nous apercevons ici que le modèle *marginalisé* ainsi considéré où $\Pi_i = \Phi(\frac{x_i'\beta}{\sqrt{1+\sigma_i^2}})$ et le modèle *marginal* comme défini par Breslow et Clayton [1] par $\Pi_i = \Phi(x_i'\beta)$ ne coïncident pas. La présence de l'effet multiplicatif $\frac{1}{\sqrt{1+\sigma_i^2}}$, indique que le modèle *marginalisé* a su prendre en compte des perturbations introduites par les effets aléatoires tandis que le modèle *marginal* dans sa définition oublie leur présence.

D'autre part, il est important de noter que la méthode GAR a été introduite dans le cas où $\forall i \in \{1, ..., N\}\ \sigma_i^2 = \sigma^2$. Avec cette hypothèse, il est possible d'envisager le changement d'échelle suivant : $\tilde{\beta} = \frac{\beta}{\sqrt{1+\sigma^2}}$, et de poursuivre les calculs sur cette nouvelle échelle. C'est ce qu'ont fait Foulley et Manfredi [6].
Cependant, ici on abandonne cette interprétation puisqu'on supprime l' hypothèse d'homogénéité des variances pour considérer le cas plus général de variances hétérogènes.

Calcul de la matrice de variance-covariance marginale: V

Intéressons nous tout d'abord au calcul des éléments diagonaux de V : variances des variables F_i.

On a $\quad \mathrm{var}(F_i) = \frac{1}{n_i^2}\left[\sum_{r=1}^{n_i}\mathrm{var}(Y_{ir}) + \sum_{r\neq r'}\mathrm{cov}(Y_{ir}, Y_{ir'})\right]$.

La formule classique $\mathrm{var}(Y_{ir}) = E(\mathrm{var}(Y_{ir}|U)) + \mathrm{var}(E(Y_{ir}|U))$ nous permet d'obtenir :

$$\mathrm{var}(Y_{ir}) = \Pi_i(1-\Pi_i) \ .$$

Et, à l'aide des variables normales sous-jacentes L_{ir}, on a :

$$\mathrm{cov}(Y_{ir}, Y_{ir'}) = \Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi^2(\eta_i^*) \ ,$$

où Φ_2 : fonction de répartition de la loi normale bivariée [3].

Ainsi,

$$\begin{aligned} \mathrm{var}(F_i) &= \frac{\Pi_i(1-\Pi_i) + (n_i-1)[\Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi^2(\eta_i^*)]}{n_i} \\ &= \frac{\Pi_i(1-\Pi_i) - [\Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi^2(\eta_i^*)]}{n_i} + [\Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi^2(\eta_i^*)] \ . \end{aligned}$$

D'autre part, les covariances des variables F_i, F_j pour $i \neq j$ (éléments non diagonaux de V) s'expriment de la façon suivante :

$$\begin{aligned} \mathrm{cov}(F_i, F_j) &= \mathrm{cov}(Y_{ir}, Y_{jr'}) \qquad (i \neq j) \\ &= \Phi_2(\eta_i^*, \eta_j^*, t_{ij}) - \Phi(\eta_i^*)\Phi(\eta_j^*) \ , \end{aligned}$$

On peut ainsi décomposer V en somme de deux matrices (dont l'une est diagonale) :

$$V = A + B$$

$$\begin{aligned} \text{où} \quad A &= \mathrm{diag}\left\{ \frac{\Pi_i(1-\Pi_i) - [\Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi^2(\eta_i^*)]}{n_i} \right\}_{i=1,\ldots,N} \\ \text{et} \quad B &= \left\{ \Phi_2(\eta_i^*, \eta_j^*, t_{ij}) - \Phi(\eta_i^*)\Phi(\eta_j^*) \right\}_{i,j=1,\ldots,N} \ . \end{aligned}$$

Estimation à l'aide de la fonction de quasi-vraisemblance

Ayant ainsi obtenu l'expression des deux premiers moments des variables $F_1, \ldots, F_N$, on est en mesure de définir la fonction de quasi-vraisemblance associée. Et c'est l'estimation par maximisation de cette fonction que l'on va considérer. En effet, comme nous l'avons déjà signalé, il apparait impossible d'obtenir une expression analytique de la fonction de vraisemblance, l'utilisation de la notion de quasi-vraisemblance nous permet dans ce cas

3. On note $\Phi_2(y_1, y_2, \rho) = P([Y_1 \leq y_1] \cap [Y_2 \leq y_2])$ où Y_1, Y_2 identiquement distribuées de loi $\mathcal{N}(0,1)$ avec $\mathrm{corr}(Y_1, Y_2) = \rho$.

de prendre en compte l'information contenue dans le calcul des deux premiers moments marginaux.

Pour cela, au vu de l'expression de l'espérance marginale, nous allons plonger le modèle marginal dans la structure d'un GLM. Ainsi, dans un premier temps en supposant les composantes de la variance connues (V dépend alors uniquement du paramètre β), on utilise les équations du maximum de quasi-vraisemblance dans les GLM (méthode rappelée en annexe B). On obtient le système itératif suivant, s désignant l'indice d'itération :

$$(D'V^{-1}D)\Delta\beta^{[s]} = D'V^{-1}(f - \Pi^{[s]}) \ ,$$

où f est le vecteur des fréquences observées, et $D = \dfrac{\partial \Pi}{\partial \beta'}$. D'après l'expression de Π obtenue précédemment et en définissant les matrices : $K = \text{diag}\,\{\varphi(\eta_i^*)\}_{i=1,\ldots,N}$ où φ est la fonction de densité de la loi normale centrée réduite, et $M = \text{diag}\left\{\dfrac{1}{\sqrt{(1+\sigma_i^2)}}\right\}_{i=1,\ldots,N}$, on exprimera D de la façon suivante :

$$\begin{aligned} D &= KMX \\ &= LX \qquad \text{avec} \qquad L = KM = \text{diag}\left\{\frac{\varphi(\eta_i^*)}{\sqrt{(1+\sigma_i^2)}}\right\}_{i=1,\ldots,N} . \end{aligned}$$

Notons qu'avec cette définition de la matrice M, l'élément diagonal de la matrice T peut aussi s'écrire $t_{ii} = (MZ)_i' G (MZ)_i$ où $(MZ)_i'$ est la $i^{\text{ème}}$ ligne de la matrice MZ.

Avec ces notations, on aboutit donc au système itératif en β (cf annexe B) :

$$(X'W^{[s]-1}X)\beta^{[s+1]} = X'W^{[s]-1}\zeta^{[s]} \tag{1}$$

où $\begin{aligned} \zeta^{[s]} &= X\beta^{[s]} + L^{[s]-1}(f - \Pi^{[s]}) \\ W^{[s]} &= L^{[s]-1}V^{[s]}L^{[s]-1} \end{aligned}$

Remarquons que les matrices V, L ainsi que W dépendent de la valeur courante de β.

Mais dans ce qui précède, les valeurs $\sigma_1^2, \ldots, \sigma_N^2$ dépendent des composantes de la variance $\alpha_1^2, \ldots, \alpha_K^2$ par l'expression $\sigma_i^2 = z_i' G z_i$, où $G = \text{diag}\{\alpha_j^2 A_j\}_{j=1,\ldots,K}$. Or, ces composantes de la variance sont en général des paramètres inconnus. Il est donc nécessaire de proposer une méthode d'estimation de ces paramètres, afin de pouvoir remplacer dans le système (1) les vraies valeurs de α_j^2 par leurs estimations.

2.2.2 Etape d'approximation de V et d'estimation des α_j^2

Afin de proposer une estimation des composantes de la variance, on procède à une approximation de la matrice de variance V et on reconnait alors la forme classique utilisée dans les équations de Henderson (cf annexe C). En formant ces équations, on pourra ainsi obtenir les estimations souhaitées. De plus, la solution en β à ces équations étant identique à celle du système (1), il sera possible d'obtenir simultanément une estimation de l'effet fixe. Il est aussi important de noter que ces équations permettent de prédire le vecteur d'effets aléatoires U. On utilisera pour cela les valeurs obtenues dans le processus itératif.

