Année 1998

THÈSE

pour obtenir le grade de Docteur de

l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

discipline : Mathématiques Appliquées

présentée et soutenue publiquement

par

Catherine TROTTIER

le 2 juillet 1998

ESTIMATION DANS LES MODÈLES LINÉAIRES GÉNÉRALISÉS À EFFETS ALÉATOIRES

COMPOSITION DU JURY

Jean-Louis	SOLER	Président
Michel	BONNEU	Rapporteur
Jean-Louis	FOULLEY	Rapporteur
Anestis	ANTONIADIS	Examinateur
Christian	LAVERGNE	Directeur de thèse

Thèse préparée au sein du Laboratoire de Modélisation et de Calcul et de l'INRIA Rhône-Alpes

Remerciements

Cette thèse est le fruit de quelques années de travail effectuées pour une moitié au sein de l'équipe SMS du Laboratoire de Modélisation et de Calcul, et pour l'autre au sein du projet IS2 de l'INRIA Rhône-Alpes. Même si elle porte le nom de son auteur, elle est aussi le résultat de la combinaison presque magique de diverses contributions. Et je voudrais ici remercier les nombreuses personnes qui, à des titres divers, ont participé à son élaboration.

En tout premier lieu c'est à Christian que je tiens à adresser un grand merci. Il a suggéré ce travail, l'a à la fois dirigé et accompagné me laissant une certaine liberté d'action. J'ai d'ailleurs beaucoup apprécié la confiance qu'il m'a faite. Je le remercie pour sa disponibilité, son dynamisme et le contact qui s'est établi progressivement entre nous. Les échanges que nous avons eus tant sur le plan de la thèse que plus généralement m'ont beaucoup appris et ouverte à ce milieu que je découvrais. Il a constamment été présent et je fais un clin d'œil particulier à son optimisme débordant qui a tenté à certains moments précis de contrebalancer mes doutes et baisses de moral. Entre un souci d'efficacité et un souci de perfection, finalement le travail avance.

J'adresse aussi de vifs remerciements aux autres membres du jury :

à Jean-Louis Soler d'une part, pour m'avoir accueillie au sein de l'équipe SMS mais aussi pour avoir accepté de présider ce jury. Je n'oublie pas son investissement dans la formation que j'ai reçue auparavant.

à Michel Bonneu et Jean-Louis Foulley d'autre part, pour l'accueil qu'ils ont réservé à mon travail en acceptant d'y consacrer de leur temps précieux pour une lecture attentive et l'écriture d'un rapport. Je les remercie pour les questions qu'ils ont posées, les critiques qu'ils ont formulées et les échanges que nous avons eus à ces sujets, tout cela dans un esprit constructif.

à Anestis Antoniadis enfin pour avoir examiné ce travail.

À ces personnes, je voudrais aussi associer Gilles Celeux. Il m'a accueillie au sein du projet IS2 et m'a donc donné l'opportunité de goûter et de participer à une ambiance différente de celle d'un laboratoire universitaire. Je le remercie de s'être intéressé de près à l'évolution de ce travail, d'être resté ouvert et disponible à tout type de questions et discussions et de m'avoir soutenu aux moments opportuns.

Durant ces années de thèse, j'ai aussi découvert la joie d'enseigner. Cette activité a été source d'une grande motivation qui a sans aucun doute eu des retombées sur mon travail de thèse. Je remercie donc les responsables du CIES pour la formation qu'ils nous ont apportée ainsi que Bernard Van Cutsem pour avoir été mon tuteur.

Mais puisqu'une thèse est une aventure bien particulière, il y a encore toute une série de mercis que je souhaite adresser :

aux membres du LMC et de l'INRIA, personnels administratifs et techniques, et plus particulièrement :

à mes collègues de bureau : Mhamed-Ali et Olivier (à la tour), Véronique et Henri (à l'INRIA). Les uns ont accompagné mes premiers pas, les autres ont soutenu les derniers. Ils ont tous suivi jour après jour, heure après heure même, mes péripéties de thèse et autres. Ils ont su, chacun selon leur caractère, être à l'écoute, soutenant et relançant la machine quand il le fallait. Qu'ils trouvent ici une juste récompense de cette présence quotidienne.

à Maryse, Marie-Christine, Florence et aux Olivier qui ont facilité et enrichi la préparation des enseignements. J'ai beaucoup apprécié leurs remarques judicieuses et les discussions agréables dont ils m'ont fait bénéficier.

à Claudine toujours discrètement attentive à l'état de mon travail et de ma personne.

aux thésards de ces labos, en particulier ceux des équipes SMS et IS2 mais aussi aux amis des projets BIP, SHERPA, AIRELLE, OPERA, qui ont coloré les heures passées au travail.

aux buveurs de thé de 16h30, aux joueurs de foot et de basket et aux joggers du midi sans oublier la "pique-niqueuse" qui attendait patiemment.

à tous les amis qui ont su être là aux bons moments, une oreille toujours attentive, patients aussi, supportant les baisses de régime et m'encourageant toujours. Leur amitié m'est très précieuse et que ces quelques lignes leur redisent mon sincère et profond attachement.

à mes professeurs de mathématiques qui m'ont donné le goût pour cette discipline.

aux divers chocolatiers ...

Enfin puisque je dois m'arrêter là (même si c'est frustrant), c'est un merci du fond du cœur que je lance à mes frères et sœurs, leurs "jolis associés" et leurs enfants mais aussi à la famille plus largement : ils ont été, sont et resteront les piliers qui proposent un appui sans limite. Et en particulier puisque l'occasion se présente à nouveau, je réserve le plus grand des mercis à mes parents. C'est le plus chaleureusement possible que je les remercie d'être tout simplement ce qu'ils sont !

J'ai cherché bien souvent au cours de discussions à donner une image de ce type de travail. Je n'ai jamais réussi. Une thèse demande un gros investissement sans toutefois que l'on puisse toujours en reconnaître les fruits (au moins immédiats). Je souhaite au passage bon courage à tous les futurs docteurs. Il n'existe pas de recette miracle et ce travail est somme toute soumis au hasard des rencontres, des discussions, des idées lancées et attrapées au vol ... Toute image semblait donc trop restrictive à mon goût. Bref, l'alchimie fait qu'aucune thèse ne ressemble à sa voisine. Quoi qu'il en soit, je considère la mienne moins comme un aboutissement que comme une étape. Et je remercie à nouveau toutes les personnes que j'ai citées pour tout ce qu'elles m'ont appris au cours de ces années. J'espère de tout cœur que cette expérience formatrice à divers niveaux appellera des suites ...

Table des matières

Introduction

Lorsqu'il s'agit de modéliser des phénomènes naturels, l'utilisation de la loi normale s'impose dans de nombreuses situations. Malgré cette prédominance bien justifiée, il est un certain nombre de phénomènes observés difficilement modélisables par cette fameuse loi. C'est le cas, par exemple, de relevés sur des durées de vie de matériels, de l'observation du nombre d'individus dans une population ayant telle ou telle caractéristique, ou encore du décompte d'événements rares.

Ainsi, afin de permettre une analyse satisfaisante de données manifestement non gaussiennes, les modèles linéaires classiques ont été étendus à la classe plus large de modèles que sont les modèles linéaires généralisés.

Sur un autre plan, la modélisation des effets pouvant intervenir dans l'explication du phénomène étudié s'est aussi enrichie. Il a été introduit la notion d'effet aléatoire, en la distinguant de celle d'effet fixe. La partie explicative du modèle s'est ainsi raffinée, combinant linéairement ces deux types d'effets. La modélisation à effets aléatoires est notamment très usitée dans le domaine de la génétique animale.

La combinaison de ces deux types d'extension des modèles linéaires classiques a donné naissance aux modèles linéaires généralisés à effets aléatoires.

Dans ce travail, nous nous intéressons à l'estimation des paramètres de tels modèles, et en particulier à l'estimation des composantes de la variance - paramètres de variance des effets aléatoires.

Pour répondre à la question de l'estimation de paramètres dans des modèles, la théorie statistique nous guide vers la maximisation de la fonction de vraisemblance. Or, dans notre cas précis, la distribution de la variable aléatoire modélisant le phénomène observé est difficile à décrire.

En effet, avec la présence des effets aléatoires, les modèles que nous étudions englobent deux sources d'aléa. De ce fait, les hypothèses sur la distribution de la variable à expliquer ne peuvent être posées correctement que conditionnellement aux effets aléatoires. Il est donc nécessaire, pour atteindre la distribution marginale de cette variable, de lever le conditionnement. Dans les modèles gaussiens à effets aléatoires, les règles de conditionnement de la loi normale rendent cette opération réalisable. Malheureusement ici, ce déconditionnement n'est pas chose aisée, et constitue même la difficulté principale.

Certains choix particuliers de distributions pour la variable, les effets, ou les deux simultanément, permettent cependant de contourner le problème. Mais nous avons voulu, dans notre travail, garder un point de vue global sur cette classe de modèles, ainsi que l'hypothèse classique de distribution normale des effets aléatoires.

L'élaboration de méthodes d'estimation va donc devoir faire face à ce problème du déconditionnement. 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. Notre travail s'inscrit donc dans un objectif d'étude et de mise en place de méthodes d'estimation. Ces méthodes seront illustrées au cours du travail par des résultats de simulation.

Dans un premier chapitre, nous donnons une description précise des modèles linéraires généralisés à effets aléatoires. Les différentes hypothèses concernant la modélisation du phénomène observé sont présentées. Dans ce but, nous revenons à la fois sur la particularité des modèles linéaires généralisés, ainsi que sur la notion d'effet aléatoire. Ceci sera l'occasion de préciser en quoi ce dernier type d'hypothèses diffère, à nos yeux, des modèles linéaires généralisés avec "surdispersion".

Ensuite, nous abordons dans les chapitres 2 et 3, l'étude de l'estimation de paramètres au sein de ces modèles. Comme nous l'avons signalé, selon les démarches, des approximations sont effectuées à divers stades, depuis la définition du modèle conditionnel jusqu'à celle du modèle marginal.

Dans un premier temps, au chapitre 2, c'est un point de vue conditionnel que nous adoptons. En effet, nous commençons par nous intéresser à une linéarisation du modèle, conditionnellement aux effets aléatoires. Cette linéarisation se réalise par l'introduction d'une variable dépendante - technique propre à l'estimation par maximum de vraisemblance dans les modèles linéaires généralisés simples. Le modèle linéarisé obtenu est alors traité comme un modèle linéaire mixte, avec la particularité cependant qu'une partie de la structure de variance est, dans certains cas, connue. Nous désignons d'ailleurs ces modèles particuliers par modèles linéaires mixtes non standards.

Cette méthode proposée permet, le cas échéant, de tirer partie de l'information supplémentaire apportée sur la variance par certaines lois autres que la loi normale.

Elle s'avère proche d'une méthode proposée par SCHALL (1991), et relue par divers auteurs. Cependant, elle incite à relâcher quelque peu le conditionnement. Ce constat nous amènera, dans un deuxième temps dans ce chapitre, à comparer les diverses approches existantes selon une échelle de déconditionnement que nous préciserons.

Dans le chapitre 3, nous prenons le point de vue opposé, selon cette échelle, à celui du chapitre précédent. En effet, c'est dans un raisonnement marginal que s'inscrit ce chapitre. C'est-à-dire que la linéarisation est effectuée au niveau du modèle marginal et non au niveau du modèle conditionnel.

Nous reprenons la démarche suivie par GILMOUR, ANDERSON, et RAE (1985). Nous proposons d'étendre leur méthode, mise en place dans le cadre particulier d'un modèle binomial avec lien probit. Cette méthode s'appuie principalement sur la construction, au niveau marginal, d'une fonction de quasi-vraisemblance.

Elle a déja été étendue au cas d'un modèle de Poisson avec lien logarithmique par FOULLEY et IM (1993). Nous la reprenons ici, en relâchant en particulier une hypothèse de travail concernant l'homogénéité des variables sous-jacentes. Nous proposons surtout un formalisme commun, qui permet d'envisager l'étude d'autres cas de distributions ou de fonctions de lien selon cette démarche.

Le dernier chapitre, même s'il tente toujours de répondre à la question de l'estimation, s'inscrit dans un cadre légèrement différent. En effet, nous enrichissons à nouveau les hypothèses de modélisation, pour introduire une notion d'hétérogénéité dans les modèles linéaires généralisés à effets aléatoires.

L'hétérogénéité recouvre, selon les auteurs, des sens variés. C'est pourquoi nous explicitons tout d'abord le sens que nous lui accordons ici et la modélisation à laquelle nous nous intéressons. Elle rejoint d'ailleurs celle décrite par FOULLEY et QUAAS (1995) dans le cadre des modèles linéaires à effets aléatoires, mais n'est cependant pas classique dans celui des modèles linéaires généralisés à effets aléatoires. Elle concerne une hétérogénéité des variances des effets aléatoires.

Du point de vue de l'estimation, nous commençons par nous intéresser au cas des modèles gaussiens. L'introduction de l'hétérogéité implique que certaines démarches usuelles ne sont plus directement envisageables. L'algorithme EM constitue alors un outil essentiel. Nous étudions sa mise en place dans le cas homogène, et nous verrons que, selon l'écriture du modèle adoptée, il existe deux voies pour cela. Puis, nous envisageons le cas hétérogène, pour lequel nous retrouvons la méthodologie de Foulley et Quaas (1995).

Pour finir, nous en venons à l'objectif principal de ce chapitre, en nous plaçant dans le cadre des modèles linéaires généralisés à effets aléatoires à variances hétérogènes. Une mise en place directe de l'algorithme EM se confronte aux mêmes problèmes que ceux du déconditionnement, énoncés dans les chapitres précédents. Nous proposons alors une méthode d'estimation, alliant les techniques de linéarisation du chapitre 2, et le travail précédent à l'aide de ce même algorithme dans le cas linéaire.

Chapitre 1

Les modèles étudiés

1.1 Introduction

Nous pensons indispensable de débuter la présentation de ce travail par une définition précise des modèles — ou plus exactement de la classe de modèles — que nous étudions. Les **modèles linéaires généralisés à effets aléatoires** — ou **modèles linéaires généralisés mixtes** — constituent en effet un ensemble assez vaste de modèles. Ils répondent à une démarche générale de modélisation qu'il nous semble important de décrire. Mais ils englobent aussi différents points de vue possibles de modélisation. C'est pourquoi nous tenons à énoncer les hypothèses que nous avons adoptées, et ainsi à clarifier les termes que nous emploierons dans toute la suite. Ceci nous permettra en outre de définir les notations utilisées tout au long de ce document, et surtout par la suite de positionner notre travail par rapport aux travaux déjà réalisés dans ce domaine. Nous adoptons la notation **GL2M** (**G**eneralized **L**inear **M**ixed **M**odels) pour désigner ces modèles.

Les GL2M ont vu le jour dans les années 80. Ils sont à la croisée de deux types d'extension des modèles linéaires classiques, notés **LM** (**L**inear **M**odels). La première est une extension en termes de loi et donne naissance à la classe des **modèles linéaires généralisés**, désignés classiquement par **GLM** (**G**eneralized **L**inear **M**odels). La deuxième est une extension en termes d'introduction d'effets aléatoires, qui aboutit quant à elle à la classe des **modèles linéaires mixtes**, notés **L2M** (**L**inear **M**ixed **M**odels).

Ce chapitre est donc consacré, dans un premier temps, à une description de ces deux classes de modèles : GLM et L2M. Nous nous arrêterons sur la spécificité de chacune d'elles. En prolongement des L2M, nous introduirons la notion de **modèles linéaires mixtes non standards**, notés **nsL2M** (**n**on **s**tandard **L**inear **M**ixed **M**odels). Cette nouvelle classe de modèles peut être vue comme plus générale que celle des L2M. Elle est constituée de modèles pour lesquels une partie de la structure de variance est supposée connue. Elle trouvera plus particulièrement sa justification lorsque nous aborderons la question de l'estimation des paramètres à partir du chapitre 2.

Enfin, tout ceci nous conduira naturellement à la définition des GL2M, objet d'étude principal de cette thèse. Nous illustrerons sur deux exemples ce type de modélisation. Et nous verrons pour finir des extensions qui peuvent en être données.

1.2 Retour sur les GLM

Nous revenons dans cette section sur l'origine des GLM et les hypothèses qui les caractérisent.

1.2.1 Leur origine

Les premières heures des modèles linéaires classiques remontent au début du $19^{ème}$ siècle. Les données analysées alors par Legendre et Gauss, issues du domaine de l'astrologie, sont essentiellement des mesures de quantités continues. En modélisant les erreurs de mesure par une loi normale, ils développent la méthode des moindres carrés. Gauss s'apercevra plus tard que c'est moins l'hypothèse de normalité que l'hypothèse de variance constante et de données indépendantes qui justifie cette méthode.

Dans les temps qui suivent et surtout au début du $20^{ème}$ siècle, sous l'impulsion de Fisher notamment, d'autres types de modélisation sont mises en place. En effet, la nature des données a alors changé. L'analyse de données discrètes sous forme de comptage ou de proportions se développe. Les distributions de Poisson ou binomiale s'ajoutent alors à la loi normale et enrichissent le panel des distributions disponibles à la modélisation.

La famille exponentielle permet de regrouper toutes ces lois et donne naissance à une nouvelle classe de modèles. Ce sont les **modèles linéaires généralisés** – GLM. La terminologie est introduite en 1972 par Wedderburn et Nelder. Comme son nom l'indique, cette classe de modèles généralise les modèles linéaires classiques. C'est une généralisation en termes de loi de probabilité d'une part mais aussi en termes de lien à la linéarité. Tout comme la remarque précédente de Gauss, l'hypothèse sur la distribution associée à chaque modélisation s'efface devant une propriété de linéarité commune à tous les modèles et l'importance de la relation espérance-variance sur laquelle nous reviendrons.

Cette classe des modèles linéaires généralisés permet donc l'analyse de données discrètes, mais aussi de données continues comme des durées de vie, pour lesquelles la loi normale n'est pas des plus adaptées. Elle a maintenant pris une place importante dans la modélisation statistique, trouvant son intérêt dans différents domaines d'application. McCullagh et Nelder (1989) en font une présentation complète et plus récemment Fahrmeir et Tutz (1994) élargissent cette présentation générale à l'analyse de données multivariées.

Nous revenons sur les hypothèses posées par les GLM, en soulignant les enrichissements apportés par rapport aux modèles linéaires classiques.

1.2.2 Les hypothèses du GLM

Trois hypothèses permettent de caractériser un GLM : la distribution (pour la variable à expliquer), l'expression de la linéarité (mettant en jeu les variables explicatives) et le lien à cette linéarité (donc lien entre variable à expliquer et variables explicatives).

- **Distribution du vecteur aléatoire observé**

On note y le vecteur de taille N des observations, réalisation du vecteur aléatoire Y (variable à expliquer). On suppose que les composantes Y_i $(i = 1, ..., N)$ de Y sont indépendantes et distribuées selon une loi appartenant à la famille exponentielle au sens de Nelder et Wedderburn (1972). C'est-à-dire que la fonction de densité de la variable aléatoire Y_i (par rapport à une mesure adaptée selon les cas discret ou continu) s'écrit :

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

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 : $a_i(\phi) = \dfrac{\phi}{\omega_i}$ avec ω_i un poids connu associé à l'observation i (poids différent de 1 lorsque les données ont été groupées).

Cette famille de lois regroupe un certain nombre de lois dont les lois classiques : binomiale, Poisson, normale, gamma, ... Dans le tableau 1.1 ci-après, nous décrivons pour chacune de celles-ci, l'expression du paramètre canonique θ_i en fonction des paramètres naturels de la loi, le paramètre ϕ et les fonctions b et a_i associées (la fonction c n'ayant pas un grand intérêt par la suite, nous ne la mentionnons pas). Pour simplifier la lecture du tableau, nous avons omis l'indice i.

	θ	$b(\theta)$	$a(\phi)$
$\frac{\mathcal{B}(n,\pi)}{n}$ [a]	$\theta = \ln(\frac{\pi}{1-\pi})$	$b(\theta) = \ln(1+e^{\theta})$	$\phi = 1, \quad \omega = n\ ; \quad a(\phi) = \frac{1}{n}$
$\mathcal{P}(\lambda)$	$\theta = \ln(\lambda)$	$b(\theta) = e^{\theta}$	$\phi = 1, \quad \omega = 1\ ; \quad a(\phi) = 1$
$\mathcal{E}xp(\lambda)$	$\theta = \frac{1}{\lambda}$	$b(\theta) = \ln(\theta)$	$\phi = -1, \quad \omega = 1\ ; \quad a(\phi) = -1$
$\mathcal{N}(\mu,\sigma^2)$	$\theta = \mu$	$b(\theta) = \frac{\theta^2}{2}$	$\phi = \sigma^2, \quad \omega = 1\ ; \quad a(\phi) = \sigma^2$
$\mathcal{G}(a,\lambda)$ [b]	$\theta = \frac{1}{a\lambda}$	$b(\theta) = \ln(\theta)$	$\phi = -\frac{1}{a}, \quad \omega = 1\ ; \quad a(\phi) = -\frac{1}{a}$

TAB. 1.1 – *Paramètre canonique et fonctions a et b caractérisant les lois usuelles de la famille exponentielle.*

[a] proportions

[b] la loi de densité $f(x) = \frac{1}{\lambda^a\,\Gamma(a)}\, x^{a-1}\, e^{-\frac{x}{\lambda}}\, \mathbf{1}_{\mathbb{R}^+}(x)$

Pour chacune de ces lois, l'espérance et la variance de la variable associée s'expriment à l'aide des fonctions a_i et b. En effet, soit $L(\theta; y) = \ln(f_Y(y,\theta))$ la fonction de log-vraisemblance. À partir des résultats classiques suivants :
$\left\{\begin{array}{lcl} E(\frac{\partial L}{\partial \theta}) & = & 0 \\ E(\frac{\partial^2 L}{\partial \theta^2}) + E((\frac{\partial L}{\partial \theta})^2) & = & 0 \end{array}\right.$, on obtient :

$$\begin{array}{rcl} E(Y_i) & = & b'(\theta_i) \\ \mathrm{var}(Y_i) & = & a_i(\phi) b''(\theta_i)\ . \end{array}$$

Il est donc important de souligner qu'il existe une relation directe entre l'espérance de Y_i, que l'on note μ_i, et sa variance :

$$\mathrm{var}(Y_i) = a_i(\phi) b''(b'^{-1}(\mu_i)) = \frac{\phi}{\omega_i} b''(b'^{-1}(\mu_i))\ .$$

On désignera par la suite $v = b'' \circ b'^{-1}$ cette fonction de variance. Ainsi dans le cas où ϕ est connu, la variance des observations est contrainte à être fonction de l'espérance, ou encore la connaissance de μ_i implique celle de la variance. C'est un point essentiel qui n'apparaissait pas avec la loi normale dans un modèle linéaire classique et une information supplémentaire que l'on tentera d'exploiter dans notre travail. Nous donnons, dans le tableau 1.2, l'expression de l'espérance en fonction du (des) paramètre(s) naturel(s), du paramètre canonique ainsi que la fonction de variance.

	μ		$v(\mu)$
$\dfrac{\mathcal{B}(n,\pi)}{n}$	π	$\dfrac{e^\theta}{1+e^\theta}$	$\mu(1-\mu)$
$\mathcal{P}(\lambda)$	λ	e^θ	μ
$\mathcal{E}xp(\lambda)$	λ	$\dfrac{1}{\theta}$	$-\mu^2$
$\mathcal{N}(\mu,\sigma^2)$	μ	θ	1
$\mathcal{G}(a,\lambda)$	$a\lambda$	$\dfrac{1}{\theta}$	$-\mu^2$

TAB. 1.2 – *Expression de l'espérance et de la variance des lois usuelles de la famille exponentielle.*

Notons enfin qu'étant donnée l'hypothèse d'indépendance des composantes de Y, on obtient la matrice de variance suivante[1]:

$$\mathrm{Var}(Y) = \{_d\ a_i(\phi)v(\mu_i)\ \}_{i=1,\ldots,N}\ .$$

Dans cette famille de loi, nous distinguons donc les lois où le paramètre ϕ est connu (et vaut 1 ou -1) : c'est le cas des lois binomiale, Poisson et exponentielle, de celles où il est inconnu et s'exprime à l'aide d'un paramètre supplémentaire : lois normale et gamma. Pourtant, dans certains cas, la modélisation fait intervenir les lois de la première catégorie avec un paramètre de dispersion ϕ inconnu. C'est une modélisation particulière qui prend en compte une *surdispersion* des données (WILLIAMS 1982), dans le cas précisément où, sur les données observées, la variance ne semble pas varier en fonction de la moyenne selon la relation impliquée par le modèle adopté.
Dans ce travail, nous nous intéressons plus particulièrement au cas où θ est le paramètre d'intérêt, ϕ est connu et où on ne fait pas intervenir de surdispersion.

- **Introduction de la linéarité**

Comme dans les modèles linéaires, les variables explicatives interviennent linéai-

1. On notera dans tout le document $\{_d\ a_i\ \}_{i=1,\ldots,n}$ la matrice diagonale A, de dimension $n \times n$, dont les éléments diagonaux sont définis par : $\forall i \in \{1,\ldots,n\}\ ,\ \ A_{i,i} = a_i$. Par extension, on notera aussi $\{_d\ B_i\ \}_{i=1,\ldots,k}$ la matrice diagonale par blocs (avec k blocs) dont le $i^{\text{ème}}$ bloc est la matrice B_i.

rement dans la modélisation. On définit ainsi le prédicteur linéaire :

$$\eta = X\beta \ ,$$

où β est un vecteur de paramètres inconnus de taille p, et X une matrice $N \times p$ connue, fixée par l'expérience.

• **Fonction de lien**

Le lien entre la $i^{ème}$ composante de ce prédicteur linéaire et l'espérance de Y_i s'établit par l'intermédiaire de la fonction g (monotone et deux fois dérivable) appelée *fonction de lien* :

$$\eta_i = g(\mu_i) \ .$$

Parmi toutes les fonctions de lien, celle qui permet d'égaler le prédicteur linéaire et le paramètre canonique est appelée *fonction de lien canonique.* Puisque $\eta_i = g(b'(\theta_i))$, la fonction de lien canonique associée à une distribution donnée sera $g = b'^{-1}$.
Dans le tableau 1.3, nous avons indiqué les fonctions de lien canoniques associées aux lois classiques.

$\dfrac{\mathcal{B}(n,\pi)}{n}$	$\mathcal{P}(\lambda)$	$\mathcal{E}xp(\lambda)$	$\mathcal{N}(\mu,\sigma^2)$	$\mathcal{G}(a,\lambda)$
$g(x) = \ln(\dfrac{x}{1-x})$	$g(x) = \ln(x)$	$g(x) = \dfrac{1}{x}$	$g(x) = x$	$g(x) = \dfrac{1}{x}$

TAB. 1.3 – *Fonctions de liens canoniques des lois usuelles de la famille exponentielle.*

Dans le cadre des modèles linéaires cette fonction de lien n'était pas apparue puisque la fonction de lien canonique associée à la loi normale est l'identité.

Ainsi les **modèles linéaires généralisés** sont rapidement décrits par ces deux fonctions :
- celle spécifiant l'introduction de la linéarité : fonction de lien,
- celle spécifiant la relation espérance-variance : fonction de variance.

Nous abondonnons quelques temps cette catégorie de modèles, pour nous intéresser à un autre type de généralisation des modèles linéaires : les L2M.

1.3 Définition des L2M et L2M non standards

Nous nous intéressons dans un premier temps à la notion d'effets aléatoires comme outil de modélisation. Puis nous verrons comment leur introduction dans les modèles linéaires classiques donne naissance aux L2M. Nous prolongerons enfin cette définition des L2M à celle de ce que nous appelons L2M non standard.

1.3.1 La modélisation avec effets aléatoires

Dans tout relevé d'expérience, les données présentent une certaine *variabilité*. L'intérêt d'une étude statistique réside justement dans l'analyse de celle-ci. On aimerait pouvoir en déterminer la nature, l'importance, les sources, les facteurs, ... C'est Fisher qui, au début de ce siècle, par sa méthodologie de l'analyse de la variance, a réalisé des avancées dans ce domaine. Cette méthodologie tente en effet de cloisonner les différentes sources de variation, et de répondre notamment à des questions sur la significativité de différences observées entre moyennes de sous-groupes de données.

Les **modèles à effets aléatoires** constituent un moyen plus élaboré d'étudier cette variabilité. Par l'introduction de ces effets dans la modélisation, on arrive à préciser les diverses sources de variation. Mais qu'est-ce qu'un effet aléatoire? Répondre à cette question nous oblige à opposer les deux natures possibles des effets : effet *fixe* / effet *aléatoire*. Au cours d'une expérience, différents facteurs sont soupçonnés affecter les résultats de l'expérience, donc les valeurs de la variable observée. Les données relevées peuvent alors être classées selon les différents niveaux de ces facteurs. On distingue deux types de facteurs :

- les facteurs à *effets fixes* avec un nombre fini de niveaux. Les données se répartissent sur ces différents niveaux. Et on aimerait justement en retirer une information sur l'effet de chaque niveau sur la variable d'intérêt.

- les facteurs à *effets aléatoires* avec un nombre infini de niveaux. Bien entendu les données observées (en nombre fini) ne peuvent pas se répartir sur tous les niveaux. Un échantillon de ces niveaux seulement est représenté. Dans ce cas, la façon dont chacun des niveaux influe sur le résultat ne présente pas d'intérêt. Mais de l'échantillon aléatoire des niveaux, on aimerait connaître la part de variabilité induite par cet effet.

Ainsi l'introduction d'effets aléatoires permet de séparer la *variation totale* en deux parties : la variation dûe aux effets aléatoires et celle que l'on affecte aux erreurs. On est donc plus précis quant à son origine puisqu'on a introduit différentes *composantes* à la variance.

Tentons d'illustrer ceci sur un exemple purement fictif. Imaginons que l'on s'intéresse à l'effet de 3 types de médicaments sur des maux de tête sévères. On dispose pour cela d'un échantillon de 12 personnes souffrant régulièrement de ces maux de

tête. Et on donne à chacune un type de médicament de façon à ce que chaque type soit administré à 4 personnes différentes. Pour chaque personne, on relève, après chacune de 4 prises du médicament (en 4 occasions différentes), le temps de disparition des maux de tête. On dispose donc d'un ensemble de 48 données concernant les temps de disparition des maux de tête après prise de médicament. Toutes ces personnes souffrent du même mal mais chacune peut avoir une influence sur le résultat. On a donc mentionné deux facteurs pouvant avoir effet : le médicament administré et la personne concernée. Rappelons que ce qui nous intéresse ce sont les effets des médicaments. Ainsi, chaque niveau du facteur *médicament* apparaît important et l'on aimerait en mesurer l'effet sur le soulagement des maux du malade. Ce facteur est donc considéré comme facteur à effets fixes. Au contraire, l'effet de chacune des personnes sur le résultat ne nous importe peu. Ces 12 personnes ne sont qu'un échantillon de l'ensemble de toutes les personnes souffrant de ces maux. Ce qui est alors intéressant c'est de mesurer la variabilité des données induites par ces personnes. Ceci représentera une des composantes de la variation totale. Le facteur *personne* est donc considéré comme facteur à effets aléatoires.
On peut très bien imaginer une suite à cette étude en comparant la variabilité induite par un certain groupe de personnes par rapport à un autre groupe (par exemple des groupes selon le type d'activité professionnelle).

Bien entendu, les modélisations sont en réalité bien souvent plus compliquées, croisant différents effets entre eux. De plus, il n'est pas toujours évident de savoir si un effet doit entrer dans la modélisation avec une nature fixe ou aléatoire (se référer à la discussion en p.15-16 de SEARLE, CASELLA, et MC CULLOCH (1992)). Mais cet exemple très simple tentait uniquement d'éclairer les notions d'effets fixes et aléatoires.

Les **modèles mixtes** contiennent donc ces deux types d'effets. Dans ces modèles, nous nous intéressons à la fois à l'estimation de l'effet fixe ainsi qu'à celle des composantes de la variance.

Il est important de souligner d'ores et déja que, dans tout ce travail, nous sommes concernés par la mise en évidence des sources de variation, et l'identification des variations induites par la présence d'effets aléatoires. Cette *extra-variation* correspond donc à des hypothèses de modélisation de certains effets. Elle n'est pas de même nature que ce qui est pris en compte dans les modèles avec *surdispersion*. Dans ces derniers en effet, on augmente la variance totale des données sans pour autant avoir identifié la source de cette augmentation. Nous reviendrons plus précisément sur cette remarque dans la section 1.4.1.

1.3.2 Les hypothèses des L2M

Grâce à cette notion d'effet aléatoire, les modèles linéaires classiques ne contenant que des effets fixes ont pu être enrichis et élargis. On les appelle alors **modèles linéaires mixtes** – L2M. Une partie aléatoire vient s'ajouter à la partie fixe et l'on peut alors leur associer le formalisme suivant :

$$Y = \underbrace{X\beta}_{\text{partie effets fixes}} + \underbrace{U\xi}_{\text{partie effets aléatoires}} + \varepsilon \tag{1.1}$$

où

- Y : vecteur aléatoire à expliquer, de taille N, dont le vecteur des observations $y = (y_1, ..., y_N)'$ est une réalisation,

- β : vecteur de paramètres inconnus des effets fixes, de taille p, et X, de dimension $N \times p$, sa matrice d'incidence supposée fixe et connue,

- ξ : vecteur d'effets aléatoire de taille q. En toute généralité, ce vecteur se décompose en K parties $\xi = (\xi_1', ..., \xi_K')'$ où K est le nombre d'effets aléatoires considérés dans le modèle. Chaque composante ξ_j est un vecteur aléatoire de dimension q_j. Il est consitué des q_j réalisations du $j^{ème}$ effet aléatoire, observées au sein des données ($\sum_{j=1}^{K} q_j = q$).

  Remarque : On confond ici, et on le fera par commodité dans tout ce document, le vecteur aléatoire et sa réalisation.

  Dans l'exemple des médicaments de la section précédente, nous n'avons introduit qu'un seul effet aléatoire ($K = 1$): l'effet *personne*. Il sera modélisé par un vecteur de taille 12 où l'on affectera une réalisation de l'effet à chaque personne. Cela introduit une dépendance entre les 4 données relevées sur la personne.
  On suppose en général une distribution normale centrée réduite des effets aléatoires, c'est-à-dire : $\forall j \in \{1, ..., K\}$, $\xi_j \sim \mathcal{N}_{q_j}(0, \sigma_j^2 A_j)$. D'autre part, $\forall i, j \in \{1, ..., K\}^2$, ξ_i et ξ_j sont indépendants. Donc $\xi \sim \mathcal{N}_N(0, D)$ où D est une matrice diagonale par blocs : $D = \{_d\ \sigma_j^2 A_j\ \}_{j=1,...,K}$.
  La matrice d'incidence U, connue, est formée des différentes matrices d'incidence U_j de chaque effet aléatoire : $U = [U_1 \vdots ... \vdots U_K]$. Elle se compose le plus souvent de 0 et de 1. Dans les cas les plus simples, chaque ligne contient un 1 et des 0, indiquant ainsi que la mesure concernée a été prise pour telle réalisation (niveau) du facteur. Il arrive pourtant que cette matrice U soit plus compliquée, les effets pouvant s'accumuler avec une intensité diminuant, par exemple.

- ε : vecteur aléatoire d'erreurs de taille N. Puisque le modèle considéré est linéaire, la distribution de ε est : $\varepsilon \sim \mathcal{N}_N(0, \sigma_0^2 V_0)$. On notera aussi $R = \sigma_0^2 V_0$. On suppose que $\forall j \in \{1, ..., K\}$, ε et ξ_j sont indépendants.

Sous ces différentes hypothèses, on notera aussi :

$$\begin{aligned} \mathrm{var}(Y) = \Gamma &= R + UDU' , \\ &= \sigma_0^2 V_0 + \sum_{j=1}^{K} \sigma_j^2 U_j A_j U_j' . \end{aligned}$$

Ce qui, avec $\forall j \in \{1, \ldots, K\}$, $V_j = U_j A_j U_j'$, donne :

$$\Gamma = \sum_{j=0}^{K} \sigma_j^2 V_j .$$

La variation totale est donc scindée en plusieurs composantes que l'on appelle *composantes de la variance*.
Enfin, par les propriétés de conditionnement de la loi normale, on dispose aussi des distributions suivantes [2] :

- $Y|\xi \sim \mathcal{N}_N (X\beta + U\xi, R)$,
- $Y \sim \mathcal{N}_N (X\beta, \Gamma)$,
- $\begin{pmatrix} Y \\ \xi \end{pmatrix} \sim \mathcal{N}_{N+q} \left(\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} R + UDU' & UD \\ DU' & D \end{pmatrix} \right)$,
- $\xi|Y \sim \mathcal{N}_q (DU'\Gamma^{-1}(y - X\beta), D - DU'\Gamma^{-1}UD)$.

Dans ce modèle, le vecteur des effets fixes β ainsi que le vecteur des $K + 1$ paramètres de variance $\sigma^2 = (\sigma_0^2, ..., \sigma_K^2)'$ sont inconnus et nous nous intéresserons plus particulièrement à leur estimation.

1.3.3 Les L2M non standards

Nous venons de donner la définition d'un L2M. Dans ce type de modèles, il peut arriver que certaines des composantes de la variance soient connues ; soit lors de la

2. On écrira abusivement dans tout le document $Y|X$ pour signifier $Y|X = x$; mais nous utiliserons malgré cela la réalisation x de la variable aléatoire X dans l'expression associée.

réalisation de l'expérience, soit par la modélisation retenue. En effet, nous l'avons fait remarquer dans le cadre des GLM, dans certains cas, la modélisation de l'espérance impose celle de la variance. Cette idée propre aux GLM ressurgit ici dans le cadre des L2M.
C'est une information supplémentaire importante qu'il est intéressant de prendre en compte. C'est pourquoi nous définissons une nouvelle classe de modèle : les **modèles linéaires mixtes non standards** – nsL2M.