Approximation de la matrice V

En supposant que les éléments de la matrice $T_{(N\times N)}$ soient petits (ce qui équivaut à considérer que $\forall i,j \in \{1,...,N\}$ les éléments $z_i' G z_j$ sont petits ou encore bien souvent que les composantes de la variance elles-même sont petites), on peut alors procéder à une approximation au premier ordre de la fonction de répartition Φ_2, pour ρ proche de 0 :

$$\Phi_2(y_1, y_2, \rho) \approx \Phi(y_1)\Phi(y_2) + \rho\varphi(y_1)\varphi(y_2) \ ,$$

et on obtient :

$$\begin{array}{rcccl} A & \approx & V_0 & = & \mathrm{diag}\left\{ \dfrac{\Pi_i(1-\Pi_i) - t_{ii}\varphi^2(\eta_i^*)}{n_i} \right\}_{i=1,\dots,N} \\ B & \approx & V_1 & = & \left\{ t_{ij}\varphi(\eta_i^*)\varphi(\eta_j^*) \right\}_{i,j=1,\dots,n} \\ & & & = & \left\{ \dfrac{\varphi(\eta_i^*)}{\sqrt{1+\sigma_i^2}} z_i' G z_j \dfrac{\varphi(\eta_j^*)}{\sqrt{1+\sigma_j^2}} \right\}_{i,j=1,\dots,N} \end{array}$$

Ainsi, on a :

$$\begin{array}{rcccl} V & \approx & V_0 & + & V_1 \\ & \approx & V_0 & + & LZGZ'L \ , \end{array}$$

ce qui donne comme approximation de W :

$$\begin{array}{rcccl} W & \approx & L^{-1}V_0L^{-1} & + & ZGZ' \\ & \approx & R & + & ZGZ' \end{array}$$

où R est donc de la forme : $R = \mathrm{diag}\left\{ \dfrac{1}{n_i}\left[\dfrac{\Pi_i(1-\Pi_i)\ (1+\sigma_i^2)}{\varphi^2(\eta_i^*)} - \sigma_i^2 \right] \right\}_{i=1,\dots,N}$.

Les équations de Henderson

La matrice des poids W (ou matrice de variance de ζ) étant ainsi approchée et écrite sous la forme : $R + ZGZ'$, cela nous permet d'obtenir une estimation de β par résolution itérative du système d'équations de Henderson :

$$\begin{pmatrix} X'R^{[s]-1}X & X'R^{[s]-1}Z \\ Z'R^{[s]-1}X & Z'R^{[s]-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X'R^{[s]-1}\zeta^{[s]} \\ Z'R^{[s]-1}\zeta^{[s]} \end{pmatrix}. \quad (2)$$

Cependant, la matrice G n'étant pas connue, il nous est aussi nécessaire d'obtenir une estimation des composantes de la variance afin de pouvoir définir complètement ce système.
Un avantage que l'on trouve à utiliser ces équations, est qu'elles nous fournissent non seulement une estimation de β mais aussi une solution en u.
Ainsi, selon la méthode de Henderson (cf annexe C), le sous-produit $\hat{u}$ obtenu est interprété comme une prédiction du vecteur d'effets aléatoires dans le modèle linéaire mixte associé (L2M).

Sachant que $G = \text{diag}\{\alpha_j^2 A_j\}_{j=1,\ldots,K}$, on peut alors estimer le vecteur $(K \times 1)$ des composantes de la variance $\alpha^2 = (\alpha_1^2, .., \alpha_K^2)'$ par la méthode itérative :

$$\forall j \in \{1, ..., K\}\ \hat{\alpha}_j^{2\,[s+1]} = \frac{\hat{u}_j' A_j^{-1} \hat{u}_j}{q_j - \dfrac{tr(A_j^{-1} C_{jj})}{\hat{\alpha}_j^{2\,[s]}}}$$

où C_{jj} est le $j^{\text{ème}}$ bloc correspondant au $j^{\text{ème}}$ effet aléatoire (u_j) dans l'inverse de la matrice du système.

Ainsi, on pourra itérer la résolution du système de Henderson précédent à partir des valeurs courantes de β, α^2 pour obtenir une estimation de ces deux vecteurs de paramètres ainsi qu'une prédiction de u.

2.3 Remarques

Les auteurs de la méthode GAR analysent les valeurs de u obtenues à convergence de l'algorithme comme une prédiction des effets aléatoires du modèle initial. Autant, l'interprétation donnée au sous-produit $\hat{u}$ paraît pertinente dans le L2M associé, autant on peut discuter le lien établi avec les effets aléatoires du modèle initial. En effet, pour procéder à l'estimation, on s'est justement placé dans le modèle marginalisé dans lequel les effets

aléatoires n'interviennent plus. Et, l'étape de "marginalisation" a déjà pris en compte, par l'intermédiaire de l'effet multiplicatif, la présence initiale de ces effets.

D'autre part, revenons sur le type des données considérées, qui sont, dans cette section, des données binomiales. Comme nous l'avons déja mentionné, ces données sont issues de somme de données binaires à seuil. En effet, à partir d'une variable normale sous-jacente non observée, les données révèlent si un seuil τ a été atteint ou non (τ ayant été fixé à 0 dans la présentation). De même, on peut imaginer une classification en plusieurs catégories ordonnées associées à plusieurs seuils. C'est ce à quoi se sont intéressés les auteurs dans leur second article : Gilmour, Anderson et Rae [11]. D'autres auteurs ont aussi envisagé ce cas. Nous ne présentons pas ici l'extension de la méthode GAR à ce type de données. La démarche est tout à fait semblable au cas binomial. Cependant, tout en plaçant le premier seuil à l'origine, ce type de données introduit des paramètres supplémentaires à estimer, à savoir les autres seuils.

3 GAR - Données poissonniennes - Lien logarithme

Nous nous intéressons toujours à des données discrètes mais qui sont, conditionnellement aux effets aléatoires, distribuées selon une loi de Poisson. Foulley, Gianola et Im [3] justifient l'utilisation de cette distribution pour modéliser des données de reproduction.
Dans leur note [5], Foulley et Im ont décrit comment la méthode GAR, initialement développée pour des données binomiales avec lien probit, peut être adaptée à ce type de données.
L'adaptation aux données poissonniennes qu'ils proposent, reste fidèle à la méthodologie GAR, dans la mesure où elle repose sur l'estimation par quasi-vraisemblance, la fonction de quasi-vraisemblance étant construite à partir des expressions marginales de l'espérance et d'une approximation de la matrice de variance-covariance. La démarche reste donc identique à celle du cas binomial mais c'est l'étape de calcul et d'approximation de cette matrice qui diffère.
Nous présentons cette adaptation mais dans des termes similaires à ceux du paragraphe précédent et nous nous attacherons à considérer de nouveau le cas où les variances des variables sous-jacentes ne sont pas homogènes.

3.1 Le modèle et les notations

Nous considérons à présent que les composantes du vecteur Y (dont le vecteur des observations y est une réalisation) sont, conditionnellement à U, indépendantes et de loi :

$$\forall i \in \{1, ..., N\} \quad Y_i \mid U = u \ \sim \ \mathcal{P}(\lambda_{u,i})$$

Le paramètre $\lambda_{u,i}$ de cette loi est une réalisation de la variable aléatoire $\lambda_{U,i}$ liée à la $i^{\text{ème}}$ composante du prédicteur linéaire. Nous envisageons ici le cas du lien canonique :

$\ln(\lambda_{U,i}) = x_i'\beta + z_i'U = \eta_{U,i}$.
On a donc pour cette distribution :

$$E(Y_i|U) = \lambda_{U,i} = \exp(x_i'\beta + z_i'U) = \mu_{U,i}.$$

On émet toujours l'hypothèse selon laquelle les effets aléatoires sont distribués selon une loi normale : $U \sim \mathcal{N}(0, G)$ avec $G = \text{diag}\{\alpha_j^2 A_j\}_{j=1,\ldots,K}$ matrice de variance des K effets aléatoires.

3.2 La méthode

3.2.1 Etape de "marginalisation" et estimation de β

Le premier objectif est la construction de la fonction de quasi-vraisemblance sur laquelle se basera l'estimation. Pour cela, il est nécessaire d'obtenir une expression de l'espérance et de la matrice de variance-covariance marginales.

Calcul de l'espérance marginale: μ
On intègre les différents effets aléatoires :

$$\begin{array}{rcccl} E(Y_i) & = & \mu_i & = & E(E(Y_i|U)) \\ & & & = & E(\lambda_{U,i}) \\ & & & = & E(\exp(x_i'\beta + z_i'U)) \end{array}$$

Du fait de la distribution normale de U et en utilisant l'expression de la fonction génératrice de cette distribution, on obtient :

$$\begin{array}{rcl} \mu_i & = & \exp(x_i'\beta + \dfrac{z_i'Gz_i}{2}) \\ & = & \exp(\eta_i*) \end{array}$$

où $\eta_i* = x_i'\beta + \dfrac{z_i'Gz_i}{2}$ que nous nommerons prédicteur linéaire marginalisé.

Une fois de plus, on peut constater que la "marginalisation" a conservé la fonction de lien inverse (ici la fonction exponentielle) reliant l'espérance et le prédicteur linéaire : c'est donc la même fonction tant au niveau conditionnel qu'au niveau marginal. Comme le remarquent Breslow et Clayton [1], cette "marginalisation" introduit uniquement un décalage dans le prédicteur mais, au contraire du cas binomial au paragraphe précédent, n'a aucun effet multiplicatif sur le paramètre β. On peut tout de même noter que, comme précédemment, la présence des effets aléatoires entraine une dépendance entre les données.

Calcul de la matrice de variance-covariance marginale : V

En utilisant les propriétés standards des lois de Poisson et log-normales et la formule de conditionnement, on obtient :

$$\begin{aligned} \mathrm{var}(Y_i) &= E(\lambda_{U,i}) + \mathrm{var}(\lambda_{U,i}) \\ &= \mu_i + \mu_i^2\,[\exp(z_i' G z_i) - 1] \end{aligned}$$

$$\begin{aligned} \text{et}\quad \mathrm{cov}(Y_i, Y_j) &= E(\lambda_{U,i}\lambda_{U,j}) - E(\lambda_{U,i})E(\lambda_{U,j}) \\ &= \mu_i\mu_j\,[\exp(z_i' G z_j) - 1] \qquad (i \neq j) \end{aligned}$$

On peut donc, dans ce cas aussi, lire la variance de Y, V, comme une somme de matrices (dont l'une est diagonale) :

$$V = A + B$$

où $A = \mathrm{diag}\{\mu_i\}_{i=1,\ldots,N}$
et $B = \{\mu_i\mu_j\,(\exp(z_i' G z_j) - 1)\}_{i,j=1,\ldots,N}$.

Estimation à l'aide de la fonction de quasi-vraisemblance

On procède alors à la construction de la fonction de quasi-vraisemblance et à sa maximisation, pour obtenir les équations suivantes :

$$D' V^{-1}(y - \mu) = 0$$

où $D = \dfrac{\partial \mu}{\partial \beta'} = KX$ avec $K = \mathrm{diag}\{\exp(\eta_i^*)\}_{i=1,\ldots,N}$,
V : matrice de variance-covariance de Y.

Ce qui conduit à :

$$(X' W^{[s]-1} X)\beta^{[s+1]} = X' W^{[s]-1} \zeta^{[s]}$$

avec $\zeta^{[s]} = X\beta^{[s]} + K^{[s]-1}(y - \mu^{[s]})$
$W^{[s]} = K^{[s]-1} V^{[s]} K^{[s]-1}$

Ceci serait suffisant si l'on connaissait les composantes de la variance. Comme dans le cas binomial, nous allons procéder à leur estimation.

3.2.2 Etape d'approximation de V et d'estimation des α_j^2

De même que l'on utilise un développement au premier ordre de la fonction Φ_2 pour le cas binomial lien probit, on utilise dans le cas Poisson lien logarithme l'approximation de la fonction exponentielle au voisinage de 0 pour obtenir, pour tout $i, j \in \{1, \ldots N\}$ où $z_i' G z_j$ proche de 0 :

$$\begin{array}{rcl} A & = & V_0 = \mathrm{diag}\{\mu_i\}_{i=1,\ldots,N} \\ B & \approx & V_1 = \{\mu_i\mu_j \; z_i' G z_j\}_{i,j=1,\ldots,N} \; . \end{array}$$

Ainsi, en remarquant que : $V_0 = \mathrm{diag}\{\mu_i\}_{i=1,\ldots,N} = \mathrm{diag}\{\exp(\eta_i^*)\}_{i=1,\ldots,N} = K$, on a :

$$\begin{array}{rcccl} V & \approx & V_0 & + & V_1 \\ & \approx & V_0 & + & KZGZ'K \; , \end{array}$$

d'où on approche la matrice des poids W par :

$$\begin{array}{rcccl} W & \approx & K^{-1}V_0K^{-1} & + & ZGZ' \\ & \approx & R & + & ZGZ' \end{array}$$

où R a la forme suivante : $R = \mathrm{diag}\{\dfrac{1}{\mu_i}\}_{i=1,\ldots,N} = \mathrm{diag}\{\exp(-\eta_i^*)\}_{i=1,\ldots,N}$.

On résout alors itérativement les équations de Henderson. Les valeurs courantes des paramètres $\beta^{[s]}$ et $\alpha^{2[s]}$ obtenues, interviennent à chaque itération pour former $\eta_i^{*[s]}$ et donc obtenir $R^{[s]}$, $\zeta^{[s]}$ et approcher $W^{[s]} \approx R^{[s]} + ZG^{[s]}Z'$.

3.3 Remarques

Dans ce cas aussi, il a donc été possible de donner une approximation de W permettant de reconnaître une structure de variance identique à celle des équations de Henderson. Pour pouvoir utiliser cette approximation, il est donc nécessaire de vérifier que les éléments $z_i'Gz_j$ sont proches de 0 ou encore que les composantes de la variance sont petites. Notons aussi que lors du calcul de l'espérance marginale la fonction de lien a été conservée.
D'autre part, comme dans le cas binomial, on peut émettre des doutes quant à la prédiction de u obtenue.

4 GAR - Données exponentielles - Lien logarithme

Dans tout ce paragraphe, nous allons maintenant considérer des données qui, conditionnellement aux effets aléatoires, sont distribuées selon une loi exponentielle (notée [4] $\mathcal{E}(\lambda)$). Ces données continues sont de nature complètement différente de celle des cas précédents. Elles ne relèvent pas du même cadre non plus. Alors que dans le cas des données binomiales ou de

4. la loi désignée par $\mathcal{E}(\lambda)$ est la loi exponentielle de densité définie sur $\mathbb{R}^+$ par $f(x) = \lambda e^{-\lambda x}$.

Poisson, de nombreuses applications relèvent du domaine de la génétique animale (Ducrocq [2]), on peut trouver des applications de modèles de loi exponentielle à effets aléatoires dans le domaine de la fiabilité des logiciels (Gaudoin, Lavergne & Soler [7]), où les effets aléatoires sont interprétés comme des effets de correction sur les logiciels.

Nous proposons ici une adaptation de la méthode GAR pour ce type de modèle de loi exponentielle à effets aléatoires, dans le cas d'un lien (non canonique) logarithmique. Notons que pour la distribution exponentielle, le lien canonique associé est la fonction inverse.

La démarche consiste toujours à exprimer tout d'abord l'espérance et la matrice de variance-covariance marginale V (l'introduction d'effets aléatoires ayant induit une dépendance entre les données), puis, par l'intermédiraire de la fonction de quasi-vraisemblance et après approximation de V, à procéder à la phase d'estimation.

4.1 Le modèle et les notations

Dans le cas présent, les composantes de Y sont, conditionnellement à U, indépendantes et de loi :

$$\forall i \in \{1, ..., N\}\ Y_i \mid U = u \ \sim\ \mathcal{E}(\lambda_{u,i}).$$

Ceci implique donc notamment $E(Y_i|U) = \dfrac{1}{\lambda_{U,i}} = \mu_{U,i}$.

Chacun des $\lambda_{U,i}$ est relié à la $i^{\text{ème}}$ composante du prédicteur linéaire $\eta_{U,i} = x_i'\beta + z_i'U$ par la fonction de lien logarithme :

$$\eta_{U,i} = \ln(\mu_{U,i}) \Longleftrightarrow \mu_{U,i} = \exp(x_i'\beta + z_i'U) = \frac{1}{\lambda_{U,i}}\ .$$

Une raison qui justifie le choix de ce lien est qu'il permet d'assurer la positivité du paramètre de la loi exponentielle, ce qui n'est pas le cas du lien inverse.

On garde la même distribution normale pour les effets aléatoires.

4.2 La méthode

4.2.1 Etape de "marginalisation" et d'estimation de β

Le choix de la fonction de lien est le même que dans le cadre du paragraphe précédent pour des données poissonniennes, et l'espérance conditionnelle s'exprime de la même manière à l'aide des effets fixes et aléatoires.
Ainsi, on peut envisager le calcul des élément marginaux en utilisant de nouveau les propriétés inhérentes à la distribution log-normale ou à la fonction génératrice de la loi normale.

Calcul de l'espérance marginale: μ

$$\begin{array}{rcccl} E(Y_i) & = & \mu_i & = & E(E(Y_i|U)) \\ & & & = & E(\mu_{U,i}) \\ & & & = & E(\exp(x_i'\beta + z_i'U) \\ & & & = & \exp(x_i'\beta + \frac{z_i'Gz_i}{2}) \\ & & & = & \exp(\eta_i*) \end{array} .$$

avec $\eta_i* = x_i'\beta + \frac{z_i'Gz_i}{2}$: prédicteur linéaire marginalisé.

Comme dans le cas poissonnien, on observe la conservation du lien et le fait que la "marginalisation" n'a pas introduit d'effet multiplicatif sur β mais uniquement un décalage dans le prédicteur linéaire.