Leur définition ne diffère de celle du L2M que dans la séparation des composantes de la variance connues et inconnues. La matrice de variance V_ψ contient les premières et les secondes constituent le vecteur $\sigma^2 = (\sigma_1^2, ..., \sigma_K^2)'$.
Le modèle s'écrit alors :

$$Y = X\beta + \sum_{j=1}^{K} U_j \xi_j + \psi$$

$$\begin{array}{llllll} \text{où} \quad \forall j \in \{1, ..., K\}\,, & \xi_j & \sim & \mathcal{N}_{q_j}(0, \sigma_j^2 A_j) & \text{avec} & \sigma_j^2 \quad \text{inconnu,} \\ \qquad\qquad \text{et} & \psi & \sim & \mathcal{N}_N(0, V_\psi) & \text{avec} & V_\psi \quad \text{connue.} \end{array}$$

En prenant $V_\psi = 0$, on retrouve la définition du L2M classique, qui apparaît alors comme cas particulier d'un L2M non standard.

Comme nous l'avons signalé, ces modèles trouveront plus particulièrement leur justification lorsque nous aborderons la question de l'estimation des paramètres.

1.4 Les modèles étudiés : les GL2M

Il est maintenant temps de conjuguer ces deux classes de modèles (GLM et L2M) pour définir les GL2M, notre sujet d'étude. Les hypothèses qui leur sont associées seront énoncées dans la première sous-section. Nous donnerons ensuite deux exemples correspondant à une telle modélisation, avant d'en voir des extensions possibles.

1.4.1 Les hypothèses des GL2M

De même que les effets aléatoires ont été introduits dans les LM, ils peuvent l'être tout naturellement au sein des GLM pour donner naissance au modèle linéaire généralisé mixte. C'est alors dans l'expression du prédicteur linéaire, qu'une partie aléatoire vient s'ajouter à la partie fixe. On additionne effets fixes et aléatoires sur une même échelle. Ainsi, en gardant les notations de la section précédente concernant le vecteur des paramètres d'effets fixes β et sa matrice associée X ainsi que le vecteur des effets aléatoires ξ et sa matrice associée U, le prédicteur s'exprime de la façon suivante :

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

Pour bien insister sur l'introduction des effets aléatoires dans ce prédicteur, nous l'avons indicé par ξ. Nous conservons la même hypothèse de normalité sur ξ : $\xi \sim \mathcal{N}(0, D)$ où $D = \{_d \ \sigma_j^2 A_j \ \}_{j=1,\ldots,K}$.

Remarque : De même que nous confondons dans ξ le vecteur aléatoire et sa réalisation, η_ξ sera de nature aléatoire ou non selon les cas.

Nous avons insisté lors de la définition du GLM dans la section 1.2.2 sur l'importance des deux fonctions de lien et de variance. Elles restent primordiales pour les GL2M, mais cette fois-ci dans un raisonnement conditionnel à ξ. En effet, la fonction de lien relie ici le prédicteur linéaire à l'espérance conditionelle (que nous indiçons également par ξ) :

$$\eta_\xi = g(\mu_\xi) \qquad \text{où} \quad \mu_\xi = E(Y|\xi) \ .$$

Quant à la fonction de variance v, elle intervient ici dans l'expression de la variance conditionnelle :

$$\forall i \in \{1, \ldots, N\} \ , \qquad \mathrm{var}(Y_i|\xi) = a_i(\phi) v(\mu_{\xi,i}) \ .$$

Enfin, c'est sur la loi de Y conditionnelle à ξ qu'est formulée l'hypothèse de distribution. D'une part on suppose que, conditionnellement à ξ, les composantes de Y sont indépendantes, on obtient donc la matrice de variance conditionnelle suivante :

$$\mathrm{Var}(Y|\xi) = \{_d \ a_i(\phi) v(\mu_{\xi,i}) \ \}_{i=1,\ldots,N} \ .$$

D'autre part, $\forall i \in \{1, \ldots, N\}$, $Y_i|\xi$ est supposée distribuée selon une loi issue de la famille exponentielle.

Ainsi, conditionnellement à ξ, le GL2M conserve toutes les propriétés du GLM. Par conséquent le GL2M se trouve principalement défini dans un raisonnement conditionnel à ξ. C'est pourquoi il nous arrivera par la suite de le nommer *modèle conditionnel.* Il peut être résumé de la façon suivante :

- les composantes de Y sont, conditionnellement à ξ, indépendantes et de loi appartenant à la famille exponentielle,
- le prédicteur linéaire s'écrit : $\eta_\xi = X\beta + U\xi$,
- l'espérance conditionnelle de Y est reliée au prédicteur linéaire par la fonction de lien : $\eta_\xi = g(\mu_\xi)$.

Dans le cas de la loi normale (une des lois de la famille exponentielle), on retrouve bien la définition précédente du L2M avec un lien identité. Ainsi, comme le LM était un cas particulier des GLM, le L2M en est un des GL2M. Cependant il est important de noter, comme nous l'avons vu en section 1.3.2, que pour les L2M, il y a conservation de loi lors du passage des lois de $Y|\xi$ et ξ, à celle marginale de Y. Cette propriété est spécifique à la loi normale. Elle ne se retrouve pas pour d'autres lois. Le L2M est donc un cas bien particulier de GL2M.

Le schéma suivant résume les quatre types de modèles présentés :
- les flèches horizontales représentent une généralisation en termes d'effets aléatoires,
- les flèches verticales représentent une généralisation en termes de loi et de fonction de lien.

$$\begin{tikzcd}
\text{LM} \arrow[r, "\text{effets aléatoires}"] \arrow[d, "\text{loi}"'] & \text{L2M} \arrow[d] \\
\text{GLM} \arrow[r] & \text{GL2M}
\end{tikzcd}$$

Nous revenons sur la notion de surdispersion et la remarque faite au 1.3.1. Comme nous l'avons dit au 1.2.2, la surdispersion permet de modéliser une éventuelle variation inattendue des données par rapport au choix de loi réalisé (la loi imposant une relation entre espérance et variance). Elle est introduite dans les GLM en permettant au paramètre ϕ de prendre d'autres valeurs que celle imposée par la loi considérée. Si l'on voulait traduire cette hypothèse en termes d'effets aléatoires, cela reviendrait à introduire un effet (que nous nommons ci-dessous effet surdispersion) avec autant de réalisations que de données : une par donnée. Nous tenons alors à souligner la différence d'une telle modélisation avec les effets aléatoires que nous avons introduits dans les modèles. Cette différence tient principalement à deux raisons. D'une part, en introduisant cet effet surdispersion nous n'avons pas identifié la source de la variation supplémentaire. Cela n'a pas répondu à un choix de modélisation d'un certain facteur comme facteur à effet aléatoire. D'autre part, et c'est ce que nous retiendrons surtout, les composantes de ξ étant supposées indépendantes, le fait d'en introduire une par donnée, conserve marginalement une indépendance des composantes Y_i de Y. Au contraire, les effets aléatoires que nous introduisons induisent une dépendance entre ces composantes. C'est ce que nous avions évoqué sur l'exemple des médicaments en disant que le fait d'associer une réalisation à chaque personne traduisait une dépendance des 4 données relevées pour cette personne.

Nous illustrons cette modélisation GL2M sur deux exemples.

1.4.2 Deux exemples

Permettant de modéliser des situations où les données sont discrètes notamment, les GLM se sont grandement développés. Leur extension aux GL2M par l'introduction d'effets aléatoires a suivi naturellement. Nous présentons dans ce paragraphe deux exemples issus de domaines d'application bien distincts.

1.4.2.1 Un exemple en génétique animale

En biométrie, les applications où les GLM ont trouvé leur utilité sont multiples. L'exemple que nous présentons est issu du domaine de la génétique quantitative. Des données (non réelles), présentées par Shaeffer et Wilton (1976) à l'origine puis réutilisées par Gianola et Foulley (1983), illustrent l'étude de l'évaluation génétique pour la facilité de vêlage. Nous reprenons ici cette expérience en modifiant légèrement les termes de la description.

Dans 2 troupeaux, des génisses et des vaches sont croisées à 4 pères (taureaux) donnant naissances à 20 veaux mâles et femelles. Pour chacun, on enregistre la difficulté de vêlage selon 2 catégories : facile/difficile. On note aussi l'âge de la mère, le numéro du troupeau et le sexe du veau. La variable d'intérêt est donc l'information liée à la difficulté de vêlage qui peut prendre 2 valeurs possibles (elle est donc discrète).

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 plutot à estimer σ^2, la variabilité dûe à cet effet.
En considérant :

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

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 :

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

pour laquelle on a alors : $\qquad E(Y_i|\xi) = p_{\xi,i}$.
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 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 par exemple.

Ce type de plan d'expérience sera utilisé en de nombreuses occasions dans les chapitres suivants pour des simulations.

1.4.2.2 Un exemple en fiabilité des logiciels

On s'intéresse à la modélisation de l'amélioration stochastique d'un logiciel par des corrections apportées à chaque défaillance.
Pour cela, on note $\forall i \geq 1$:

- Y_i : la variable aléatoire *temps inter-défaillance*. C'est le temps séparant la panne $i-1$ de la panne i.
- Λ_i : la variable aléatoire positive représentant le *taux de panne* juste avant la panne i.

On observe les temps inter-défaillances. La loi adoptée pour modéliser ces temps inter-défaillances est la loi exponentielle :

$$\forall i \in \{1, \ldots, N\} \ , \quad Y_i|\Lambda_i \sim \mathcal{E}xp(\lambda_i) \ .$$

De plus, on fait l'hypothèse que les composantes Y_i conditionnellement à Λ_i sont indépendantes.

À chaque panne, les corrections vont modifier le taux de panne pour la défaillance suivante. Ici, on considère des corrections stochastiques ; ce qui entraîne un taux de panne de nature aléatoire. Nous supposons que les corrections à la panne i se traduisent par l'application d'un facteur multiplicatif. On obtient alors un modèle où les taux de défaillance sont proportionnels :

$$\Lambda_{i+1} = C_i \Lambda_i \ .$$

Lorsque $C_i < 1$, la correction est efficace, il y a accroissement de la fiabilité.
Ce coefficient multiplicatif est plus précisément écrit sous la forme :

$$C_i = \exp(-Q_i) \qquad \text{où} \quad Q_i = \theta + \xi_i \quad \text{et} \quad \xi_i \sim \mathcal{N}(0, \sigma^2) \ .$$

Ainsi, l'effet Q_i des corrections est réparti de façon normale autour d'un effet moyen θ avec une variance σ^2 inconnue.

Finalement, en prenant $\Lambda_1 = \lambda$, on aboutit à :

$$\forall i \geq 2 \ , \qquad \Lambda_i = \exp[\ln(\lambda) - (i-1)\theta - \sum_{j=1}^{i-1} \xi_j] \ .$$

Ce modèle est appelé LPM (Lognormal Proportional Model) par ses auteurs Gaudoin, Lavergne, et Soler (1994). À l'aide de N observations de pannes successives, on estimera les paramètres $\beta = (\ln(\lambda), -\theta)'$ et σ^2. L'effet aléatoire introduit dans ce cas correspond donc à la correction. On en observe $N-1$ réalisations : $\xi = (-\xi_1, ..., -\xi_{N-1})'$.
Les matrices X et U s'écrivent :

$$X = \begin{pmatrix} 1 & 0 \\ & 1 \\ & 2 \\ \vdots & \vdots \\ 1 & (N-1) \end{pmatrix} \qquad\qquad U = \begin{pmatrix} 0 & & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 1 & \dots & \dots & 1 & 0 \end{pmatrix}$$

Remarquons que bien que que l'on ait introduit autant (excepté pour Y_1) de réalisations de l'effet aléatoire que de données observées, nous ne sommes pas pour autant dans un cas similaire à celui de la surdispersion. Il existe ici en effet une forte dépendance marginale des composantes Y_i entre elle par le cumul progressif des corrections. À ce sujet, il est possible de diminuer l'impact des corrections passées par un cumul dégressif. Il suffit pour cela de considérer un ensemble décroissant

$(a_i)_{i=1,\ldots,N-1}$ d'entrées dans la matrice U pour obtenir :

$$U = \begin{pmatrix} 0 & & \ldots & \ldots & 0 \\ a_1 & 0 & \ldots & \ldots & 0 \\ a_2 & a_1 & 0 & \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ a_{N-1} & \ldots & a_2 & a_1 & 0 \end{pmatrix}.$$

Pour résumer les hypothèses de ce modèle, on a : $\forall i \in \{1, ..., N\}$,

- $Y_i|\Lambda_i \sim \mathcal{E}xp(\lambda_i)$,
- $\eta_i = x_i'\beta + u_i'\xi$ (où x_i' et u_i' sont respectivement les $i^{\text{èmes}}$ lignes de X et U),
- $\Lambda_i = exp(\eta_i)$.

C'est donc bien un modèle linéaire généralisé à un effet aléatoire (avec lien non canonique : lien logarithme).

1.4.3 Des extensions possibles

Pour une grande part dans ce travail, nous avons choisi de nous intéresser aux GL2M tels que nous les avons décrits en section 1.4.1. D'un point de vue pratique, ils offrent déjà de nombreuses possibilités de modélisation. Cependant, il en existe des extensions tout à fait intéressantes. Et nous tenons dans ce paragraphe à évoquer trois d'entre elles.

La première fait appel à la notion de *surdispersion*. Nous en avons déjà parlé dans les sections précédentes. L'idée consiste à permettre au paramètre de surdispersion ϕ de prendre d'autres valeurs que celle qui lui était imposée. Nous avons souligné en quoi cette modélisation répondait à une démarche différente. Les modèles avec surdispersion ont fait et font l'objet de nombreuses études. Nous ne les étudions pas spécifiquement dans ce travail même s'il nous arrivera, au cours de l'exposé de méthodes d'estimation, d'envisager l'éventuelle estimation du paramètre ϕ.

La deuxième consiste en une remise en question de l'hypothèse de *normalité des effets aléatoires*. En effet, depuis quelques années, certains auteurs ont commencé à envisager d'autres distributions pour ces effets. Cependant, ils ne l'ont pas fait de façon générale mais dans le cadre bien particulier d'un modèle précis avec des distributions spécifiques. Cela consistait bien souvent à prendre pour ξ la distribution conjuguée à celle de $Y|\xi$, et permettait par là même des manipulations mathématiques plus aisées. Récemment, Lee et Nelder (1996) ont permis une avancée

majeure dans ce domaine en définissant une nouvelle catégorie de modèles : les GLM hiérarchiques (HGLM). À travers eux, dans un cadre général de GLM, ils offrent de nombreuses possibilités de modélisation des effets aléatoires, regroupant celles évoquées ci-dessus ainsi que l'hypothèse gaussienne (réservée aux GL2M).
Ce choix concernant la distribution des effets aléatoires est donc un sujet actuel de débat. L'hypothèse gaussienne reste largement répandue, peut-être parce que nous y sommes plus familiés. Néanmoins, à nos yeux, elle offre surtout d'un point de vue pratique l'avantage d'interprétations plus faciles.

Enfin, une autre extension possible conduit à la catégorie des GL2M *hétérogènes*. Ces modèles constitueront le sujet d'étude du dernier chapitre de cette thèse. Il existe plusieurs façon de faire entrer une hypothèse d'hétérogénéité dans les GL2M. En ce qui nous concerne, c'est sur les paramètres de variance des GL2M classiques (GL2M homogènes par opposition aux GL2M hétérogènes) que nous formulons cette nouvelle hypothèse. Nous reportons donc au chapitre 4 la description plus précise de ces modèles.

Chapitre 2

Une méthode simple d'estimation dans les GL2M

2.1 Introduction

Comme nous l'avons présenté au chapitre précédent, ces GL2M que nous étudions sont à la fois un prolongement des GLM et des L2M. L'aspect modélisation ayant été décrit précisément, la question naturelle qui en découle est celle de l'estimation des paramètres inconnus. Ils sont de deux types. Il y a d'une part les paramètres d'effets fixes : le vecteur β, et d'autre part les paramètres de variance : le vecteur σ^2.

Dans ce chapitre, nous commençons par revenir sur les méthodes d'estimation :
- des effets fixes dans les GLM,
- des composantes de la variance dans les L2M.

En combinant ces deux types de techniques, nous serons naturellement conduits vers une méthode d'estimation des paramètres dans les GL2M. C'est une méthode simple (dans sa présentation et dans son implémentation) que nous proposons et qui utilise la double nature de ces modèles. Nous présenterons des résultats de simulations.

Bien entendu, d'autres travaux ont déjà été réalisés concernant l'estimation dans les GL2M. En abordant le problème sous des angles différents, des auteurs ont proposé diverses méthodes d'estimation. Nous reviendrons sur celles-ci, en exhibant les difficultés plus spécifiques des GL2M à ce sujet. Ceci permettra aussi de situer la méthode que nous proposons par rapport à ces travaux.

2.2 Estimation dans les GLM

Dans un premier temps, nous nous replaçons donc dans le cadre des GLM et nous nous intéressons à l'estimation de β. Pour cela, nous décrivons la procédure usuelle permettant d'atteindre l'estimation par maximum de vraisemblance. Nous donnerons quelques propriétés de cette estimation. Et enfin, nous nous arrêterons un instant sur la notion de quasi-vraisemblance introduite par Wedderburn (1974), que nous utiliserons par ailleurs dans ce travail.

2.2.1 Estimation maximum de vraisemblance

Considérons un modèle linéaire généralisé tel que nous l'avons décrit à la section 1.2.2. Au sein de ce modèle, nous voulons estimer le vecteur des paramètres β, de dimension p, coefficients de la combinaison linéaire des covariables permettant d'expliquer le vecteur Y. Avec l'hypothèse d'indépendance des coordonnées de Y, la log-vraisemblance du vecteur des paramètres canoniques θ, au vu du vecteur d'observations y, est :

$$L(\theta; y) = \sum_{i=1}^{N} [\frac{y_i\theta_i - b(\theta_i)}{\phi/\omega_i} + c(y_i, \phi)] = \sum_{i=1}^{N} L_i(\theta_i; y_i) \ .$$

Le lien entre β et θ est décrit par la relation :

$$X\beta = g(b'(\theta)) \ .$$

Ainsi, cette fonction L, lue comme log-vraisemblance du vecteur de paramètres β, peut être dérivée par rapport à ses diverses composantes. On a alors :
$\forall i \in \{1, ..., N\}$, $\forall j \in \{1, ..., p\}$,

$$\begin{aligned} \frac{\partial L_i}{\partial \beta_j} &= \frac{\partial \eta_i}{\partial \beta_j} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial L_i}{\partial \theta_i} = X_{ij} \frac{1}{g'(\mu_i)} \frac{1}{b''(\theta_i)} \frac{y_i - \mu_i}{\phi/\omega_i} \ , \\ \text{d'où} \quad \frac{\partial L}{\partial \beta_j} &= \sum_{i=1}^{N} X_{ij} \frac{1}{g'(\mu_i)^2 \mathrm{var}(Y_i)} \ g'(\mu_i) \ (y_i - \mu_i) \ . \end{aligned}$$

Ainsi, en considérant la matrice :

$$W_\beta = \{_d \ \mathrm{var}(Y_i) g'(\mu_i)^2 \ \}_{i=1,\dots,N} = \{_d \ \frac{\phi}{\omega_i} v(\mu_i) g'(\mu_i)^2 \ \}_{i=1,\dots,N} \ ,$$

les équations du maximum de vraisemblance pour β s'écrivent :

$$X' W_\beta^{-1} \frac{d\eta}{d\mu} (y - \mu) = 0 \ . \qquad (2.1)$$

où $\qquad \dfrac{d\eta}{d\mu} = \{_d \ \dfrac{d\eta_i}{d\mu_i} \ \}_{i=1,\ldots,N} = \{_d \ g'(\mu_i) \ \}_{i=1,\ldots,N}.$

Ce système d'équations n'est pas linéaire en β, qui intervient à la fois dans les matrices W_β, $\dfrac{d\eta}{d\mu}$ et dans le vecteur μ. C'est pourquoi, il en est envisagé une résolution itérative. L'algorithme itératif usuel mis en place pour cela, est l'algorithme des scores de Fisher. Il procède aux itérations suivantes :

$$\begin{aligned}
\beta^{[t+1]} &= \beta^{[t]} - \left(E\left[\left\{ \frac{\partial^2 L}{\partial\beta\partial\beta'} \right\} \right]^{[t]} \right)^{-1} \frac{\partial L}{\partial\beta}^{[t]} \\
&= \beta^{[t]} + \left(X'W^{-1}_{\beta^{[t]}} X \right)^{-1} X'W^{-1}_{\beta^{[t]}} \frac{d\eta}{d\mu}^{[t]} \left(y - \mu^{[t]} \right) \\
&= \left(X'W^{-1}_{\beta^{[t]}} X \right)^{-1} X'W^{-1}_{\beta^{[t]}} \ z^{[t]}
\end{aligned}$$

où $z^{[t]} = X\beta^{[t]} + \dfrac{d\eta}{d\mu}^{[t]} \left(y - \mu^{[t]} \right)$.

Remarque : Dans le cas d'un lien canonique, l'algorithme des scores de Fisher est identique à celui de Newton-Raphson.

Cet algorithme itératif peut être relu de la façon suivante. Si l'on introduit le vecteur dépendant défini par :

$$z_\beta = \eta + \frac{d\eta}{d\mu}(y - \mu) = X\beta + \frac{d\eta}{d\mu}(y - \mu) \ ,$$

les équations (2.1) deviennent alors :

$$X'W_\beta^{-1}(z_\beta - X\beta) = 0 \ . \tag{2.2}$$

où l'on a pris soin d'indicer par β les quantités qui en dépendent.
Ainsi, le même algorithme est décrit en résolvant itérativement les équations (2.2) comme des équations normales. À chaque itération, la valeur courante de β permet le calcul de la matrice des poids W_β et du vecteur dépendant z_β, et alors l'obtention, par résolution de ce système *linéarisé*, d'une nouvelle valeur de β.

Cette réécriture (2.2) permet une interprétation de type linéaire, que nous exploiterons plus particulièrement dans le cadre des GL2M. À β fixé, en considérant z_β comme un nouveau vecteur de données et W_β comme une matrice de poids fixés, on reconnait alors dans le système (2.2) les équations classiques des moindres carrés généralisés associées au modèle :

$$Z_\beta = X\beta + e$$

où $E(e) = 0$ et $\mathrm{Var}(e) = W_\beta$. Puisque la $i^{ème}$ composante du vecteur aléatoire $Z_\beta = X\beta + g'(\mu)(Y - \mu)$ a pour variance : $\mathrm{var}(Z_{\beta,i}) = g'(\mu_i)^2\mathrm{var}(Y_i)$, W_β correspond bien à la matrice de variance de Z_β.
Ce modèle présente deux particularités. D'une part, le vecteur à expliquer Z_β n'est pas observé. On en obtient des valeurs à chaque nouvelle valeur de β et à l'aide des observations y. Le vecteur z_β est donc recalculé à chaque itération. D'autre part, la matrice des poids dépend aussi du paramètre à estimer. Ainsi, à l'itération $[t]$, pour obtenir l'estimation $[t+1]$ de β, on utilise en réalité la méthode des moindres carrés généralisés dans le modèle que nous notons $\mathcal{M}^{[t]}$: $Z_{\beta^{[t]}} = X\beta + e^{[t]}$ où $E(e^{[t]}) = 0$ et $\mathrm{Var}(e^{[t]}) = W_{\beta^{[t]}}$. Ce qui est équivalent à estimer β dans le modèle $\mathcal{M}^{[t]}$ par maximum de vraisemblance après avoir supposé une distribution normale et l'indépendance des variables aléatoires $Z_{\beta,i}$. Par la suite, nous appellerons ce modèle $\mathcal{M}^{[t]}$: *modèle linéarisé.*

Notons d'une part qu'un développement au premier ordre de la fonction de lien g en μ donne $g(y) \approx g(\mu) + g'(\mu)(y - \mu) = z$. C'est pourquoi z_β peut être vue comme une approximation au premier ordre de $g(y)$ en μ_β. Cette remarque peut être utilisée pour obtenir une valeur initiale à β en appliquant une procédure de moindres carrés ordinaires sur $g(y)$. Cette linéarisation est aussi présentée par certains (cf. Engel et Keen 1994) comme un développement de μ autour de β.
Cet algorithme a fait l'objet de divers commentaires et éclairages dont celui de Hillis et Davis (1994).

En conclusion, nous retenons que l'estimation du maximum de vraisemblance de β dans le GLM écrit sous la forme :

$$Y = g^{-1}(\eta) + \varepsilon \quad \text{où} \quad \begin{array}{lcl} E(\varepsilon) & = & 0 \\ \mathrm{Var}(\varepsilon) & = & \{_d\; a(\phi)v(\mu_i)\;\}_{i=1,\dots,N} \end{array}$$

est équivalent à l'estimation successive du maximum de vraisemblance dans le LM $\mathcal{M}^{[t]}$ défini à l'étape $[t]$ par :

$$Z_{\beta^{[t]}} = \eta + e^{[t]} \qquad \text{où} \;\; E(e^{[t]}) = 0 \;\; \text{et} \;\; \mathrm{Var}(e^{[t]}) = W_{\beta^{[t]}}\ .$$

2.2.2 Propriétés asymptotiques

Dans cette section, nous nous intéressons brièvement aux propriétés asymptotiques de cet estimateur du maximum de vraisemblance. Pour cela, différentes hypothèses de travail peuvent être envisagées qui dépendent du type de propriétés désiré et des cas particuliers étudiés. Dans le cadre général des GLM, Fahrmeir et Kaufmann (1985) démontrent différents résultats dont nous retenons ici le théorème sur la normalité asymptotique de $\hat{\beta}_n$, solution des équations du maximum de vraisemblance pour un jeu de données de taille n. Ce théorème repose sur des hypothèses concernant les matrices hessiennes $H_n(\beta)$ et d'information de Fisher $I_n(\beta)$. Nous le présentons ici dans le cadre d'un GLM avec lien quelconque, pour lequel l'unicité du maximum de vraisemblance n'est plus assuré, excepté dans certains cas.

On note ici :

- $L_n(\beta; y)$: la log-vraisemblance du paramètre β associée au vecteur de données y (de taille n).
- $H_n(\beta)$: la matrice $H_n(\beta) = -\dfrac{\partial^2 L_n(\beta; y)}{\partial\beta\partial\beta'}$.
- $I_n(\beta)$: la matrice d'information de Fisher $I_n(\beta) = E(H_n(\beta)) = X'W_\beta^{-1}X$ où $W_\beta = \{_d\ a(\phi)b''(\theta_i)g'(\mu_i)^2\ \}_{i=1,\ldots,N}$.
- $I_n^{1/2}(\beta)$: la matrice vérifiant $I_n^{1/2}(\beta)\ I_n^{1/2}(\beta)' = I_n(\beta)$ et que l'on suppose inversible d'inverse $I_n^{-1/2}(\beta)$.
- β_0 : vraie valeur inconnue du paramètre.

Considérons les deux conditions suivantes :

(C1) :

$\lambda_{min}(I_n(\beta_0)) \stackrel{n\to+\infty}{\longrightarrow} +\infty$
où $\lambda_{min}(I_n(\beta_0))$ est la plus petite valeur propre de la matrice d'information $I_n(\beta_0)$

(C2) :

$\forall\delta > 0, \forall\lambda \in \mathbb{R}^p /\ ||\lambda|| = 1$
$\max_{\beta\in V_n(\delta)} ||I_n^{-1/2}(\beta_0)\ H_n(\beta)\ I_n^{-1/2}(\beta_0)' - Id_p|| \stackrel{n\to+\infty}{\longrightarrow} 0$, *en probabilité sous la vraie loi P et sous P_{β_n},*
où $\beta_n = \beta_0 + \delta\ I_n^{-1/2}(\beta_o)'\ \lambda$,
et $V_n(\delta) = \{\beta \in \mathbb{R}^p / ||I_n^{1/2}(\beta_0)'\ (\beta_0 - \beta)|| \leq \delta\}$

Dans le cas du lien canonique, cette condition se réécrit :

(C2)* :

$\forall\delta > 0,\ \max_{\beta\in V_n(\delta)} ||I_n^{-1/2}(\beta_0)\ I_n(\beta)\ I_n^{-1/2}(\beta_0)' - Id_p|| \stackrel{n\to+\infty}{\longrightarrow} 0$

Fahrmeir et Kaufmann (1985) démontrent alors le théorème suivant :

Théorème : *Si les hypothèses (C1) et (C2) sont vérifiées, la suite des estimateurs de maximum de vraisemblance $\hat{\beta}_n$ est asymptotiquement gaussienne :*

$$I_n^{1/2}(\beta_0)'(\hat{\beta}_n - \beta_0) \overset{n \to +\infty}{\longrightarrow} \mathcal{N}(0, Id_p), \text{ en loi.}$$

L'estimateur du maximum de vraisemblance $\hat{\beta}_n$ est donc asymptotiquement gaussien : $\mathcal{N}(\beta_0, (X'W_{\beta_0}^{-1}X)^{-1})$.

2.2.3 La notion de quasi-vraisemblance

Jusqu'à présent, nous nous sommes concentrés sur l'estimation par maximum de vraisemblance. En effet lorsqu'il s'agit d'inférence, la théorie statistique accorde une importance primordiale à cette fonction de vraisemblance. Cependant, tout en relâchant les hypothèses contraignantes des lois de probabilités, et en ne retenant que les deux premiers moments et la fonction qui les relie, Wedderburn (1974) propose de réaliser cette inférence grâce à la fonction de *quasi-vraisemblance*. McCullagh et Nelder (1989) consacrent tout un chapitre de leur ouvrage à cette notion de quasi-vraisemblance. Nous y revenons succintement pour souligner son intérêt au sein des GLM. Et nous l'utiliserons particulièrement au chapitre 3.

Relâchant les hypothèses des GLM, nous supposons ici que :

- les observations sont indépendantes,
- le paramètre d'intérêt intervient dans la modélisation de l'espérance du vecteur aléatoire observé : $E(Y) = \mu(\beta)$,
- la matrice de variance de Y s'écrit : $\text{Var}(Y) = \{_d \; a(\phi)v(\mu_i) \; \}_{i=1,\ldots,N}$, où l'on retrouve une fonction de variance v, de même que dans la définition d'un GLM. En toute généralité, nous faisons apparaître dans cette expression un paramètre ϕ, qui lui ne dépend en aucun cas de β.

On retient donc essentiellement les deux hypothèses fortes au sein des GLM, concernant les deux premiers moments. En effet, dans certaines expériences, il peut s'avérer difficile de décrire le mécanisme probabiliste mis en jeu. L'information nécessaire à la construction de la fonction de vraisemblance fait alors défaut. C'est à ce moment que la fonction de quasi-vraisemblance peut jouer un rôle intéressant.

Cette fonction de quasi-vraisemblance est construite à partir des hypothèses ci-dessus de la façon suivante.
Soit U_i la fonction associée à la $i^{ème}$ composante de y :

$$U_i(\mu_i; y_i) = \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \ .$$

On définit alors le logarithme de la fonction de quasi-vraisemblance, la *log-quasi-vraisemblance*, par :

$$\begin{aligned} Q(\mu; y) &= \sum_{i=1}^{N} \int_{y_i}^{\mu_i} U_i(t; y_i) \, dt \\ &= \sum_{i=1}^{N} \int_{y_i}^{\mu_i} \frac{y_i - t}{a(\phi)v(t)} \, dt \ . \end{aligned}$$

Avec cette définition, puisque $E(U_i) = 0$, $\text{var}(U_i) = \dfrac{1}{a(\phi)v(\mu_i)}$ et $-E(\dfrac{\partial U_i}{\partial \mu_i}) = \dfrac{1}{a(\phi)v(\mu_i)}$, la log-quasi-vraisemblance va se comporter de la même façon qu'une log-vraisemblance.

Pour estimer le paramètre β, on cherche à annuler les dérivées de Q. Autrement dit, en notant :

$$\mathcal{U}(\beta) = G'V^{-1}\left(\frac{y-\mu}{a(\phi)}\right) \ , \quad \text{où } \ G = \frac{\partial \mu}{\partial \beta'} \ , \quad \text{et} \quad V = \{_d \ v(\mu_i) \ \}_{i=1,\dots,N} \ ,$$

on cherche $\hat{\beta}$ tel que $\mathcal{U}(\hat{\beta}) = 0$. Cette fonction $\mathcal{U}(\beta)$ est la fonction *quasi-score*.
En utilisant un algorithme de scores de Fisher pour résoudre ces équations, on obtient la procédure itérative :

$$(G'V^{-1}G)\Delta\beta^{[t]} = G'V^{-1}(y - \mu^{[t]}) \ .$$

Comme dans le cadre des GLM, en décrivant de manière plus précise le lien entre μ et β, on peut alors préciser l'expression de la matrice G. Pour cela, on note g la fonction de lien considéré et h sa réciproque. On a $\mu = h(\eta)$. Et on obtient :

$$G = KX \quad \text{où } \ K = \{_d \ h(x_i'\beta) \ \} \ ,$$

les matrices K, et par conséquent G, dépendent de β.
Le schéma itératif précédent peut alors s'écrire :

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

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

On retrouve le même type de procédure itérative que dans les GLM, pour l'estimation par maximum de vraisemblance. On remarquera 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.

Il est important de constater que, si le choix de modélisation de la fonction de variance v correspond à la fonction de variance naturelle associée à une loi de la famille exponentielle, cette fonction quasi-score correspond alors à la fonction score, dans le GLM défini par cette loi et la fonction de lien g. On retrouve ainsi les équations (2.1). Et la solution $\hat{\beta}$ du maximum de quasi-vraisemblance correspondra à l'estimation maximum de vraisemblance dans ce modèle. Cependant, et c'est là un intérêt primordial, grâce à cette notion de quasi-vraisemblance, il est aussi possible d'envisager d'autres fonctions de variance que celles précisément associées aux lois classiques. On peut alors poursuivre une inférence sans s'appuyer sur une hypothèse de loi de probabilité.

Enfin, cette notion s'étend au cas de variables dépendantes (cf. McCullagh et Nelder 1989). Elle a été étudiée par différents auteurs pour son extension à l'estimation du paramètre de dispersion (cf. Davidian et Carroll 1988, Godambe et Thompson 1989, ou encore Nelder et Pregibon 1987). C'est dans ce cadre que Nelder et Lee (1992) citent différentes situations pour son utilisation.

2.3 Estimation dans les L2M et nsL2M

Après ce retour sur l'estimation dans les GLM, nous nous intéressons dans cette troisième section à l'estimation des paramètres au sein des L2M et nsL2M. Alors que la section précédente était consacrée à l'estimation des effets fixes, ici c'est davantage l'estimation des composantes de la variance qui sera étudiée.

Depuis le début du $20^{ème}$ siècle, de nombreux travaux sur les L2M ont été réalisés sous des formes et selon des approches différentes. Searle, Casella, et Mc Culloch (1992) y consacrent leur ouvrage. Nous présentons ici quelques unes de ces méthodes ou algorithmes d'estimation : maximum de vraisemblance (ML – Maximum Likelihood), maximum de vraisemblance restreint (REML – Restricted Maximum Likelihood), l'algorithme EM (Expectation Maximisation) et une méthode dite "de Henderson". Ces méthodes s'appuient particulièrement sur les hypothèses de lois du modèle. Ce qui dans d'autres approches n'est pas forcément nécessaire.

Une particularité cependant dans cette section : nous verrons également si et comment ces méthodes peuvent être adaptées au cas où des composantes de la variance sont connues, autrement dit dans un nsL2M.
Enfin nous aborderons les propriétés asymptotiques de ces estimations.

Rappelons brièvement les hypothèses du L2M décrit en section 1.3.2 :

$$Y = X\beta + U\xi + \varepsilon \tag{2.3}$$

où

- $\varepsilon \sim \mathcal{N}_N(0, \sigma_0^2 V_0)$,
- $\forall j \in \{1, ..., K\}\ , \quad \xi_j \sim \mathcal{N}_{q_j}(0, \sigma_j^2 A_j)$.
  Les ξ_j sont indépendants entre eux et indépendants de ε.
  Alors $\xi \sim \mathcal{N}_q(0, D)$ avec D diagonale par blocs $D = \{_d\ \sigma_j^2 A_j\ \}_{j=1,...,K}$.

On a donc :

- $$\begin{array}{rcl} E(Y|\xi) & = & X\beta + U\xi\ , \\ \mathrm{var}(Y|\xi) & = & \mathrm{var}(\varepsilon|\xi) = \mathrm{var}(\varepsilon) = R = \sigma_0^2 V_0\ , \end{array}$$
- $$\begin{array}{rcl} E(Y) & = & X\beta\ , \\ \mathrm{var}(Y) & = & R + UDU' \\ & = & \sum_{j=0}^{K} \sigma_j^2 V_j \qquad \text{avec} \qquad \forall j \in \{1, ..., K\}\ , \quad V_j = U_j A_j U_j', \\ & = & \Gamma \end{array}$$

On note $\sigma^2 = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_K^2)'$ le vecteur des composantes de la variance.

La fonction de vraisemblance, étant donné le vecteur y des observations, s'écrit :
$f(\beta, \sigma^2; y) = \dfrac{1}{(2\pi)^{N/2}\ |\Gamma|^{1/2}}\ \exp\{-\frac{1}{2}\ (y - X\beta)'\ \Gamma^{-1}\ (y - X\beta)\}\ ,$

et la log-vraisemblance :
$l(\beta, \sigma^2; y) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Gamma|) - \frac{1}{2}(y - X\beta)'\ \Gamma^{-1}\ (y - X\beta)\ .$

À l'aide de cette fonction de vraisemblance, nous envisageons dans un premier temps l'estimation par maximum de vraisemblance.

2.3.1 Estimation par maximum de vraisemblance

2.3.1.1 Dérivation directe des équations

En dérivant la log-vraisemblance ci-dessus, on obtient facilement[1] :

$$\left\{\begin{array}{lcl} \dfrac{\partial l(\beta,\sigma^2;y)}{\partial\beta} & = & X'\,\Gamma^{-1}\,(y - X\beta) \\ \forall j \in \{0,\ldots,K\}\,, & & \\ \dfrac{\partial l(\beta,\sigma^2;y)}{\partial\sigma_j^2} & = & -\frac{1}{2}\,\mathrm{tr}(\Gamma^{-1}V_j) \,+\, \frac{1}{2}(y - X\beta)'\,\Gamma^{-1}\,V_j\,\Gamma^{-1}\,(y - X\beta)\;. \end{array}\right.$$

Les estimateurs du maximum de vraisemblance (ML) de β et σ^2 doivent donc non seulement être solutions du système précédent lorsqu'on égalise chacune des lignes à zéro, mais aussi vérifier $\sigma_0^2 > 0$ et $\forall j \in \{1,...,K\}\;,\;\; \sigma_j^2 \geq 0$.
Le système d'équations que l'on cherche à résoudre est le suivant :

$$\left\{\begin{array}{lcll} X'\,\Gamma^{-1}\,X\;\beta & = & X'\,\Gamma^{-1}\,y & \\ \mathrm{tr}(\Gamma^{-1}V_j) & = & y'\,P\,V_j\,P\,y & \qquad j = 0,...,K, \end{array}\right. \tag{2.4}$$

où $P = \Gamma^{-1}(I - X(X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1})$.
Les équations de ce système sont résolues simultanément. La matrice Γ intervient dans la première équation concernant β. Autrement dit, l'estimation de β est relativement transparente au fait que les composantes de la variance soient connues ou estimées. Si une composante est connue, on reportera dans Γ cette vraie valeur, sinon on la remplacera par son estimation. D'autre part, les $K+1$ autres équations ne sont pas linéaires en σ^2. On devra donc envisager une solution itérative au système (2.4).

Cependant, même itérativement, la résolution directe de ce système n'est pas aisée. Plusieurs transformations peuvent en être envisagées (dont celle de Hartley-Rao (cf. Searle et al.)). L'une rapide et très lisible conduit au système (2.5) équivalent à (2.4) :

$$\left\{\begin{array}{lcl} X'\,\Gamma^{-1}\,X\;\beta & = & X'\,\Gamma^{-1}\,y \\ \left(\,\mathrm{tr}(\Gamma^{-1}V_i\Gamma^{-1}V_j)\,\right)_{i,j=0,\ldots,K} \begin{pmatrix} \sigma_0^2 \\ \vdots \\ \sigma_K^2 \end{pmatrix} & = & (y'\,P\,V_j\,P\,y)_{j=0,\ldots,K} \end{array}\right. \tag{2.5}$$

1. Rappelons pour cela les résultats (cf. Searle et al. (1992) p.456-475) : $\dfrac{\partial log|A|}{\partial\sigma_j^2} = \mathrm{tr}(A^{-1}\,\dfrac{\partial A}{\partial\sigma_j^2})$, et $\dfrac{\partial A^{-1}}{\partial\sigma_j^2} = -A^{-1}\dfrac{\partial A}{\partial\sigma_j^2}A^{-1}$.

On a utilisé pour cela le fait que
$$\begin{aligned} \mathrm{tr}(\Gamma^{-1}V_i) &= \mathrm{tr}(\Gamma^{-1}V_i\Gamma^{-1}\Gamma) \\ &= \sum_{j=0}^{K} \sigma_j^2 \,\mathrm{tr}(\Gamma^{-1}V_i\Gamma^{-1}V_j)\ . \end{aligned}$$
Malgré cette nouvelle présentation, les équations ne sont toujours pas linéaires en σ^2, du fait notamment de la présence de Γ dans les membres des équations. Elles permettent cependant de mettre en place très rapidement un algorithme itératif (qui ne sera pas forcément optimum). À partir de valeurs initiales de σ^2, on itère la résolution des équations concernant les composantes de la variance jusqu'à convergence, en résolvant le système linéaire à chaque étape. Puis à l'aide des valeurs alors obtenues, on résoud la première équation de (2.5) pour trouver l'estimation de β. Cependant rien n'assure la positivité des estimations des composantes. Pour cela, si lors d'une itération, l'estimation obtenue est négative, on forcera cette composante à 0 (ou à une petite valeur positive) ; ce qui revient à annuler l'effet du facteur correspondant.

Nous retenons donc que les estimations ML $\hat{\beta}$ et $\hat{\sigma}^2$ dans un L2M vérifient :

$$\left\{ \begin{array}{ccc} X'\,\hat{\Gamma}^{-1}\,X\,\hat{\beta} & = & X'\,\hat{\Gamma}^{-1}\,y \\ \Big(\mathrm{tr}(\hat{\Gamma}^{-1}V_i\hat{\Gamma}^{-1}V_j)\Big)_{i,j=0,\ldots,K} \begin{pmatrix} \hat{\sigma}_0^2 \\ \vdots \\ \hat{\sigma}_K^2 \end{pmatrix} & = & \Big(y'\,\hat{P}\,V_j\,\hat{P}\,y\Big)_{j=0,\ldots,K} \end{array} \right. \tag{2.6}$$

dans lequel $\hat{\Gamma}$ désigne Γ où l'on a remplacé σ^2 par son estimation, et $\hat{P} = \hat{\Gamma}^{-1}(y - X\hat{\beta})$ d'où $\hat{P}y = \hat{\Gamma}^{-1}(y - X\hat{\beta})$.

Regardons maintenant ce qui se passe dans le cas d'un nsL2M.

2.3.1.2 Adaptation au nsL2M

On suppose pour cela que σ_0^2 est connu et que, parmi les K composantes de la variance, L seulement sont inconnues. On réécrit alors le modèle sous la forme d'un nsL2M :
$$Y = X\beta + \sum_{j=1}^{L} U_j\xi_j + \psi\ ,$$
avec $\mathrm{Var}(\psi) = V_\psi$ connue et $\Gamma = \sum_{j=1}^{L} \sigma_j^2 V_j + V_\psi$.

Dans une démarche d'estimation par maximum de vraisemblance, en annulant les dérivées de la vraisemblance, on aboutit alors au système (2.7), qui n'est autre que le système (2.4) où les lignes correspondant aux composantes connues ont été supprimées :

$$\left\{\begin{array}{lcl} X'\,\Gamma^{-1}\,X\,\beta & = & X'\,\Gamma^{-1}\,y \\ \forall j \in \{1,\ldots,L\}\,, & & \\ \mathrm{tr}(\Gamma^{-1}V_j) & = & y'\,P\,V_j\,P\,y\;. \end{array}\right. \tag{2.7}$$

C'est dans le passage équivalent à celui de (2.4) à (2.5), que l'on va prendre en compte le fait que V_ψ est connue. On obtient alors le système :

$$\left\{\begin{array}{lcl} X'\,\Gamma^{-1}\,X\,\beta & = & X'\,\Gamma^{-1}\,y \\ (\,\mathrm{tr}(\Gamma^{-1}V_i\Gamma^{-1}V_j)\,)_{i,j=1,\ldots,L}\begin{pmatrix}\sigma_1^2\\ \vdots \\ \sigma_L^2\end{pmatrix} & = & (y'\,P\,V_i\,P\,y - \mathrm{tr}(\Gamma^{-1}V_i\Gamma^{-1}V_\psi))_{i=1,\ldots,L} \end{array}\right. \tag{2.8}$$

Ainsi, ce système tient compte de l'information détenue sur une partie de la variance. Comme précédemment, il sera résolu itérativement de façon à obtenir les estimations ML des paramètres d'un nsL2M.

Pour terminer, remarquons que dans le cas où $V_\psi = 0$ et $L = 1$ (cas d'un L2M avec une seule composante de la variance), on a alors :

$$\left\{\begin{array}{l} \Gamma = \sigma_1^2 V_1\;, \\ \mathrm{tr}(\Gamma^{-1}V_1\Gamma^{-1}V_1) = \dfrac{N}{\sigma_1^4}\;, \\ y'PV_1Py = \dfrac{y'P_{V_1^{-1}}y}{\sigma_1^2}\;, \end{array}\right.$$

et l'on retrouve $\hat{\sigma}_1^2 = \dfrac{y'\hat{P}_{V_1^{-1}}y}{N} = \dfrac{||y - X\hat{\beta}||^2_{V_1^{-1}}}{N}$, l'estimateur du maximum de vraisemblance classique.

2.3.2 Estimation par maximum de vraisemblance restreint

Comme son nom l'indique, cette démarche d'estimation reste apparentée à celle précédente du maximum de vraisemblance. Mais ici, on se focalise davantage sur l'estimation des composantes de la variance. Pour cela, on fait disparaître momentanément les effets fixes pour ne maximiser que la partie de la vraisemblance concernant les composantes et indépendante des effets fixes.
Cette méthode (REML), sous-jacente depuis le début des années 50, a été mise au point de façon générale par Patterson et Thompson dans les années 70. On peut la justifier selon plusieurs points de vue. Nous en évoquerons deux ici. Et nous verrons ce à quoi elle aboutit dans le cas d'un nsL2M.

2.3.2.1 Dérivation des équations

Afin d'éliminer la partie effet fixe, ce n'est pas directement sur le vecteur Y que l'on va travailler mais sur une transformation de ce vecteur. On s'intéresse à des combinaisons linéaires $k'Y$ indépendantes et d'espérance nulle, aussi appelées *contrastes*. Si rang$(X) = p$, il existe $N-p$ combinaisons linéaires indépendantes de la sorte. Elles sont regroupées dans la matrice K'. Puisque $K'X = 0$, on a alors $E(K'Y) = K'X\beta = 0$, $\mathrm{Var}(K'Y) = K'\Gamma K$, et $K'Y \sim \mathcal{N}_{N-p}(0, K'\Gamma K)$.
Cela revient à projeter le modèle sur l'orthogonal du sous-espace vectoriel engendré par les colonnes de X noté $X^\perp$. Ainsi, la matrice K est la matrice dont les $N-p$ vecteurs colonnes constituent une base de $X^\perp$ alors $K'X = 0$.
Le modèle projeté est donc :

$$\begin{array}{llllllll} & K'\,Y & = & K'\,X\,\beta & + & K'\,U\,\xi & + & K'\,\varepsilon \quad , \\ \text{d'où} & K'\,Y & = & K'\,U\,\xi & + & K'\,\varepsilon \quad , & & \\ \text{et} & \mathrm{Var}(K'\,Y) & = & K'\,\Gamma\,K \quad . & & & & \end{array}$$

Alors, l'estimation par maximum de vraisemblance restreint n'est autre que l'estimation par maximum de vraisemblance dans le modèle projeté. Pour établir ces équations du maximum de vraisemblance pour les composantes, il suffit de reprendre les équations (2.4), de supprimer la première ligne (celle des effets fixes) et d'effectuer le changement de notation :

$$\begin{array}{llll} Y & \to & K'Y & \\ U_i & \to & K'U_i & i = 1, ..., K \\ V_i & \to & K'V_iK & i = 0, ..., K \\ \Gamma & \to & K'\Gamma K\ , & \end{array}$$

en sachant que $K(K'\Gamma K)^{-1}K' = P$. On obtient alors le système :

$$\left\{ \mathrm{tr}(PV_i) \;=\; y'\,P\,V_i\,P\,y \qquad i = 0, ..., K. \right. \tag{2.9}$$

De même que dans la section précédente, on envisage plutôt une résolution itérative du système équivalent suivant :

$$\left\{ \left(\mathrm{tr}(PV_iPV_j)\right)_{i,j=0,\dots,K} \begin{pmatrix} \sigma_0^2 \\ \vdots \\ \sigma_K^2 \end{pmatrix} = (y'\,P\,V_i\,P\,y)_{i=0,\dots,K} \right. \tag{2.10}$$

Ce système permet donc d'obtenir les estimations REML dans un L2M. Cette méthode d'estimation a l'avantage sur la méthode ML de tenir compte de la perte de degrés de liberté occasionnée par l'estimation des effets fixes. Après avoir estimé les composantes de la variance, il suffit alors de former la matrice $\hat{\Gamma}$ pour obtenir directement l'estimation de β.
Remarquons, et ceci sera utilisé en d'autres occasions, que pour passer du système ML au système REML, il a suffi de remplacer Γ^{-1} par P.

2.3.2.2 Justification bayésienne de ces équations

Ces équations REML peuvent aussi trouver une justification dans un raisonnement bayésien. Pour cela, on reprend les hypothèses du L2M de départ mais en donnant aux paramètres β et σ^2 une nature aléatoire. On suppose des connaissances a priori sur ces paramètres, qui se traduisent par des hypothèses sur leur distribution. Dans tout ce paragraphe, on adopte la notation générique p pour désigner les lois de probabilité.
En supposant a priori :
- les paramètres indépendants,
- et une distribution uniforme (non informative) pour β,
on a :

$$p(\beta, \sigma^2) = p(\beta)p(\sigma^2) \propto p(\sigma^2) .$$

D'où on obtient :

$$p(y|\beta, \sigma^2) \propto p(\beta|y)p(y|\sigma^2) .$$

La fonction de vraisemblance est ainsi décomposée en produit de la loi a posteriori de β et de la vraisemblance de σ^2.
D'autre part, en remarquant que : $||y-X\beta||^2_{\Gamma^{-1}} = ||y-X\hat{\beta}||^2_{\Gamma^{-1}}+||\hat{\beta}-\beta||^2_{\mathrm{Var}(\hat{\beta})^{-1}}$ avec $\hat{\beta} = (X'\,\Gamma^{-1}\,X)^{-1}\,X'\,\Gamma^{-1}\,y$, on peut alors écrire la fonction de vraisemblance sous la forme suivante (où l'on reconnait les deux parties de la décomposition précédente) :

$$L(\beta, \sigma^2; y) =$$

$$\frac{|X'\Gamma^{-1}X|^{-\frac{1}{2}}}{(2\pi)^{\frac{N-p}{2}}|\Gamma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}||y - X\hat{\beta}||^2_{\Gamma^{-1}}\} \times \underbrace{\frac{1}{(2\pi)^{\frac{p}{2}}|X'\Gamma^{-1}X|^{-\frac{1}{2}}} \exp\{||\beta - \hat{\beta}||^2_{\mathrm{Var}(\hat{\beta})^{-1}}\}}_{\text{densité a posteriori de } \beta|y}$$

La vraisemblance "marginale" (celle de σ^2) après intégration sur β (puisque l'idée est toujours de faire disparaître les effets fixes) est alors :

$$L(\sigma^2; y) = \frac{1}{(2\pi)^{\frac{N-p}{2}}|\Gamma|^{\frac{1}{2}}|X'\Gamma^{-1}X|^{\frac{1}{2}}} \exp\{-\frac{1}{2}||y - X\hat{\beta}||^2_{\Gamma^{-1}}\} ,$$

et donc la log-vraisemblance :

$$l(\sigma^2; y) = -\frac{N-p}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Gamma|) - \frac{1}{2}\ln(|X'\Gamma^{-1}X|) - \frac{1}{2}(y - X\hat{\beta})'\,\Gamma^{-1}\,(y - X\hat{\beta}) .$$

Par dérivation, on obtient :

$$\frac{\partial l(\sigma^2; y)}{\partial \sigma_j^2} = -\tfrac{1}{2}\mathrm{tr}(PV_j) \ + \ \tfrac{1}{2}y'\,P\,V_j\,P\,y \qquad j = 0, ..., K .$$

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

On retrouve donc le système d'équations (2.9) et le (2.10) qui s'en déduit.

2.3.2.3 Adaptation au nsL2M

De même que dans 2.3.1.2, on se place maintenant dans un nsL2M où une partie des composantes de la variance est connue. On peut alors adapter le système (2.10) au cas où L composantes de la variance uniquement restent inconnues. Pour cela, dans le passage de (2.9) à (2.10), on utilise la propriété $P\Gamma P = P$. On obtient alors le système (2.11) :

$$\left\{ \left(\operatorname{tr}(PV_iPV_j)\right)_{i,j=1,\ldots,L} \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_L^2 \end{pmatrix} = \left(y'\,P\,V_i\,P\,y - \operatorname{tr}(PV_iPV_\psi)\right)_{i=1,\ldots,L} \right. \tag{2.11}$$

Si l'on envisage comme au 2.3.1.2, le cas où $V_\psi = 0$ et $L = 1$, on obtient cette fois-ci : $\hat{\sigma}_1^2 = \dfrac{||y - X\hat{\beta}||^2_{V_1^{-1}}}{N-p}$, l'estimateur du maximum de vraisemblance restreint classique (estimateur non biaisé). On retrouve au dénominateur le fait que l'estimation REML prend en compte la perte de degrés de liberté dans l'estimation des effets fixes.

Un débat a eu lieu pour l'utilisation de l'un ou l'autre de ces estimateurs ML ou REML. Nous ne reprenons pas ici ces arguments.
Dans la section suivante, nous nous intéressons plutôt à l'utilisation de l'algorithme EM pour atteindre ces estimations ML ou REML.

2.3.3 Utilisation de l'algorithme EM

Les systèmes (2.4) et (2.9) envisagés précédemment pour l'estimation ML (ou REML) des composantes de la variance sont des systèmes non linéaires avec contraintes. Outre le fait que rien ne nous assure la positivité des estimations pas à pas, nous ne sommes pas non plus certains qu'ils mènent à un maximum global de la fonction de vraisemblance. D'autres alternatives à la résolution itérative de ces systèmes ont été proposées. L'algorithme EM (Espérance - Maximisation) en fait partie. Cette méthodologie a été mise en place par Dempster, Laird, et Rubin (1977). Appliquée aux L2M, elle constitue un outil permettant d'atteindre également ces estimations ML ou REML.

Nous présentons ici rapidement cet algorithme dans notre cadre bien précis des L2M. C'est dans le chapitre 4 que nous en ferons une description plus détaillée et dans des termes plus généraux. Cependant, dans cette section, nous suivrons davantage une démarche utilisant les notions de statistiques exhaustives.

Pour les L2M, l'idée est donc la suivante. Supposons qu'au cours de l'expérience

on observe non seulement le vecteur y mais aussi le vecteur d'effets aléatoires ξ. On forme alors le vecteur $x = (y', \xi')'$, que l'on appelle vecteur des *données complètes* et qui regroupe toutes les observations. À l'aide de x, il s'agit ensuite d'obtenir des statistiques exhaustives $t(x)$ permettant de réaliser l'estimation de chacun des paramètres d'intérêt que l'on note ici de façon générique θ.
À ce stade, les effets aléatoires n'étant en réalité pas observés, on calculera l'espérance des expressions obtenues conditionnellement au vecteur des données réellement observées y. L'estimateur du maximum de vraisemblance, ou maximum de vraisemblance restreint, de θ sera alors une fonction de l'espérance conditionnelle des statistiques $t(x)$: $E(t(x)|y, \theta)$.
En pratique, cet algorithme se déroulera de façon itérative, chaque itération étant constituée de deux étapes. A l'itération $[t]$, partant de la valeur $\theta^{[t]}$ du paramètre :

- étape E : calculer l'espérance conditionnelle $E(t(x)|y, \theta^{[t]})$,
- étape M : calculer la nouvelle valeur du paramètre $\theta^{[t+1]}$ en remplaçant les statistiques exhaustives par leur espérance conditionnelle obtenue à l'étape précédente.

Appliquons cela plus précisément pour l'estimation ML et REML.

2.3.3.1 Algorithme EM pour ML

Selon la démarche décrite précédemment, on complète le vecteur des observations y par $\xi_1, ..., \xi_K$. Si l'on observait une réalisation du vecteur aléatoire ξ_j, on estimerait naturellement σ_j^2 par $\hat{\sigma}_j^2 = \dfrac{\xi_j' A_j^{-1} \xi_j}{q_j}$, et β par $\hat{\beta} = (X'V_0^{-1}X)^{-1}X'V_0^{-1}(y - \sum_{j=1}^K U_j \xi_j)$. Cela correspond bien aux estimations obtenues en maximisant la vraisemblance complétée. En effet, les propriétés usuelles de conditionnement de la loi normale nous donne la distribution du vecteur des données complètes $x = [y', \xi_1', ..., \xi_K']'$. C'est une loi normale de moyenne $\mu = \begin{bmatrix} X\beta \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ et de matrice de variance[2] : $\Sigma = \begin{bmatrix} \Gamma & \left\{_r\ \sigma_j^2 U_j A_j\ \right\}_{j=1,...,K} \\ \left\{_c\ \sigma_j^2 A_j U_j'\ \right\}_{j=1,...,K} & \left\{_d\ \sigma_j^2 A_j\ \right\}_{j=1,...,K} \end{bmatrix}$.

2. On désigne par $\{_c\ a_j\ \}$ la matrice A obtenue en superposant en colonne les éléments a_j et par $\{_r\ a_i\ \}$ la matrice A obtenue en juxtaposant en ligne les éléments a_i.

Le calcul du déterminant et d'une décomposition de l'inverse de Σ (cf SEARLE et al. 1992 p 450) aboutissent à :

- $$\begin{aligned} |\Sigma| &= \prod_{j=1}^{K}[(\sigma_j^2)^{q_j}|A_j|] \times |\Gamma - \sum_{j=1}^{K}\sigma_j^2 U_j A_j U_j'| \\ &= \prod_{j=1}^{K}[(\sigma_j^2)^{q_j}|A_j|] \times |\sigma_0^2 V_0| \\ &= \prod_{j=0}^{K}[(\sigma_j^2)^{q_j}|A_j|] \quad \text{ en notant } q_0 = N,\ A_0 = V_0 \text{ et } U_0 = I_N, \end{aligned}$$

- $$\begin{aligned} \Sigma^{-1} &= \begin{bmatrix} 0 & 0 \\ 0 & \{_d\ \frac{A_j^{-1}}{\sigma_j^2}\ \}_{j=1,\dots,K} \end{bmatrix} \\ &+ \begin{bmatrix} I \\ -\{_c\ U_j'\ \}_{j=1,\dots,K} \end{bmatrix} (\Gamma - \sum_{j=1}^{K}\sigma_j^2 U_j A_j U_j')^{-1} \begin{bmatrix} I & -\{_r\ U_j\ \}_{j=1,\dots,K} \end{bmatrix}. \end{aligned}$$

La fonction de vraisemblance des données complètes est donc :

$$l(\beta, \sigma^2; x) = -\frac{1}{2}(\sum_{j=0}^{K} q_j)\ln(2\pi) - \frac{1}{2}\sum_{j=0}^{K}(q_j \ln(\sigma_j^2) + \ln(|A_j|)) - \frac{1}{2}\sum_{j=0}^{K}\frac{\xi_j' A_j^{-1}\xi_j}{\sigma_j^2}$$

avec $\xi_0 = \varepsilon = y - X\beta - \sum_{j=1}^{K} U_j \xi_j$.

Ainsi les estimateurs du maximum de vraisemblance basée sur les données complètes sont :

$$\begin{cases} \hat{\sigma}_j^2 = \dfrac{\xi_j' A_j^{-1}\xi_j}{q_j} & \forall j \in \{0, ..., K\}, \\ (X'V_0^{-1}X)\hat{\beta} = X'V_0^{-1}(y - \sum_{j=1}^{K} U_j \xi_j) \end{cases}.$$

Or, comme on ne dispose que d'observations de y, EM propose de remplacer les statistiques exhaustives $\xi_j' A_j^{-1}\xi_j$ et $y - \sum_{j=1}^{K} U_j\xi_j$ par leurs espérances conditionnelles sachant y et les valeurs courantes des paramètres.

D'après les résultats sur le conditionnement des variables aléatoires normales, on a : $\forall j \in \{0, \dots, K\}$:

$$\begin{aligned} E(\xi_j|y) &= \sigma_j^2 A_j\, U_j'\, \Gamma^{-1}\,(y - X\beta) \\ E(\xi_j' A_j^{-1}\xi_j|y) &= \sigma_j^4\,(y - X\beta)'\,\Gamma^{-1}\,V_j\,\Gamma^{-1}\,(y - X\beta) + \mathrm{tr}(\sigma_j^2\, I_{q_j} - \sigma_j^4\, U_j'\,\Gamma^{-1}\,U_j A_j) \\ &= \sigma_j^4\,(y - X\beta)'\,\Gamma^{-1}\,V_j\,\Gamma^{-1}\,(y - X\beta) + q_j\sigma_j^2 - \sigma_j^4 \mathrm{tr}(\Gamma^{-1}V_j) \end{aligned}$$

$$\begin{aligned} \text{et} \quad E(Y - \textstyle\sum_{j=1}^{K} U_j\xi_j|y) &= y - \textstyle\sum_{j=1}^{K}\sigma_j^2 V_j \Gamma^{-1}(y - X\beta) \\ &= y - (\Gamma - \sigma_0^2 V_0)\Gamma^{-1}(y - X\beta) \\ &= X\beta + \sigma_0^2 V_0 \Gamma^{-1}(y - X\beta) \end{aligned}$$

Deux schémas itératifs peuvent alors être retenus : le premier donne à chaque pas des nouvelles valeurs à chacun des paramètres, le deuxième itère uniquement pour des estimations successives des σ_j^2, estimant β à la convergence. À partir des estimations obtenues à l'étape m :

- Version 1 :

$$\left\{\begin{array}{lcl} q_j\sigma_j^{2(m+1)} & = & \sigma_j^{4(m)}(y - X\beta^{(m)})'\,\Gamma^{-1(m)}\,V_j\,\Gamma^{-1(m)}\,(y - X\beta^{(m)}) \\ & & +q_j\sigma_j^{2(m)} - \sigma_j^{4(m)}\mathrm{tr}(\Gamma^{-1(m)}\,V_j) \\ X\beta^{(m+1)} & = & X(X'V_0^{-1}X)^{-1}X'V_0^{-1}\,(X\beta^{(m)} + \sigma_0^{2(m)}V_0\,V^{-1(m)}\,(y - X\beta^{(m)})) \end{array}\right.$$

- Version 2 :

$$\left\{\begin{array}{lcl} q_j\sigma_j^{2(m+1)} & = & \sigma_j^{4(m)}y'\,P^{(m)}\,V_j\,P^{(m)}\,y + q_j\sigma_j^{2(m)} - \sigma_j^{4(m)}\mathrm{tr}(\Gamma^{-1(m)}\,V_j) \\ \multicolumn{3}{l}{\text{à la convergence :} \quad X\hat{\beta} = X\,(X'\,\hat{\Gamma}^{-1}\,X)^{-1}\,X'\,\hat{\Gamma}^{-1}\,y} \end{array}\right.$$

En considérant la situation de convergence où $\hat{\beta} = \beta^{(m+1)} = \beta^{(m)}$ et $\hat{\sigma}^2 = \sigma^{2(m+1)} = \sigma^{2(m)}$, on peut montrer que l'on retrouve les équations du système (2.4). L'algorithme EM fournit bien des solutions aux équations d'estimation du maximum de vraisemblance.

Cet algorithme constitue donc une alternative à la résolution du système (2.4). Pourtant, certaines difficultés persistent. D'une part, cet algorithme peut s'avérer très lent. D'autre part, même si on est sûr de faire croître pas à pas la valeur de la fonction de vraisemblance et de rester dans l'espace des paramètres, cet algorithme peut en pratique resté coincé en un maximum local de la fonction. Cela constitue un problème important. C'est pourquoi, des auteurs ont proposé différentes extensions à cet algorithme, qui tentent de contourner cette difficulté en lui permettant notamment des sauts aléatoires (cf. McLachlan et Krishnan 1997). Il existe aussi des conditions théoriques sur la fonction de vraisemblance pour assurer la convergence vers le maximum global mais qui sont difficiles à vérifier en pratique.

2.3.3.2 Algorithme EM pour REML

Il s'agit donc de reprendre l'algorithme EM pour ML mais cette fois-ci dans le modèle projeté. Mettant de coté l'estimation de β, on utilisera donc la version 2 décrite précédemment. Dans le schéma itératif d'estimation, en utilisant la remarque faite à la fin du 2.3.2.1, on remplace Γ^{-1} par P. On obtient ainsi l'algorithme EM adapté au maximum de vraisemblance restreint. Cet algorithme à la convergence atteint les solutions du système (2.9) sous les mêmes réserves que précédemment.

2.3.3.3 Adaptation au nsL2M

Certaines composantes sont maintenant supposées connues. Nous reprenons le modèle décrit au 2.3.1.2 où la variance se décompose en L composantes inconnues regroupées dans le vecteur $\sigma^{*2} = (\sigma_1^2, ..., \sigma_L^2)'$, et une matrice V_ψ connue.
Dans cette situation, adaptant la méthodologie EM, nous envisageons de compléter le vecteur des observations par les vecteurs d'effets aléatoires ξ_j correspondant aux composantes inconnues uniquement. Nous réduisons ainsi le vecteur des données complètes par rapport à la section précédente. Et la log-vraisemblance complétée s'écrit :

$$\begin{aligned} l(\beta, \sigma^{*2}; y) = \; & -\tfrac{1}{2}(N + \textstyle\sum_{j=1}^L q_j) \ln(2\pi) - \tfrac{1}{2}[\sum_{j=1}^L (q_j \ln(\sigma_j^2) + \ln(|A_j|)) + \ln(|V_\psi|)] \\ & -\tfrac{1}{2} \textstyle\sum_{j=1}^L \dfrac{\xi_j' A_j^{-1} \xi_j}{\sigma_j^2} - \tfrac{1}{2} \xi_0' V_\psi^{-1} \xi_0 \end{aligned}$$

avec $\xi_0 = \psi = y - X\beta - \sum_{j=1}^L U_j \xi_j$.

En dérivant cette log-vraisemblance par rapport à β et aux différentes composantes de la variance inconnues σ_j^2 , $j \in \{1, ..., L\}$, on obtient les estimations du maximum de vraisemblance pour les données complètes :

$$\begin{aligned} & (X'V_\psi^{-1}X)\hat{\beta} = X'V_\psi^{-1}(y - \textstyle\sum_{j=1}^L U_j \xi_j) \, , \\ \forall j \in \{1, ..., L\} \, , \quad & q_j \hat{\sigma}_j^2 = \xi_j' A_j^{-1} \xi_j \, . \end{aligned}$$

L'étape suivante consiste à en calculer les espérances conditionnelles aux données observées y. Elles s'expriment ici de façon similaire à la section précédente :

$$\begin{aligned} & \forall j \in \{1, ..., L\} \, , \\ & \quad E(\xi_j' A_j^{-1} \xi_j | y) = \sigma_j^4 (y - X\beta)' \Gamma^{-1} V_j \Gamma^{-1} (y - X\beta) + q_j \sigma_j^2 - \sigma_j^4 \mathrm{tr}(\Gamma^{-1} V_j) \, , \end{aligned}$$

$$\begin{aligned} E(Y - \textstyle\sum_{j=1}^L U_j \xi_j | y) &= y - \textstyle\sum_{j=1}^L \sigma_j^2 V_j \Gamma^{-1} (y - X\beta) \\ &= y - (\Gamma - V_\psi) \Gamma^{-1} (y - X\beta) \\ &= X\beta + V_\psi \Gamma^{-1} (y - X\beta) \, . \end{aligned}$$

On obtient alors l'algorithme :

$$\begin{aligned} \forall j \in \{1, ..., L\} \, , \quad q_j \sigma_j^{2(m+1)} &= \sigma_j^{4(m)} (y - X\beta^{(m)})' \Gamma^{-1(m)} V_j \Gamma^{-1(m)} (y - X\beta^{(m)}) \\ & \quad + q_j \sigma_j^{2(m)} - \sigma_j^{4(m)} \mathrm{tr}(\Gamma^{-1(m)} V_j) \, , \end{aligned}$$

$$(X'V_\psi^{-1}X)\beta^{(m+1)} = X'V_\psi^{-1}(X\beta^{(m)} + V_\psi \Gamma^{-1(m)} (y - X\beta^{(m)})) \, .$$

De la même façon, on adapte très rapidement cet algorithme à l'estimation REML.

Remarquons que dans le nsL2M, le vecteur des résidus ε a été remplacé par Ψ. Et alors que la matrice de variance résiduelle contenait précédemment un paramètre inconnu à estimer, la matrice V_Ψ ici est entièrement connue. Pourtant, précédemment à l'aide de y et de ξ, on réalisait aussi l'estimation de σ_0^2. Ainsi, avec $x = [y', \xi_1', \ldots, \xi_K']'$, il est possible d'estimer $K+1$ composantes, ou plus exactement K composantes et une résiduelle. Cela signifie aussi que lorsqu'une seule composante de la variance est connue σ_0^2, le vecteur des données complètes n'est pas réduit. Il reste le même.
Enfin, comme pour les autres cas, en considérant la situation de convergence où $\hat\beta = \beta^{(m+1)} = \beta^{(m)}$ et $\hat\sigma^2 = \sigma^{2(m+1)} = \sigma^{2(m)}$, ces valeurs sont solutions des équations de MV du nsL2M avec de nouveau les mêmes réserves algorithmiques.

2.3.4 La méthode de Henderson

La dernière méthode d'estimation des composantes de la variance que nous envisageons ici se présente comme un sous-produit de la résolution des équations de Henderson. La démarche n'est pas tout à fait similaire aux précédentes. En effet, Henderson, au cours de ses nombreux travaux, s'est plus particulièrement intéressé à l'estimation du mérite génétique de certains animaux. Lors d'un processus de sélection de taureaux par exemple, on cherche à déterminer, au vu des productions laitières des filles, le père idéal pour la prochaine génération. Pour cela, on est amené à prédire des réalisations non observées d'un effet aléatoire à l'intérieur d'un modèle mixte. Ainsi la prédiction de ξ devient un élément important et indispensable à l'étude. Cette prédiction sera seulement ensuite utilisée pour l'estimation des composantes de la variance.

2.3.4.1 Les équations de Henderson

Plusieurs prédictions de ξ sont envisageables. Celle que nous considérons ici est nommée BLUP (Best Linear Unbiased Predictor). Cette prédiction $\tilde\xi$, fonction linéaire des données sera non biaisée ($E(\tilde\xi) = E(\xi)$) et la meilleure au sens des carrés moyens (pour toute matrice A symétrique, définie, positive $E((\tilde\xi - \xi)'A(\tilde\xi - \xi))$ est minimum).

Nous reprenons le L2M décrit au début de la section 2.3 comprenant à la fois effets fixes (β paramètre) et effets aléatoires (ξ vecteur aléatoire). Dans ce modèle, Henderson, Kempthorne, Searle, et VonKrosig (1959) ont proposé des équations qui permettent d'obtenir simultanément la prédiction BLUP de ξ et l'estimation BLUE (Best Linear Unbiased Estimator) de β (estimation, notée ici $\hat\beta$, équivalente au maximum de vraisemblance sous des hypothèses de normalité adéquates). Pour former ce système d'équations, la distribution jointe de Y et ξ est maximisée en β et ξ. Ainsi après avoir utilisé sa distribution pour construire la fonction de vraisemblance, ξ joue ensuite le rôle de paramètre. Cette distribution jointe s'écrivant :

$$f(y,\xi) = \frac{1}{(2\pi)^{\frac{N+q}{2}}|R|^{\frac{1}{2}}|D|^{\frac{1}{2}}} \exp\{-\frac{1}{2}[(y - X\beta - U\xi)'R^{-1}(y - X\beta - U\xi) + \xi' D^{-1}\xi]\} \ ,$$

on en déduit le système d'équations :

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}U \\ U'R^{-1}X & U'R^{-1}U + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \xi \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ U'R^{-1}y \end{pmatrix} . \tag{2.12}$$

Ces équations sont souvent appelées *équations du modèle mixte* (MME - Mixed Model Equation). Les résoudre nécessite l'inversion des matrices R et D (souvent diagonales) et de la matrice du système (d'ordre p+q). Remarquons que sans la présence de D^{-1} dans la partie inférieure droite de ce système, il correspondrait aux équations du maximum de vraisemblance lorsque l'on traite ξ comme un effet fixe. Donc par l'introduction de D^{-1}, on prend en compte en partie la nature aléatoire de ξ.
Ce système est équivalent à :

$$\left\{ \begin{array}{rcl} X'\Gamma^{-1}X\,\beta & = & X'\Gamma^{-1}\,y \\ \xi & = & DU'\Gamma^{-1}\,(y - X\beta) \ = \ E(\xi|y) \ . \end{array} \right. \tag{2.13}$$

Puisque $\hat{\beta}$ et $\tilde{\xi}$ sont solutions de (2.12), elles le sont donc aussi de (3.2). Cependant le système (2.13) nécessite l'inversion de Γ non diagonale et d'ordre N souvent plus grand que $p + q$.
Ainsi, pour obtenir l'estimation BLUE de β et la prédiction BLUP de ξ, les équations de Henderson constituent une alternative à l'inversion de Γ et à la résolution directe de (2.13).

À l'aide de $\hat{\beta}$ et $\tilde{\xi}$, il nous reste maintenant à estimer les composantes de la variance.