Calcul de la matrice de variance-covariance marginale: V

D'après les hypothèses de loi et d'indépendance des variables aléatoires, on a :

$$\begin{array}{lcccl} \mathrm{var}(Y_i|U) & = & \frac{1}{\lambda_{U,i}^2} & = & \exp(2(x_i'\beta + z_i'U)) \quad = \quad \mu_{U,i}^2 \\ \mathrm{cov}(Y_i, Y_j|U) & = & 0 & & (i \neq j) \ . \end{array}$$

Ce qui nous permet d'obtenir :

$$\begin{array}{rcl} \mathrm{var}(Y_i) & = & E(\mathrm{var}(Y_i|U)) + \mathrm{var}(E(Y_i|U)) \\ & = & 2 * E(\frac{1}{\lambda_{U,i}^2}) - E(\frac{1}{\lambda_{U,i}})^2 \\ & = & \mu_i^2[2 * \exp(z_i'Gz_i) - 1] \ , \end{array}$$

$$\begin{array}{rcll} \text{et} \quad \mathrm{cov}(Y_i, Y_j) & = & E(\frac{1}{\lambda_{U,i}}\frac{1}{\lambda_{U,j}}) - E(\frac{1}{\lambda_{U,i}})E(\frac{1}{\lambda_{U,j}}) & \\ & = & \mu_i\mu_j[\exp(z_i'Gz_j) - 1] & (i \neq j) \ , \end{array}$$

et on écrit V comme somme de deux matrices (dont l'une est diagonale) :

$$V = A + B$$

$$\begin{array}{lcl} \text{où} \quad A & = & \mathrm{diag}\{\mu_i^2 \exp(z_i'Gz_i)\}_{i=1,\ldots,N} \\ \text{et} \quad B & = & \{\mu_i\mu_j \ (\exp(z_i'Gz_j) - 1)\}_{i,j=1,\ldots,N} \ . \end{array}$$

Estimation à l'aide de la fonction de quasi-vraisemblance

Les deux premiers moments étant calculés, on maximise la fonction de quasi-vraisemblance. Et on est amené, dans ce cas, à résoudre le système itératif :

$$(X'W^{[s]-1}X)\beta^{[s+1]} = X'W^{[s]-1}\zeta^{[s]} \tag{3}$$

$$\begin{array}{lrcl} \text{avec} & \zeta^{[s]} & = & X\beta^{[s]} + K^{[s]-1}(y-\mu^{[s]}) \ , \\ & W^{[s]} & = & K^{[s]-1}V^{[s]}K^{[s]-1} \ , \\ \text{et} & K^{[s]} & = & \mathrm{diag}\{\exp(\eta_i^*)^{[s]}\}_{i=1,\ldots,N} \ . \end{array}$$

4.2.2 Etape d'approximation de V et d'estimation des α_j^2

Faisant l'hypothèse que $\forall i, j \in \{1, ..., N\}^2$ les éléments $z_i' G z_j$ sont proches de 0, on utilise, de même que précédemment, une approximation de la fonction exponentielle au voisinage de 0 pour obtenir :

$$\begin{array}{rcl} A & \approx & V_0 = \mathrm{diag}\{\mu_i^2(1 + z_i'Gz_i)\}_{i=1,\ldots,N} \\ B & \approx & V_1 = \{\mu_i\mu_j \ z_i'Gz_j\}_{i,j=1,\ldots,N} \ . \end{array}$$

Ainsi, on a :

$$\begin{array}{rcccc} V & \approx & V_0 & + & V_1 \\ & \approx & V_0 & + & KZGZ'K \ . \end{array}$$

Ce qui nous conduit à l'approximation de W suivante :

$$\begin{array}{rcccc} W & \approx & K^{-1}V_0K^{-1} & + & ZGZ' \\ & \approx & R & + & ZGZ' \end{array}$$

où R a la forme suivante : $R = \mathrm{diag}\{1 + z_i'Gz_i\}_{i=1,\ldots,N}$.

W étant approchée sous cette forme, on ne résout pas le système (3) mais on en obtient des solutions grâce aux équations de Henderson (cf (2)), pour lesquelles on a donné les expressions de R et de ζ.
De même, les valeurs de u obtenues par résolution du système permettent une estimation des composantes de la variance.

4.3 Remarques

Remarquons que la condition d'approximation repose de nouveau sur la proximité des $z_i'Gz_j$ à 0. Les remarques sont similaires à celles des 2 cas précédents, la démarche l'étant aussi.

5 Une formalisation commune

5.1 Introduction

Notre objectif ici est de proposer une écriture commune permettant de regrouper les différents cas envisagés jusqu'à présent. Cette formalisation nous conduira à développer une nouvelle démarche. Elle permettra par la suite d'étudier l'adaptation de la méthode GAR à d'autres situations.

Tout d'abord résumons rapidement les objets communs utilisés lors de l'étude des trois cas. Pour cela, dans tout ce paragraphe, nous référerons par cas 1, 2 et 3 respectivement les trois situations :
- Cas 1 : Données binomiales - Lien probit.
- Cas 2 : Données poissonniennes - Lien logarithme.
- Cas 3 : Données exponentielles - Lien logarithme.

Il est essentiel de remarquer que dans tous ces cas, la matrice de variance V de Y obtenue après approximation se présente sous la forme : $V_0 + LZGZ'L$ où $L = KM$. C'est ce qui permet l'utilisation des équations de Henderson pour obtenir les estimations.
Les matrices K et M sont définies comme suit.
D'une part pour M, on a :

- cas 1 : $M = \text{diag}\left\{ \dfrac{1}{\sqrt{1 + z_i' G z_i}} \right\}_{i=1,\ldots,N}$
- cas 2 et 3 : $M = I_N$ (matrice identité de dimension N).

Cette matrice M correspond à la matrice des effets multiplicatifs que la "marginalisation" a pu introduire sur β.

Ainsi on a pu écrire $\eta^* = M\eta + C$ où :

η^*	:	prédicteur linéaire marginalisé introduit dans les sections précédentes,
$\eta = X\beta$	:	prédicteur linéaire marginal en référence à Breslow & Clayton (1993) [1] qui interprètent η comme le prédicteur linéaire dans le modèle marginal,
C	:	vecteur de décalage.

D'autre part, la matrice K s'écrit dans les trois cas de la façon suivante :

$$K = \text{diag}\{h'(\eta_i^*)\}_{i=1,\ldots,N} ,$$

où $h = g^{-1}$: inverse de la fonction de lien.
En effet : - cas 1 : $K = \text{diag}\{\varphi(\eta_i^*)\}_{i=1,\ldots,N}$
- cas 2 et 3 : $K = \text{diag}\{\exp(\eta_i^*)\}_{i=1,\ldots,N}$.

Bien que l'approximation de la matrice V repose sur une approximation dans un cas de la fonction de répartition de la loi normale centrée réduite bivariée, et dans les autres de la fonction exponentielle, nous allons voir qu'une nouvelle démarche permet d'aboutir à la même approximation finale.

5.2 Une nouvelle démarche

5.2.1 Introduction

Rappelons que le principe de la méthode GAR repose sur une estimation par maximisation de la fonction de quasi-vraisemblance marginale. Cette fonction est construite à partir des deux premiers moments marginaux. Cependant, le calcul exact de la matrice de variance marginale V peut dans certains cas s'avérer difficile. D'autre part, les composantes de la variance, intervenant notamment dans l'expression de V, constituent des paramètres inconnus du modèle et qu'il est nécessaire d'estimer. C'est pourquoi nous sommes amenés à considérer une approximation de V dont la forme permet d'utiliser les équations de Henderson.
Dans la nouvelle démarche que nous proposons ici, nous conservons le schéma général en deux étapes observé lors de l'étude précédente : "marginalisation" puis utilisation des équations de Henderson pour l'estimation. Notre démarche s'appuie sur la définition d'un modèle approché pour un nouveau vecteur aléatoire $\tilde{Y}$. Dans ce modèle, la fonction de quasi-vraisemblance marginale est identique à celle construite après approximation de V.
Pour cela, il est nécessaire d'imposer une hypothèse supplémentaire au modèle initial.

5.2.2 Le modèle initial

Nous considérons le modèle linéaire généralisé mixte défini succintement par :

- soit Y le vecteur $(N \times 1)$ à expliquer et y son observation,
- on suppose que conditionnellement aux effets aléatoires, les composantes de Y sont indépendantes et distribuées selon une loi de la famille exponentielle pour laquelle on a :

$$\begin{array}{rrcl} & E(Y_i|U) & = & \mu_{U,i} \\ \text{et} & \text{var}(Y_i|U) & = & v(\mu_{U,i}) \quad \text{où } v \text{ est la fonction de variance,} \end{array}$$

- on considère le prédicteur linéaire suivant (contenant effets fixes et aléatoires):

$$\eta_U = X\beta + ZU \ ,$$

- on relie ce prédicteur linéaire à l'espérance conditionnelle par la fonction de lien g $(h = g^{-1})$: $\eta_U = g(\mu_U)$.

L'hypothèse supplémentaire au modèle est la suivante. Nous supposons que le calcul de l'espérance marginale est réalisable et que l'on peut écrire :

$$\begin{aligned} E(Y_i) &= E(E(Y_i|U)) \\ &= E(h(\eta_{U,i})) \\ &= h(\eta_i*) \end{aligned}$$

avec $\eta^* = MX\beta + C$: prédicteur linéaire marginalisé (prédicteur au niveau marginal). M est la matrice des effets multiplicatifs, C est un vecteur de décallage indépendant de β.

Cette hypothèse de travail impose donc une conservation de la fonction de lien inverse. Notons qu'elle est vérifiée dans les trois cas étudiés.

5.2.3 Le modèle approché

Comme nous l'avons déjà mentionné, nous allons considérer que le vecteur des observations y est réalisation du vecteur aléatoire $\tilde{Y}$ défini ci-dessous. L'espérance marginale de ce vecteur sera identique à celle de Y et sa matrice de variance marginale $\tilde{V}$ s'avèrera être l'approximation de V déja rencontrée. Ainsi, la construction de la quasi-vraisemblance ne reposant que sur les deux premiers moments, nous obtiendrons la même fonction. Nous poursuivrons l'estimation en utilisant les équations de Henderson pour aboutir donc aux mêmes résultats.

Définition de $\tilde{Y}$

Considérons le vecteur aléatoire $\tilde{Y}$ dont les composantes $\tilde{Y}_i$ sont indépendantes.
Au sein de la famille exponentielle, définissons l'espérance conditionnelle de $\tilde{Y}_i$ par :

$$\begin{aligned} E(\tilde{Y}_i|U) &= \tilde{\mu}_{U,i} \\ &= h(\eta_i^*) + h'(\eta_i^*)\tilde{z}_i' U \ , \end{aligned}$$

où $\tilde{z}_i'$ est la $i^{\text{ème}}$ ligne de la matrice $\tilde{Z} = MZ$. Grâce à l'hypothèse émise sur le modèle initial, nous connaissons η^* et M.

La variance conditionnelle de $\tilde{Y}_i$ est définie par :

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

avec la même fonction de variance v que la variable aléatoire Y_i.

Avant de calculer l'espérance et la matrice de variance marginales de $\tilde{Y}$, remarquons qu'une réalisation de l'espérance conditionnelle : $\tilde{\mu}_{u,i}$ peut être vue comme un développement limité au premier ordre de $h(\tilde{\eta}_{u,i})$ en η_i^* (pour u proche de 0). Le prédicteur linéaire

conditionnel $\tilde{\eta}_U$ sur une nouvelle échelle est défini par : $\tilde{\eta}_U = \eta^* + MZU$ où l'on rajoute à η^* une partie aléatoire à laquelle on applique l'effet multiplicatif introduit lors de la marginalisation.
Alors que Foulley et Manfredi [6] interprétaient (dans le cas de variances homogènes) cet effet multiplicatif comme un changement d'échelle sur β, on peut ici le lire comme un changement d'échelle dans l'expression des régresseurs.
En effet, on a aussi : $\begin{array}{rcl} \tilde{\eta}_U & = & M\eta_U + C \\ & = & \tilde{X}\beta + \tilde{Z}U + C \quad \text{où} \quad \tilde{X} = MX \\ & & \qquad\qquad\qquad\qquad\quad \tilde{Z} = MZ. \end{array}$

Espérance et variance marginales

Ainsi, l'espérance marginale de $\tilde{Y}_i$ est :

$$E(\tilde{Y}_i) = E(\tilde{\mu}_{U,i}) = h(\eta_i^*) = E(\mu_{U,i}) = E(Y_i) \ .$$

Les variables Y_i et $\tilde{Y}_i$ ont donc même espérance marginale.

Et pour la matrice de variance marginale $\tilde{V}$, on obtient :

$$\begin{array}{lrcl} \forall i \in \{1,...,N\} & \text{var}(\tilde{Y}_i) & = & E(\text{var}(\tilde{Y}_i|U)) + \text{var}(E(\tilde{Y}_i|U)) \\ & & = & E(v(\tilde{\mu}_{U,i})) + \text{var}(\tilde{\mu}_{U,i}) \\ & & = & E(v(\tilde{\mu}_{U,i})) + h'(\eta_i^*)^2 \tilde{z_i}' G \tilde{z_i} \end{array}$$

Et,

$$\begin{array}{lrcl} \forall i,j \in \{1,...,N\} & \text{cov}(\tilde{Y}_i, \tilde{Y}_j) & = & E(\text{cov}(\tilde{Y}_i, \tilde{Y}_j|U)) + \text{cov}(E(\tilde{Y}_i|U), E(\tilde{Y}_j|U)) \\ & & = & 0 + \text{cov}(\tilde{\mu}_{U,i}, \tilde{\mu}_{U,j}) \\ & & = & h'(\eta_i^*)h'(\eta_j^*)\tilde{z_i}' G \tilde{z_j} \end{array}$$

D'où $\tilde{V}$ s'exprime comme suit :

$$\begin{array}{rcl} \tilde{V} = \text{var}(\tilde{Y}) & = & E(v(\tilde{\mu}_U)) + K\tilde{Z}G\tilde{Z}'K \\ & = & V_0 + KMZGZ'MK \ , \end{array}$$

où K est toujours la matrice définie par : $K = \text{diag}\{h'(\eta_i^*)\}$.

Vérifions que $\tilde{V}$ est bien la même matrice que l'approximation obtenue dans les sections précédentes. Dans l'expression de $\tilde{V}$ la deuxième partie de la somme correspond bien à ce qui avait été obenu. Qu'en est-il de la première partie : la matrice V_0 ? Reprenons les trois cas :

• Cas 1 : Pour un GLM avec loi binomiale, la fonction de variance associée est : $v(\mu_i) = \dfrac{\mu_i(1-\mu_i)}{n_i}$.

D'où : $V_{0_{ii}} = E(\dfrac{\tilde{\mu}_{U,i}(1-\tilde{\mu}_{U,i})}{n_i})$

$= E(\dfrac{1}{n_i}[h(\eta_i^*) + h'(\eta_i^*)\tilde{z_i}'U][1 - h(\eta_i^*) - h'(\eta_i^*)\tilde{z_i}'U])$

$= \dfrac{1}{n_i}[h(\eta_i^*)(1-h(\eta_i^*)) + h'(\eta_i^*)^2\tilde{z_i}'G\tilde{z_i}].$

Notons que cette expression établie pour des données binomiales reste vraie quelle que soit la fonction de lien. On en verra l'utilisation dans la section suivante.
Or, pour le lien probit, $h = \Phi$ et $h' = \varphi$, on obtient donc :

$$V_0 = \text{diag}\left\{\frac{\Phi(\eta_i^*)(1-\Phi(\eta_i^*)) + \varphi(\eta_i^*)^2(MZ)_i'G(MZ)_i}{n_i}\right\}_{i=1,\ldots,N}.$$

Ce qui, étant donné que l'on peut écrire $\Pi_i = \Phi(\eta_i^*)$ et $t_{ii} = (MZ)_i'G(MZ)_i$ (cf paragraphe 2.2), nous redonne bien la même expression de $\tilde{V}$.

• Cas 2 : Pour un GLM avec loi de Poisson, la fonction de variance associée est : $v(\mu_i) = \mu_i$.
D'où : $V_{0_{ii}} = E(\tilde{\mu}_{U,i})$

$= h(\eta_i^*) = \mu_i$.

Donc on vérifie encore :

$$V_0 = \text{diag}\{\mu_i\}_{i=1,\ldots,N}.$$

• Cas 3 : Pour un GLM avec loi exponentielle, la fonction de variance associée est : $v(\mu_i) = \mu_i^2$.
D'où : $V_{0_{ii}} = E(\tilde{\mu}_{U,i}^2)$

$= E([h(\eta_i^*) + h'(\eta_i^*)\tilde{z_i}'U]^2)$

$= h(\eta_i^*)^2 + h'(\eta_i^*)^2\tilde{z_i}'G\tilde{z_i}$

$= \mu_i^2 + \mu_i^2\tilde{z_i}'G\tilde{z_i}.$

Or, dans ce cas $M = I_N$ et on vérifie une fois encore :

$$V_0 = \text{diag}\left\{\mu_i^2(1 + z_i'Gz_i)\right\}_{i=1,\ldots,N}.$$

Dans les trois cas, la matrice de variance marginale $\tilde{V}$ est donc bien la même que l'approximation de V calculée pour l'utilisation des équations de Henderson.

5.2.4 Conclusion

L'hypothèse forte de conservation de la fonction de lien inverse lors du calcul de l'espérance marginale est vérifiée dans les trois cas étudiés et le sera dans d'autres cas envisagés ultérieurement. Elle semble être la clé principale à cette méthodologie GAR. Sous cette hypothèse, on a donc défini un vecteur aléatoire $\tilde{Y}$ tel que :

- son espérance marginale est égale à celle de Y,
- sa matrice de variance marginale $\tilde{V}$ correspond à la matrice de variance approchée de Y.

La quasi-vraisemblance ne reposant que sur les deux premiers moments marginaux, la maximisation de cette fonction par résolution des équations de Henderson aboutira donc aux mêmes estimations (en prenant y comme une réalisation de $\tilde{Y}$).

Notons que dans la démarche initiale, l'hypothèse de composantes de la variance petites intervenait directement lors de l'approximation de V. Ce n'est plus le cas dans cette nouvelle démarche. Pourtant , le modèle approché ne sera justifié que parce qu'il permet d'aboutir notamment à l'expression de la variance marginale qui n'est autre que l'approximation précédente. Ainsi, sans intervenir explicitement, cette hypothèse de σ^2 proche de 0 est sous-jacente.

Ce détour par $\tilde{Y}$ nous a permis d'éviter le calcul exact de la matrice de variance V mais aussi de prendre en compte des approximations de nature différentes réalisées sur la fonction de lien.