2.3.4.2 Estimation ML et REML par Henderson

Dans le système (2.12), les matrices R et D dépendent respectivement des valeurs σ_0^2 et $\sigma_1^2, ..., \sigma_K^2$ toutes inconnues. L'estimation de ces composantes est donc une nécessité. Pour cela, les valeurs de ξ et β obtenues par résolution du système de Henderson, vont permettre de calculer les estimations ML et REML dans un schéma itératif (Harville 1977). C'est ce qui constitue un intérêt supplémentaire aux équations (2.12). En effet, à partir des systèmes (2.4) pour ML et (2.9) pour REML, on peut construire les procédures itératives suivantes :

ML

$$\left[\begin{array}{lcl} \sigma_j^{2(m+1)} & = & \dfrac{\xi_j'^{(m)}\, A_j^{-1}\, \xi_j^{(m)}}{q_j - \mathrm{tr}(P_{jj}^{(m)})} \qquad , \quad j = 1, \ldots, K \\ \sigma_0^{2(m+1)} & = & \dfrac{y'\, V_0^{-1}\, (y - X\beta^{(m)} - U\xi^{(m)})}{N} \end{array}\right. \tag{2.14}$$

où $P = (I + U'R^{-1}UD)^{-1}$
P_{jj} : $j^{\text{ème}}$sous matrice de P.

REML

$$\left[\begin{array}{lcl} \sigma_j^{2(m+1)} & = & \dfrac{\xi_j'^{(m)}\, A_j^{-1}\, \xi_j^{(m)}}{q_j - \mathrm{tr}(Q_{jj}^{(m)})} \qquad , \quad j = 1, \ldots, K \\ \sigma_0^{2(m+1)} & = & \dfrac{y'\, V_0^{-1}\, (y - X\beta^{(m)} - U\xi^{(m)})}{N - \mathrm{rg}(X)} \end{array}\right. \tag{2.15}$$

où $Q = (I + U'SUD)^{-1}$
Q_{jj} : $j^{\text{ème}}$sous matrice de Q,
et $S = R^{-1}(I - X(X'R^{-1}X)^{-1}X'R^{-1})$

Ainsi la procédure d'estimation alterne entre :
- pour des valeurs de σ_j^2 (fixant les valeurs de R et D), la résolution de (2.12), et
- pour des valeurs de β et ξ, la résolution de (2.14) ou (2.15).

Notons que le système d'équations de Henderson peut être légèrement transformé en posant $\bar{\xi} = D\xi$ de façon à éviter les problèmes numériques liés à l'inversion de D dans le cas d'estimation de σ_j^2 très petits.

De façon équivalente, on peut aussi utiliser les schémas itératifs suivants :

ML

$$\left[\begin{array}{lcl} \sigma_j^{2(m+1)} & = & \dfrac{\xi_j'^{(m)}\, A_j^{-1}\, \xi_j^{(m)}}{q_j - \frac{\mathrm{tr}(C_{jj}^{*(m)})}{\sigma_j^{2(m)}}} \\ \sigma_0^{2(m+1)} & = & \dfrac{(y - X\beta^{(m)} - U\xi^{(m)})'\, V_0^{-1}\, (y - X\beta^{(m)} - U\xi^{(m)})}{N - \sum_{j=1}^{K}(q_j - \frac{\mathrm{tr}(A_j^{-1}C_{jj}^{*(m)})}{\sigma_j^{2(m)}})} \end{array}\right. \tag{2.16}$$

où C^* : est l'inverse de la matrice formée par les q dernières lignes et colonnes de la matrice des coefficients du système de Henderson (2.12),
C_{jj}^* : $j^{\text{ème}}$ sous matrice de C^*, correspondant au $j^{\text{ème}}$ effet aléatoire.

REML

$$\left[\begin{array}{lcl} \sigma_j^{2(m+1)} & = & \dfrac{\xi_j'^{(m)}\, A_j^{-1}\, \xi_j^{(m)}}{q_j - \frac{\mathrm{tr}(C_{jj}^{(m)})}{\sigma_j^{2(m)}}} \\ \sigma_0^{2(m+1)} & = & \dfrac{(y - X\beta^{(m)} - U\xi^{(m)})'\, V_0^{-1}\, (y - X\beta^{(m)} - U\xi^{(m)})}{N - \mathrm{rg}(X) - \sum_{j=1}^{K}(q_j - \frac{\mathrm{tr}(A_j^{-1}C_{jj})}{\sigma_j^{2(m)}})} \end{array}\right. \tag{2.17}$$

où C : est la matrice formée des q dernières lignes et colonnes de l'inverse de la matrice des coefficients du système de Henderson (2.12),
C_{jj} : $j^{ème}$sous matrice de C, correspondant au $j^{\text{ème}}$ effet aléatoire.

Ce schéma itératif peut s'avérer plus utile d'un point de vue pratique. En effet, lors de la résolution du système (2.12), les matrices C et C^* s'obtiennent facilement. Il est notamment utilisé par SCHALL (1991).

2.3.4.3 Adaptation au nsL2M

Il peut être pour cette méthode plus délicat d'envisager une adaptation au nsL2M. En effet, la prédiction de ξ étant un point central, la question se pose de savoir si l'on désire obtenir des prédictions des effets aléatoires dont les composantes sont connues. En supposant que non, la démarche est alors identique à celle de l'algorithme EM. Le vecteur ξ intervenant dans le système est réduit aux ξ_j dont les composantes sont inconnues. Les matrices U et D sont naturellement adaptées à ce nouveau vecteur ξ^*. On s'intéresse donc à la vraisemblance jointe de (y, ξ^*). V_ψ remplace R et le système s'écrit :

$$\begin{pmatrix} X'V_\psi^{-1}X & X'V_\psi^{-1}U \\ U'V_\psi^{-1}X & U'V_\psi^{-1}U + D^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \tilde{\xi}^* \end{pmatrix} = \begin{pmatrix} X'V_\psi^{-1}y \\ U'V_\psi^{-1}y \end{pmatrix}.$$

Ensuite, lorsqu'il s'agit d'obtenir les estimations ML et REML de $\sigma^{*2} = (\sigma_1^2, ..., \sigma_L^2)'$, on reprend les systèmes (2.14) et (2.15), en les réduisant aux L composantes de la variance inconnues :

$$\textbf{ML :} \quad \forall j \in \{1, ..., L\}\ ,\ \ \sigma_j^{2(m+1)} = \frac{\xi_j'^{*(m)}\, A_j^{-1}\, \xi_j^{*(m)}}{q_j - \mathrm{tr}(P_{jj}^{(m)})}\ ,$$

$$\textbf{REML :} \quad \forall j \in \{1, ..., L\}\ ,\ \ \sigma_j^{2(m+1)} = \frac{\xi_j'^{*(m)}\, A_j^{-1}\, \xi_j^{*(m)}}{q_j - \mathrm{tr}(Q_{jj}^{(m)})}\ .$$

La matrice V_ψ étant connue, il n'y a plus de variance résiduelle (comme σ_0^2 précédemment) à estimer.

2.3.5 Propriétés asymptotiques

Lorsqu'on s'intéresse aux propriétés asymptotiques des estimateurs des composantes de la variance, il faut être vigilant vis-à-vis de la notion de grands échantillons. En effet, à cause de la présence des effets aléatoires, les questions de limite doivent être manipulées avec précaution : $N \to +\infty$ ne dit pas comment ces données supplémentaires se répartissent (nombre de niveau des facteurs, nombre de réalisations des effets aléatoires ...). Bien souvent d'ailleurs, lorsque $N \to +\infty$, p et q augmentent aussi. Ici, on se place à nombre de réalisations des effets fixé.

Sous des conditions de régularité concernant la fonction de vraisemblance, les estimateurs de maximum de vraisemblance et maximum de vraisemblance restreint, possèdent des propriétés de convergence presque sûre et de normalité asymptotique. L'énoncé de ces résultats, les différentes conditions nécessaires et les démonstrations, pourront être trouvés dans Rao et Kleffe (1988), ou de façon plus spécifique au cas de l'estimateur ML dans Sweeting (1980) et de l'estimateur REML dans Cressie et Lahiri (1993) (selon les conditions adoptées, différentes nuances peuvent être apportées à ces résultat). On donne ici uniquement, dans les deux cas, les matrices de variance asymptotiques, qui sont :

ML :

$$\begin{aligned}
\mathrm{Var}\begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{bmatrix} &\simeq \left(I \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} (X'\Gamma^{-1}X)^{-1} & 0 \\ 0 & 2[\{\mathrm{tr}(\Gamma^{-1}V_i\Gamma^{-1}V_j)\}_{i,j=1,\dots,K}]^{-1} \end{bmatrix}
\end{aligned}$$

REML :

$$\begin{aligned}
\mathrm{Var}\begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{bmatrix} &\simeq \left(I \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} (X'\Gamma^{-1}X)^{-1} & 0 \\ 0 & 2[\{\mathrm{tr}(PV_iPV_j)\}_{i,j=1,\dots,K}]^{-1} \end{bmatrix}
\end{aligned}$$

2.4 Estimation dans les GL2M

Après ce retour sur l'estimation dans les GLM en section 2.2 et dans les L2M (ou nsL2M) en section 2.3, il est tout naturel maintenant de nous intéresser à la question de l'estimation dans les GL2M. Pour cela, nous reprenons les hypothèses du GL2M décrites au 1.4.1. C'est plus particulièrement sur l'estimation des composantes de la variance que nous portons notre attention dans toute cette section.

Nous rappelons qu'un GL2M peut se résumer brièvement par :

- conditionnellement à ξ, les Y_i sont indépendantes et de loi appartenant à la famille exponentielle,
- le prédicteur linéaire contient effets fixes et effets aléatoires : $\eta_\xi = X\beta + U\xi$, et l'on maintient l'hypothèse de normalité sur ξ faite au 1.3.2.
- on relie l'espérance mathématique conditionnelle $\mu_\xi = E(Y|\xi)$ au prédicteur linéaire par une fonction de lien g : $\eta_\xi = g(\mu_\xi)$.

On dispose d'observations y_i des Y_i mais on n'observe pas les effets aléatoires ξ réalisés au cours de l'expérience.

Dans cette section, nous présentons une méthodologie simple d'estimation tirant profit de la double nature de ces modèles et utilisant à chaque étape les outils propres aux GLM puis aux L2M. Nous reviendrons ensuite sur d'autres travaux déjà réalisés par divers auteurs. Et nous tenterons de comparer les différentes approches sur une échelle que nous appelons *échelle de déconditionnement*.

2.4.1 Méthode d'estimation proposée

La méthode que nous proposons se décrit en deux étapes. L'une s'inspire du fait que le GL2M peut être considéré comme l'extension d'un GLM, l'autre le regarde davantage comme l'extension d'un L2M. Ainsi, la première étape consiste en une linéarisation conditionnelle à ξ et analogue à celle utilisée dans les GLM. Le modèle alors obtenu est un nsL2M. La deuxième étape procède ensuite à l'estimation dans ce modèle linéarisé. Cette méthode découle donc tout naturellement de ce qui a été présenté aux sections précédentes.

2.4.1.1 Étape de linéarisation

Dans un premier temps, on regarde donc davantage le GL2M comme extension d'un GLM. Pour cela, on se place conditionnellement à ξ. Si ξ était un paramètre fixé, le modèle considéré serait alors un GLM que l'on pourrait écrire :

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

avec les hypothèses de lois conditionnelles adéquates. Sachant ξ, en reprenant alors la démarche décrite au paragraphe 2.2.1 pour les GLM, on introduit la variable dépendante $z_{\beta,\xi} = X\beta + U\xi + (y - \mu_\xi)g'(\mu_\xi)$. Et on est ensuite amené à considérer le *modèle linéarisé* que l'on note $\mathcal{M}_\xi$:

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

où
$$\begin{array}{lllll} E(Z_{\beta,\xi}|\xi) & = & X\beta + U\xi & & \\ \mathrm{Var}(Z_{\beta,\xi}|\xi) & = & \mathrm{Var}(e|\xi) & = & W_{\beta,\xi} \ . \end{array}$$

En se plaçant conditionnellement à ξ, on a fait perdre momentanément à l'effet aléatoire sa nature aléatoire. Considéré comme une paramètre, on pourrait alors l'estimer par moindres carrés généralisés itérés selon la technique des GLM. Alors, de même qu'au 2.2.1, dans ce modèle $\mathcal{M}_\xi$, la variable dépendante $Z_{\beta,\xi}$ n'est pas observée mais calculée pour les valeurs courantes de β et ξ par :

$$\begin{array}{lccccc} z_{\beta,\xi} & = & g(\mu_\xi) & + & (y - \mu_\xi)g'(\mu_\xi) & \approx \ g(y) \\ & = & \eta_\xi & + & \varepsilon \ g'(\mu_\xi) & \\ & = & X\beta + U\xi & + & \varepsilon \ g'(\mu_\xi) \ , & \end{array}$$

pouvant toujours être vu comme un développement limité au premier ordre de g en μ_ξ. La procédure de moindres carrés itérés utiliserait alors la matrice des poids ou matrice de variance du vecteur des erreurs conditionnelle à ξ :

$$\begin{array}{lll} W_{\beta,\xi} & = & \mathrm{Var}(\varepsilon g'(\mu_\xi)|\xi) \\ & = & \{_d \ g'(\mu_{\xi,i})^2 \mathrm{var}(\varepsilon_i|\xi) \ \} \ . \end{array}$$

Ce qui dans le cas d'un lien canonique n'est autre que $W_{\beta,\xi} = \{_d \ a(\phi)g'(\mu_{\xi,i}) \ \}$. Ainsi, on nommera plus précisément $\mathcal{M}_\xi$ *modèle linéarisé conditionnel*. Dans le tableau 2.1, on trouvera l'expression de cette matrice de variance, pour les lois classiques (mentionnées au premier chapitre) de la famille exponentielle et en considérant le lien canonique associé.

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.

2.4.1.2 Étape d'estimation

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

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

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

TAB. 2.1 – *Matrice de variance du modèle linéarisé conditionnel associé aux lois usuelles de la famille exponentielle.*

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

$$\begin{array}{lcl} E(Z_{\beta,\xi}|\xi) & = & X\beta + U\xi \\ \mathrm{Var}(Z_{\beta,\xi}|\xi) & = & W_{\beta,\xi} \end{array}$$

et donc

$$\begin{array}{lclcl} E(Z_{\beta,\xi}) & = & X\beta & & \\ \mathrm{Var}(Z_{\beta,\xi}) & = & UDU' & + & E(W_{\beta,\xi}) \\ & = & UDU' & + & W_\beta \\ & = & \Gamma\,. & & \end{array}$$

La matrice de variance des erreurs de ce modèle linéaire mixte est donc $W_\beta = E(W_{\beta,\xi}) = E(\{_d \ g'(\mu_{\xi,i})^2 \mathrm{var}(\varepsilon_i|\xi) \ \})$. Ce qui, dans le cas d'un lien canonique aboutit à $W_\beta = E(\{_d \ a(\phi)g'(\mu_{\xi,i}) \ \})$. 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. Nous décrivons l'algorithme dans la section suivante. À chaque itération, les valeurs de $z_{\beta,\xi}$ et W_β sont calculées et on estime les paramètres du modèle alors défini.
Cependant, il est important d'observer aussi que la matrice W_β n'introduit pas de nouvelle composante inconnue à la variance, excepté pour le cas de la loi normale, dont l'étude a déja été envisagée au 2.3, ainsi que le cas de la loi gamma où le paramètre a inconnu intervient. Dans les autres cas (où le paramètre ϕ est connu),

pour des valeurs fixées de β et σ^2, la matrice est totalement déterminée. Le modèle $\mathcal{M}_\xi$ s'écrit donc sous la forme d'un nsL2M à K composantes inconnues où $\psi = e$ et $V_\psi = W_\beta$. Pour procéder à l'estimation des paramètres dans ce modèle, on utilise alors les méthodes d'estimation adaptées au nsL2M. Le système REML adapté s'écrit :

$$\left\{ \left(\operatorname{tr}(PV_iPV_j) \right)_{i,j=1,\ldots,K} \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_K^2 \end{pmatrix} = \Big(z'_{\beta,\xi}\, P\, V_i\, P\, z_{\beta,\xi} - \operatorname{tr}(PV_iPW_\beta) \Big)_{i=1,\ldots,K} \right. \tag{2.18}$$

que l'on pourra résoudre soit de manière directe, soit par l'algorithme EM ou encore par les méthodes de Henderson. Dans tous les cas, on utilisera donc les versions adaptées au nsL2M.
L'estimation de β est alors donnée par :

$$\hat{\beta} = (X'\hat{\Gamma}^{-1}X)^{-1}X'\hat{\Gamma}^{-1}z_{\beta,\xi} \tag{2.19}$$

Dans ses travaux, Schall (1991) utilise la structure de variance : $\Gamma_\xi = UDU' + W_{\beta,\xi}$ du modèle linéarisé conditionnel. Ici, on remplace $W_{\beta,\xi}$ par $W_\beta = E(W_{\beta,\xi}) = E(\{_d\ g'(\mu_{\xi,i})^2 \operatorname{var}(\varepsilon_i|\xi)\ \})$. On relâche donc un peu le conditionnement. Nous approfondirons ce point à la section 2.4.2.
Dans le cas lien canonique et sous l'hypothèse $\xi \sim \mathcal{N}(0, D)$ avec D diagonale par blocs : $D = \{_d\ \sigma_j^2 A_j\ \}$, ceci conduit au calcul de $E(\{_d\ \frac{\phi}{\omega_i}\ g'(\mu_{\xi,i})\ \})$. Ce qui donne les résultats exhibés dans le tableau 2.2 pour les lois classiques de la famille exponentielle.

Malheureusement, ce calcul de $E(W_{\beta,\xi})$ n'est pas toujours réalisable analytiquement. Nous voyons dans le tableau 2.2 qu'il ne pose pas de problème particulier dans le cas de lien canonique. Il est aussi possible de calculer cette matrice W_β 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 et nous y reviendrons que dans le cas de la loi normale (lien identité), on a $W_{\beta,\xi} = W_\beta = \sigma_0^2\, I_N$.

Ainsi, il apparaît donc possible de tirer partie de l'information supplémentaire apportée dans les GL2M par la fonction de variance. Elle implique en effet ici que dans le modèle linéarisé la matrice de variance des résidus est connue et donc une composante de la variance est connue. Cependant, notons tout de même que, si jamais le paramètre ϕ est inconnu, il nous est aussi laissé la possibilité de l'estimer au même titre que les autres composantes de la variance.

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

TAB. 2.2 – *Matrice de variance du modèle linéarisé marginal associé aux lois usuelles de la famille exponentielle.*

[a] On note U_{ji} la $i^{ème}$ ligne de la matrice U_j.

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

2.4.1.3 L'algorithme

L'algorithme va bien entendu suivre les deux étapes de linéarisation puis estimation. Dans le système (2.18), la dépendance de P, W_β et $z_{\beta,\xi}$ aux valeurs courantes des paramètres suggère immédiatement un algorithme itératif.
Soient $\beta^{[t]}$ et $\sigma^{2[t]} = (\sigma_1^{2[t]}, ..., \sigma_K^{2[t]})'$, les valeurs obtenues à l'étape t, l'algorithme effectue alors les pas suivants :

<u>Pas 1</u> : réactualisation des données : on calcule $z^{[t+1]}$,
<u>Pas 2</u> : calcul de $W_\beta^{[t]}, \Gamma^{[t]}$ et $P^{[t]}$,
le modèle $\mathcal{M}^{[t]} : Z^{[t]} = X\beta + U\xi + e^{[t]}$ est alors défini,
<u>Pas 3</u> : résolution du système (2.18) pour l'obtention de $\sigma^{2[t+1]}$,
<u>Pas 4</u> : à l'aide de (2.19), calcul de $(\beta^{[t+1]})$,
puis incrémentation de t et retour au pas 1.

On itère ce processus jusqu'à la convergence de β et σ^2.
Différentes modifications peuvent être envisagées à cette procédure. Par exemple :

- changer l'ordre des étapes,
- remplacer le pas 3 par : "on itère la résolution de (2.18) jusqu'à convergence de σ^2".

Quoiqu'il en soit, on effectue toujours un va-et-vient incessant entre la constitution du modèle $\mathcal{M}^{[t]}$ (pour les données réactualisées $z^{[t]}$ et avec la matrice de variance $W_\beta^{[t]}$) et la résolution du système REML adapté. À chaque itération, le modèle linéarisé est donc modifié.

Cependant, cet algorithme n'est pas encore complet. En effet, l'étape de réactualisation des données implique le calcul de :

$$z^{[t+1]} = X\beta^{[t]} + U\xi^{[t]} + (y - \mu_\xi^{[t]})g'(\mu_\xi^{[t]}) \qquad \text{où} \qquad \mu_\xi^{[t]} = g^{-1}(X\beta^{[t]} + U\xi^{[t]}) \ .$$

Il nécessite donc de donner non seulement des valeurs à $\beta^{[t]}$ mais aussi à $\xi^{[t]}$. Or, si dans l'algorithme précédent, à chaque étape on dispose de nouvelles valeurs de β et σ^2, rien implique que l'on donne aussi des valeurs à ξ.
Dans le cas où le système (2.18) est résolu par l'intermédiaire des équations de Henderson, le problème ne se pose plus puisque l'algorithme fournit naturellement des valeurs pour ξ. Comme mentionné au paragraphe 2.3.4.1, ces valeurs correspondent à $E(\xi|z) : \tilde{\xi} = DU'\Gamma^{-1}(z - X\beta)$ donnant à l'étape t : $\xi^{[t+1]} = DU'\Gamma^{[t]-1}(z^{[t]} - X\beta^{[t]})$.
Sinon, plusieurs démarches sont envisageables :

- dans le modèle $Z = X\beta + U\xi + e$ où $\mathrm{Var}(e) = W_\beta$, considérons β connu et ξ paramètre à estimer. Alors en réécrivant le modèle : $Z - X\beta = U\xi + e$, on aboutit à l'estimation (même si ici des précautions doivent être prises quant à l'emploi du mot *estimation*) usuelle : $\tilde{\xi} = (U'W_\beta^{-1}U)^{-1}U'W_\beta^{-1}(z - X\beta)$,

- un raisonnement similaire peut être mené dans le modèle linéarisé conditionnel à ξ : $Z - X\beta = U\xi + e$ où $\mathrm{Var}(e|\xi) = W_\xi$.
Alors $\tilde{\xi} = (U'W_{\beta,\xi}^{-1}U)^{-1}U'W_{\beta,\xi}^{-1}(z - X\beta)$,

- on reprend, conditionnellement à ξ, le raisonnement menant aux équations de Henderson : $\tilde{\xi} = DU'\Gamma_\xi^{-1}(z - X\beta) \qquad \text{où} \qquad \Gamma_\xi = UDU' + W_{\beta,\xi}$,

- on peut aussi tenter d'approcher le calcul de $E(\xi|y)$, où y sont les données d'origine. Mais ce calcul semble délicat.

Notons enfin, à ce sujet, que l'on peut aussi envisager de remplacer les données réactualisées par d'autres valeurs ne nécessitant pas la connaissance de ξ. En effet, nous avons remarqué précédemment que $z_{\beta,\xi}$ apparaissait comme un développement limité de g en μ_ξ. Ainsi, une valeur approchée de $z_{\beta,\xi}$ peut être $g(y)$.
Enfin, en affinant ce raisonnement, on peut apporter une correction correspondante au calcul de l'espérance du premier terme négligé dans le développement limité : $E\left(\dfrac{(Y - \mu_\xi)^2}{2} g''(\mu_\xi)\right)$.

Autant de façon de résoudre le problème de la réactualisation. Nous garderons dans notre implémentation la résolution par les équations de Henderson puisque cette méthode semble être rapide et source de moindres problèmes numériques.

2.4.1.4 Des résultats de simulation

Nous procédons à quelques simulations dans le cas des lois binomiales, Poisson et exponentielle. Les plans d'expériences utilisés sont non équilibrés pour les effets fixes comme pour les effets aléatoires. On inclut 4 réalisations de l'effet aléatoire et 3 niveaux d'effet fixe. Le vecteur des données simulées est de taille 36. On simule 200 jeux de données et l'on trouve dans les trois tableaux ci-après les résultats des moyennes et écarts-types des estimations obtenues pour différentes valeurs de σ^2, la valeur de β ayant été fixée en simulation à $\beta = (0, -1, 2)'$.

- **Cas binomial - lien logit**

Tableau 2.3 suivant.

- **Cas Poisson - lien log**

Tableau 2.4 suivant.

- **Cas exponentiel - lien log**

Tableau 2.5 suivant.

On constate dans ces trois situations un bon comportement général de cette méthode, que ce soit pour l'estimation des paramètres d'effets fixes ou pour celle des paramètres de variance. Bien entendu, plus la valeur simulée de la composante de la variance est importante, plus la précision de ces estimations diminue en ce sens où les écarts-types augmentent. On peut noter une légère surestimation de la variance dans le cas exponentiel et en particulier pour une valeur faible de $\sigma^2 = 0.05$. Dans ce cas précis, c'est pour une distribution binomiale que les estimations s'avèrent les meilleures avec les écarts-types plus faibles.

	Valeurs estimées				
Valeurs simulées		$\hat{\sigma}^2$	$\hat{\beta}$		
$\sigma^2 = 0.05$	moy.	0.0543	-0.0194	-1.0215	2.0001
	e.t.	0.0487	0.1166	0.1316	0.1567
$\sigma^2 = 0.5$	moy.	0.4763	0.0017	-0.9956	2.0177
	e.t.	0.4214	0.3627	0.3680	0.3775
$\sigma^2 = 1$	moy.	1.0505	-0.0376	-1.0463	1.9499
	e.t.	0.9429	0.4740	0.4803	0.4945
$\sigma^2 = 1.5$	moy.	1.5376	-0.0428	-1.0513	1.9666
	e.t.	1.2639	0.6320	0.6450	0.6479
$\sigma^2 = 2$	moy.	1.9848	-0.0061	-1.0234	2.0157
	e.t.	1.7356	0.7341	0.7501	0.7478

TAB. 2.3 – *Résultats d'estimation par la méthode proposée sur 200 simulations : modèle binomial - lien logit.*

	Valeurs estimées				
Valeurs simulées		$\hat{\sigma}^2$	$\hat{\beta}$		
$\sigma^2 = 0.05$	moy.	0.0931	-0.0517	-1.1033	1.9833
	e.t.	0.0991	0.2887	0.5986	0.1841
$\sigma^2 = 0.5$	moy.	0.6062	-0.0647	-1.1207	1.9502
	e.t.	0.6435	0.4888	0.6983	0.4421
$\sigma^2 = 1$	moy.	1.0675	-0.0907	-1.1382	1.9494
	e.t.	1.0052	0.6203	0.8268	0.5901
$\sigma^2 = 1.5$	moy.	1.6859	-0.0957	-1.1855	1.9492
	e.t.	1.4196	0.6859	0.8892	0.6396
$\sigma^2 = 2$	moy.	1.8989	0.0948	-0.9267	2.1591
	e.t.	1.4093	0.7672	0.8940	0.7088

TAB. 2.4 – *Résultats d'estimation par la méthode proposée sur 200 simulations : modèle Poisson - lien log.*

	Valeurs estimées				
Valeurs simulées		$\hat{\sigma}^2$	$\hat{\beta}$		
$\sigma^2 = 0.05$	moy.	0.1363	-0.0562	-1.0633	1.9509
	e.t.	0.1289	0.2640	0.3314	0.3641
$\sigma^2 = 0.5$	moy.	0.6049	-0.0856	-1.1343	1.8952
	e.t.	0.5310	0.4497	0.4499	0.4533
$\sigma^2 = 1$	moy.	1.0928	-0.1228	-1.1205	1.8266
	e.t.	0.9353	0.5716	0.5493	0.6410
$\sigma^2 = 1.5$	moy.	1.5282	-0.1300	-1.1316	1.8292
	e.t.	1.2857	0.6660	0.6800	0.7140
$\sigma^2 = 2$	moy.	2.0458	-0.1014	-1.0839	1.8535
	e.t.	1.7155	0.7009	0.6877	0.7984

TAB. 2.5 – *Résultats d'estimation par la méthode proposée sur 200 simulations : modèle exponentiel - lien log.*

2.4.2 D'autres démarches pour l'estimation

Nous donnons un aperçu des travaux déjà réalisés avant de nous arrêter sur deux de ces méthodes : Schall et GAR (du nom de leurs auteurs). Nous tenterons ensuite de les comparer à la méthode que nous venons de proposer.

2.4.2.1 Un aperçu des travaux déjà réalisés

La question de l'estimation des effets fixes et des composantes de la variance dans les GL2M a été considérée par de nombreux auteurs dans des travaux très dispersés. Elle continue de susciter beaucoup d'intérêt et de faire l'objet de nombreuses publications. D'une part, peut être, parce que l'intérêt pratique de ce type de modélisation est grandissant. D'autre part, semble-t-il, il n'est pas une méthode, s'il devait en exister une, qui fasse réellement l'unanimité. À notre connaissance, aucun ouvrage à ce jour n'a permis d'unifier et de comparer les différents travaux réalisés et leurs enjeux.
C'est pourquoi il semble difficile de faire un descriptif global et général de ces travaux. Et nous n'en donnons ici qu'un bref aperçu.

Parmi ces travaux, il en est un certain nombre qui ne concernent souvent qu'un cas particulier de loi au sein de la famille exponentielle (souvent le cas de la loi binomiale pour des données binaires) ou une modélisation particulière des effets aléatoires (surdispersion ou effets aléatoires emboîtés par exemple). Moins nombreux sont ceux qui s'intéressent à ces modèles de façon générale. Citons tout de même Schall (1991), Breslow et Clayton (1993) et plus récemment encore McGilchrist (1994) ou Engel et Keen (1994). Et c'est plus dans cette optique générale que nous nous plaçons ici.

Comme nous l'avons déja évoqué à la section 1.4.1, un GL2M est correctement défini conditionnellement aux effets aléatoires ξ. C'est ce qui constitue l'obstacle principal à la mise en place de procédures d'estimation dans la mesure où ces effets aléatoires se réalisent au cours de l'expérience et ne sont pas observés. Cet obstacle est d'autant plus important que l'on cherche à estimer des paramètres (les composantes de la variance) de leur distribution. Les approches adoptées par les divers auteurs peuvent alors se différencier selon leur façon de lever ce conditionnement, ou encore par leur degré de déconditionnement comme nous le verrons plus particulièrement au 2.4.2.4.

Une démarche de déconditionnement complet consiste en l'obtention de la fonction de vraisemblance marginale et en sa maximisation. Puisque l'on ne connait que la loi des observations conditionnellement aux effets aléatoires, la fonction de log-vraisemblance des paramètres β et σ^2 s'obtient par intégration :

$$L(\beta, \sigma^2; y_1, .., y_n) = \int_{\mathbb{R}^q} \prod_{i=1}^{n} f(y_i|\xi) f(\xi). \tag{2.20}$$

Ce calcul d'intégrale multiple est numériquement très exigeant. Et malgré le développement des capacités informatiques, il est difficilement pratiquable en toute généralité. Nous écartons ici les modélisations ne correspondant pas aux hypothèses décrites au chapitre 1, en particulier celles où la distribution des effets aléatoires est choisie conjuguée à la distribution conditionnelle de Y de sorte que ce calcul intégral ne pose pas de problème. De façon générale, ce calcul est donc peu envisageable. Dans certains cas cependant (surdispersion, effets emboités), cette intégrale multiple peut être scindée et on peut alors en envisager une approximation numérique telle que la quadrature gaussienne par exemple. Cette démarche a été adoptée notamment par Anderson et Aitkin (1985) pour des données binaires.
En dehors de ces cas particuliers, d'autres approximations de la vraisemblance marginale (2.20) ont été pratiquées : approximation de Solomon et Cox (1992), approximation de Laplace pour la définition d'une quasi-vraisemblance pénalisée (Breslow et Clayton 1993). Cette dernière démarche ne correspond pas réellement à un déconditionnement au même titre que celui de l'approximation par quadrature gaussienne. En effet, plutôt que de faire disparaître les effets aléatoires, la quasi-vraisemblance pénalisée en définitive rajoute un terme les concernant.

Puisque la distribution marginale des observations est difficile à atteindre, une autre façon de raisonner sur un modèle marginal associé est la suivante. Sans aller jusqu'au calcul de la loi marginale, on s'arrête au calcul des deux premiers moments marginaux : l'espérance et la variance "déconditionnées". Cela permet ensuite l'utilisation de méthodes de quasi-vraisemblance. C'est ce que font , dans le cas de données binomiales Gilmour, Anderson, et Rae (1985), dans le cas de données poissoniennes Foulley et Im (1993), ou encore Breslow et Clayton (1993) en définissant une quasi-vraisemblance marginale. Nous reviendrons sur cette méthodologie que nous proposons d'étendre à d'autres types de lois et d'autres fonctions de lien au chapitre 3. Mais c'est un calcul qui s'avère aussi délicat dans un cadre très général.

La méthode proposée par Schall (1991), quant à elle, débute par un raisonnement conditionnel pour effectuer une linéarisation du modèle. Ainsi replongé dans un cadre linéaire, le problème du calcul intégral est alors contourné. Nous revenons plus longuement sur cette méthode dans la section suivante. La démarche menée par McGilchrist (1994) s'inscrit aussi davantage dans un raisonnement conditionnel puisque c'est la vraisemblance jointe que l'on cherche à maximiser après approximation. On n'aborde pas directement le déconditionnement. C'est le même type de démarche que celle menée par Henderson dans le cas linéaire (cf. 2.3.4) pour construire ses équations.

D'autres approches encore ont été développées. L'algorithme EM a été utilisé dans des cas particuliers de modélisation des effets aléatoires (Anderson et Hinde 1988) ou de lois (Hinde 1982). Au chapitre 4, nous décrirons plus précisément les limites de cet algorithme dans le cadre des GL2M.
Des raisonnements bayésiens ont été mis en place, notamment pour des données binaires ou à catégories ordonnées (Gianola et Foulley 1983, Harville et Mee 1984, Albert et Chib 1993). Pour cela, le modèle est relu sous un angle bayésien. On suppose une distribution a priori pour β, et les hypothèses sur la distribution de ξ constituent l'a priori le concernant (avec éventuellement un a priori sur les hyper-paramètres que sont les paramètres de sa distribution). S'en suivent alors le calcul des modes a posteriori ou la mise en place de procédure d'échantillonage, tel l'échantillonage de Gibbs, de la distribution a posteriori.

Nous revenons maintenant plus précisément sur les méthodes que nous nommons Schall (Schall 1991) et GAR (Gilmour, Anderson, et Rae 1985). Si nous avons choisi ces deux méthodes c'est parce qu'elles nous semblent illustrer clairement des façons opposées d'aborder le déconditionnement du modèle. Cela nous servira à les comparer à la méthode que nous avons proposée au 2.4.1.

2.4.2.2 La méthode Schall

Schall (1991) propose une méthode d'estimation des paramètres dans un GL2M analysé de manière globale, c'est-à-dire sans spécification particulière de loi ou de modélisation des effets. Sa démarche consiste en une linéarisation du modèle conditionnellement à ξ, puis en l'estimation des paramètres par utilisation des équations de Henderson pour les modèles linéaires mixtes. Finalement, on retrouve les deux étapes, décrites au 2.4.1, de linéarisation et d'estimation et la mise en place d'une procédure itérative. À ce niveau, notre méthode est tout à fait similaire ainsi que celle d'Engel et Keen (1994). Lors de la linéarisation, la variable dépendante est introduite de façon identique. Cependant, le modèle linéarisé adopté par la suite, diffère. En effet, le point de vue de Schall est de se placer dans le modèle que nous avons appelé *modèle linéarisé conditionnel* $\mathcal{M}_\xi : Z = X\beta + U\xi + \varepsilon$ où la matrice de variance des résidus est $W_{\beta,\xi} = \mathrm{Var}(\varepsilon|\xi)$. 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. À aucun moment il ne considère le calcul de $W_\beta = E(W_{\beta,\xi})$, la matrice de variance marginale des résidus. Cela le prive de la prise en compte, dans un nsL2M, d'une éventuelle information détenue sur une partie de la variance.

Ainsi, la différence entre ces deux méthodes réside essentiellement dans le modèle linéarisé considéré : *conditionnel* ($\mathcal{M}_\xi$) ou *marginal* ($\mathcal{M}$).

Avec le choix de $\mathcal{M}_\xi$, Schall est alors amené au cours de son raisonnement à écrire : $\text{Var}(Z) = UDU' + W_{\beta,\xi} = \Gamma_\xi$ (cf. Schall 1991 p. 721). Cette écriture semble peu cohérente puisque les deux termes de cette somme sont de nature différente. Le premier prend en compte la nature aléatoire de ξ (d'où la présence de la matrice D de variance de ce vecteur), alors que le deuxième est conditionnel à ξ. Pourtant, ce raisonnement n'est pas totalement injustifié. En effet, Schall utilise les équations de Henderson pour obtenir les estimations dans le L2M associé. Or, nous l'avons évoqué dans la section 2.3.4.1, la construction de ces équations se base sur la loi du couple (Z, ξ) comme produit de la loi conditionnelle de Z 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 dérivation 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 (2.4) ou REML (2.9), 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 par un modèle gaussien du modèle linéarisé alors que $W_{\beta,\xi} \neq W_\beta$. Ce qui n'est jamais le cas dans un réel L2M.
Cependant, ce choix de $\mathcal{M}_\xi$ lui offre l'avantage de ne pas être limité dans l'utilisation des fonctions de lien. A contrario, nous l'avons vu au 2.4.1.2, le calcul de W_β n'est pas réalisable quel que soit la fonction de lien associé au modèle.
Enfin, en utilisant les équations de Henderson, il dispose à chaque étape de valeurs de ξ lui permettant à la fois de réactualiser les données et de calculer W_ξ. Mais au vu des remarques au 2.4.1.3, la question se pose de savoir si ces valeurs de ξ sont les plus appropriées.