5.3 Comparaison avec la démarche de Breslow & Clayton

Dans leur article, Breslow et Clayton [1] envisagent une autre démarche pour définir un modèle marginal et utiliser la quasi-vraisemblance marginale. Zeger, Liang et Albert [17] ont adopté une démarche similaire. Nous y avons déjà fait référence mais revenons-y à titre de comparaison.

En gardant les mêmes notations, le modèle initial (le vrai modèle), défini conditionnellement aux effets aléatoires est un GLM. On peut l'écrire sous la forme :

$$\begin{array}{rcccl} Y_i & = & \mu_{U,i} & + & \varepsilon_i \\ & = & h(x_i'\beta + z_i'U) & + & \varepsilon_i \end{array} \qquad (4)$$

avec $E(Y_i|U) = \mu_{U,i} = h(\eta_{U,i}) = h(x_i'\beta + z_i'U)$
$V(Y_i|U) = V(\varepsilon_i|U) = v(\mu_{U,i})$.
Les auteurs spécifient alors le GLM marginal en termes de l'espérance marginale par :

$$E(Y_i) = \mu_i = h(x_i'\beta)$$

utilisant le prédicteur linéaire marginal $\eta = x_i'\beta$.
L'espérance marginale ainsi définie, à moins d'un lien identité (c'est le cas du L2M), ne coïncide pas avec le vrai calcul de l'espérance marginale.
Cependant, ils remarquent aussi que ce modèle marginal peut être dérivé du modèle initial par une approximation au premier ordre de l'équation (4), lorsque les composantes de dispersion tendent vers 0. En effet, on obtient alors :

$$Y_i \approx h(x_i'\beta) + h'(x_i'\beta) z_i' U + \varepsilon_i$$

Ce qui conduit à :

$$V = V(Y) = V_0 + KZGZ'K$$

avec $V_0 = \text{diag}\{v(\mu_i)\}_{i=1,\ldots,N}$,
$K = \text{diag}\{h'(\eta_i)\}_{i=1,\ldots,N}$.

Cette écriture, au contraire de la méthode GAR, ne tient pas du tout compte des effets multiplicatifs introduits lors de la "marginalisation". D'autre part, cela revient aussi à écrire de façon peu justifiée que :

$$\begin{array}{rcccl} V_{0_{ii}} & = & V(\varepsilon_i) & = & E(V(\varepsilon_i|U)) \\ & = & v(\mu_i) & = & E(v(\mu_{U,i})) \, . \end{array}$$

Ainsi, la méthode GAR semble mieux adaptée puisqu'elle prend en compte la présence initiale des effets aléatoires.

6 GAR - Données binomiales - Lien logit

Utilisant le formalisme de la section (5), nous envisageons maintenant une adaptation de la méthode GAR au cas où les données sont distribuées selon une loi binomiale et où l'on considère un modèle avec lien logit.
La modélisation avec lien logistique adaptée à des données binaires est très répandue et est beaucoup plus utilisée que le lien probit, notamment dans le milieu médical. Remarquons aussi que le lien logistique correspond au lien canonique associé à la loi binomiale.

Afin de poursuivre la démarche précédente, nous allons tout d'abord déterminer la nouvelle échelle marginale (en identifiant l'effet multiplicatif) pour pouvoir ensuite approcher la matrice de variance-covariance V et procéder à l'estimation.

6.1 Modèle et notations

Comme dans le cadre du paragraphe 1, la distribution des composantes Y_i est, conditionnellement aux effets aléatoires, la loi binomiale. On a :

$$\forall i \in \{1, ..., N\}\ Y_i \mid U = u \sim \mathcal{B}in(n_i, p_{u,i}) \ ,$$

et on s'intéresse toujours aux fréquences : $F_i = \dfrac{Y_i}{n_i}$.

Mais ici, le lien entre $p_{U,i}$ et le prédicteur linéaire $\eta_{U,i} = x_i'\beta + z_i'U$ se fait par l'intermédiaire de la fonction logistique :

$$\ln(\frac{p_{U,i}}{1-p_{U,i}}) = \eta_{U,i} \Longleftrightarrow p_{U,i} = \frac{\exp(\eta_{U,i})}{1+\exp(\eta_{U,i})} \ .$$

et l'on note $g(x) = \ln(\dfrac{x}{1-x})$ et $h(x) = \dfrac{\exp(x)}{1+\exp(x)}$ $\qquad (h = g^{-1})$.

On suppose toujours une distribution gaussienne pour les effets aléatoires : $U \sim \mathcal{N}(0, G)$.

6.2 Calcul de η^*

On s'intéresse tout d'abord au calcul du prédicteur linéaire marginalisé η^* et à l'identification de M. Ainsi, essayons de réaliser le calcul exact de l'espérance marginale.

$$\begin{aligned} E(F_i) \quad = \quad E(E(F_i|U)) \quad &= \quad E(p_{U,i}) \\ &= \quad E(\frac{\exp(\eta_{U,i})}{1+\exp(\eta_{U,i})}) \ . \end{aligned}$$

Malheureusement, ce calcul exact est difficilement réalisable. Nous utilisons alors l'approximation usuelle de la fonction logistique (Zeger, Liang & Albert [17]) :

$$\frac{\exp(x)}{1+\exp(x)} \approx \Phi(cx) \quad \text{où} \quad c = \frac{16\sqrt{3}}{15\pi} \ .$$

Ce passage par la fonction Φ est réalisé successivement dans un sens puis dans l'autre. C'est un artifice de calcul permettant d'utiliser les propriétés de cette fonction pour le calcul de l'espérance marginale. Cela signifie que momentanément on se place dans un modèle avec lien probit.

$$\begin{array}{rrcl} \text{Ainsi} & E(F_i) & \approx & E(\Phi(c\eta_{U,i})) \\ & & \approx & E(\Phi(cx_i'\beta + cz_i'U)) \\ & & \approx & \Phi(\dfrac{cx_i'\beta}{\sqrt{1+c^2z_i'Gz_i}}) \ . \end{array}$$

Utilisant alors l'approximation inverse, on a :

$$E(F_i) \approx \frac{\exp(\dfrac{x_i'\beta}{\sqrt{1+c^2z_i'Gz_i}})}{1+\exp(\dfrac{x_i'\beta}{\sqrt{1+c^2z_i'Gz_i}})} = h(\eta_i^*) = \mu_i \ ,$$

$$\begin{array}{rrcl} \text{avec} & \eta_i^* & = & \dfrac{x_i'\beta}{\sqrt{1+c^2z_i'Gz_i}} \ , \\ \text{d'où} & M & = & \text{diag}\left\{\dfrac{1}{\sqrt{1+c^2z_i'Gz_i}}\right\}_{i=1,\ldots,N} \ . \end{array}$$

Maintenant que l'on a su identifier η^* et M , on peut définir :

$$\begin{array}{rcl} \tilde{\eta}_U & = & \eta^* + MZU \\ & = & \eta^* + \tilde{Z}U \ , \end{array}$$

ainsi que

$$\begin{array}{rcl} K & = & \text{diag}\{h'(\eta_i^*)\}_{i=1,\ldots,N} \\ & = & \text{diag}\left\{\dfrac{\exp(\eta_i^*)}{(1+\exp(\eta_i^*))^2}\right\}_{i=1,\ldots,N} \ ; \\ & = & \text{diag}\{\mu_i(1-\mu_i)\}_{i=1,\ldots,N} \ . \end{array}$$

6.3 Calcul de $\tilde{V}$

Pour calculer $\tilde{V}$, les matrices K et M étant connues, il s'agit maintenant de déterminer V_0. D'après l'expression obtenue précédemment dans le cas binomial avec une fonction de lien quelconque ; on a ici, avec $h(x) = \dfrac{\exp(x)}{1+\exp(x)}$:

$$\begin{array}{rcl} V_{0_{ii}} & = & \dfrac{1}{n_i}[h(\eta_i^*)(1-h(\eta_i^*)) + h'(\eta_i^*)^2\tilde{z_i}'G\tilde{z_i}] \\ & = & \dfrac{\exp(\eta_i^*)}{n_i(1+\exp(\eta_i^*))^2}[1 + \dfrac{\exp(\eta_i^*)}{(1+\exp(\eta_i^*))^2}z_i'MGMz_i] \ . \end{array}$$

La matrice $\tilde{V}$ s'écrit alors:

$$\begin{aligned}\tilde{V} &= V_0 + KMZGZ'MK \\ &= V_0 + LZGZ'L \qquad \text{avec} \quad L = KM = \text{diag}\left\{\frac{h'(\eta_i^*)}{\sqrt{1+c^2 z_i' G z_i}}\right\} .\end{aligned}$$

6.4 Estimation.

De manière identique aux paragraphes précédents, on procède à l'estimation par maximisation de la fonction de quasi-vraisemblance. Pour cela, G étant inconnue, on utilise le modèle approché et la matrice de variance $\tilde{V}$, et l'on résout le système itératif suivant:

$$(X'\tilde{W}^{[s]-1}X)\beta^{[s+1]} = X'\tilde{W}^{[s]-1}\zeta^{[s]}$$

$$\begin{array}{lrcl} \text{avec} & \zeta^{[s]} & = & X\beta^{[s]} + L^{[s]-1}(f-\mu^{[s]}) \\ \text{et} & \tilde{W}^{[s]} & = & L^{[s]-1}\tilde{V}^{[s]}L^{[s]-1} \\ & & = & L^{[s]-1}V_0^{[s]}L^{[s]-1} + ZGZ' \\ & & = & R^{[s]} + ZGZ' , \end{array}$$

où R a ici la forme: $$\begin{aligned} R &= \text{diag}\left\{\frac{1}{n_i}[\frac{(1+exp(\eta_i^*))^2(1+c^2 z_i' G z_i)}{exp(\eta_i^*)} + z_i' G z_i]\right\}_{i=1,\ldots,N} \\ &= \text{diag}\left\{\frac{1}{n_i}[\frac{1+c^2 z_i' G z_i}{\mu_i(1-\mu_i)} + \sigma_i^2]\right\}_{i=1,\ldots,N} . \end{aligned}$$

On obtient une solution à ces équations par résolution des équations de Henderson. Pour cela, on utilise le sous-produit obtenu comme prédiction de U ainsi que pour l'estimation des composantes de la variance.

7 Conclusion

Comme nous venons de le voir, la méthode GAR, développée à l'origine dans le cadre bien précis d'un modèle probit pour données binomiales, peut s'étendre à d'autres types de modélisations. Foulley et Im ont les premiers envisagé cette extension pour des données poissoniennes. Le formalisme que nous proposons ici permet d'aborder cette démarche de façon plus générale. L'utilisation de la quasi-vraisemblance marginale à partir des deux premiers moments marginaux devrait pouvoir s'appliquer à de nombreux cas.

Cependant notre présentation se limite à l'hypothèse forte de conservation du lien inverse. Notons pourtant que dans tous les modèles avec lien identité ($h = Id$), cette hypothèse est vérifiée puisqu'alors $\mu_{U,i} = \eta_{U,i} = x_i'\beta + z_i'U$ et $\mu_i = \eta_i^* = x_i'\beta$.

Prenons pour exemple le cas très simple de données poissonniennes où l'espérance conditionnelle s'écrit : $\mu_{U,i} = x_i'\beta + z_i'U$ et la fonction de variance : $v(x) = x$. L'échelle marginale alors est la même que celle d'origine: $M = I_N$, $\tilde{\eta}_{U,i} = \eta_{U,i}$ et $K = I_N$. Aussi, $\tilde{\mu}_{U,i} = \mu_{U,i}$ donc $\tilde{V} = V$, la matrice de variance s'écrit directement sous la forme $V = V_0 + KMZGZ'MK$ avec $V_0 = \text{diag}\{x_i'\beta\}_{i=1,\ldots,N}$. Dans ce cas très simple, nous vérifions encore la validité du formalisme.

D'autre part, certaines applications utilisent la modélisation exponentielle avec lien inverse (lien canonique : $h(x) = \frac{1}{x}$). Outre le fait que ce lien ne permette pas d'assurer la positivité du paramètre de la loi, le calcul de $\eta_i^* = E(\mu_{U,i}) = E(\frac{1}{\eta_{U,i}}) = E(\frac{1}{x_i'\beta + z_i'U})$ pour $U \sim \mathcal{N}(0, G)$ lorsqu'il converge s'avère délicat.
Pourtant, il est possible d'envisager un développement limité de l'espérance conditionnelle (pour u proche de 0) :

$$\frac{1}{x_i'\beta + z_i'U} = \frac{1}{x_i'\beta}[1 - \frac{z_i'U}{x_i'\beta} + (\frac{z_i'U}{x_i'\beta})^2 + o(u^2)] \; .$$

Selon que l'on se restreint à un développement limité au premier ordre ou au second ordre, on aura : $\mu_i = h(\eta_i^*)$ avec $\eta_i^* = x_i'\beta$ (d'où $M = I_N$) dans un cas ; et $\eta_i^* = x_i'\beta/(1 + \frac{z_i'Gz_i}{(x_i'\beta)^2})$ (d'où $M = \text{diag}\left\{1/(1 + \frac{z_i'Gz_i}{(x_i'\beta)^2})\right\}_{i=1,\ldots,N}$) dans l'autre. Il est alors possible de poursuivre avec $\tilde{\mu}_{U,i} = \frac{1}{\tilde{x}_i'\beta} - \frac{\tilde{z}_i'U}{(\tilde{x}_i'\beta)^2}$ où $\tilde{X} = MX$ et $\tilde{Z} = MZ$. On aboutit à $\tilde{V} = V_0 + KMZGZ'MK$

$$\begin{array}{lrcl} \text{où} & K & = & \{-\frac{1}{(\tilde{x}_i'\beta)^2}\}_{i=1,\ldots,N} \\ \text{et} & V_{0_{ii}} & = & E(v(\tilde{\mu}_{U,i})) \\ & & = & E(\tilde{\mu}_{U,i}^2) \\ & & = & E((\frac{1}{\tilde{x}_i'\beta} - \frac{\tilde{z}_i'U}{(\tilde{x}_i'\beta)^2})^2) \\ & & = & \frac{1}{(\tilde{x}_i'\beta)^2}(1 + \frac{\tilde{z}_i'G\tilde{z}_i}{(\tilde{x}_i'\beta)^2}) \end{array}$$

Ainsi, même s'il est en apparence limité, ce formalisme permet d'étendre la méthode GAR à de nombreuses autres situations.

De façon générale, GAR est donc une méthode d'estimation des effets fixes et composantes de la variance qui repose sur un raisonnement marginal. D'autres méthodes ont été développées (cf [15]) avec des raisonnements conditionnels et pour lesquelles la marginalisation n'est intervenue que partiellement ou pas du tout. Notre objectif ici n'est pas de

comparer ces méthodes entre elles (ce qui sera illustré dans un travail ultérieur) mais plutot de présenter le formalisme propre à GAR.
Notons tout de même que l'utilisation de cette méthode est réservée au domaine de validité de l'approximation de la matrice de variance V par $\tilde{V}$; c'est-à-dire pour des petites valeurs des composantes de la variance. Nous pouvons illustrer ceci par les résultats de simulations suivants.
Nous avons simulé un vecteur de données binomiales de taille 20 issu d'un modèle avec lien logit. Un seul effet aléatoire avec 4 réalisations a été introduit. Après avoir imposé un effet fixe nul, nous avons observé, pour chaque valeur de σ^2 considérée, 200 simulations et résultats d'estimation que l'on peut résumer ainsi :

σ^2 simulé	σ^2=0.01	σ^2=0.1	σ^2=0.5	σ^2=1	σ^2=2	$\sigma^2 = 4$
σ^2 estimé	0.013	0.115	0.536	1.336	11.378	384.2
écart type	0.015	0.112	0.543	1.459	114.878	3973.163
β estimé	-0.005	-0.010	0.014	-0.011	0.059	0.930
écart type	0.065	0.151	0.304	0.501	1.278	7.838

Ainsi à partir de $\sigma^2 = 1$ et surtout pour $\sigma^2 = 2$ et $\sigma^2 = 4$, les estimations obtenues n'ont aucune signification.

Enfin, nous avons eu l'occasion dans ce document d'émettre des doutes quant à la prévision de U obtenue. Des simulations effectuées ne sont pas venues diminuer ces doutes.

ANNEXE A

Propriété : Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $E(\Phi(X)) = \Phi(\frac{\mu}{\sqrt{1+\sigma^2}})$,
avec Φ : fonction de répartition de la loi normale centrée réduite.

Démonstration :

$$E(\Phi(X)) = \int_{-\infty}^{+\infty} \Phi(x) f_X(x) dx$$

avec f_X : fonction de densité de la loi normale $\mathcal{N}(\mu, \sigma^2)$.

Soit U variable aléatoire de loi $\mathcal{N}(0,1)$ indépendante de X, alors :

$$\begin{aligned} E(\Phi(X)) &= \int_{-\infty}^{+\infty} P(U \leq x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} (\int_{-\infty}^{x} f_U(u) du) f_X(x) dx \\ &= \iint_{u \leq x} f_U(u) f_X(x) du dx \\ &= P(V \in D) \end{aligned}$$

où : $V = (U, X)$ couple de composantes indépendantes
D est le domaine défini par $D = \{(u, x) \in \mathbb{R}^2 / u \leq x\}$.

$$\begin{aligned} E(\Phi(X)) &= P(U \leq X) \\ &= P(U - X \leq 0) \qquad \text{avec} \quad U - X \sim \mathcal{N}(-\mu, 1+\sigma^2) \\ &= P(\frac{U - X + \mu}{\sqrt{1+\sigma^2}} \leq \frac{\mu}{\sqrt{1+\sigma^2}}) \\ &= \Phi(\frac{\mu}{\sqrt{1+\sigma^2}}) \end{aligned}$$

CQFD.