Depuis sa mise en place, cette méthode Schall a reçu divers éclairages. En effet, les 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 Breslow et Clayton (1993). Nous l'avons déjà évoqué précédemment, elle s'appuie sur le fait que les paramètres maximisant la vraisemblance marginale approchée (par approximation de Laplace) se révèlent être ceux qui maximisent la quasi-vraisemblance pénalisée de Green (1987). Or, l'écriture des équations de maximisation en β et ξ de cette quasi-vraisemblance pénalisée, conduit (cf. Breslow et Clayton 1993 p.11) au système (résolu itérativement) de Henderson dans le modèle linéarisé $\mathcal{M}_\xi$ de Schall :

$$\begin{pmatrix} X'W_{\beta,\xi}^{-1}X & X'W_{\beta,\xi}^{-1}U \\ U'W_{\beta,\xi}^{-1}X & U'W_{\beta,\xi}^{-1}U + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \xi \end{pmatrix} = \begin{pmatrix} X'W_{\beta,\xi}^{-1}z \\ U'W_{\beta,\xi}^{-1}z \end{pmatrix}, \tag{2.21}$$

où z est la variable dépendante définie par : $z = X\beta + U\xi + g'(\mu_\xi)(y - \mu_\xi)$. En poursuivant ensuite d'autres approximations, l'estimation des composantes proposée par Breslow et Clayton est identique à l'estimation REML par Henderson dans un modèle gaussien défini sur z.
Cette approximation de Laplace a été reprise dans le cadre plus général des HGLM (Hierarchical Generalized Linear Models qui englobent les GL2M) par Lee et Nelder (1996). Dans ces modèles, ils définissent la *h-vraisemblance* comme la vraisemblance jointe définie à partir de la loi jointe comme produit de la loi de $Y|\xi$ et de celle de 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 Schall (1991), Breslow et Clayton (1993) ou McGilchrist (1994).
Il est important de souligner ici, que ces deux types de travaux se veulent basés sur le fait que la miximisation de la vraisemblance jointe correspond, en β, à la maximisation de l'approximation de Laplace de la vraisemblance marginale. Ce qui tendrait à relier un point de vue conditionnel à un point de vue marginal approché pour β.
Enfin, cette méthode trouve aussi une justification du point de vue bayésien adopté par Stiratelli, Laird, et Ware (1984), qui s'intéressent au mode a posteriori lorsque l'on suppose un a priori non informatif uniforme sur β. En effet, comme l'a suggéré Schall, en notant $p(\beta|G)$ la distribution a priori normale centrée, de variance G du vecteur de paramètre β, et indépendante de $p(\xi|D)$ la distribution normale des effets aléatoires considérée comme distribution a priori, on a alors :

$$\begin{aligned} f(\beta, \xi|Y; G, D) &\propto f(Y|\beta, \xi)\, p(\beta|G)\, p(\xi|D) \\ &\propto f(Y, \xi|\beta; D)\, p(\beta|G)\ . \end{aligned}$$

Ainsi, en prenant un a priori diffus ($G^{-1} \longrightarrow 0$), on a $f(\beta, \xi|Y; G, D) \propto f(Y, \xi|\beta; D)$. La maximisation de la densité a posteriori de β et ξ correspond alors à la maximisation de la vraisemblance jointe de (Y, ξ). Ce raisonnement bayésien aboutit donc encore aux même équations (2.21).

Ainsi, toutes ces démarches s'inscrivent dans un raisonnement conditionnel puisqu'elles conduisent à la maximisation de la vraisemblance jointe de (Y, ξ). Et même si elles se justifient par une approximation de la vraisemblance marginale, cela concerne surtout le paramètre β et moins les composantes de la variance (qui sont estimées dans une étape ultérieure), et elles évitent en tout cas le calcul intégral.

2.4.2.3 La méthode GAR

Gilmour, Anderson, et Rae (1985) présentent une méthode d'estimation dans un modèle à effets aléatoires pour des données binomiales et avec un lien probit (qui n'est pas le lien canonique). Cette méthode concerne donc un modèle spécifié par une loi et une fonction de lien particulières. Cependant, Foulley et Im (1993) l'ont étendue au cas d'une distribution de Poisson et nous proposons au chapitre 3 de l'adapter encore à d'autres cas. Ce qui nous intéresse ici, c'est davantage l'idée qui la fonde, la démarche adoptée (notamment vis-à-vis du conditionnement).

Cette méthode se base, dans un premier temps, sur le calcul des deux premiers moments marginaux : l'espérance et la variance. Ceux-ci permettent ensuite la définition d'une fonction de quasi-vraisemblance, que l'on cherchera à maximiser pour procéder à l'estimation. 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. En effet, après approximation, on reconnaît la forme de la structure de variance d'un L2M. Il est alors possible d'utiliser les méthodes d'estimation propres aux L2M. Et c'est par résolution des équations de Henderson que les auteurs obtiennent finalement leurs estimations. Ainsi, 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 décontionné*) qu'intervient la linéarisation, contrairement à la méthode Schall où la première étape consiste à effectuer cette linéarisation. Une fois l'approximation de la variance marginale faite, l'utilisation des équations de Henderson implique de façon indirecte la réintroduction d'effets aléatoires. On ne connaît pas pour autant leur lien réel avec les effets aléatoires du modèle d'origine, si ce n'est qu'ils possèdent la même distribution et en particulier la même variance. On peut donc s'interroger quant à la pertinence de la prédiction de ξ obtenue.

Breslow et Clayton (1993) ont suivi la même démarche, dans la deuxième méthode qu'ils proposent, en définissant une fonction de *quasi-vraisemblance marginale*. Cependant, l'approximation des moments marginaux, utilisés pour la construction de cette quasi-vraisemblance, est sensiblement différente. Nous revenons de manière plus précise sur ce point et sur la méthode GAR au chapitre suivant.

2.4.2.4 Des lectures différentes du même modèle initial

Nous reprenons quelques unes des remarques faites pour tenter de jeter un regard en parallèle sur ces trois méthodes.

Comme nous venons de le signaler, l'idée première de la méthode GAR est de se libérer du conditionnement pour procéder ensuite à l'estimation dans le modèle décontionné associé. Ce modèle, au niveau marginal, est linéarisé par approximation de la variance. Ainsi, cette méthode s'oppose totalement 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.* Notre méthode apparaît alors comme intermédiaire entre ces deux méthodes puisqu'on relâche ce conditionnement. De même que pour Schall, la linéarisation se fait conditionnellement à ξ, cependant dans le L2M associé $\mathcal{M}$, c'est cette fois-ci avec la matrice de variance déconditionné W_β que nous travaillons.
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 REML 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 REML. 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 notre 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 intervenant ensuite. Nous représentons cette échelle à la figure 2.1.

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.

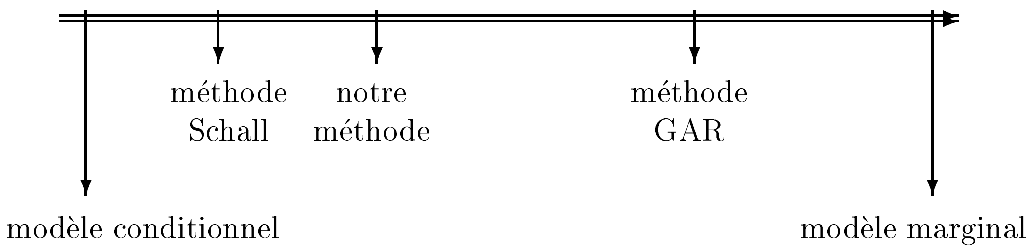

FIG. 2.1 – *Échelle de déconditionnement pour une comparaison des trois méthodes.*

Il est important de constater que notre méthode, 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 (2.6) les différents cas en indiquant :
- les équivalences,
- le cas où, faute de savoir calculer W_β, notre méthode 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).

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

2.4.3 Simulations comparées

Afin de faire tourner en parallèle ces trois méthodes d'estimation, de nombreuses simulations numériques ont été effectuées. Nous avons considéré différentes lois (binomiale, Poisson, exponentielle) et différentes fonctions de lien pour chacune d'elles.

Dans le tableau (2.7), nous présentons les résultats de simulation obtenus dans un modèle avec loi binomial et lien logit. Nous avons pris un plan d'expérience avec un 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 4), 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.

Remarque : la méthode GAR dans ce cas est en fait l'extension (présentée au chapitre suivant) de la méthode d'origine dans le cas du lien logit.

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

		Schall		Notre méthode		GAR
Binomial	probit	×		?		×
	logit	×		×		×
Poisson	logarithme	×		×		×
	identité	×		×	=	×
Exponentielle	logarithme	×	=	×		×
	inverse	×		×		?
	identité	×		×	=	×
Normale	identité	×	=	×	=	×

TAB. 2.6 – *Définitions et équivalences des trois méthodes.*

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 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.

- On ne constate pas de différence significative entre notre méthode et celle de Schall. Ce qui est un peu décevant mais pas complètement 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 moyenner 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 estimées					
		Schall		**Notre méthode**		**GAR**	
Valeurs simulées		$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\beta}$
$\beta = 1 \quad \sigma^2 = 0.05$	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
$\beta = 1 \quad \sigma^2 = 0.5$	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
$\beta = 1 \quad \sigma^2 = 1$	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
$\beta = 1 \quad \sigma^2 = 2$	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. 2.7 – *Résultats simulation comparée des trois méthodes : modèle binomial - lien logit.*

valeurs de σ^2. De plus, on observe qu'à N/q fixé, les résultats se détériorent avec N diminuant.

- En termes du nombre d'itérations nécessaire à la convergence (pour un test d'arrêt sur les valeurs des paramètres β et σ^2), là encore, notre méthode et celle de Schall se tiennent mais la méthode GAR nécessite un nombre d'itérations plus important.

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

Dans les algorithmes que nous avons implémentés, nous avons toujours retenu l'estimation par résolution du système de Henderson (plutot que le système REML ou autre), source de moindres problèmes numériques. Mais ceci n'est pas forcément le meilleur moyen de tirer partie de la connaissance d'une composante de la variance.

Ce type de résolution par système de Henderson ne permet pas l'estimation dans des modèles voulant prendre en compte la surdispersion. En effet, dans ces cas là, en introduisant une réalisation de l'effet aléatoire par donnée (cf 1.3.1), on a alors (avec p la taille du vecteur d'effet fixe) $p+q > N$. Il est alors difficile d'envisager de résoudre un système à $p+q$ inconnues et N équations.

Notons pour finir que ces trois méthodes, pour σ^2 petit, donnent de très bons résultats. La méthode Schall et la notre continuent de bien se comporter pour des plus grandes valeurs de σ^2. C'est ce qui est primordial.

2.4.4 Point de vue asymptotique

Dans cette section, nous évoquons la question du comportement asymptotique des estimateurs $\hat{\beta}$ et $\hat{\sigma}^2$ dans les GL2M.

Nous avons vu dans le cadre des GLM et dans celui des L2M que, sous des conditions bien précises, des résultats de normalité asymptotique ont été démontrés. Malheureusement, comme le mentionnent Breslow et Clayton (1993), il n'existe pas de telles justifications formelles dans le cadre des GL2M. Les seuls résultats qui peuvent être énoncés le sont de façon approchés et sans mesurer réellement le degré de cette approximation.

Ainsi, pour l'estimateur $\hat{\beta}$ de la méthode Schall, il est donné la matrice de variance approchée (avec nos notations) :

$$\mathrm{Var}(\hat{\beta}) = \left(X'\Gamma_{\beta,\xi}^{-1}X\right)^{-1} . \tag{2.22}$$

où $\Gamma_{\beta,\xi} = UDU' + W_{\beta,\xi}$.
Ce résultat se trouve dans les différents travaux Schall (1991), Breslow et Clayton (1993) et McGilchrist (1994), qui, sous des arguments d'approximation de nature différente, réfèrent au même estimateur. Il copie les résultats exacts dans la théorie usuelle gaussienne. Notons que cette matrice de variance (2.22) correspond aux p premières lignes et colonnes de l'inverse de la matrice du système (2.21) (système de Henderson adapté à l'estimation dans les GL2M).
En ce qui concerne les composantes de la variance, en appliquant les résultats de la théorie normale à la variable dépendante Z, il est donné la matrice de variance approchée :

$$\mathrm{Var}(\hat{\sigma}^2) = \left(\frac{1}{2}\left\{\mathrm{tr}(PV_iPV_j)\right\}\right)^{-1} , \tag{2.23}$$

où $P = \Gamma_{\beta,\xi}^{-1}(I - X(X'\Gamma_{\beta,\xi}^{-1}X)^{-1}X'\Gamma_{\beta,\xi}^{-1})$.

Sans plus de justification théorique, nous pouvons reprendre le même type d'arguments approchés dans notre méthode pour donner la matrice de variance approchée de $\hat{\beta}$:

$$\mathrm{Var}(\hat{\beta}) = \left(X'\Gamma_\beta^{-1}X\right)^{-1} ,$$

où $\Gamma_\beta = UDU' + W_\beta$,

et celle de $\hat{\sigma}^2$:

$$\mathrm{Var}(\hat{\sigma}^2) = \left(\frac{1}{2}\left\{\mathrm{tr}(PV_iPV_j)\right\}\right)^{-1} ,$$

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

Chapitre 3

Extension de la méthode GAR

3.1 Introduction

Développée par Gilmour, Anderson, et Rae (1985), la méthode que nous étudions tout au long de ce chapitre, a été initialement conçue pour estimer les effets fixes dans un modèle mixte adapté à des données binomiales. Elle permet aussi une prédiction des effets aléatoires ainsi qu'une estimation de leur variance. Nous la désignons par GAR dans tout le document.

Les données binomiales pour lesquelles elle a été 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 (1980) constate l'inadaptation des méthodes d'analyse développées pour des données continues distribuées selon une loi normale. Pour pallier cela, divers auteurs (Gianola et Foulley 1983, Harville et Mee 1984, Stiratelli et al. 1984) développent des méthodes d'estimation dans des modèles non gaussiens plus adaptés, des GL2M. Ces méthodes utilisent des arguments différents (bayésiens notamment) mais sont en finalité équivalentes. La méthode GAR, quant à elle, apparaît alors comme originale. Elle se base sur l'utilisation de la fonction de quasi-vraisemblance marginale (cf. 2.2.3). Cette fonction est construite à partir du calcul de l'espérance et de la variance marginales. Comme nous l'avons décrit au 2.4.2, parmi les différents points de vue permettant d'aborder l'estimation dans les modèles mixtes, cette approche GAR est davantage marginale. Nous ne reprenons pas ici la discussion à ce sujet.

L'article fondateur de cette méthode a fait l'objet de diverses relectures : Foul-

LEY, GIANOLA, et IM (1990), FOULLEY et MANFREDI (1991). Une extension aux données multicatégories en a été proposée par les auteurs eux-mêmes (GILMOUR, ANDERSON, et RAE 1987). Et FOULLEY et IM (1993) en suggèrent une adaptation dans le cas de données poissonniennes.

Dans ce chapitre, nous présentons dans un premier temps, une nouvelle lecture de cette méthode, dans le cas de données binomiales, en relâchant l'hypothèse d'origine selon laquelle les variances des variables sous-jacentes sont homogènes. Puis, la deuxième section montre comment une démarche similaire peut être adoptée pour traiter le cas de données poissonniennes. Ces deux situations sont bien entendu traitées en lien étroit avec les travaux respectivement de GILMOUR et al. (1985) et de FOULLEY et IM (1993).
Ensuite, nous proposons une adaptation de la méthode à des données exponentielles dans un modèle avec lien logarithmique (lien non canonique).
Ces trois cas ayant été étudiés, nous proposons un formalisme 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. Cette formalisation 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 fait l'objet de la dernière section.

3.2 Données binomiales - Lien probit

Dans cette section, nous relisons donc la méthode d'origine, en commençant par une présentation du modèle envisagé par GILMOUR et al. (1985), puis une description précise de la méthode, avant de finir par quelques remarques.

3.2.1 Le modèle et les notations

Vu le type de données auquel on s'intéresse (cf. 3.1), et étant donnée la distribution normale sous-jacente, le modèle adopté est, dans la famille des GL2M, un modèle binomial mixte avec lien probit.
Remarquons brièvement que ce n'est donc pas le lien canonique qui est envisagé ici, la fonction de lien canonique associée à la loi binomiale étant la fonction de lien logit.

Nous rappelons ci-dessous les hypothèses de ce modèle, cas particulier de GL2M décrit précisément au 1.4.1, et les notations adoptées.

Soit y le vecteur $(N \times 1)$ des observations, réalisation du vecteur aléatoire Y. Conditionnellement au vecteur d'effets aléatoires ξ, on suppose que les composantes

Y_i sont indépendantes et que :[1]

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

ou encore de façon équivalente :

$$\forall i \in \{1, ..., N\}\ , \quad Y_i = \sum_{r=1}^{n_i} Y_{ir} \quad \text{où} \quad Y_{ir} \mid \xi \ \sim \ \mathcal{B}(1, p_{\xi,i}) \ \text{ et indépendantes.}$$

On s'intéresse au vecteur $(N \times 1)$ des fréquences

$$f_i = \frac{y_i}{n_i}\ ,$$

et l'on note F_i les variables aléatoires dont elles sont issues, composantes du vecteur F (de dimension $N \times 1$).

Le prédicteur linéaire intervenant dans le modèle comporte les parties fixe et aléatoire (cf. 1.4.1) :

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

où les dimensions des différents objets X, Z, β, ξ sont respectivement $N \times p$, $N \times q$, $p \times 1$, et $q \times 1$. Le vecteur ξ regroupe K effets aléatoires : $\xi = (\xi_1', \ldots, \xi_K')'$. On suppose

$$\forall j \in \{1, \ldots, K\}\ , \qquad \xi_j \sim \mathcal{N}_{q_j}(0, \sigma_j^2\ A_j)\ ,$$

les matrices A_j, de dimension $q_j \times q_j$ (avec $q = \sum_{j=1}^{K} q_j$) étant connues. Ainsi,

$$\xi \sim \mathcal{N}(0, D)\ ,$$

où la matrice de variance D est diagonale par blocs : $D = \{_d\ \sigma_j^2\ A_j\ \}_{j=1,\ldots,K}$.
Le vecteur d'effets fixes β et le vecteur des composantes de la variance $\sigma^2 = (\sigma_1^2, \ldots, \sigma_K^2)'$ sont les paramètres du modèle à estimer.

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 :

$$\begin{aligned} E(F_i|\xi) \ = \ p_{\xi,i} \ &= \ \Phi(\eta_{\xi,i}) \\ &= \ \Phi((X\beta + U\xi)_i) \\ &= \ \Phi(x_i'\beta + u_i'\xi) \end{aligned}$$

où x_i' et u_i' sont les $i^{\text{èmes}}$ lignes de X et U respectivement.
Comme nous l'avons déjà mentionné, ce lien est inhérent à l'hypothèse d'existence

1. Comme indiqué au chapitre 1, nous confondons dans ξ, le vecteur aléatoire et sa réalisation. De plus, nous indiçons par ξ, tous les objets qui en dépendent.

de variables normales sous-jacentes. En effet, pour chacune des variables Y_{ir}, notons L_{ir} la variable latente associée : $Y_{ir} = 1 \Longleftrightarrow L_{ir} > 0$ (on fixe ici le seuil τ à 0). Supposons ensuite que :

$$\forall i \in \{1, \dots, N\}, \forall r \in \{1, \dots, n_i\}, \quad L_{ir} = x_i'\beta + u_i'\xi + \varepsilon_{ir} \quad \text{où} \quad \varepsilon_{ir} \sim \mathcal{N}(0,1) \, ,$$

et qu'elles sont toutes indépendantes, on obtient alors :

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

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

$$\forall i \in \{1, ..., N\} \, , \quad \forall r \neq r' \in \{1, ..., n_i\} \, , \qquad \operatorname{cov}(L_{ir}, L_{ir'}) = u_i' D u_i \, ,$$

et $\forall i \neq j \in \{1, ..., N\} \, , \forall r \in \{1, ..., n_i\} \, , \forall r' \in \{1, ..., n_j\} \, ,$

$$\operatorname{cov}(L_{ir}, L_{jr'}) = u_i' D u_j \, .$$

Ce qui, en notant $\alpha_i^2 = u_i' D u_i$, nous permet d'obtenir :

$$\begin{aligned} \operatorname{corr}(L_{ir}, L_{ir'}) &= \frac{\alpha_i^2}{1 + \alpha_i^2} &&= t_{ii} \\ \operatorname{corr}(L_{ir}, L_{jr'}) &= \frac{u_i' D u_j}{\sqrt{1 + \alpha_i^2}\sqrt{1 + \alpha_j^2}} &&= t_{ij} \, . \end{aligned}$$

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

Remarquons ici que le paramètre introduit $\alpha_i^2 = u_i' D u_i$ coïncide bien souvent avec la composante de la variance σ_i^2. En effet, dans de nombreux cas comme par exemple dans des modèles liés à la sélection animale (Gianola et Foulley 1983), le vecteur ligne u_i' n'est composé que d'un seul 1 et de 0.

Le modèle étant défini et les notations prises, nous pouvons maintenant présenter la méthode d'estimation.

3.2.2 La méthode d'estimation

La méthode d'estimation proposée par Gilmour et al. (1985), 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 raisonner au niveau d'un modèle marginal, où Y vecteur à expliquer est observé. Pour cela, on procède au calcul de l'espérance et de la matrice de variance marginales. Ensuite, après approximation de la matrice de variance de Y, on pourra procéder à l'estimation des composantes de la variance par l'intermédiaire des équations de Henderson. C'est en fait plus spécialement avec F que l'on va travailler.

3.2.2.1 Étape de "marginalisation" et estimation de β

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 peut alors estimer le paramètre β.

Calcul de l'espérance marginale : Π

$$\begin{array}{rcccl} E(F_i) & = & \Pi_i & = & E(E(F_i|\xi)) \\ & & & = & E(p_{\xi,i}) \\ & & & = & E\left(\Phi(x_i'\beta + u_i'\xi)\right) \end{array}$$

on obtient (le calcul est donné à la fin de ce chapitre) :

$$\Pi_i = \Phi\left(\frac{x_i'\beta}{\sqrt{1+\alpha_i^2}}\right) .$$

On notera par la suite $\eta_i^* = \dfrac{x_i'\beta}{\sqrt{1+\alpha_i^2}}$, 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 $\Pi_i = \Phi\left(\dfrac{x_i'\beta}{\sqrt{1+\alpha_i^2}}\right)$, et le modèle *marginal* défini par Breslow et Clayton (1993) par $\Pi_i = \Phi(x_i'\beta)$, ne coïncident pas. La présence de l'effet multiplicatif $\frac{1}{\sqrt{1+\alpha_i^2}}$, 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 $\forall i \in \{1, ..., N\}$, $\alpha_i^2 = \alpha^2$. Avec cette hypothèse, il est possible d'envisager le changement d'échelle suivant : $\tilde{\beta} = \dfrac{\beta}{\sqrt{1+\alpha^2}}$, et de poursuivre les calculs sur cette nouvelle échelle. C'est ce qu'ont fait Foulley et Manfredi (1991). Cependant, on abandonne ici 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 : $\operatorname{var}(F_i) = \dfrac{1}{n_i^2}\left[\sum_{r=1}^{n_i}\operatorname{var}(Y_{ir}) + \sum_{r\neq r'}\operatorname{cov}(Y_{ir}, Y_{ir'})\right]$.

Or, d'une part, l'utilisation de la formule classique :

$$\operatorname{var}(Y_{ir}) = E(\operatorname{var}(Y_{ir}|\xi)) + \operatorname{var}(E(Y_{ir}|\xi)) \ ,$$

nous permet d'obtenir :

$$\begin{aligned}\operatorname{var}(Y_{ir}) &= E(p_{\xi,i}(1-p_{\xi,i})) + \operatorname{var}(p_{\xi,i}) \\ &= \Pi_i(1-\Pi_i) \ .\end{aligned}$$

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

$$\begin{aligned}\operatorname{cov}(Y_{ir}, Y_{ir'}) &= P\left((L_{ir}>0)\cap(L_{ir'}>0)\right) - P\left(L_{ir}>0\right)^2 \\ &= \Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi(\eta_i^*)^2 \ ,\end{aligned}$$

où Φ_2 désigne la fonction de répartition de la loi normale bivariée [2].

Ainsi,

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

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

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

2. 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 sont identiquement distribuées de loi $\mathcal{N}(0,1)$ avec $\operatorname{corr}(Y_1, Y_2) = \rho$.

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

$$V = A + B \ ,$$

$$\begin{array}{lrcl} \text{où} & A & = & \left\{_d \ \dfrac{\Pi_i(1-\Pi_i) - [\Phi_2(\eta_i^*, \eta_i^*, t_{ii}) - \Phi(\eta_i^*)^2]}{n_i} \right\}_{i=1,\dots,N} \\ \text{et} & B & = & \left\{\Phi_2(\eta_i^*, \eta_j^*, t_{ij}) - \Phi(\eta_i^*)\Phi(\eta_j^*)\right\}_{i,j=1,\dots,N} \ . \end{array}$$

Estimation par maximum de quasi-vraisemblance

Ayant ainsi obtenu l'expression des deux premiers moments marginaux du vecteur $F = (F_1, \dots, F_N)'$, 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 quasi-vraisemblance nous permet dans ce cas, de prendre en compte l'information contenue dans le calcul des deux premiers moments marginaux.

Pour cela, en supposant dans un premier temps que les composantes de la variance sont connues (V dépend alors uniquement du paramètre β), on maximise la quasi-vraisemblance comme décrit au 2.2.3. On obtient le système itératif suivant (t désignant l'indice d'itération) :

$$(G'V^{[t]-1}G)\Delta\beta^{[t]} = G'V^{[t]-1}(f - \Pi^{[t]}) \ ,$$

où f est le vecteur des fréquences observées, et $G = \dfrac{\partial \Pi}{\partial \beta'}$.

D'après l'expression de Π obtenue précédemment, et en définissant les matrices :

- $K = \{_d \ \varphi(\eta_i^*) \ \}_{i=1,\dots,N}$, où φ est la fonction de densité de la loi normale centrée réduite,
- $M = \left\{_d \ \dfrac{1}{\sqrt{(1+\alpha_i^2)}} \right\}_{i=1,\dots,N}$, qui contient donc l'effet multiplicatif introduit par l'intégration des effets aléatoires,

on exprimera alors G de la façon suivante :

$$\begin{array}{rcl} G & = & KMX \\ & = & LX \qquad \text{avec} \qquad L = KM = \left\{_d \ \dfrac{\varphi(\eta_i^*)}{\sqrt{(1+\alpha_i^2)}} \right\}_{i=1,\dots,N} \ . \end{array}$$

Notons qu'avec cette définition de la matrice M, l'élément diagonal de la matrice T peut aussi s'écrire $t_{ii} = u_i' M D M u_i$.

Avec ces notations, on aboutit donc au système résolu itérativement en β (cf. 2.2.3) :

$$(X' W^{[t]\,-1} X)\beta^{[s+1]} = X' W^{[t]\,-1} \zeta^{[t]} \,, \tag{3.1}$$

$$\text{où} \quad \begin{array}{rcl} \zeta^{[t]} & = & X\beta^{[t]} + L^{[t]\,-1}(f - \Pi^{[t]}) \\ W^{[t]} & = & L^{[t]\,-1} V^{[t]} L^{[t]\,-1} \,. \end{array}$$

Remarquons que les matrices V, L ainsi que W dépendent de la valeur courante de β. Comme pour la procédure d'estimation dans les GLM (cf. 2.2.1), on reconnaît dans le système (3.1), une procédure de moindres carrés pondérés itérés. Cette procédure peut être associée au modèle linéaire :

$$\zeta = X\beta + e \quad \text{où} \qquad \mathrm{Var}(e) = W \,.$$

Ainsi, la matrice W apparaît comme la matrice de variance de ζ.

On a donc obtenu un algorithme itératif pour réaliser l'estimation par maximum de quasi-vraisemblance de β. Malheureusement, dans ce qui précède, les valeurs $\alpha_1^2, ..., \alpha_N^2$ dépendent des composantes de la variance $\sigma_1^2, ..., \sigma_K^2$ par l'expression $\alpha_i^2 = u_i' D u_i$, où $D = \{_d \ \sigma_j^2 A_j \ \}_{j=1,...,K}$. Or, ces composantes de la variance sont des paramètres inconnus. Il est donc nécessaire de proposer une méthode d'estimation de ces paramètres, afin de pouvoir remplacer dans le système (3.1) les valeurs des σ_j^2 par leurs estimations.

3.2.2.2 Étape d'approximation de V et estimation des σ_j^2

Afin de proposer une estimation des composantes de la variance, on approche la matrice de variance V, et de ce fait, la matrice W. On reconnait alors dans le résultat obtenu pour W, la forme classique de la structure de variance d'un L2M (cf. 1.3.2). Ainsi, l'estimation de β peut aussi se réaliser par résolution des équations de Henderson. En effet (cf. 2.3.4), la solution en β correspond à celle du système (3.1) précédent, prenant en compte l'approximation de V. L'avantage est qu'à l'aide de ces équations, il sera possible d'obtenir simultanément une estimation des composantes de la variance (cf. 2.3.4). Notons qu'elles permettent aussi de donner une prédiction du vecteur d'effets aléatoires ξ.

Approximation de la matrice V

On suppose que les éléments de la matrice T (définie au 3.2.1) sont petits. Ce qui équivaut à considérer que $\forall i, j \in \{1, ..., N\}$, les éléments $u_i' D u_j$ 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 :

$$\begin{array}{rcl} A \approx V_0 & = & \left\{_d \; \dfrac{\Pi_i(1-\Pi_i) - t_{ii}\varphi^2(\eta_i^*)}{n_i} \right\}_{i=1,\ldots,N} \\ B \approx V_1 & = & \left\{ t_{ij}\varphi(\eta_i^*)\varphi(\eta_j^*) \right\}_{i,j=1,\ldots,n} \\ & = & \left\{ \dfrac{\varphi(\eta_i^*)}{\sqrt{1+\alpha_i^2}} u_i' D u_j \dfrac{\varphi(\eta_j^*)}{\sqrt{1+\alpha_j^2}} \right\}_{i,j=1,\ldots,N} \end{array} .$$

Ainsi, on a :

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

Ce qui donne comme approximation de W :

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

où R est donc de la forme :

$$R = \left\{_d \; \frac{1}{n_i}\left[\frac{\Pi_i(1-\Pi_i)\;(1+\sigma_i^2)}{\varphi^2(\eta_i^*)} - \sigma_i^2\right] \right\}_{i=1,\ldots,N} .$$

Les équations de Henderson

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

$$\begin{pmatrix} X'R^{[t]\,-1}X & X'R^{[t]\,-1}U \\ U'R^{[t]\,-1}X & U'R^{[t]\,-1}U + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \xi \end{pmatrix} = \begin{pmatrix} X'R^{[t]\,-1}\zeta^{[t]} \\ U'R^{[t]\,-1}\zeta^{[t]} \end{pmatrix} . \tag{3.2}$$

Sachant que $D = \{_d \ \sigma_j^2 A_j \ \}_{j=1,...,K}$, on peut alors estimer le vecteur des composantes de la variance $\sigma^2 = (\sigma_1^2, .., \sigma_K^2)'$, en adoptant une démarche REML, par la méthode itérative (cf. 2.17) :

$$\forall j \in \{1, ..., K\} \ , \qquad \hat{\sigma}_j^{2\,[t+1]} = \frac{\hat{\xi}_j' A_j^{-1} \hat{\xi}_j}{q_j - \dfrac{\mathrm{tr}(A_j^{-1} C_{jj})}{\hat{\sigma}_j^{2\,[t]}}}$$

où C_{jj} est le $j^{\text{ème}}$ bloc correspondant au $j^{\text{ème}}$ effet aléatoire (ξ_j) dans l'inverse de la matrice du système.
Le sous-produit $\hat{\xi}$ obtenu est interprété comme une prédiction du vecteur d'effets aléatoires dans le L2M associé.

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

3.2.3 Remarques

Nous venons de le mentionner, les auteurs de la méthode GAR analysent les valeurs de ξ obtenues à la convergence de l'algorithme comme une prédiction des effets aléatoires du modèle initial. Autant l'interprétation donnée au sous-produit $\hat{\xi}$ 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, l'approximation effectuée sur V (pour des petites valeurs des composantes de la variance), implique une linéarisation du modèle *marginalisé*. Cette linéarisation se traduit par le plongement de ζ dans un L2M, avec par conséquent les hypothèses gaussiennes adéquates.

Enfin, 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 (1987). 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 0, ce type de données introduit des paramètres supplémentaires à estimer, à savoir les autres seuils.

3.3 Données poissonniennes - Lien logarithme

Après l'analyse de données binomiales, nous nous intéressons à présent à des données toujours discrètes, mais distribuées, conditionnellement aux effets aléatoires, selon une loi de Poisson. Foulley, Gianola, et Im (1987) justifient l'utilisation de cette distribution pour modéliser des données de reproduction.
Foulley et Im (1993) ont décrit comment la méthode GAR, initialement développée pour des données binomiales avec lien probit, pouvait être adaptée à ce type de données. Cette adaptation qu'ils proposent, reste fidèle à la méthodologie GAR, dans la mesure où elle repose sur l'estimation par maximum de quasi-vraisemblance. La fonction de quasi-vraisemblance est 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 dans des termes similaires à ceux du paragraphe précédent, et nous nous attachons à considérer de nouveau le cas où les variances des variables sous-jacentes ne sont pas homogènes.

3.3.1 Le modèle et les notations

Nous reprenons succintement les hypothèses du cas particulier, dans la famille des GL2M, d'un modèle de Poisson avec lien logarithmique. Pour cela, nous considérons que les composantes du vecteur Y (dont le vecteur des observations y est une réalisation) sont, conditionnellement à ξ, indépendantes et de loi :

$$\forall i \in \{1, ..., N\}\ , \qquad Y_i \mid \xi \ \sim\ \mathcal{P}(\lambda_{\xi,i})$$

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

$$\ln(\lambda_{\xi,i}) = x_i'\beta + u_i'\xi = \eta_{\xi,i}\ .$$

On a donc pour cette distribution :

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

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

3.3.2 La méthode d'estimation

3.3.2.1 Étape 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 effets aléatoires :

$$\begin{aligned} E(Y_i) = \mu_i &= E(E(Y_i|\xi)) \\ &= E(\lambda_{\xi,i}) \\ &= E(\exp(x_i'\beta + u_i'\xi)) \,. \end{aligned}$$

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

$$\begin{aligned} \mu_i &= \exp(x_i'\beta + \frac{u_i'Du_i}{2}) \\ &= \exp(\eta_i *) \end{aligned}$$

où $\eta_i * = x_i'\beta + \dfrac{u_i'Du_i}{2}$, 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 (1993), 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 :

$$\begin{aligned} \forall i \in \{1,\ldots,N\}\,, \quad \mathrm{var}(Y_i) &= E(\lambda_{\xi,i}) + \mathrm{var}(\lambda_{\xi,i}) \\ &= \mu_i + \mu_i^2\,[\exp(u_i'Du_i) - 1]\,, \\ \text{et} \quad \forall i \neq j \in \{1,\ldots,N\}\,, \quad \mathrm{cov}(Y_i,Y_j) &= E(\lambda_{\xi,i}\lambda_{\xi,j}) - E(\lambda_{\xi,i})E(\lambda_{\xi,j}) \\ &= \mu_i\mu_j\,[\exp(u_i'Du_j) - 1]\,. \end{aligned}$$

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

$$V = A + B \ ,$$

$$\begin{array}{lrcl} \text{où} & A & = & \{_d \ \mu_i \ \}_{i=1,\dots,N} \\ \text{et} & B & = & \{\mu_i \mu_j \ (\exp(u_i' D u_j) - 1)\}_{i,j=1,\dots,N} \ . \end{array}$$

Estimation par maximum de quasi-vraisemblance

On procède alors à la construction de la fonction de quasi-vraisemblance et à sa maximisation. Et l'on obtient les équations suivantes :

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

où $\quad G = \dfrac{\partial \mu}{\partial \beta'} = KX \quad$ avec $\quad K = \{_d \ \exp(\eta_i^*) \ \}_{i=1,\dots,N}$.

Remarque : Comparativement à la section précédente, il n'y a pas ici de matrice M, puisque comme nous l'avons signalé, aucun effet multiplicatif n'est introduit à l'intégration des effets aléatoires.

La résolution de ces équations (cf. 2.2.3), conduit à l'algorithme itératif :

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

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

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

3.3.2.2 Étape d'approximation de V et estimation des σ_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 $i, j \in \{1, \dots N\}$ où $u_i' D u_j$ est proche de 0 :

$$\begin{array}{rcl} A & = & V_0 = \{_d \ \mu_i \ \}_{i=1,\dots,N} \\ B & \approx & V_1 = \{\mu_i \mu_j \ u_i' D u_j\}_{i,j=1,\dots,N} \ . \end{array}$$

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

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

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

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

où R prend la forme suivante :

$$R = K^{-1} = \{_d \ \frac{1}{\mu_i} \ \}_{i=1,\ldots,N} = \{_d \ \exp(-\eta_i^*) \ \}_{i=1,\ldots,N} \ .$$

Cette approximation étant réalisée, la procédure est ensuite tout à fait identique à celle de la section 3.2.2.2. On résout itérativement les équations de Henderson pour obtenir l'estimation de β (avec W approchée) et l'estimation des composantes de la variance. Les valeurs courantes des paramètres $\beta^{[t]}$ et $\sigma^{2\,[t]}$ obtenues, interviennent à chaque itération pour former $\eta_i^{*\,[t]}$ et donc obtenir $R^{[t]}$, $\zeta^{[t]}$ et approcher $W^{[t]} \approx R^{[t]} + UD^{[t]}U'$.

3.3.3 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 d'un L2M, nous permettant d'utiliser les équations de Henderson. Pour pouvoir réaliser cette approximation, il est nécessaire de vérifier que les éléments $u_i'Du_j$ sont proches de 0, ou encore que les composantes de la variance sont petites.

De même que dans le cas binomial :

- lors du calcul de l'espérance marginale la fonction de lien a été conservée,
- l'approximation de V implique une linéarisation du modèle marginal qui se traduit par des hypothèses gaussiennes sur ζ,
- enfin, on peut aussi émettre des réserves quant à la prédiction de ξ obtenue.

3.4 Données exponentielles - Lien logarithme

Dans toute cette section, nous allons maintenant considérer des données distribuées, conditionnellement aux effets aléatoires, selon une loi exponentielle (notée[3] $\mathcal{E}(\lambda)$). Ces données continues sont de nature tout à fait 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 Poisson, de nombreuses applications existent dans le domaine de la génétique animale (Ducrocq 1990), on peut trouver des applications de modèles de loi exponentielle à effets aléatoires dans le domaine de la fiabilité des logiciels (Gaudoin, Lavergne, et Soler 1994). Ces deux situations sont illustrées par les deux exemples cités au 1.4.2.

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

Pour cela, nous reprenons la démarche des deux sections précédentes. Elle 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.

3.4.1 Le modèle et les notations

Nouveau cas particulier de GL2M, nous retrouvons les trois hypothèses définissant le modèle.

- Les composantes de Y sont, conditionnellement à ξ, indépendantes et de loi :

$$\forall i \in \{1, ..., N\} \ , \qquad Y_i \mid \xi \ \sim \ \mathcal{E}(\lambda_{\xi,i}).$$

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

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

$$\eta_{\xi,i} = \ln(\mu_{\xi,i}) \iff \mu_{\xi,i} = \exp(x_i'\beta + u_i'\xi) = \lambda_{\xi,i} \ .$$

  Une raison qui justifie le choix de ce lien, est notamment le fait que ce lien peut 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.

3. Comme au premier chapitre, la loi désignée par $\mathcal{E}(\lambda)$ désigne la loi exponentielle de densité définie par : $f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \ \mathbf{1}_{\mathbb{R}^+}(x)$.

3.4.2 La méthode d'estimation

3.4.2.1 Étape de "marginalisation" et estimation de β

Le choix de la fonction de lien est le même que celui de la section précédente 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

$$\begin{array}{rcl} E(Y_i) = \mu_i & = & E(E(Y_i|\xi)) \\ & = & E(\mu_{\xi,i}) \\ & = & E(\exp(x_i'\beta + u_i'\xi) \quad . \\ & = & \exp(x_i'\beta + \dfrac{u_i'Du_i}{2}) \\ & = & \exp(\eta_i *) \end{array}$$

avec $\eta_i * = x_i'\beta + \dfrac{u_i'Du_i}{2}$: 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, conditionnellement à ξ, on a :

$$\begin{array}{llcccccc} \forall i \in \{1,\ldots,N\} \, , & \mathrm{var}(Y_i|\xi) & = & \dfrac{1}{\lambda^2_{\xi,i}} & = & \exp(2(x_i'\beta + u_i'\xi)) & = & \mu^2_{\xi,i} \\ \forall i \neq j \in \{1,\ldots,N\} \, , & \mathrm{cov}(Y_i,Y_j|\xi) & = & 0 & . & & & \end{array}$$

Ce qui nous permet d'obtenir :

$$\begin{array}{lllcr} & \forall i \in \{1,\ldots,N\} \, , & \mathrm{var}(Y_i) & = & E(\mathrm{var}(Y_i|\xi)) + \mathrm{var}(E(Y_i|\xi)) \\ & & & = & 2 * E(\lambda^2_{\xi,i}) - E(\lambda_{\xi,i})^2 \\ & & & = & \mu_i^2[2 * \exp(u_i'Du_i) - 1] \, , \\ \text{et,} & \forall i \neq j \in \{1,\ldots,N\} \, , & \mathrm{cov}(Y_i,Y_j) & = & E(\lambda_{\xi,i}\lambda_{\xi,j}) - E(\lambda_{\xi,i})E(\lambda_{\xi,j}) \\ & & & = & \mu_i\mu_j[\exp(u_i'Du_j) - 1] \, . \end{array}$$

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

$$V = A + B \ ,$$

$$\begin{array}{lrcl} \text{où} & A & = & \{_d \ \mu_i^2 \exp(u_i' D u_i) \ \}_{i=1,\dots,N} \\ \text{et} & B & = & \{\mu_i \mu_j \ (\exp(u_i' D u_j) - 1)\}_{i,j=1,\dots,N} \ . \end{array}$$

Estimation par maximum de quasi-vraisemblance

La démarche reste la même. Les deux premiers moments étant calculés, on construit et on maximise la fonction de quasi-vraisemblance. On est amené, dans ce cas, à résoudre le système itératif :

$$(X'W^{[t]\,-1}X)\beta^{[t+1]} = X'W^{[t]\,-1}\zeta^{[t]} \ , \tag{3.3}$$

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

Remarque: La matrice K est la même qu'au 3.3.2. En effet, l'espérance s'exprimant de la même façon, la matrice $G = \dfrac{\partial \mu}{\partial \beta'} = KX$, ne change pas. Il n'y a toujours pas d'effet multiplicatif.

3.4.2.2 Étape d'approximation de V et estimation des σ_j^2

Faisant l'hypothèse que $\forall i, j \in \{1, ..., N\}^2$, les éléments $u_i' D u_j$ sont proches de 0, on utilise, de même que pour le cas Poisson, une approximation de la fonction exponentielle au voisinage de 0 pour obtenir :

$$\begin{array}{rcl} A & \approx & V_0 = \{_d \ \mu_i^2(1 + u_i' D u_i) \ \}_{i=1,\dots,N} \\ B & \approx & V_1 = \{\mu_i \mu_j \ u_i' D u_j\}_{i,j=1,\dots,N} \ . \end{array}$$

Ainsi, on a :

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

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

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

où R s'écrit :

$$R = \{_d \ 1 + u_i' D u_i \ \}_{i=1,\dots,N} \ .$$

W étant approchée sous cette forme d'une structure de variance d'un L2M, on ne résout pas directement le système (3.3) (avec W approchée), mais on en obtient des solutions grâce aux équations de Henderson (3.2). Cela nous permet, grâce aux valeurs de ξ obtenues par résolution de ce système, d'estimer les composantes de la variance. Les expressions de R et de ζ nécessaires pour (3.2) sont décrites ci-dessus.

3.4.3 Remarques

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

3.5 Une formalisation commune

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

3.5.1 Des objets communs

Tout d'abord, résumons rapidement les objets utilisés dans les sections précédentes. Pour cela, dans toute cette section, nous référerons aux cas 1, 2 et 3 respectivement les trois situations :
- <u>Cas 1</u> : Données binomiales - Lien probit.
- <u>Cas 2</u> : Données poissonniennes - Lien logarithme.
- <u>Cas 3</u> : Données exponentielles - Lien logarithme.

Il est essentiel de remarquer que dans tous les cas, la matrice de variance V de Y, obtenue après approximation se présente sous la forme :

$$V_0 + LZDZ'L \quad \text{où} \quad L = KM \ .$$

C'est ce qui permet ensuite 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 : $M = \left\{ {}_d \dfrac{1}{\sqrt{1 + u_i'Du_i}} \right\}_{i=1,\ldots,N}$
- cas 2 et 3 : $M = I_N$.

Cette matrice M correspond à la matrice des effets multiplicatifs que la "marginalisation" (ou l'intégration des effets aléatoires) a pu introduire sur β.
Ainsi on a pu écrire :

$$\eta^* = M\eta + C$$

où

η^*	:	prédicteur linéaire *marginalisé*, introduit aux sections précédentes,
$\eta = X\beta$	:	prédicteur linéaire *marginal*, en référence à Breslow et Clayton (1993), 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 :

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

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

- cas 2 et 3 : $K = \{_d \; \exp(\eta_i^*) \; \}_{i=1,\ldots,N}$.

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.

3.5.2 Une nouvelle démarche

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, pour maximiser la quasi-vraisemblance approchée et estimer les composantes de la variance.

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 $\tilde{Y}$, dont y serait issu. Dans ce modèle approché, la fonction de quasi-vraisemblance marginale est identique à celle construite, dans le modèle initial, après approximation de V. Pour cela, il est nécessaire d'imposer une hypothèse supplémentaire au modèle initial.

3.5.2.1 Le modèle initial

Nous considérons le GL2M dont nous rappelons brièvement les hypothèses :

- soit Y le vecteur 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 :

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

- on considère le prédicteur linéaire :

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

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

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

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 :

$$\begin{array}{rcl} E(Y_i) & = & E(E(Y_i|\xi)) \\ & = & E(h(\eta_{\xi,i})) \\ & = & h(\eta_i *) \ , \end{array}$$

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.

3.5.2.2 Le modèle approché

Comme nous l'avons déjà mentionné, nous allons modéliser le vecteur des observations y par le vecteur aléatoire $\tilde{Y}$.

Définition de $\tilde{Y}$

Considérons le vecteur aléatoire $\tilde{Y}$ dont les composantes $\tilde{Y}_i$ sont indépendantes. L'espérance conditionnelle de $\tilde{Y}_i$ est donnée par :

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

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

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

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

avec la même fonction de variance v que pour Y_i, dans le modèle initial.

Définissons un nouveau *prédicteur linéaire conditionnel* par :

$$\tilde{\eta}_\xi = \eta^* + MU\xi \ ,$$

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 (1991) 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 : $$\begin{aligned} \tilde{\eta}_\xi &= M\eta_\xi + C \\ &= \tilde{X}\beta + \tilde{U}\xi + C \quad \text{où} \quad \tilde{X} = MX \\ & \qquad\qquad\qquad\qquad\quad\; \tilde{U} = MU. \end{aligned}$$

À l'aide de ce *prédicteur linéaire conditionnel*, remarquons qu'une réalisation de l'espérance conditionnelle $\tilde{\mu}_{\xi,i}$, peut être vue comme un développement limité au premier ordre de $h(\tilde{\eta}_{\xi,i})$ en η_i^* (pour ξ proche de 0).

Intéressons nous maintenant au calcul de l'espérance et de la matrice de variance marginales de $\tilde{Y}$.

Espérance et variance marginales de $\tilde{Y}$

Tout d'abord, l'espérance marginale de $\tilde{Y}_i$ est :

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

Ainsi, les variables Y_i et $\tilde{Y}_i$ ont même espérance marginale.

Pour la matrice de variance marginale $\tilde{V}$, on a :

$$\forall i \in \{1,...,N\}\,, \qquad \begin{array}{rcccl} \mathrm{var}(\tilde{Y}_i) & = & E(\mathrm{var}(\tilde{Y}_i|\xi)) & + & \mathrm{var}(E(\tilde{Y}_i|\xi)) \\ & = & E(v(\tilde{\mu}_{\xi,i})) & + & \mathrm{var}(\tilde{\mu}_{\xi,i}) \\ & = & E(v(\tilde{\mu}_{\xi,i})) & + & h'(\eta_i^*)^2\tilde{u}_i{}'D\tilde{u}_i\,, \end{array}$$

et,

$$\forall i \neq j \in \{1,...,N\}\,, \quad \begin{array}{rcccl} \mathrm{cov}(\tilde{Y}_i,\tilde{Y}_j) & = & E(\mathrm{cov}(\tilde{Y}_i,\tilde{Y}_j|\xi)) & + & \mathrm{cov}(E(\tilde{Y}_i|\xi),E(\tilde{Y}_j|\xi)) \\ & = & 0 & + & \mathrm{cov}(\tilde{\mu}_{\xi,i},\tilde{\mu}_{\xi,j}) \\ & = & h'(\eta_i^*)h'(\eta_j^*)\tilde{u}_i{}'D\tilde{u}_j\,. & & \end{array}$$

D'où $\tilde{V}$ s'écrit sous la forme :

$$\begin{array}{rcccl} \tilde{V} = \mathrm{Var}(\tilde{Y}) & = & E(v(\tilde{\mu}_\xi)) & + & K\tilde{U}D\tilde{U}'K \\ & = & V_0 & + & KMUDU'MK\,, \end{array}$$

où K est toujours la matrice définie par : $K = \{_d\ h'(\eta_i^*)\ \}_{i=1,...,N}$.

Vérifions que $\tilde{V}$ est bien la même matrice que celle obtenue, après approximation, dans les sections précédentes. En ce qui concerne le deuxième terme de la somme, on retrouve bien la même expression. Mais qu'en est-il du premier terme : 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 :

$$\begin{array}{rcl} V_{0_{ii}} & = & E(\dfrac{\tilde{\mu}_{\xi,i}(1-\tilde{\mu}_{\xi,i})}{n_i}) \\ & = & E(\dfrac{1}{n_i}[h(\eta_i^*)+h'(\eta_i^*)\tilde{u}_i{}'\xi][1-h(\eta_i^*)-h'(\eta_i^*)\tilde{u}_i{}'\xi]) \\ & = & \dfrac{1}{n_i}[h(\eta_i^*)(1-h(\eta_i^*))-h'(\eta_i^*)^2\tilde{u}_i{}'D\tilde{u}_i]. \end{array}$$

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

$$V_0 = \left\{_d\ \frac{\Phi(\eta_i^*)(1-\Phi(\eta_i^*)) - \varphi(\eta_i^*)^2 u_i'MDMu_i}{n_i}\right\}_{i=1,...,N}.$$

Ce qui, étant donné que l'on peut écrire $\Pi_i = \Phi(\eta_i^*)$ et $t_{ii} = u_i'MDMu_i$ (cf paragraphe 3.2.2), nous redonne bien la même expression de V_0, et donc 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 :

$$\begin{aligned} V_{0_{ii}} &= E(\tilde{\mu}_{\xi,i}) \\ &= h(\eta_i^*) = \mu_i \ . \end{aligned}$$

Donc on vérifie encore :

$$V_0 = \{_d \ \mu_i \ \}_{i=1,\dots,N}.$$

• Cas 3 : Pour un GLM avec loi exponentielle, la fonction de variance associée est : $v(\mu_i) = \mu_i^2$.
D'où :

$$\begin{aligned} V_{0_{ii}} &= E(\tilde{\mu}_{\xi,i}^2) \\ &= E([h(\eta_i^*) + h'(\eta_i^*)\tilde{u}_i{}'\xi]^2) \\ &= h(\eta_i^*)^2 + h'(\eta_i^*)^2\tilde{u}_i{}'D\tilde{u}_i \\ &= \mu_i^2 + \mu_i^2\tilde{u}_i{}'D\tilde{u}_i. \end{aligned}$$

Or, dans ce cas, $M = I_N$, d'où $\tilde{u}_i = u_i$, et on vérifie une fois encore :

$$V_0 = \left\{_d \ \mu_i^2(1 + u_i'Du_i) \ \right\}_{i=1,\dots,N}.$$

Dans les trois cas, la matrice de variance marginale $\tilde{V}$ de $\tilde{Y}$ dans le modèle approché, est donc bien la même que l'approximation de V du modèle initial, calculée pour l'utilisation des équations de Henderson dans les démarches directes.

3.5.2.3 La méthode d'estimation

Dans un modèle approché, 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 marginale approchée de Y.

La quasi-vraisemblance étant construite à partir des deux premiers moments marginaux, nous obtenons la même fonction. Ainsi, sa maximisation par résolution des équations de Henderson aboutira donc aux mêmes estimations (en considérant bien entendu y comme réalisation de $\tilde{Y}$).

Le détour par $\tilde{Y}$ que nous proposons, permet 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.

3.5.3 Commentaires

À l'aide du formalisme que nous venons de présenter, 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. 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, lors du calcul de l'espérance marginale. Elle est vérifiée dans les trois cas étudiés auparavant, et le sera aussi de façon approchée à la section 3.6. Elle semble être la clé principale à cette méthodologie GAR.

Notons que dans tous les modèles avec lien identité ($h = Id$), cette hypothèse est vérifiée puisqu'alors $\mu_{\xi,i} = \eta_{\xi,i} = x_i'\beta + u_i'\xi$ et $\mu_i = \eta_i^* = x_i'\beta$. Prenons pour exemple le cas très simple de données poissonniennes où l'espérance conditionnelle s'écrit : $\mu_{\xi,i} = x_i'\beta + u_i'\xi$ et la fonction de variance : $v(x) = x$. L'échelle marginale alors est la même que celle d'origine : $M = I_N$, $\tilde{\eta}_{\xi,i} = \eta_{\xi,i}$ et $K = I_N$. Aussi, $\tilde{\mu}_{\xi,i} = \mu_{\xi,i}$ donc $\tilde{V} = V$, la matrice de variance s'écrit directement sous la forme $V = V_0 + KMUDU'MK$ avec $V_0 = \{_d\ x_i'\beta\ \}_{i=1,\dots,N}$. 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 : $h(x) = \dfrac{1}{x}$). Outre le fait que ce lien ne permette pas d'assurer la positivité du paramètre de la loi, le calcul de $\eta_i^* = E(\mu_{\xi,i}) = E\left(\dfrac{1}{\eta_{\xi,i}}\right) = E\left(\dfrac{1}{x_i'\beta + u_i'\xi}\right)$ pour $\xi \sim \mathcal{N}(0, D)$, lorsqu'il converge, s'avère délicat. Pourtant, il est possible d'envisager un développement limité de l'espérance conditionnelle (pour ξ proche de 0) :

$$\frac{1}{x_i'\beta + u_i'\xi} = \frac{1}{x_i'\beta}[1 - \frac{u_i'\xi}{x_i'\beta} + (\frac{u_i'\xi}{x_i'\beta})^2 + o(\xi^2)] \ .$$

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 $\eta_i^* = \dfrac{x_i'\beta}{1 + \dfrac{u_i'Du_i}{(x_i'\beta)^2}}$ (d'où $M = \left\{_d\ \dfrac{1}{1 + \dfrac{u_i'Du_i}{(x_i'\beta)^2}}\right\}_{i=1,\dots,N}$) dans l'autre.

Il est alors possible de poursuivre avec :

$$\tilde{\mu}_{\xi,i} = \frac{1}{\tilde{x}_i'\beta} - \frac{\tilde{u}_i'\xi}{(\tilde{x}_i'\beta)^2} \ ,$$

où $\tilde{X} = MX$ et $\tilde{U} = MU$. On aboutit alors à

$$\tilde{V} = V_0 + KMUDU'MK$$

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

Ainsi, même s'il est limité par l'hypothèse prise, ce formalisme permet d'étendre la méthode GAR à d'autres situations.

D'autre part, notons que dans la démarche initiale, l'approximation de V était réalisé sous l'hypothèse de composantes de la variance petites. Cette hypothèse intervenait donc directement, et ce ne semble plus être le cas dans la nouvelle démarche. Pourtant, le modèle approché ne sera justifié que parce qu'il permet d'aboutir notamment à l'expression de la variance marginale approchée. Ainsi, sans intervenir explicitement, cette hypothèse de σ^2 proche de 0 est sous-jacente.

Enfin, dans leur article, Breslow et Clayton (1993) envisagent une autre démarche pour définir un modèle marginal et utiliser la quasi-vraisemblance marginale. Zeger, Liang, et Albert (1988) 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, défini conditionnellement aux effets aléatoires, est un GLM, que l'on peut écrire sous la forme :

$$\begin{array}{rclcl}
Y_i & = & \mu_{\xi,i} & + & \varepsilon_i \\
 & = & h(x_i'\beta + u_i'\xi) & + & \varepsilon_i \ ,
\end{array} \tag{3.4}$$

$$\begin{array}{lrcl}
\text{avec} & E(Y_i|\xi) & = & \mu_{\xi,i} = h(\eta_{\xi,i}) = h(x_i'\beta + u_i'\xi) \\
 & V(Y_i|\xi) & = & V(\varepsilon_i|\xi) = v(\mu_{\xi,i}) \ .
\end{array}$$

Les auteurs spécifient alors un modèle marginal en termes de l'espérance marginale par :

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

utilisant le prédicteur linéaire marginal $\eta_i = x_i'\beta$. 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 (3.4), lorsque les composantes de dispersion tendent vers 0. En effet, on obtient alors :

$$Y_i \approx h(x_i'\beta) + h'(x_i'\beta)u_i'\xi + \varepsilon_i$$

Tout ceci les conduit à considérer la matrice de variance marginale :

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

$$\text{avec} \quad \begin{array}{rcl} V_0 & = & \{_d\ v(\mu_i)\ \}_{i=1,\ldots,N}\ , \\ K & = & \{_d\ h'(\eta_i)\ \}_{i=1,\ldots,N}\ . \end{array}$$

Cette écriture, au contraire de la méthode GAR, ne tient pas du tout compte des effets multiplicatifs introduits lors de la "marginalisation". Et cela implique aussi, de façon peu justifiée, que :

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

3.6 GAR - Données binomiales - Lien logit

Utilisant le formalisme de la section 3.5, nous envisageons maintenant une adaptation de la méthode GAR au cas où les données sont distribuées selon une loi binomiale, dans un modèle où l'on considère le lien logit. Ce type de modélisation est très répandu, et est beaucoup plus utilisée que le lien probit, notamment dans le milieu médical. Remarquons aussi que le lien logit 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.

3.6.1 Modèle et notations

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

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

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

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

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

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 : $\xi \sim \mathcal{N}(0, D)$.

3.6.2 Calcul de η^*

On s'intéresse tout d'abord au calcul du prédicteur linéaire marginalisé η^* et à l'identification de M.

$\forall i \in \{1, \ldots, N\}$,

$$\begin{array}{rcccl} E(F_i) & = & E(E(F_i|\xi)) & = & E(p_{\xi,i}) \\ & & & = & E(h(\eta_{\xi,i})) \\ & & & = & E\left(\dfrac{\exp(\eta_{\xi,i})}{1+\exp(\eta_{\xi,i})}\right) \ . \end{array}$$

Malheureusement, ce calcul exact est difficilement réalisable. Nous envisageons alors l'approximation usuelle de la fonction logistique (ZEGER et al. 1988) :

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

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.
Ainsi,

$$\begin{array}{rcl} E(F_i) & \approx & E(\Phi(c\eta_{\xi,i})) \\ & \approx & E(\Phi(cx_i'\beta + cu_i'\xi)) \\ & \approx & \Phi\left(\dfrac{cx_i'\beta}{\sqrt{1+c^2u_i'Du_i}}\right) \ . \end{array}$$

À l'aide de l'approximation inverse, on a alors :

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

$$\begin{aligned} \text{avec} \quad \eta_i^* &= \frac{x_i'\beta}{\sqrt{1+c^2u_i'Du_i}} \ , \\ \text{d'où} \quad M &= \left\{_d \ \frac{1}{\sqrt{1+c^2u_i'Du_i}} \right\}_{i=1,\dots,N} \ . \end{aligned}$$

Maintenant que l'on a su identifier η^* et M, on peut définir *le prédicteur linéaire conditionnel* sur la nouvelle échelle :

$$\begin{aligned} \tilde{\eta}_\xi &= \eta^* + MU\xi \\ &= \eta^* + \tilde{U}\xi \ , \end{aligned}$$

ainsi que la matrice K :

$$\begin{aligned} K &= \{_d \ h'(\eta_i^*) \ \}_{i=1,\dots,N} \\ &= \left\{_d \ \frac{\exp(\eta_i^*)}{(1+\exp(\eta_i^*))^2} \right\}_{i=1,\dots,N} \\ &= \{_d \ \mu_i(1-\mu_i) \ \}_{i=1,\dots,N} \ . \end{aligned}$$

3.6.3 Calcul de $\tilde{V}$

Pour calculer $\tilde{V}$, les matrices K et M étant connues, il reste à déterminer V_0. D'après l'expression obtenue au 3.5.2.2, dans le cas binomial avec une fonction de lien quelconque ; on a ici, avec $h(x) = \dfrac{\exp(x)}{1+\exp(x)}$:

$$\begin{aligned} V_{0_{ii}} &= \frac{1}{n_i}[h(\eta_i^*)(1-h(\eta_i^*)) - h'(\eta_i^*)^2\tilde{u}_i{}'D\tilde{u}_i] \\ &= \frac{\exp(\eta_i^*)}{n_i(1+\exp(\eta_i^*))^2}[1 - \frac{\exp(\eta_i^*)}{(1+\exp(\eta_i^*))^2}u_i'MDMu_i] \ . \end{aligned}$$

La matrice $\tilde{V}$ s'écrit alors :

$$\begin{aligned} \tilde{V} &= V_0 + KMUDU'MK \\ &= V_0 + LUDU'L \end{aligned}$$

avec $L = KM = \left\{_d \dfrac{h'(\eta_i^*)}{\sqrt{1 + c^2 u_i' D u_i}} \right\}$.

3.6.4 Méthode d'estimation

On procède donc à l'estimation par maximisation de la fonction de quasi-vraisemblance. Pour cela, si D est connue, on résout le système itératif suivant :

$$(X'W^{[t]-1}X)\beta^{[t+1]} = X'W^{[t]-1}\zeta^{[t]} \,, \tag{3.5}$$

$$\begin{array}{llcl} \text{avec} & \zeta^{[t]} & = & X\beta^{[t]} + L^{[t]-1}(f - \mu^{[t]}) \\ \text{et} & W^{[t]} & = & L^{[t]-1}V^{[t]}L^{[t]-1} \,. \end{array}$$

D étant inconnue, on approche V par $\tilde{V}$ et on considère :

$$\begin{aligned} \tilde{W}^{[t]} &= L^{[t]-1}\tilde{V}^{[t]}L^{[t]-1} \\ &= L^{[t]-1}V_0^{[t]}L^{[t]-1} + UDU' \\ &= R^{[t]} + UDU' \,, \end{aligned}$$

où R est alors de la forme :

$$\begin{aligned} R &= \left\{_d \frac{1}{n_i}\left[\frac{(1 + exp(\eta_i^*))^2(1 + c^2 u_i' D u_i)}{exp(\eta_i^*)} - u_i' D u_i\right] \right\}_{i=1,\dots,N} \\ &= \left\{_d \frac{1}{n_i}\left[\frac{1 + c^2 u_i' D u_i}{\mu_i(1 - \mu_i)} - u_i' D u_i\right] \right\}_{i=1,\dots,N} . \end{aligned}$$

On obtient donc une solution aux équations (3.5) avec $\tilde{W}$, par résolution des équations de Henderson. Pour cela, on utilise le sous-produit $\hat{\xi}$ obtenu, comme prédiction de ξ ainsi que pour l'estimation des composantes de la variance.

3.6.5 Quelques simulations

Dans le tableau ci-dessous, nous présentons les résultats de quelques simulations. Si nous avons choisi ce cadre particulier du modèle binomial - lien logit pour réaliser des simulations, c'est qu'une méthodologie GAR a pu être envisagée dans ce type de modèle grâce au formalisme proposé.

Dans ce modèle, nous avons simulé des vecteurs de données de taille 20. Un seul effet aléatoire avec 4 réalisations a été introduit. Après avoir imposé un effet fixe nul, nous avons réalisé, pour chaque valeur de σ^2 considérée, 200 simulations. Les résultats des moyennes et écart-types pour β et σ^2 obtenues sont les suivantes :

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

TAB. 3.1 – *Résultats d'estimation des paramètres par la méthode GAR adaptée à un modèle binomial - lien logit.*

Ainsi à partir de $\sigma^2 = 1$ et surtout pour $\sigma^2 = 2$ et $\sigma^2 = 4$, les estimations obtenues se détériorent. Nous retrouvons là le fait 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.

Enfin, nous avons eu l'occasion d'émettre des doutes quant à la prévision de ξ obtenue. Des simulations effectuées ne sont pas venues diminuer ces doutes, en comparant le ξ simulé et le ξ prédit.

Annexe au chapitre 3

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 :

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

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

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

□

Chapitre 4

Une hétérogénéité dans les modèles mixtes

4.1 Introduction

Dans ce chapitre, nous nous intéressons à une nouvelle catégorie de modèles que sont les **modèles linéaires généralisés mixtes à variances hétérogènes**. Comme leur nom l'indique, nous restons dans le cadre d'une modélisation liant effets aléatoires et modèles linéaires généralisés. Notre attention se porte toujours sur l'estimation des paramètres de variance de ces effets aléatoires – les composantes de la variance. Cependant, nous introduisons ici de nouveaux paramètres de variance afin de modéliser l'hétérogénéité, et le but de ce chapitre est d'en proposer une méthode d'estimation.

Si la notion de *modèle linéaire généralisé mixte*, sujet central de cette thèse, a été largement décrite dans les chapitres précédents, il n'en est pas de même en ce qui concerne celle d'*hétérogénéité*. Cette notion d'hétérogénéité peut recouvrir des sens différents selon les auteurs et les contextes, et peut donner lieu à des modélisations variées. Aussi, dans un premier temps, nous précisons le sens que nous lui accordons ici, et nous donnons une définition des modèles étudiés en soulignant les difficultés particulières qu'ils soulèvent pour la mise en place de procédure d'estimation.

Dans ce chapitre et pour atteindre notre objectif, l'algorithme EM va s'avérer un outil essentiel. Ainsi l'objet de la section 4.3 est de revenir tout d'abord sur son utilisation dans le cadre général des modèles à effets aléatoires. Puis, après avoir décrit cet algorithme dans le cas particulier d'un GLM avec surdispersion, nous verrons qu'il semble peu envisageable de prolonger, sans autre détour, cette démarche pour un GL2M quelconque. Malgré tout, nous nous arrêtons dans la section 4.4 sur sa mise en place dans les modèles linéaires mixtes. Ce qui donne alors naissance à différents algorithmes que nous présentons, et souligne le bien-fondé de la démarche EM pour traiter l'hétérogénéité.

C'est pourquoi nous réutilisons cette idée, de façon indirecte cette fois-ci, en section 4.5 pour en arriver à proposer une méthode d'estimation des composantes de la variance dans un GL2M hétérogène. Cette méthode combine à la fois la linéarisation du modèle linéaire généralisé à effets aléatoires, déjà rencontrée au chapitre 2, et l'utilisation des algorithmes proposés dans la section précédente au sein des modèles linéaires mixtes à variances hétérogènes. Nous verrons comment cette méthode, dans le cas particulier de l'homogénéité, nous ramène aux travaux présentés au deuxième chapitre. Des résultats de simulations seront présentés pour observer son comportement.

4.2 Introduction de l'hétérogénéité dans les modèles mixtes

4.2.1 La notion d'hétérogénéité et les L2M à variances hétérogènes

La notion d'hétérogénéité et, à l'opposé, celle d'homogénéité sont utilisées par de nombreux auteurs dans des contextes très variés. Cette multiplicité donne lieu à des définitions très vagues dans le cadre général. D'autant plus vagues qu'à ces deux notions viennent parfois se greffer, voire se confondre, celles d'hétéroscédasticité et d'homoscédasticité. Si bien qu'une différenciation précise de ces termes n'est pas toujours très claire dans la littérature, mais ce n'est pas notre sujet. Nous nous intéressons ici aux modèles *à variances hétérogènes* et nous précisons dans cette section le sens que nous donnons à ces termes.

L'étude de la nature des variations observées dans des données relevées sur une population, est l'un des objectifs principaux de la statistique. Lorsqu'une population se divise en sous-groupes, il est intéressant de savoir si les observations se comportent de manière identique d'un sous-groupe à l'autre. Si c'est le cas, cette répartition en sous-population sera dite dans le langage usuel *homogène*. A contrario, constater une hétérogénéité peut constituer une source d'information importante lors de l'analyse de données. Nous allons voir comment une modélisation peut justement prendre en compte ces différences de comportement d'un sous-groupe à un autre.

Tout d'abord, le découpage en sous-groupes (ou classes ou sous-population) peut être précis ou au contraire flou. En effet, dans certaines situations une hétérogénéité des données est pressentie sans bien savoir à quel découpage elle correspond précisément ; autrement dit, si telle donnée appartient ou non à tel sous-groupe. Les modèles de mélange apparaissent alors comme un outil naturel permettant de prendre en compte ce type d'hétérogénéité. Des lois sont supposées pour chaque classe, et en affectant à chaque donnée une certaine probabilité d'appartenance aux différentes classes, ces modèles permettent de respecter la non connaissance exacte du découpage. Dans ce cadre, on peut se référer notamment aux travaux de DIETZ (1992), DIETZ et BÖHNING (1995), qui étudient justement les mélanges de GLM.

Ici, nous nous intéressons davantage au cas où les sous-populations sont parfaitement définies par l'expérience : on connait le groupe auquel appartient chaque donnée. Ce découpage est indicé par i. Au cours de la modélisation, un vecteur aléatoire est introduit, dont plusieurs réalisations interviennent (non exclusivement) dans la $i^{\text{ème}}$ sous-population. Alors pour traduire des différences de comportement d'une sous-population à l'autre, on suppose que la composante de ce vecteur, associée à la $i^{\text{ème}}$ sous-population, est distribuée normalement avec un paramètre de variance égal à σ_i^2. L'hétérogénéité s'exprime alors par une différence entre ces paramètres de variance σ_i^2 à travers les sous-groupes.

Il est possible d'insérer cette description de l'hétérogénéité à différents niveaux. Dans le cadre des modèles linéaires, le vecteur aléatoire introduit précédemment, et sur la distribution duquel s'exprime l'hétérogénéité, peut correspondre à deux sources d'aléa. Il représente, en effet, soit la composante résiduelle du modèle, soit le vecteur d'effets aléatoires. Nous décrivons ces deux situations.

- Cas 1 : l'hétérogénéité se situe au niveau de la distribution des *erreurs*.
  Dans cette situation, le vecteur aléatoire est de même taille que celui des données. Chaque réalisation est associée à une et une seule donnée. Le vecteur peut alors être découpé selon les sous-populations. Dans ce cadre, on pourra considérer un modèle linéaire avec ou sans effets aléatoires. AITKIN (1987) envisage le cas du modèle linéaire à variances hétérogènes lorsqu'il reprend la modélisation de l'hétérogénéité proposée par COOK et WEISBERG (1983). Dans son travail, chaque donnée correspond à un sous-groupe, et de plus un modèle log-linéaire pour les différents paramètres de variance est supposé. D'autres auteurs, comme FOULLEY, SAN CRISTOBAL, GIANOLA, et IM (1992), ont considéré le cas des modèles linéaires mixtes à variances hétérogènes, en supposant de même un modèle pour la variance.

- Cas 2 : l'hétérogénéité se situe au niveau de la distribution des *effets aléatoires*.
  Le vecteur aléatoire n'est alors plus de même taille que celui des données. Une même réalisation peut intervenir auprès de plusieurs données et le vecteur ne peut alors plus se découper selon les sous-populations. Ainsi, pour décrire cela, on suppose qu'un deuxième découpage de la population, indicé par j celui-là,

est effectué, emboîté ou croisé au découpage précédent. On oublie un instant le premier découpage et l'on suppose qu'à la $j^{\text{ème}}$ sous-population est associé le niveau d'un facteur ξ_j. Ce facteur est ensuite considéré au cours de la modélisation comme facteur à effet aléatoire, c'est-à-dire que les ξ_j sont des réalisations indépendantes d'une loi normale centrée de variance σ^2. Le vecteur $\xi = (\xi_1, ..., \xi_j, ...)'$ rassemble ces réalisations, que nous n'observons pas directement au cours de l'expérience. En reprenant maintenant le découpage selon les classes, on renouvelle les hypothèses précédentes et l'on suppose que les ξ_j dans la $i^{\text{ème}}$ classe sont des réalisations indépendantes d'une loi normale centrée de variance σ_i^2, cette fois-ci dépendante (indice i) de la classe. Dans cette situation alors, l'hypothèse d'hétérogénéité se traduit encore par les différences entre les σ_i^2. Cette hétérogénéité concernant les effets aléatoires, avec ou sans association à une hétérogénéité résiduelle, correspond au type de modélisation considéré notamment par Gianola (1986), ou encore Foulley et Quaas (1995).

Notons que certains auteurs ont adopté d'autres définitions de la notion d'hétérogénéité. Selon Jacqmin-Gadda et Commenges (1995) par exemple, cette notion se rapporte davantage à la présence ou non des effets aléatoires. Dans notre travail, nous conservons cependant la description précédente du cas 2. Cela répond à de nombreuses situations pratiques en génétique animale. Garrick, Pollak, Quaas, et Van Vleck (1989) le soulignent par des exemples dans le domaine de la production animale. Hill (1984) a mis en évidence qu'il était important de ne pas oublier de prendre en compte cette hétérogénéité dans des procédures d'évaluation génétique. Depuis, divers travaux ont eu lieu dans ce domaine.

Ainsi, c'est cette définition, selon laquelle l'hétérogénéité se traduit par des différences sur les paramètres de variance d'un vecteur aléatoire, et plus précisément des effets aléatoires, que nous reprenons dans ce travail en l'élargissant au cadre des modèles linéaires généralisés mixtes. Nous décrivons plus précisément dans la sous-section suivante les modèles étudiés.