ANNEXE B : La quasi-vraisemblance dans les GLM

Lorsqu'on ne dispose pas d'une information suffisante pour construire une fonction de vraisemblance, il est possible de faire de l'inférence à partir d'expériences par construction d'une fonction de quasi-vraisemblance (cf [13]).
Pour cela, on suppose que l'on dispose de N observations indépendantes de variables aléatoires, composantes du vecteur Y. La seule information disponible concerne les moments d'ordre 1 et 2 de Y. Son espérance est notée μ (dépendant du paramètre β), et sa matrice de variance $V = \text{diag}\{V_i(\mu_i)\}$.
Avec ces hypothèses, la log-quasi-vraisemblance est alors définie par :

$$Q(\mu; y, V) = \sum_{i=1}^{N} Q_i(\mu_i; y_i, V_i) \qquad \text{où} \qquad Q_i(\mu_i; y_i) = \int_{y_i}^{\mu_i} \frac{y_i - t}{V_i(t)} dt \ .$$

On utilise cette fonction pour obtenir les équations d'estimation pour β en annulant la fonction quasi-score associée : $\dfrac{\partial Q(\mu; y, V)}{\partial \beta}$.
Ces équations sont alors :

$$D' V^{-1} (y - \mu) = 0 \qquad \text{où} \qquad D = \frac{\partial \mu}{\partial \beta'} \ .$$

Utilisant la méthode des scores de Fisher, on obtient le système itératif :

$$(D' V^{-1} D) \Delta\beta^{[s]} = D' V^{-1} (y - \mu^{[s]}) \ .$$

Ces équations se généralisent au cas de variables dépendantes où V est la matrice de variance non diagonale.

Dans le cadre des GLM, on décrit de manière plus précise le lien entre μ et β. On note g la fonction de lien du GLM considéré et h sa réciproque, on a $\mu = h(\eta)$. Ainsi, on obtient : $D = KX$ où $K = \text{diag}\{h(x_i'\beta)\}$ (K et donc D dépendront dans le schéma itératif de la valeur courante de β).
Le système précédent peut alors s'écrire :

$$(X' W^{-1} X) \beta^{[s+1]} = X' W^{-1} \zeta^{[s]} \ ,$$

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

On prendra soin de remarquer que ces équations correspondent aux équations classiques d'estimation dans le modèle linéaire $\zeta = X\beta + \varepsilon$ avec $\varepsilon \sim \mathcal{N}(0, W)$. Dans ce modèle, les données sont réactualisées à chaque nouvelle valeur de β.

ANNEXE C: Les équations de Henderson

Les équations de Henderson (ou équations du modèle mixte: MME) permettent d'obtenir une estimation des effets fixes ainsi qu'une prédiction des effets aléatoires dans un modèle linéaire mixte (L2M).
Considérons le modèle suivant:

$$Y = X\beta + ZU + \varepsilon$$

où $U \sim \mathcal{N}_q(0, G)$
$\varepsilon \sim \mathcal{N}_N(0, R)$.

Notons qu'alors $V(Y) = R + ZGZ' = V$.

Les équations de Henderson s'obtiennent par maximisation de la densité jointe des vecteurs aléatoires Y et U.
On a $f_{Y,U}(y,u) = f_{Y|U}(y) * f_U(u)$

$$= \frac{1}{(2\Pi)^{\frac{N}{2}}|R|^{\frac{1}{2}}} exp(-\tfrac{1}{2}[(y - X\beta - Zu)'R^{-1}(y - X\beta - Zu)])$$
$$* \frac{1}{(2\Pi)^{\frac{q}{2}}|G|^{\frac{1}{2}}} exp(-\tfrac{1}{2}[u'G^{-1}u])$$

En annulant les dérivées, par rapport à β et u, du logarithme de cette fonction, on obtient le système d'équations:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}$$

Quelques remarques:

• La solution $\hat{\beta}$ à ces équations est aussi la solution des équations des moindres carrés généralisés: $(X'V^{-1}X)\beta = X'V^{-1}y$
dans le modèle $Y = X\beta + \varepsilon'$ avec $\varepsilon' \sim \mathcal{N}(0, V)$. Quand V est connue, l'estimateur ainsi défini correspond au meilleur (i.e. variance minimum) estimateur linéaire sans biais de β (cf [14] annexe S-2).

• La solution $\hat{u}$ à ces équations peut aussi s'écrire:

$$\hat{u} = GZ'V^{-1}(y - X\hat{\beta})$$

correspondant au meilleur prédicteur linéaire sans biais de u (cf [14] p277).

• Si l'on traite u comme un effet fixe et que l'on supprime alors G^{-1} de ces équations, ce système correspond alors aux équations du maximum de vraisemblance pour l'estimation de β et u.
En effet, ce que l'on avait noté densité conditionnelle correspond dans ce cas à la densité de Y.

D'autre part, ces équations permettent aussi d'obtenir des estimations du maximum de vraisemblance ou du maximum de vraisemblance restreint des composantes de la variance. On suppose G de la forme : $G = \{{}_d\alpha_j^2 A_j\}$. Alors, à l'aide des prédictions $\hat{u}$ obtenues, les estimations du maximum de vraisemblance restreint des composantes de la variance α_j^2 peuvent s'óbtenir itérativement par :

$$\hat{\alpha}_j^{2\,[s+1]} = \frac{\hat{u}_j' A_j^{-1} \hat{u}_j}{q_j - \dfrac{tr(A_j^{-1} C_{jj})}{\hat{\alpha}_j^{2\,[s]}}}$$

où C_{jj} est le bloc correspondant au $j^{\text{ème}}$ effet aléatoire dans l'inverse de la matrice du système des équations de Henderson.

Références

[1] Breslow (N.E.) et Clayton (D.G.). – Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, vol. 88, n° 421, 1993, pp. 9–25.

[2] Ducrocq (V.). – Estimation of genetic parameters arising in non linear models. *In : 4th World Congr. Genet. Appl. Livestock Prod.* pp. 419–28. – W.G. Hill and R. Thompson and J.A. Wooliams, Edinburgh.

[3] Foulley (J.L.), Gianola (D.) et Im (S.). – Genetic evaluation of traits distributed as poisson-binomial with reference to reproductive characters. *Theor. Applied Genet.*, vol. 73, 1987, pp. 870–7.

[4] Foulley (J.L.), Gianola (D.) et Im (S.). – Genetic Evaluation for Discrete Polygenic Traits in Animal Breeding. *In : Advances in Statistical Methods for Genetic Improvement of Livestock*, éd. par Gianola (D.) et K. Hammond, Springer-Verlag (Heidelberg), pp. 361–409.

[5] Foulley (J.L.) et Im (S.). – A marginal quasi-likelihood approach to the analysis of poisson variables with generalized linear mixed models. *Genetics, Selection, Evolution*, vol. 25, 1993, pp. 101–7.

[6] Foulley (J.L.) et Manfredi (E.). – Approches statistiques de l'évaluation génétique des reproducteurs pour des cractères binaires à seuils. *Genetics, Selection, Evolution*, vol. 23, 1991, pp. 309–38.

[7] Gaudoin (O.), Lavergne (C.) et Soler (J.L.). – A generalized geometric de-eutrophication software reliability model. *IEEE Trans. on Reliability*, vol. 43(4), 1994, pp. 536–41.

[8] Gianola (D.). – Genetic evaluation of animals for traits with categorical responses. *J. Anim. Sci.*, vol. 51, 1980b, pp. 1272–6.

[9] Gianola (D.) et Foulley (J.L.). – Sire evaluation for ordered categorical data with a threshold model. *Genetics, Selection, Evolution*, vol. 15, 1983, pp. 201–24.

[10] Gilmour (A.R.), Anderson (R.D.) et Rae (A.L.). – The analysis of binomial data by a generalized linear mixed model. *Biometrika*, vol. 72, 1985, pp. 593–9.

[11] Gilmour (A.R.), Anderson (R.D.) et Rae (A.L.). – Variance components on an underlying scale for ordered multiple threshold categorical data using a generalized linear mixed model. *J. Anim. Breedg. Genet.*, vol. 104, 1987, pp. 149–55.

[12] Harville (D.A.) et Mee (R.W.). – A mixed model procedure for analyzing ordered categorical data. *Biometrics*, vol. 40, 1984, pp. 393–408.

[13] McCullagh (P.) et Nelder (J.). – *Generalized Linear Models.* – Chapman and Hall, London, 1989, seconde édition.

[14] S. R. Searle (G. Casella) et McCulloch (C. E.). – *Variance Components.* – John Wiley & Sons, inc, 1992.

[15] Schall (R.). – Estimation in generalized linear models with random effects. *Biometrika*, vol. 78, 1991, pp. 719–27.

[16] Stiratelli (R.), Laird (N.) et Ware (J.H.). – Random effects models for serial observations with binary response. *Biometrics*, vol. 40, 1984, pp. 961–71.

[17] Zeger (S.L.), Liang (K.Y.) et Albert (P.S.). – Models for longitudinal data : a generalized estimating equation approach. *Biometrics*, vol. 44, 1988, pp. 1049–60.

Table des matières

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