4.2.2 Définition des GL2M à variances hétérogènes

La modélisation avec variances hétérogènes présente un intérêt certain en pratique. Mais jusqu'à présent, les travaux réalisés (notamment par Gianola (1986), Aitkin (1987), Gianola et al. (1992), Foulley et Quaas (1995)) concernent principalement l'hétérogénéité dans les modèles linéaires ou modèles linéaires mixtes. Nous reprenons ici le point de vue selon lequel l'hétérogénéité se traduit par des hypothèses sur la distribution des effets aléatoires, et nous l'étendons au sein des modèles linéaires généralisés mixtes.

Pour cela, nous supposons une certaine stratification des données et c'est relativement à celle-ci que s'énoncera une fois de plus l'hypothèse d'hétérogénéité. Dans le contexte de la génétique animale, les différents niveaux de cette stratification correspondent souvent à différents environnements (pouvant être des troupeaux, des régions ...). Nous supposons disposer de I environnements et nous cherchons à mettre en évidence des différences de comportements entre les environnements. En ce qui concerne les effets aléatoires, nous ne faisons pas ici d'hypothèses supplémentaires : ils peuvent être croisés, emboîtés ou correspondre à une surdispersion. Nous présentons dans un premier temps les modèles avec un seul effet aléatoire, puis nous en donnerons ultérieurement l'extension à un nombre K d'effets.

Pour définir les modèles linéaires généralisés mixtes à variances hétérogènes, nous reprenons les trois hypothèses classiques de définition des modèles linéaires généralisés mixtes. Parmi elles, les hypothèses d'une part sur la *distribution conditionnelle* aux effets aléatoires du vecteur réponse, que l'on suppose appartenir à la famille exponentielle, d'autre part sur la *fonction de lien* reliant l'espérance conditionnelle et le prédicteur linéaire, restent identiques. C'est par contre dans l'expression du *prédicteur linéaire* qu'intervient la modélisation de l'hétérogénéité. En effet, pour tout environnement i, on définit le prédicteur linéaire associé au sous-vecteur Y_i (de taille n_i) du vecteur réponse (observations pour le $i^{\text{ème}}$ environnement), de la façon suivante :

$$\forall i \in \{1, ..., I\}\ , \qquad \eta_i = X_i\beta + \sigma_i U_i \xi$$

où X_i et U_i (respectivement $n_i \times p$ et $n_i \times q$) sont les sous-matrices de X et U fixées par l'expérience, et β le vecteur des paramètres fixes inconnus de taille p. Le vecteur aléatoire ξ, à q composantes, est le vecteur des effets aléatoires. Il est distribué selon la loi normale multivariée centrée réduite. Il se réalise en $\xi = (\xi_1, \ldots, \xi_j, \ldots, \xi_q)'$ au cours de l'expérience, réalisations qui ne sont pas observées directement.

Remarque : Nous rappelons que, pour des commodités d'écriture, et dans tout le document, nous désignons par ξ à la fois le vecteur aléatoire et sa réalisation.

En notant x'_{im} et u'_{im} les $m^{\text{èmes}}$ lignes de X_i et U_i, le prédicteur linéaire pour la $m^{\text{ème}}$ observation de l'environnement i est donc :

$$\eta_{im} = x'_{im}\beta + \sigma_i u'_{im}\xi\ .$$

Dans cette expression du prédicteur linéaire, il est important de noter les deux points suivants.

- Les paramètres de variance σ_i^2 dépendent de chaque environnement (indice i). L'hétérogénéité naît donc dans cette différence exprimée d'un environnement à l'autre. Ils ont été sortis de la distribution du vecteur d'effets aléatoires ξ et interviennent comme paramètre d'échelle. La loi normale supposée a alors une variance unité. Si ξ était un vecteur fixe, ces paramètres de variance pourraient être vus comme de nouveaux paramètres de régression.

- Le vecteur ξ est le seul non indicé par i. Autrement dit, c'est le même vecteur aléatoire qui intervient dans les différents environnements. Ce qui signifie qu'un découpage de ξ, en séparant les différents niveaux du facteur à effets aléatoires selon les environnements, n'est pas toujours réalisable. Il se peut qu'une même réalisation d'un des niveaux du facteur intervienne dans plusieurs environnements. Ce sont ensuite les éléments de la matrice U qui en donneront l'indication. En pratique, cela se traduit de la façon suivante. Reprenons l'exemple classique en génétique animale de modélisation des mesures sur les veaux et où à chaque taureau père, on fait correspondre une réalisation d'un facteur à effet aléatoire. On imagine alors qu'un même taureau peut être transféré dans plusieurs environnements. La réalisation de l'effet aléatoire qui lui sera affectée interviendra alors dans deux environnements différents i et i', associée à deux prédicteurs différents η_i et $\eta_{i'}$ avec deux paramètres d'échelle différents σ_i et $\sigma_{i'}$. Cette hypothèse ne pouvait alors se traduire autrement que dans une écriture où le paramètre σ_i est sorti de la distribution de ξ, sans quoi une même réalisation n'aurait pu être réalisation de 2 lois différentes.

Ces deux points traduisent la délicate introduction de l'hétérogénéité dans ces modèles. À cause du deuxième point mentionné, alors que les paramètres σ_i avaient été sortis de la distribution de ξ au début par commodité d'écriture, il semble maintenant impossible de les y replacer : on doit les maintenir comme facteur d'échelle.

Ces remarques auront toute leur importance ensuite lors de la mise en place des procédures d'estimation. En effet, sans l'hétérogénéité, dans les modèles linéaires mixtes, le paramètre σ^2 peut être maintenu comme paramètre de variance dans la distribution de l'effet aléatoire. On a vu alors dans les chapitres précédents que la variance empirique des valeurs prédites (i.e. des valeurs estimées des réalisations non observées : $\hat{\xi}_k,\ \ k = 1, ..., q$) de l'effet aléatoire permet d'en estimer la variance. Ici, on ne peut plus se ramener à cela puisqu'une même réalisation peut être associée à deux paramètres de variance différents. Gianola (1986) explique cependant que la procédure BLUP usuelle est adaptable au cas hétérogène. Mais il envisage pour cela une modélisation différente. Pour tout j, il répète la réalisation de l'effet aléatoire ξ_j autant de fois qu'elle intervient dans des environnements différents. Puis il affecte chacune de ces répétitions aux environnements concernés. Enfin, bien entendu, pour modéliser le fait que ces répétitions sont fortement liées entre elles (puisqu'issues d'une même réalisation), il introduit des coefficients de corrélation. Il nous semble cependant plus naturel de garder une seule et même réalisation au sein du modèle lorsque c'est un même taureau par exemple qui intervient dans des environnements différents. Mais cela implique de nouvelles difficultés pour définir des procédures d'estimation. C'est ce que nous étudions dans les sections 4.4 et 4.5.

Pour finir, l'extension à la modélisation avec plusieurs effets aléatoires est immédiate. Supposons K effets aléatoires : $\xi = (\xi_1', ..., \xi_j', ..., \xi_K')'$ où ξ_j $(n_j \times q_j)$ contient maintenant les q_j réalisations du $j^{\text{ème}}$ effet aléatoire. Le prédicteur linéaire s'exprime alors de la façon suivante :

$$\eta_i = X_i\beta + \sum_{j=1}^{K} \sigma_{ij} U_{ij} \xi_j$$

où σ_{ij}^2 est la variance du $j^{\text{ème}}$ effet aléatoire dans le $i^{\text{ème}}$ environnement.

4.3 Utilisation directe de l'algorithme EM dans les modèles mixtes

Nous venons d'exposer les hypothèses de modélisation que nous adoptons afin de traduire une certaine hétérogénéité des données. Ces hypothèses concernent les paramètres de variance des effets aléatoires. La question naturelle qui en découle est bien entendu celle de l'estimation de ces paramètres : nous rejoignons alors le point d'ancrage principal de cette thèse. Pour cette estimation, nous l'avons déjà brièvement évoqué et nous y reviendrons, la plupart des algorithmes usuels n'ont plus cours. L'algorithme EM présente, quant à lui, de nombreux atouts. Nous consacrons la section 4.4 à son utilisation au sein des L2M, et la section 4.5 à l'utilisation que nous en proposons au sein des GL2M hétérogènes. Mais avant cela, nous décrivons dans cette section d'une part l'idée qui anime cet algorithme lorsqu'il s'agit d'un modèle quelconque à effets aléatoires et d'autre part, dans les GL2M, les limites de son utilisation directe au cas de données surdispersées.

4.3.1 La démarche EM dans les modèles à effets aléatoires

Depuis sa mise en place par Dempster, Laird, et Rubin (1977), l'algorithme EM a permis, dans des contextes variés, de résoudre de nombreux problèmes liés à l'estimation de paramètres. En effet, cet algorithme constitue un outil conceptuellement simple pour obtenir des estimations du maximum de vraisemblance. Il permet, dans diverses situations, de contourner la difficulté d'obtention de la vraisemblance des observations lorsque la distribution "marginale" de ces observations est délicate à spécifier. Il réalise cela par l'introduction de données manquantes, que l'on n'observe pas au cours de l'expérience, mais dont on connaît la vraisemblance jointe aux données observées. Cet algorithme est à la base de nombreux travaux et diverses extensions en ont été proposées (cf. McLachlan et Krishnan 1997).

Les modèles à effets aléatoires sont un cadre naturellement propice à l'utilisation de cette méthodologie. En effet, dans ces modèles, la distribution conditionnelle du vecteur réponse conditionnellement aux effets aléatoires, et la distribution marginale des effets aléatoires sont connues. Ainsi, la distribution jointe du vecteur réponse Y et des effets aléatoires ξ s'obtient immédiatement. Ce qui est loin d'être le cas de la distribution marginale de Y.
Puisque nous n'observons pas les effets aléatoires, ils joueront logiquement le rôle des données manquantes. Et la distribution jointe précédente constitue alors la distribution des données complètes. En notant Φ le vecteur des paramètres à estimer, ceci se décrit par : $f(Y,\xi|\Phi) = f(Y|\xi,\Phi).f(\xi|\Phi)$ (on adopte ici la notation générique de f comme fonction de densité des lois des variables indiquées).
L'algorithme est itératif et se découpe à chaque pas en deux étapes. Soit $\Phi^{(t)}$ la valeur des paramètres au pas t, nous décrivons ces deux étapes de la façon suivante :

- l'étape E (Expectation) : les effets aléatoires n'étant pas observés, on remplace la log-vraisemblance des données complètes par son espérance selon la distribution conditionnelle des effets aléatoires sachant les données observées et l'on s'intéresse à la fonction de Φ suivante :
$$Q(\Phi|\Phi^{(t)}) = E[\ln(f(Y,\xi|\Phi))|y,\Phi^{(t)}]$$
- l'étape M (Maximization) : on maximise $Q(\Phi|\Phi^{(t)})$ pour obtenir $\Phi^{(t+1)}$:
$$\Phi^{(t+1)} = \text{argmax}(Q(\Phi|\Phi^{(t)}))\ .$$

Pour finir, on itère ces deux étapes jusqu'à convergence. De façon générale, de nombreux travaux théoriques, pour étudier les conditions de convergence et de convergence vers le maximum de vraisemblance de cet algorithme (ou de ses extensions), ont été réalisés (cf. Wu 1983). Nous n'insistons pas ici sur ce point.

Dans le cadre des modèles linéaires mixtes, cet algorithme en a rejoint d'autres permettant d'obtenir des estimations du maximum de vraisemblance ou maximum de vraisemblance restreint. Il a notamment l'avantage de pouvoir éventuellement s'étendre à des hypothèses de lois sur les effets aléatoires autres que la loi normale, à partir du moment où l'on dispose de la loi de ces effets conditionnellement aux observations. Dans la section suivante, nous verrons comment cet algorithme s'adapte très bien à l'introduction de l'hétérogénéité dans les L2M.

Pour les modèles linéaires généralisés mixtes, de nombreux travaux ont été réalisés. La difficulté principale est alors liée au calcul de l'intégrale à l'étape E : que ce soit pour le calcul de l'espérance de la log-vraisemblance complète ou pour celui de sa dérivée (après échange des signes de somme et de dérivation puisque l'on cherche à maximiser la fonction Q). Pour réaliser ce calcul, Bock et Aitkin (1981) ont proposé l'utilisation d'une approximation numérique : la quadrature gaussienne. Cette idée a été reprise par de nombreux auteurs par la suite. Citons notamment Hinde (1982) pour des données Poissonniennes (il programme la méthode en GLIM), Anderson et Aitkin (1985), Jansen et Hoekstra (1993) pour des données binaires (ou multicatégories), et dans un cadre général Anderson et Hinde (1988). Nous décrivons plus précisément cette démarche dans la sous-section suivante. Cependant, il est important de noter d'ores et déjà que ces travaux concernent essentiellement les cas de données surdispersées, tout en s'adaptant plus ou moins facilement au cas d'effets aléatoires emboîtés.

D'autres développements ont été réalisés à ce sujet. Im et Gianola (1988) ont comparé cette méthode avec une procédure du simplex. A l'instar des travaux de Stiratelli, Laird, et Ware (1984) qui réalisent à la fois une approximation numérique et analytique au premier ordre de l'intégrale, Steele (1996) propose de remplacer la quadrature gaussienne par l'approximation analytique de Laplace. Enfin, en levant les hypothèses faites sur la distribution des effets aléatoires, Aitkin (1996) propose une estimation maximum de vraisemblance non paramétrique de cette distribution. On retombe alors dans le cadre des modèles de mélange.

Notons enfin, même si cela s'écarte de notre sujet, que l'algorithme EM a suscité de même de nombreux travaux dans le cadre des modèles non linéaires à effets aléatoires. Walker (1996) a dernièrement proposé un algorithme où l'étape E est réalisée par une méthode de Monte Carlo.

4.3.2 Limites de l'algorithme EM au cas du GLM surdispersé

Si l'on applique directement le raisonnement EM dans le cadre des modèles linéaires généralisés mixtes, on butte sur l'obstacle du calcul de l'espérance, réalisable avec la loi normale grâce aux règles de conditionnement, mais plus difficile pour d'autres lois. Devant la difficulté de ce calcul intégral, divers auteurs ont pris le parti, dans certains cas, d'utiliser une approximation par quadrature gaussienne. Nous décrivons ici plus précisément cette démarche. Cependant cette description est réservée au cas d'un GLM surdispersé, la surdispersion étant modélisée par l'introduction d'un effet aléatoire, dont on affecte à chaque donnée une réalisation différente. Ceci nous donnera l'occasion de souligner pourquoi cette démarche d'approximation semble peu adaptée au cas d'effets aléatoires quelconques (et notamment non emboîtés).

On note toujours Y le vecteur à expliquer, et ξ le vecteur d'effets aléatoires exprimant la surdispersion. Selon l'hypothèse de modélisation de la surdispersion décrite ci-dessus, Y et ξ sont de même taille et la $i^{\text{ème}}$ composante du prédicteur linéaire s'écrit : $\eta_i = x_i'\beta + \xi_i$. On suppose : $\xi_i \sim \mathcal{N}(0, \sigma^2)$ et pour tout i différent de j, ξ_i et ξ_j indépendants. D'autre part, conditionnellement à ξ, les composantes Y_i de Y sont indépendantes, et distribuées selon une loi de la famille exponentielle. On peut réécrire le prédicteur linéaire en sortant le paramètre σ sous la forme :

$$\eta_i = x_i'\beta + \sigma\xi_i \quad \text{où} \quad \xi_i \sim \mathcal{N}(0,1)\ . \tag{4.1}$$

Cette écriture, très utile en pratique, est souvent adoptée. Elle est indispensable, comme nous l'avons vu dans la section précédente, pour décrire les modèles avec variances hétérogènes. Les ξ_i vont donc jouer le rôle des données manquantes et l'on note $\Phi = (\beta', \sigma)'$, le vecteur des paramètres inconnus.

Remarque : Dorénavant, on conserve l'hypothèse d'une distribution normale centrée réduite des composantes de ξ. On désignera par f les différentes fonctions de densité des lois mises en jeu.

La première étape de l'algorithme EM nous conduit à nous intéresser à la fonction :

$$Q(\Phi|\Phi^{(t)}) = E[\ln(f(y,\xi|\Phi))|y, \Phi^{(t)}]\ .$$

Ici, de par l'hypothèse de surdispersion, les différents couples (Y_i, ξ_i) sont indépendants, mais ce n'est pas le cas autrement. Ainsi, on peut écrire : $\ln[f(y,\xi|\Phi)] = \sum_{i=1}^{N} \ln[f(y_i,\xi_i|\Phi)]$.

$$\begin{aligned} \text{D'où}\quad Q(\Phi|\Phi^{(t)}) &= \sum_{i=1}^{N} \int_{\mathbf{R}} \ln[f(y_i,\xi_i|\Phi)] f(\xi_i|y_i, \Phi^{(t)}) d\xi_i \\ &= \sum_{i=1}^{N} k_i^{-1} \int_{\mathbf{R}} \ln[f(y_i,\xi_i|\Phi)] f(y_i|\xi_i, \Phi^{(t)}) f(\xi_i) d\xi_i\ . \end{aligned}$$

On désigne par k_i la constante de normalisation résultant de l'application de la règle de Bayes : $k_i = \int_{\mathbf{R}} f(y_i|\xi_i, \Phi^{(t)}) f(\xi_i) d\xi_i$.

On s'aperçoit donc que l'hypothèse de surdispersion se révèle importante puisque non seulement elle n'a pas introduit de dépendance entre les différents couples (Y_i, ξ_i), mais elle a aussi permis de réduire l'intégrale multiple à N dimensions (sur toutes les composantes des effets aléatoires) en une somme d'intégrales simples.

Lors de la seconde étape, c'est la maximisation en Φ de cette fonction $Q(\Phi|\Phi^{(t)})$ qui nous intéresse. Par inversion des signes intégral et dérivée, on obtient pour la dérivée de Q par rapport à la $r^{\text{ème}}$ composante de Φ :

$$\frac{\partial Q(\Phi|\Phi^{(t)})}{\partial \Phi_r} = \sum_{i=1}^{N} k_i^{-1} \int_{\mathbf{R}} \frac{\partial \ln[f(y_i, \xi_i|\Phi)]}{\partial \Phi_r} f(y_i|\xi_i, \Phi^{(t)}) f(\xi_i) d\xi_i \ .$$

De plus, $\ln[f(y_i, \xi_i|\Phi)] = \ln[f(y_i|\xi_i, \Phi)] + \ln[f(\xi_i)]$. Or, grâce à l'écriture (4.1) précédente du prédicteur linéaire, le paramètre de variance a été sorti de la distribution des effets. Ainsi $f(\xi_i)$ ne dépend pas de Φ. On aboutit alors à :

$$\frac{\partial Q(\Phi|\Phi^{(t)})}{\partial \Phi_r} = \sum_{i=1}^{N} k_i^{-1} \int_{\mathbf{R}} \frac{\partial \ln[f(y_i|\xi_i, \Phi)]}{\partial \Phi_r} f(y_i|\xi_i, \Phi^{(t)}) f(\xi_i) d\xi_i \ .$$

Au vu de cette expression, c'est donc la distribution conditionnelle du vecteur réponse sachant les effets aléatoires que l'on va dériver par rapport aux différentes composantes de β et par rapport à σ. Et, c'est cette intégrale de dimension 1 que l'on approche par quadrature gaussienne. Cette technique consiste à discrétiser l'intégrale. C'est à dire qu'elle est remplacée par une somme finie en l points que l'on appelle les points de quadrature (cf. Stroud et Secrest 1966) et que l'on note ici ξ_m, $m = 1, \ldots, l$. Dans cette somme en pratique, les valeurs $k_i^{-1} f(y_i|\xi_m, \Phi^{(t)}) f(\xi_m)$ sont remplacées par des poids $w_{m,i}$, avec $\sum_{m=1}^{l} w_{m,i} = 1$. Le nombre de points de quadrature l est à fixer au préalable. On obtient ainsi :

$$\frac{\partial Q(\Phi|\Phi^{(t)})}{\partial \Phi_r} = \sum_{i=1}^{N} \sum_{m=1}^{l} w_{m,i} \frac{\partial \ln[f(y_i|\xi_m, \Phi)]}{\partial \Phi_r} \ .$$

Pour poursuivre le développement de cette expression, on reprend les notations du premier chapitre, où b désigne l'application qui relie l'espérance μ au paramètre canonique θ d'une loi de la famille exponentielle ($\mu_i = b'(\theta_i)$). Et on a :

$$\frac{\partial \ln[f(y_i|\xi_m, \Phi)]}{\partial \Phi_r} = \frac{\partial \theta_i}{\partial \Phi_r} [y_i - b'(\theta_i)] \ .$$

On cherche donc, pour chaque composante de Φ, à résoudre l'équation :

$$\sum_{m=1}^{l} \sum_{i=1}^{N} w_{m,i} \frac{\partial \theta_i}{\partial \Phi_r} [y_i - b'(\theta_i)] = 0 \quad \text{où} \quad b'(\theta_i) = g^{-1}(x_i'\beta + \sigma \xi_m) \ .$$

La solution du système ainsi obtenue nous donne l'estimation recherchée du vecteur des paramètres Φ. Sous cette forme, le paramètre σ a rejoint β comme paramètre de régression associé au nouveau vecteur de régression constitué par les points de quadratures ξ_m. Ainsi, pour résoudre ces équations, on peut reprendre la procédure des moindres carrés généralisés itérés, utilisée pour l'estimation des paramètres de régression dans les modèles linéaires généralisés classiques. A chaque itération, les poids $w_{m,i}$ devront être recalculés. Ceci combine donc les techniques EM et GLM.

Cette procédure, décrite dans le cas d'un modèle avec surdispersion, s'adapte plus ou moins facilement au cas d'effets aléatoires emboîtés. Anderson et Aitkin (1985), Im et Gianola (1988) ont envisagé le cas d'un emboîtement à deux niveaux. Dans un cadre général toutefois (effets aléatoires croisés par exemple), Steele (1996) notamment souligne la délicate adaptation de cette méthode pour des intégrales multiples. De plus, en génétique animale en général, les bases de données sont souvent très grandes et exigent des calculs informatiques importants. Cette complexité, combinée à celle d'un algorithme aussi exigeant à la base, ne facilite en rien l'utilisation de cette méthode.

D'autre part, il reste que la rapidité de convergence de l'algorithme est mal maitrisée. Les remarques faites à ce sujet dans le cas linéaire par Thompson et Meyer (1986), nous incitent ici à dire que plus la valeur de l'estimation de σ^2 est petite (tend vers zéro), plus la vitesse de convergence va avoir tendance à diminuer. Pour accélerer cette convergence, certaines techniques peuvent être proposées comme l'accélération d'Aitken.

Enfin, un autre inconvénient à noter pour cette méthode est que cet algorithme ne fournit pas d'erreur standard pour les estimations (alors qu' Im et Gianola (1988) souligne que la méthode du simplex peut être adaptée pour donner la variance asymptotique et offre d'autre part une convergence plus rapide).

Ceci dit, comme nous allons le voir, même si l'utilisation de l'approximation gaussienne nous semble peu attrayante pour une mise en place directe de l'algorithme EM dans les GL2M, ceci n'enlève en rien son intérêt pour appréhender l'hétérogénéité dans les modèles mixtes.

4.4 Estimation des composantes de la variance dans les L2M hétérogènes

Nous nous limitons dans cette section au cas linéaire gaussien. Nous présentons dans un premier temps un exemple d'une telle modélisation, pour ensuite nous intéresser plus particulièrement au problème de l'estimation. Nous expliquons pourquoi certaines démarches usuelles, même dans ce cas particulier de la loi normale, semblent limitées face à l'hétérogénéité. Ce n'est cependant pas le cas de la démarche EM et nous décrivons différentes procédures (soit existantes, soit que nous proposons) basées sur cette démarche. Pour des raisons de clarté de présentation, nous considérons uniquement, dans toute cette section, des modèles avec un seul effet aléatoire. Notons néanmoins que ceci peut se réécrire dans le cas de plusieurs effets aléatoires.

4.4.1 Un exemple simple

Le petit exemple suivant est issu de Foulley et Quaas (1995) et concerne le domaine de la génétique animale, mais peut être repris dans différents cadres. C'est un exemple simplifié de modèle linéaire mixte avec un seul effet aléatoire à variances hétérogènes. Les différents algorithmes évoqués au cours de cette section seront appliqués à cet exemple. Et le même plan d'expérience sera utilisé en simulations.

Il s'agit d'un relevé (tableau 4.1) de 36 observations sur des animaux provenant de 3 environnements différents. Un effet fixe est associé à chacun de ces environnements et un effet aléatoire est associé au père géniteur. Les 36 sujets sont issus de 4 pères différents qui interviennent tous (sauf pour le père 1) dans les 3 environnements. L'effet aléatoire a donc 4 réalisations croisées avec les 3 environnements.

On associe alors à cette expérience le modèle linéaire mixte à variances hétérogènes suivant :
$\forall i \in \{1,2,3\}$, (environnement)
$\forall j \in \{1,2,3,4\}$, (père)
$\forall k \in \{1,\dots,n_{ij}\}$, (n_{ij} nombre d'individus de père j dans l'environnement i)

$$Y_{ijk} = \beta_i + \sigma_i \xi_j + \varepsilon_{ijk} \tag{4.2}$$

où $\quad \xi \sim \mathcal{N}(0, I_4) \quad$ et $\quad \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_{e_i}^2 I_{36})$.

De cette façon, il est possible de prendre en compte les deux types d'hétérogénéité :
- celle sur l'effet aléatoire : avec des σ_i différents pour les 3 environnements,
- celle sur les erreurs : avec des σ_{e_i} différents pour les 3 environnements.

No Indiv.	Observ.	Environ.	Père	No Indiv.	Observ.	Environ.	Père
1	470	1	1	19	385	2	1
2	510	1	1	20	450	2	2
3	345	1	1	21	605	2	2
4	395	1	1	22	575	2	3
5	450	1	2	23	530	2	4
6	345	1	2	24	310	2	4
7	495	1	2	25	415	2	4
8	410	1	3	26	370	2	4
9	335	1	3	27	805	3	2
10	362	1	3	28	475	3	2
11	480	1	3	29	875	3	3
12	410	1	4	30	850	3	3
13	330	1	4	31	510	3	3
14	300	1	4	32	310	3	3
15	330	1	4	33	565	3	4
16	530	2	1	34	330	3	4
17	880	2	1	35	410	3	4
18	575	2	1	36	480	3	4

TAB. 4.1 – *Exemple de modélisation par un modèle linéaire mixte à variances hétérogènes. Données présentées dans* FOULLEY *et* QUAAS *(1995)* .

Notons sur cet exemple qu'il a été nécessaire d'adopter l'écriture du modèle avec σ_i sorti de la distribution de ξ. Dans le cas des erreurs par contre, vu qu'à chaque donnée est associée (par définition) une réalisation, il reste possible de maintenir le paramètre de variance à l'intérieur de la distribution de ε_{ijk}.
Dans la suite, cet exemple sera utilisé à diverses occasions.

4.4.2 Limites de certaines démarches usuelles

Dans les L2M hétérogènes, notre objectif reste l'estimation des paramètres et en particulier des différents paramètres de variance. Comme nous l'avons évoqué brièvement dans la section 4.2, les procédures usuelles d'estimation existantes pour les L2M homogènes semblent mal s'adapter à l'introduction de l'hétérogénéité. C'est ce dont nous discutons ci-dessous.

Considérons le cas simple du modèle linéaire mixte suivant avec un seul effet aléatoire (à q réalisations) et deux environnements (dans lesquels nous disposons respectivement de n_1 et n_2 observations, $N = n_1 + n_2$), une hétérogénéité de l'effet aléatoire, mais pas des erreurs :

$$\begin{array}{lcl} \text{environnement 1} & : & Y_1 = X_1\beta + \sigma_1 U_1 \xi + \varepsilon_1 \\ \text{environnement 2} & : & Y_2 = X_2\beta + \sigma_2 U_2 \xi + \varepsilon_2 \ , \end{array}$$

où :

$$\begin{array}{ccc} \varepsilon = (\varepsilon_1', \varepsilon_2')' & \sim & \mathcal{N}(0, \sigma_0^2 I_N) \\ \xi & \sim & \mathcal{N}(0, I_q) \ . \end{array}$$

Dans un premier temps, intéressons nous à une méthode directe de résolution des équations de maximum de vraisemblance. Pour cela, nous écrivons ces équations. La distribution de Y est :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} X_1\beta \\ X_2\beta \end{pmatrix}, \begin{pmatrix} \sigma_1^2 U_1 U_1' + \sigma_0^2 I_{n_1} & \sigma_1\sigma_2 U_1 U_2' \\ \sigma_1\sigma_2 U_1' U_2 & \sigma_2^2 U_2 U_2' + \sigma_0^2 I_{n_2} \end{pmatrix} \right) \ .$$

Notons V sa matrice de variance. Remarquons dès à présent que les paramètres de variance σ_i^2 de cette matrice ne peuvent pas être factorisés, au contraire du cas où $\sigma_1^2 = \sigma_2^2$. La vraisemblance de $\Phi = (\beta, \sigma_0^2, \sigma_1^2, \sigma_2^2)'$ au vu des observations y s'écrit alors :

$$l(\Phi; y) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln(|V|) - \frac{(y - X\beta)'V^{-1}(y - X\beta)}{2}$$

où $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

Il s'agit ensuite de dériver cette vraisemblance par rapport aux composantes de Φ. Pour cela, on utilise les résultats :

$$\frac{\partial V^{-1}}{\partial \sigma^2} = -V^{-1}\frac{\partial V}{\partial \sigma^2}V^{-1} \quad \text{et} \quad \frac{\partial \ln(|V|)}{\partial \sigma^2} = \mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \sigma^2}\right) \ .$$

Ce qui conduit à :

$$\frac{\partial l(\Phi; y)}{\partial \beta} = 0 \iff X'V^{-1}(y - X\beta) = 0 \ ,$$

$\forall i \in \{0, 1, 2\}$,

$$\frac{\partial l(\Phi; y)}{\partial \sigma_i^2} = 0 \iff -\frac{1}{2}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \sigma_i^2}\right) + \frac{1}{2}(y - X\beta)'V^{-1}\frac{\partial V}{\partial \sigma_i^2}V^{-1}(y - X\beta) = 0 \ ,$$

avec :

$$
\begin{aligned}
\frac{\partial V}{\partial \sigma_1^2} &= \begin{pmatrix} U_1U_1' & \frac{\sigma_2}{\sigma_1}U_1U_2' \\ \frac{\sigma_2}{\sigma_1}U_1'U_2 & 0 \end{pmatrix} \quad \text{et} \quad \frac{\partial V}{\partial \sigma_2^2} = \begin{pmatrix} 0 & \frac{\sigma_1}{\sigma_2}U_1U_2' \\ \frac{\sigma_1}{\sigma_2}U_1'U_2 & U_2U_2' \end{pmatrix} \\
\frac{\partial V}{\partial \sigma_0^2} &= I_N
\end{aligned}
$$

Comme dans le cas classique, on aboutit donc à un système d'équations non linéaire mais dont il est difficile cette fois-ci de proposer une transformation permettant l'élaboration d'un algorithme itératif simple. En effet, la présence des rapports $\frac{\sigma_1}{\sigma_2}$ et $\frac{\sigma_2}{\sigma_1}$ dans les matrices de variance dérivées en constitue un frein. À la différence du cas homogène, il n'y a pas de simplification naturelle. Ainsi aboutir directement à un système ML (ou REML) à résolution itérative simple comme dans le cas du L2M homogène semble donc peu envisageable dans le cas du L2M hétérogène.

Intéressons nous maintenant à un autre type de démarche et essayons de poursuivre un raisonnement type Henderson. Pour cela, nous cherchons tout d'abord à écrire la loi du couple (Y, ξ). Nous l'obtenons comme produit de la loi conditionnelle à ξ par la loi de ξ :

$$
\begin{aligned}
f_{Y,\xi}(y,\xi) &= f_{Y|\xi}(y) * f_\xi(\xi) \\
&= -\frac{N}{2}\ln(\sigma_0^2) - \frac{1}{2}\sum_{i=1}^{2}\frac{(y_i - X_i\beta - \sigma_iU_i\xi)'(y_i - X_i\beta - \sigma_iU_i\xi)}{\sigma_0^2} - \frac{1}{2}\xi'\xi \;.
\end{aligned}
$$

La démarche consiste ensuite à dériver cette expression par rapport à β et ξ :

$$
\begin{aligned}
\frac{\partial f_{Y,\xi}(y,\xi)}{\partial \beta} &= -\frac{1}{\sigma_0^2}\sum_{i=1}^{2}X_i'X_i\beta - X_i'(y_i - \sigma_iU_i\xi) \;, \\
\frac{\partial f_{Y,\xi}(y,\xi)}{\partial \xi} &= -\frac{1}{\sigma_0^2}\sum_{i=1}^{2}\sigma_i^2U_i'U_i\xi - \sigma_iU_i'(y_i - X_i\beta) - \xi \;.
\end{aligned}
$$

Le système, correspondant aux équations du modèle mixte, est donc le suivant :

$$
\begin{pmatrix} X_1'X_1 + X_2'X_2 & \sigma_1X_1'U_1 + \sigma_2X_2'U_2 \\ \sigma_1X_1U_1' + \sigma_2X_2U_2' & \sigma_1^2U_1U_1' + \sigma_2^2U_2U_2' + \sigma_0^2I_q \end{pmatrix} \begin{pmatrix} \beta \\ \xi \end{pmatrix} = \begin{pmatrix} X_1'y_1 + X_2'y_2 \\ \sigma_1U_1'y_1 + \sigma_2U_2'y_2 \end{pmatrix}
$$

Toute la difficulté réside alors dans l'obtention d'estimation des σ_i^2 à l'aide des valeurs de $\hat{\xi}$, solutions de ce système. Ce qui est d'autant plus difficile que ces paramètres ne correspondent plus à la variance de ξ et que des mêmes valeurs de ξ interviennent dans des environnements différents. Dans le cas homogène, les expressions des estimations obtenues à l'aide de $\hat{\xi}$ fournissaient des solutions au système d'équations ML (ou REML). Ceci ici, n'est pas facilité par le fait que ce dernier système, on l'a vu précédemment, est lui-même mal décrit.

Ces deux démarches d'estimation des composantes de la variance présentent, à nos yeux, des difficultés pour être adaptées dans le cas de variances hétérogènes. Aussi, l'algorithme EM va s'avérer être un outil très utile. En effet, la procédure EM, associée au fait que les σ_i aient été sortis de la distribution de ξ, présente un grand intérêt au vu de la démarche qu'elle adopte. Dans un premier temps, cette procédure se focalise sur la distribution conditionnelle de Y à ξ (la distribution de ξ n'apportant aucune information sur les paramètres). Les σ_i sont alors bien présents comme paramètres de régression. On peut alors obtenir une estimation de ces paramètres en fonction du vecteur des données complètes et c'est ensuite, dans le calcul de l'espérance conditionnelle, qu'intervient la distribution de ξ qui ne modifie alors en rien la démarche d'estimation. Cette procédure semble donc particulièrement adaptée. Nous la décrivons en détail dans la sous-section suivante.

4.4.3 EM dans le cas du L2M homogène

Nous nous intéressons tout d'abord au cas du L2M homogène. En effet, avant d'introduire l'hétérogénéité, nous allons décrire les procédures EM pour l'estimation maximum de vraisemblance et maximum de vraisemblance restreint des composantes de la variance. Nous envisageons pour cela deux écritures possibles du modèle (selon que le paramètre de variance se trouve à l'intérieur ou à l'extérieur de la distribution des effets aléatoires). Ce qui donne naissance à 4 algorithmes, que nous présentons et dont nous vérifierons qu'ils sont bien 2 à 2 équivalents, dans le sens où ils conduisent à la convergence aux mêmes estimations. Les 2 écritures possibles du modèle sont les suivantes :

- Ecriture 1 : $Y = X\beta + U\xi + \varepsilon$ où $\begin{array}{lcl} \xi & \sim & \mathcal{N}(0, \sigma^2 I_q) \\ \varepsilon & \sim & \mathcal{N}(0, \sigma_0^2 R)\ , \end{array}$

- Ecriture 2 : $Y = X\beta + \sigma U\xi + \varepsilon$ où $\begin{array}{lcl} \xi & \sim & \mathcal{N}(0, I_q) \\ \varepsilon & \sim & \mathcal{N}(0, \sigma_0^2 R)\ . \end{array}$

On note $\Phi = (\beta', \sigma^2, \sigma_0^2)'$ le vecteur des paramètres à estimer. On suppose la matrice R connue.

4.4.3.1 Algorithme 1

Considérons l'écriture 1 du modèle. Des hypothèses de lois et des règles de conditionnement sur la loi normale, il découle les résultats suivants :

$$\begin{array}{rcl} \begin{pmatrix} \xi \\ Y \end{pmatrix} & \sim & \mathcal{N}\left(\begin{pmatrix} 0 \\ X\beta \end{pmatrix} \begin{pmatrix} \sigma^2 I_q & \sigma^2 U' \\ \sigma^2 U & \sigma_0^2 R + \sigma^2 UU' \end{pmatrix} \right) \\ \xi | y & \sim & \mathcal{N}\left(\sigma^2 U' V^{-1}(y - X\beta), \sigma^2 I_q - \sigma^4 U' V^{-1} U\right) \end{array}$$

où $\quad V = \sigma_0^2 R + \sigma^2 UU'$.

Suivant la méthodologie EM, nous écrivons tout d'abord la vraisemblance des données complètes, autrement dit celle issue de la loi du couple (Y, ξ). La log-vraisemblance jointe est :

$$l(\Phi; y, \xi) = \text{const} - \frac{N}{2}\ln(\sigma_0^2) - \frac{(y - X\beta - U\xi)' R^{-1}(y - X\beta - U\xi)}{2\sigma_0^2} - \frac{q}{2}\ln(\sigma^2) - \frac{\xi'\xi}{2\sigma^2}\ .$$

Ce qui donne pour ses dérivées :

$$\begin{array}{rcl} \dfrac{\partial l(\Phi; y, \xi)}{\partial \sigma^2} & = & -\dfrac{q}{2\sigma^2} + \dfrac{\xi'\xi}{2\sigma^4} \\ \dfrac{\partial l(\Phi; y, \xi)}{\partial \sigma_0^2} & = & -\dfrac{N}{2\sigma_0^2} + \dfrac{(y - X\beta - U\xi)' R^{-1}(y - X\beta - U\xi)}{2\sigma_0^4} \\ \dfrac{\partial l(\Phi; y, \xi)}{\partial \beta} & = & \dfrac{-X'R^{-1}X\beta + X'R^{-1}(y - U\xi)}{\sigma_0^2}\ . \end{array}$$

Ainsi, en annulant les dérivées de $Q(\Phi|\Phi^{(m)}) = E\left(l(\Phi; y, \xi)|y, \Phi^{(m)}\right)$ par rapport aux composantes de Φ, on obtient le schéma itératif :

$$\begin{array}{rcl} \sigma^{2(m+1)} & = & \dfrac{E\left(\xi'\xi|y, \Phi^{(m)}\right)}{q} \\ \sigma_0^{2(m+1)} & = & \dfrac{E\left((y - X\beta - U\xi)' R^{-1}(y - X\beta - U\xi)|y, \Phi^{(m)}\right)}{N} \\ \beta^{(m+1)} & = & (X'R^{-1}X)^{-1}X'R^{-1}\left(y - UE(\xi|y, \Phi^{(m)}))\right)\ . \end{array}$$

D'où :

$$\begin{array}{rcl} q\ \sigma^{2(m+1)} & = & E(\xi|y, \Phi^{(m)})' E(\xi|y, \Phi^{(m)}) + \text{tr}(\text{Var}(\xi|y, \Phi^{(m)})) \\ N\ \sigma_0^{2(m+1)} & = & [y - X\beta^{(m)} - UE(\xi|y, \Phi^{(m)})]' R^{-1}[y - X\beta^{(m)} - UE(\xi|y, \Phi^{(m)})] \\ & & +\ \text{tr}[U'R^{-1}U\ \text{Var}(\xi|y, \Phi^{(m)})] \\ \beta^{(m+1)} & = & (X'R^{-1}X)^{-1}X'R^{-1}\left(y - UE(\xi|y, \Phi^{(m)})\right)\ . \end{array}$$

En utilisant les résultats précédents sur les distributions et après quelques calculs, on aboutit à :

$$\begin{aligned}
q\ \sigma^{2(m+1)} &= \sigma^{4(m)}(y - X\beta^{(m)})'V^{(m)-1}UU'V^{(m)-1}(y - X\beta^{(m)}) \\
&\quad + \mathrm{tr}[\sigma^{2(m)}I_q - \sigma^{4(m)}U'V^{(m)-1}U] \\
N\ \sigma_0^{2(m+1)} &= \sigma_0^{4(m)}(y - X\beta^{(m)})'V^{(m)-1}RV^{(m)-1}(y - X\beta^{(m)}) \\
&\quad + \mathrm{tr}[\sigma_0^{2(m)}I_N - \sigma_0^{4(m)}RV^{(m)-1}] \\
\beta^{(m+1)} &= (X'R^{-1}X)^{-1}X'R^{-1}\left(y - \sigma^{2(m)}UU'V^{(m)-1}(y - X\beta^{(m)})\right) .
\end{aligned}$$

Ce qui définit complètement l'algorithme.

A la convergence de cet algorithme, pour $\sigma^{2(m)} = \sigma^{2(m+1)} = \dot{\sigma}^2$ et $\beta^{(m)} = \beta^{(m+1)} = \dot{\beta}$, on a :

$$\begin{aligned}
q\ \dot{\sigma}^2 &= \dot{\sigma}^4(y - X\dot{\beta})'\dot{V}^{-1}UU'\dot{V}^{-1}(y - X\dot{\beta}) + \dot{\sigma}^2\mathrm{tr}[I_q - \dot{\sigma}^2U'\dot{V}^{-1}U] \\
&= \dot{\sigma}^4(y - X\dot{\beta})'\dot{V}^{-1}UU'\dot{V}^{-1}(y - X\dot{\beta}) + q\ \dot{\sigma}^2 - \dot{\sigma}^4\mathrm{tr}[\dot{V}^{-1}UU']
\end{aligned}$$

d'où

$$\mathrm{tr}[\dot{V}^{-1}UU'] = (y - X\dot{\beta})'\dot{V}^{-1}UU'\dot{V}^{-1}(y - X\dot{\beta}) \tag{4.3}$$

de même,

$$\begin{aligned}
N\ \dot{\sigma}_0^2 &= \dot{\sigma}^4(y - X\dot{\beta})'\dot{V}^{-1}R\dot{V}^{-1}(y - X\dot{\beta}) + \dot{\sigma}_0^2\mathrm{tr}[I_N - \dot{\sigma}_0^2R\dot{V}^{-1}] \\
&= \dot{\sigma}^4(y - X\dot{\beta})'\dot{V}^{-1}R\dot{V}^{-1}(y - X\dot{\beta}) + N\ \dot{\sigma}_0^2 - \dot{\sigma}_0^4\mathrm{tr}[\dot{V}^{-1}R]
\end{aligned}$$

d'où

$$\mathrm{tr}[\dot{V}^{-1}R] = (y - X\dot{\beta})'\dot{V}^{-1}R\dot{V}^{-1}(y - X\dot{\beta}) \tag{4.4}$$

et enfin,

$$\begin{aligned}
\dot{\beta} &= (X'R^{-1}X)^{-1}X'R^{-1}\left(y - \dot{\sigma}^2UU'\dot{V}^{-1}(y - X\dot{\beta})\right) \\
&= (X'R^{-1}X)^{-1}X'R^{-1}\left(y - (\dot{V} - \dot{\sigma}_0^2R)\dot{V}^{-1}(y - X\dot{\beta})\right) \\
&= (X'R^{-1}X)^{-1}X'R^{-1}\left(X\dot{\beta} + \dot{\sigma}_0^2R\dot{V}^{-1}(y - X\dot{\beta})\right)
\end{aligned}$$

d'où

$$X'\dot{V}^{-1}(y - X\dot{\beta}) = 0 . \tag{4.5}$$

Les équations (4.3), (4.4) et (4.5) sont bien les équations du maximum de vraisemblance. Les estimations obtenues à la convergence seront donc bien celles du maximum de vraisemblance.

4.4.3.2 Algorithme 2

Avec la même écriture du modèle, nous décrivons maintenant l'algorithme EM pour l'obtention des estimations maximum de vraisemblance restreint. Cette procédure REML peut être construite d'un point de vue fréquentiste par une transformation du modèle à l'aide de la matrice des contrastes. Ici, nous employons un point de vue bayésien pour reprendre la démarche notamment suivie par FOULLEY et QUAAS (1995) dans le cas hétérogène. Nous affectons à β une distribution a priori normale d'espérance β_0 et de matrice de variance B. Nous faisons ensuite tendre cette variance vers l'infini, ou plus exactement en termes matriciels, son inverse vers 0. Cette démarche bayésienne est ici davantage un artifice de calcul afin d'atteindre des estimations REML, qu'un réel souci de prendre en compte une certaine information a priori. Un raisonnement bayésien réel complet (et c'est ce que font certains auteurs) aurait aussi imposé une distribution a priori sur les paramètres de variance. Nous ne prenons pas ici ce parti.

Nous revenons aux hypothèses du modèle, en notant D la matrice de variance de ξ, $W = \sigma_0^2 R$ la matrice de variance résiduelle, d'où $V = UDU' + W$ et enfin on introduit $G = XBX' + UDU' + W$. Avec les hypothèses de lois normales pour Y, ξ et β, on obtient (cf. p. 329-332 SEARLE et al. 1992) :

- $\beta|y \sim \mathcal{N}(\text{esp}_\beta, \text{var}_\beta)$

$$\begin{aligned} \text{où} \quad \text{esp}_\beta &= (X'V^{-1}X + B^{-1})^{-1}(X'V^{-1}y + B^{-1}\beta_0) \\ \text{var}_\beta &= (X'V^{-1}X + B^{-1})^{-1} \end{aligned}$$

ce qui, lorsque $B^{-1} \longrightarrow 0$, donne :

$$\begin{aligned} \text{esp}_\beta &= (X'V^{-1}X)^{-1}(X'V^{-1}y) \\ \text{var}_\beta &= (X'V^{-1}X)^{-1} \end{aligned}$$

- $\xi|y \sim \mathcal{N}(\text{esp}_\xi, \text{var}_\xi)$

$$\begin{aligned} \text{où} \quad \text{esp}_\xi &= [U'(W^{-1} - W^{-1}X(B^{-1} + X'W^{-1}X)^{-1}X'W^{-1})U + D^{-1}]^{-1} \\ &\quad * U'(W^{-1} - W^{-1}X(B^{-1} + X'W^{-1}X)^{-1}X'W^{-1})(y - X\beta_0) \\ \text{var}_\xi &= [U'(W^{-1} - W^{-1}X(B^{-1} + X'W^{-1}X)^{-1}X'W^{-1})U + D^{-1}]^{-1} \end{aligned}$$

ce qui, lorsque $B^{-1} \longrightarrow 0$, donne :

$$\begin{aligned} \text{esp}_\xi &= DU'V^{-1}(I - X(X'V^{-1}X)^{-1}X'V^{-1})y \\ \text{var}_\xi &= [U'(W^{-1} - W^{-1}X(X'W^{-1}X)^{-1}X'W^{-1})U + D^{-1}]^{-1} \end{aligned}$$

Pour reprendre la démarche EM dans ce cadre, les données complètes sont maintenant rassemblées dans le vecteur $(Y', \xi', \beta')'$. La distribution de β n'apportant aucune information sur le vecteur de paramètres $\Phi = (\sigma_0^2, \sigma^2)'$, elle disparaîtra lors de la dérivation de $Q(\Phi|\Phi^{(m)})$. Autrement dit, la partie apportant de l'information correspond à la même vraisemblance complétée que dans le traitement non bayésien précédent. Ainsi, on obtient le même schéma itératif (en supprimant l'équation correspondant à β) :

$$\begin{aligned}
\sigma^{2(m+1)} &= \frac{E[\xi'\xi|y, \Phi^{(m)}]}{q} \\
\sigma_0^{2(m+1)} &= \frac{E[(y - X\beta - U\xi)'R^{-1}(y - X\beta - U\xi)|y, \Phi^{(m)}]}{N} .
\end{aligned}$$

En poursuivant le calcul, on a :

$$\begin{aligned}
q\ \sigma^{2(m+1)} &= E[\xi|y, \Phi^{(m)}]'E[\xi|y, \Phi^{(m)}] + \mathrm{tr}(\mathrm{Var}[\xi|y, \Phi^{(m)}]) \\
N\ \sigma_0^{2(m+1)} &= E[(y - X\beta)'R^{-1}(y - X\beta)|y, \Phi^{(m)}] \\
&\quad - 2E[\xi'U'R^{-1}(y - X\beta)|y, \Phi^{(m)}] + E[\xi'U'R^{-1}U\xi|y, \Phi^{(m)}] ,
\end{aligned}$$

avec, en notant $E_\beta = E[\beta|y, \Phi^{(m)}]$ et $E_\xi = E[\xi|y, \Phi^{(m)}]$:

$$\begin{aligned}
E[(y - X\beta)'R^{-1}(y - X\beta)|y, \Phi^{(m)}] &= (y - XE_\beta)'R^{-1}(y - XE_\beta) \\
&\quad + \mathrm{tr}(X'R^{-1}X\ \mathrm{Var}[\beta|y, \Phi^{(m)}]) \\
E[\xi'U'R^{-1}(y - X\beta)|y, \Phi^{(m)}] &= E_\xi'U'R^{-1}(y - XE_\beta) \\
&\quad - \mathrm{tr}(U'R^{-1}X\mathrm{Cov}[\beta, \xi|y, \Phi^{(m)}])
\end{aligned}$$

$$E[\xi'U'R^{-1}U\xi|y, \Phi^{(m)}] = E_\xi'U'R^{-1}UE_\xi + \mathrm{tr}(U'R^{-1}U\ \mathrm{Var}[\xi|y, \Phi^{(m)}]) .$$

Le calcul des espérances et variances conditionnelles peut alors être réalisé de deux façons :

- en utilisant les résultats sur les lois énoncés à la page précédente,
- en utilisant les valeurs fournies par le système de Henderson.

Cette deuxième solution est un artifice de calcul adopté par divers auteurs et que nous exploiterons en particulier dans le cas hétérogène pour décrire la procédure proposée par Foulley et Quaas (1995). Nous montrons ci-dessous que ces deux démarches sont bien équivalentes.

Le système de Henderson associé au modèle linéaire mixte considéré est le suivant :

$$\begin{bmatrix} X'W^{-1}X & X'W^{-1}U \\ U'W^{-1}X & U'W^{-1}U + D^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \xi \end{bmatrix} = \begin{bmatrix} X'W^{-1}y \\ U'W^{-1}y \end{bmatrix}$$

On désigne par C l'inverse de la matrice des coefficients du système que l'on découpe en quatre blocs selon les tailles de β et ξ de la façon suivante : $C = \begin{bmatrix} C_{\beta\beta} & C_{\beta\xi} \\ C_{\xi\beta} & C_{\xi\xi} \end{bmatrix}$.

La démarche consiste alors à remplacer les espérances et variances conditionnelles, respectivement par les solutions $\hat{\beta}$ et $\hat{\xi}$ de ce système, et par les éléments de la matrice C. Or, parmi les propriétés de ce système figurent le fait que $\hat{\beta}$ et $\hat{\xi}$ vérifient :

$$\begin{array}{rcl} \hat{\beta} &=& E[\beta|y,\Phi^{(m)}] \\ \hat{\xi} &=& E[\xi|y,\Phi^{(m)}] \,. \end{array}$$

Mais on a aussi, et c'est moins naturel :

$$\begin{array}{rcl} C_{\beta\beta} &=& \mathrm{Var}[\beta|y,\Phi^{(m)}] \\ C_{\xi\xi} &=& \mathrm{Var}[\xi|y,\Phi^{(m)}] \\ C_{\beta\xi} &=& \mathrm{Cov}[\beta,\xi|y,\Phi^{(m)}] \,. \end{array}$$

En effet, d'après les règles d'inversion de matrice par blocs, on a :

$$\begin{array}{rrcl} & C_{\beta\beta} &=& (X'W^{-1}X)^{-1} + (X'W^{-1}X)^{-1}X'W^{-1}UC_{\xi\xi}U'W^{-1}X(X'W^{-1}X)^{-1} \\ \text{et} & C_{\xi\xi} &=& \left((U'W^{-1}U + D^{-1}) - U'W^{-1}X(X'W^{-1}X)^{-1}X'W^{-1}U\right)^{-1} \end{array}$$

En utilisant le résultat relatif aux inversions de matrices : $(F - CA^{-1}B)^{-1} = F^{-1} + F^{-1}C(A - BF^{-1}C)^{-1}BF^{-1}$, appliqué aux matrices :

$$\begin{array}{rcl} A &=& U'W^{-1}U + D^{-1} \\ B &=& U'W^{-1}X \\ C &=& X'W^{-1}U \\ F &=& X'W^{-1}X \,, \end{array}$$

on obtient alors :

$$\begin{array}{rcl} C_{\beta\beta} &=& (X'W^{-1}X - X'W^{-1}U(U'W^{-1}U + D^{-1})^{-1}U'W^{-1}X)^{-1} \\ &=& (X'(W^{-1} - W^{-1}U(U'W^{-1}U + D^{-1})^{-1}U'W^{-1})X)^{-1} \\ &=& (X'(W + UDU')^{-1}X)^{-1} \\ &=& (X'V^{-1}X)^{-1} \end{array}$$

et,

$$C_{\xi\xi} = (U'(W^{-1} - W^{-1}X(X'W^{-1}X)^{-1}X'W^{-1})U + D^{-1})^{-1}$$

On retrouve donc bien les expressions des variances de β et ξ a posteriori données précédemment. C'est ce qui montre que l'une et l'autre des façons de calculer ces éléments a posteriori sont bien équivalentes. Il suffit à présent de replacer ces expressions dans le schéma itératif pour que l'algorithme soit complètement défini.

4.4.3.3 Algorithme 3

Nous considérons maintenant l'écriture 2 du modèle. Elle n'est pas ici nécessaire mais comme nous l'avons déjà répété, elle le sera dans le cas hétérogène. R est toujours supposée connue. On a juste sorti le paramètre σ de la distribution de ξ qui est maintenant normale centrée réduite. On obtient alors pour la loi du couple et la loi conditionnelle de ξ à Y :

$$\begin{aligned}
\begin{pmatrix} \xi \\ Y \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ X\beta \end{pmatrix} \begin{pmatrix} I_q & \sigma U' \\ \sigma U & \sigma_0^2 R + \sigma^2 UU' \end{pmatrix} \right) \\
\xi | y &\sim \mathcal{N}\left(\sigma U' V^{-1}(y - X\beta), I_q - \sigma^2 U' V^{-1} U \right)
\end{aligned}$$

où on note toujours $\quad V = \sigma_0^2 R + \sigma^2 UU'$.

Reprenons la démarche EM pour ML. La log-vraisemblance jointe du paramètre $\Phi = (\beta', \sigma^2, \sigma_0^2)'$ s'écrit alors :

$$l(\Phi; y, \xi) = \text{const} - \frac{(y - X\beta - \sigma U\xi)' R^{-1} (y - X\beta - \sigma U\xi)}{2\sigma_0^2} - \frac{N}{2} \ln(\sigma_0^2) - \frac{\xi'\xi}{2} \,.$$

Les dérivées par rapport aux composantes de Φ sont :

$$\begin{aligned}
\frac{\partial l(\Phi; y, \xi)}{\partial \sigma^2} &= \frac{-\frac{1}{\sigma}\xi' U' R^{-1}(y - X\beta) + \xi' U' R^{-1} U \xi}{2\sigma_0^2} \\
\frac{\partial l(\Phi; y, \xi)}{\partial \sigma_0^2} &= -\frac{N}{2\sigma_0^2} + \frac{(y - X\beta - \sigma U\xi)' R^{-1}(y - X\beta - \sigma U\xi)}{2\sigma_0^4} \\
\frac{\partial l(\Phi; y, \xi)}{\partial \beta} &= \frac{-X' R^{-1} X\beta + X' R^{-1}(y - \sigma U\xi)}{\sigma_0^2} \,.
\end{aligned}$$

Ainsi, on obtient cette fois-ci le schéma itératif :

$$\begin{aligned}
\sigma^{(m+1)} &= \frac{E\left(\xi' U' R^{-1}(y - X\beta) | y, \Phi^{(m)}\right)}{E\left(\xi' U' R^{-1} U \xi | y, \Phi^{(m)}\right)} \\
\sigma_0^{2(m+1)} &= \frac{E\left((y - X\beta - \sigma U\xi)' R^{-1}(y - X\beta - \sigma U\xi) | y, \Phi^{(m)}\right)}{N} \\
\beta^{(m+1)} &= (X' R^{-1} X)^{-1} X' R^{-1} \left(y - \sigma U E(\xi | y, \Phi^{(m)})\right) \,.
\end{aligned}$$

Ce qui donne, en utilisant les lois conditionnelles et en adoptant la notation

$A^{(m)} = (y - X\beta^{(m)})'V^{(m)-1}U$:

$$\begin{aligned}
\sigma^{(m+1)} &= \frac{\sigma^{(m)}A^{(m)}U'R^{-1}(y - X\beta^{(m)})}{\sigma^{2(m)}A^{(m)}U'R^{-1}UA'^{(m)} + \mathrm{tr}(U'R^{-1}U(I_q - \sigma^{2(m)}U'V^{(m)-1}U))} \\
N\sigma_0^{2(m+1)} &= \sigma_0^{4(m)}(y - X\beta^{(m)})'V^{(m)-1}RV^{(m)-1}(y - X\beta^{(m)}) \\
&\quad + \mathrm{tr}[\sigma_0^{2(m)}I_N - \sigma_0^{4(m)}RV^{(m)-1}] \\
\beta^{(m+1)} &= (X'R^{-1}X)^{-1}X'R^{-1}\left(y - \sigma^{2(m)}UU'V^{(m)-1}(y - X\beta^{(m)})\right) .
\end{aligned}$$

Plaçons nous de nouveau à la convergence de cet algorithme. Le schéma itératif pour β et σ_0^2 n'ayant pas changé, regardons uniquement l'équation dont est solution $\dot{\sigma}^2$:

$$\begin{aligned}
& \dot{\sigma} = \frac{\dot{\sigma}\dot{A}U'R^{-1}(y - X\dot{\beta})}{\dot{\sigma}^2\dot{A}U'R^{-1}U\dot{A}' + \mathrm{tr}(U'R^{-1}U(I_q - \dot{\sigma}^2U'\dot{V}^{-1}U))} \\
\iff\ & \dot{\sigma}^2\dot{A}U'R^{-1}U\dot{A}' + \mathrm{tr}(U'R^{-1}U(I_q - \dot{\sigma}^2U'\dot{V}^{-1}U)) = \dot{A}U'R^{-1}(y - X\dot{\beta}) \\
\iff\ & \dot{A}U'R^{-1}(y - X\dot{\beta} - \dot{\sigma}^2U\dot{A}') = \mathrm{tr}(R^{-1}(I_N - \dot{\sigma}^2UU'\dot{V}^{-1})UU') \\
\iff\ & (y - X\dot{\beta})'\dot{V}^{-1}UU'\dot{V}^{-1}(y - X\dot{\beta}) = \mathrm{tr}(\dot{V}^{-1}UU')
\end{aligned}$$

On retrouve donc bien une fois de plus (et heureusement) les équations du maximum de vraisemblance. Les algorithmes 1 et 3 sont donc bien équivalents. Ce dernier présente néanmoins l'avantage de pouvoir être réutilisé dans le cas hétérogène.

4.4.3.4 Algorithme 4

À ces trois algorithmes que nous venons de décrire, il est possible d'en associer un quatrième que nous n'explicitons pas ici puisqu'il sera repris en détails dans le cas du modèle à variances hétérogènes dans la section suivante. Cet algorithme 4 allie l'approche par la deuxième écriture du modèle (où σ est sorti de la distribution de ξ) et la démarche REML avec traitement bayésien de β. En effet, à l'instar de ce que nous venons de présenter pour l'algorithme 3, associant l'écriture 2 du modèle et la démarche EM-ML de l'algorithme 1, il est naturel de reprendre la démarche de l'algorithme 2 avec cette même écriture. L'algorithme obtenu est alors équivalent, à la convergence, à l'algorithme 2 ; puisque l'estimation REML est une estimation ML dans un modèle particulier (le projeté du modèle initial - cf. chapitre 2) et que nous avons prouvé l'équivalence entre les algorithmes 1 et 3.

4.4.3.5 Quelques résultats numériques

Nous résumons ces quatre algorithmes dans le tableau suivant :

	ML	REML
Ecriture 1	Algo 1	Algo 2
Ecriture 2	Algo 3	Algo 4

Notons que ces 4 algorithmes peuvent être adaptés sans aucun problème au cas de variances hétérogènes pour les erreurs. Il suffit alors, pour chaque environnement, de reprendre l'expression de $\sigma_0^{2(m+1)}$, et de découper les observations pour ne considérer que celles appartenant à l'environnement concerné.

Nous appliquons dans un premier temps ces quatre procédures sur l'exemple présenté en début de section. Voici les estimations de β et des composantes, obtenues ainsi que le nombre d'itérations nécessaires à cette estimation :

• dans le cas d'erreurs à variances homogènes :

	$\hat{\beta}$			$\hat{\sigma}^2$	$\hat{\sigma}_0^2$	*niter*
Algo 1	399.12	519.36	575.34	2383.89	17062.49	38
Algo 2	399.29	520.39	577.55	3668.42	18214.49	37
Algo 3	399.12	519.36	575.34	2383.89	17062.49	34
Algo 4	399.29	520.39	577.55	3668.42	18214.49	40

• dans le cas d'erreurs à variances hétérogènes :

	$\hat{\beta}$			$\hat{\sigma}^2$	$\hat{\sigma}_0^2$			*niter*
Algo 1	399.09	515.42	573.51	1157.29	3717.23	18650.32	36128.34	38
Algo 2	399.25	515.91	575.15	1730.24	3878.56	20041.79	39566.60	39
Algo 3	399.09	515.42	573.51	1157.29	3717.23	18650.32	36128.35	34
Algo 4	399.25	515.91	575.15	1730.24	3878.56	20041.79	39566.60	46

Nous avons utilisé comme test d'arrêt des quatre procédures, un test sur les valeurs de $\hat{\beta}, \hat{\sigma}^2$ et $\hat{\sigma}_0^2$. C'est-à-dire que pour une précision choisie (ici de 10^{-4}), on teste si les composantes des différents paramètres sont stables en valeur absolue et en valeur relative de l'itération t à l'itération $t+1$:

- $|par^{(t+1)} - par^{(t)}| < precision$
- $|\dfrac{par^{(t+1)} - par^{(t)}}{par^{(t)}}| < precision.$

Notons, qu'après ces tests d'arrêt, nous avons laissé courir l'algorithme sur 10000 itérations afin de vérifier qu'une certaine stabilité avait été atteinte. Aucun résultat n'a été modifié !

Dans les deux cas, nous constatons bien entendu l'équivalence des algorithmes 1 et 3 et celle des algorithmes 2 et 4, la différence entre ces deux catégories étant attribuée à la différence entre les estimations ML et REML. Cette équivalence peut aussi constituer un moyen pratique de vérifier que la convergence a été atteinte.

Regarder pratiquement les qualités de ces estimations sera envisagé dans la section suivante, dans le cas des modèles à variances hétérogènes pour les effets (puisque notre objectif reste celui-ci). C'est davantage sur la dernière colonne de ce tableau que nous nous arrêtons momentanément dans les simulations ci-dessous. Nous comparons plus précisément le nombre d'itérations necessaires aux procédures 1 et 3 et aux procédures 2 et 4 pour converger ; à savoir si de deux écritures différentes d'un même modèle naissent deux algorithmes de rapidité différente. Bien entendu, ceci reste inhérent au test d'arrêt choisi (et précisé ci-dessus). Nous savons en effet qu'il faut être prudent en ce qui concerne la notion de convergence, en particulier pour cet algorithme EM. Avec ce test d'arrêt, nous avons donc réalisé des simulations en conservant le même plan d'expérience que celui de l'exemple, dans les deux cas d'erreurs à variances homogènes et hétérogènes, et pour différentes grandeurs relatives des valeurs simulées des paramètres de variance (σ^2_{err} pour les erreurs et σ^2_{eff} pour les effets). Nous donnons les moyennes et écart-types des nombres d'itérations nécessaires à la convergence des algorithmes (i.e. arrêt des procédures) pour 100 simulations :

- dans le cas d'erreurs à variances homogènes :

		ML		REML	
Valeurs simulées	Nbre d'itérations	Algo 1	Algo 3	Algo 2	Algo 4
$\sigma^2_{\text{err}} = 0.5$	moy.	386.9	341.6	9.9	426.3
$\sigma^2_{\text{eff}} = 10$	e.t.	246.0	184.7	0.4	240.9
$\sigma^2_{\text{err}} = 2$	moy.	52.3	42.3	16.8	43.0
$\sigma^2_{\text{eff}} = 1.5$	e.t.	39.7	26.5	12.7	27.0
$\sigma^2_{\text{err}} = 10$	moy.	104.9	24.2	35.8	22.6
$\sigma^2_{\text{eff}} = 0.5$	e.t.	144.9	11.5	23.1	7.6

• dans le cas d'erreurs à variances hétérogènes :

		ML		REML	
Valeurs simulées	Nbre d'itérations	Algo 1	Algo 3	Algo 2	Algo 4
$\sigma^2_{\text{err}} = 1 \quad 0.5 \quad 1.5$	moy.	385.6	329.4	12.0	406.7
$\sigma^2_{\text{eff}} = 10$	e.t.	255.6	193.1	3.0	253.9
$\sigma^2_{\text{err}} = 1 \quad 1.5 \quad 2$	moy.	67.5	60.4	15.3	60.4
$\sigma^2_{\text{eff}} = 1.5$	e.t.	58.6	48.6	10.5	52.6
$\sigma^2_{\text{err}} = 6 \quad 8 \quad 9$	moy.	75.4	26.5	30.0	25.9
$\sigma^2_{\text{eff}} = 0.5$	e.t.	115.4	14.2	18.0	7.8

Que ce soit dans le cas de variances des erreurs homogènes ou hétérogènes, on constate que l'écriture 2 (σ sorti de la distribution), pour l'estimation ML, réduit le nombre d'itérations. Ceci est d'autant plus vrai que la variance des erreurs est grande devant celle des effets aléatoires. Dans ces mêmes conditions, cette remarque est aussi valable pour l'estimation REML. Cependant cela n'est plus vrai dans les autres cas et on note une différence importante entre l'algorithme 2 et l'algorithme 4 dans le cas extrême de variance des effets très grande devant celle des erreurs.
Globalement, l'algorithme 2 est celui qui est le plus rapide. Et de façon générale, l'hétérogénéité des erreurs ne semble pas agir sur le nombre d'itérations.

4.4.4 EM dans le cas du L2M hétérogène

Nous en venons maintenant au cas du L2M hétérogène. Cette hétérogénéité désigne bien sûr celle des variances des effets aléatoires. Mais nous considérerons aussi le cas échéant celle des erreurs. Pour des raisons que nous avons déjà évoquées à la section 4.2, nous adoptons alors l'écriture du modèle correspondante à l'écriture 2 précédente et nous séparons ici les I environnements :

$$\forall i \in \{1, \ldots, I\}\ , \qquad Y_i = X_i\beta + \sigma_i U_i \xi + \varepsilon_i$$

où $\quad \xi \sim \mathcal{N}(0, I_q) \quad$ et

- si les erreurs sont à variances homogènes : $\quad \varepsilon_i \sim \mathcal{N}(0, \sigma_0^2 R_i) \quad ,$

  on notera dans ce cas $W = \begin{pmatrix} \sigma_0^2 R_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_0^2 R_I \end{pmatrix} = \sigma_0^2\ R$

- si les erreurs sont à variances hétérogènes : $\quad \varepsilon_i \sim \mathcal{N}(0, \sigma_{0_i}^2 R_i) \quad ,$

  avec $W = \begin{pmatrix} \sigma_{0_1}^2 R_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_{0_I}^2 R_I \end{pmatrix}$ et $\sigma_0^2 = \begin{pmatrix} \sigma_{0_1}^2 \\ \vdots \\ \sigma_{0_I}^2 \end{pmatrix}$.

Avec cette écriture, ce sont donc les algorithmes 3 et 4 que nous allons adapter à l'hétérogénéité des effets aléatoires. Nous obtenons ainsi un premier algorithme (LINHE_ML) qui fournit une estimation ML et utilise pour le calcul des moments les distributions conditionnelles. Un deuxième algorithme (LINHE_REML), construit grâce à une hypothèse bayésienne sur β, fournit quant à lui une estimation REML utilisant le système de Henderson pour le calcul des moments. C'est l'algorithme présenté par Foulley et Quaas (1995). Il en existe un équivalent fréquentiste présenté par De Stefano (1983).

4.4.4.1 Algorithme LINHE_ML

Pour ce premier algorithme, nous reprenons l'algorithme 3 décrit précédemment et nous l'adaptons au cas de σ_i différents pour les I environnements.

Pour cela, on adopte les notations suivantes :

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_I \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_I \end{pmatrix} \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_I \end{pmatrix} \quad U = \begin{pmatrix} U_1 \\ \vdots \\ U_I \end{pmatrix} \quad \sigma^2 = \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_I^2 \end{pmatrix} ,$$

et enfin soit T la matrice diagonale par blocs définie par :

$$T = \begin{pmatrix} \sigma_1 \, I_{n_1} & & 0 \\ & \ddots & \\ & & \sigma_I \, I_{n_I} \end{pmatrix}$$

où n_i est le nombre d'observations pour l'environnement i.

Avec ces notations, le modèle peut s'écrire sous la forme :

$$Y = X\beta + TU\xi + \varepsilon \; .$$

La distribution associée est :

$$\begin{pmatrix} \xi \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ X\beta \end{pmatrix}, \begin{pmatrix} I_q & U'T \\ TU & T^2UU' + W \end{pmatrix} \right) ,$$

d'où $\quad \xi | y \sim \mathcal{N}(U'TV^{-1}(y - X\beta), I_q - U'TV^{-1}TU) \quad$ avec $\quad V = T^2UU' + W$.

Suivant la démarche EM pour ML, la log-vraisemblance jointe du paramètre

$\Phi = (\beta', \sigma^{2'}, \sigma_0^{2'})$ est :

$$\begin{aligned}
l(\Phi; y, \xi) &= \text{const} - \frac{(y - X\beta - TU\xi)'W^{-1}(y - X\beta - TU\xi)}{2} - \frac{1}{2}\ln(|W|) - \frac{\xi'\xi}{2} \\
&= \text{const} \quad -\sum_{i=1}^{I} \frac{(y_i - X_i\beta - \sigma_i U_i\xi)'R_i^{-1}(y_i - X_i\beta - \sigma_i U_i\xi)}{2\sigma_{0_i}^2} \\
&\qquad -\sum_{i=1}^{I} \frac{n_i}{2}\ln(\sigma_{0_i}^2) - \frac{\xi'\xi}{2}
\end{aligned}$$

avec dans un cas d'homogénéité des variances des erreurs : $\forall i \in \{1, \ldots, I\}$ $\sigma_{0_i}^2 = \sigma_0^2$.

Ce qui donne pour ses dérivées :

$$\forall i \in \{1, \ldots, I\}, \qquad \begin{aligned}
\frac{\partial l(\Phi; y, \xi)}{\partial \sigma_i^2} &= \frac{\frac{1}{\sigma_i}\xi'U_i'R_i^{-1}(y_i - X_i\beta) - \xi'U_i'R_i^{-1}U_i\xi}{\sigma_{0_i}^2} \\
\frac{\partial l(\Phi; y, \xi)}{\partial \beta} &= -X'W^{-1}X\beta + X'W^{-1}(y - TU\xi) ,
\end{aligned}$$

et
- dans le cas d'erreurs à variances homogènes :

$$\frac{\partial l(\Phi; y, \xi)}{\partial \sigma_0^2} = -\frac{N}{2\sigma_0^2} + \frac{(y - X\beta - TU\xi)'R^{-1}(y - X\beta - TU\xi)}{2\sigma_0^4}$$

- dans le cas d'erreurs à variances hétérogènes : $\forall i \in \{1, \ldots, I\}$,

$$\frac{\partial l(\Phi; y, \xi)}{\partial \sigma_{0_i}^2} = -\frac{n_i}{2\sigma_{0_i}^2} + \frac{(y_i - X_i\beta - \sigma_i U_i\xi)'R_i^{-1}(y_i - X_i\beta - \sigma_i U_i\xi)}{2\sigma_{0_i}^4}.$$

Pour la suite, nous distinguons les deux cas pour les erreurs.

• Cas d'erreurs à variances homogènes :

En annulant les dérivées, on obtient :

$$\forall i \in \{1, \ldots, I\}, \qquad \sigma_i^{(m+1)} = \frac{E\left(\xi'U_i'R_i^{-1}(y_i - X_i\beta)|y, \Phi^{(m)}\right)}{E\left(\xi'U_i'R_i^{-1}U_i\xi|y, \Phi^{(m)}\right)}$$

$$\begin{aligned}
\sigma_0^{2(m+1)} &= \frac{E\left((y - X\beta - TU\xi)'R^{-1}(y - X\beta - TU\xi)|y, \Phi^{(m)}\right)}{N} \\
\beta^{(m+1)} &= E\left(X'W^{-1}X|y, \Phi^{(m)}\right)^{-1} E\left(X'W^{-1}(y - TU\xi)|y, \Phi^{(m)}\right) .
\end{aligned}$$

En notant $E_c^{(m)}$ et $V_c^{(m)}$, l'espérance et la variance conditionnelles de ξ à y :

$$\begin{aligned}
E(\xi|y, \Phi^{(m)}) &= E_c^{(m)} &= U'T^{(m)}V^{(m)-1}(y - X\beta^{(m)}) \\
\text{Var}(\xi|y, \Phi^{(m)}) &= V_c^{(m)} &= I_q - U'T^{(m)}V^{(m)-1}T^{(m)}U ,
\end{aligned}$$

le schéma itératif de l'algorithme peut alors s'écrire :

$$\begin{aligned}
\forall i \in \{1,\ldots,I\}\,, \quad \sigma_i^{(m+1)} &= \frac{E_c^{(m)\prime} U_i' R_i^{-1}(y_i - X_i\beta^{(m)})}{E_c^{(m)\prime} U_i' R_i^{-1} U_i E_c^{(m)} + \mathrm{tr}(U_i' R_i^{-1} U_i V_c^{(m)})} \\
N\sigma_0^{2(m+1)} &= (y - X\beta^{(m)})' R^{-1}(y - X\beta^{(m)}) - 2(y - X\beta^{(m)})' R^{-1} T^{(m)} U E_c^{(m)} \\
&\quad + E_c^{(m)\prime} U' T^{(m)} R^{-1} T^{(m)} U E_c^{(m)} + \mathrm{tr}[U' T^{(m)} R^{-1} T^{(m)} U V_c^{(m)}] \\
\beta^{(m+1)} &= (X' W^{(m)-1} X)^{-1} X' W^{(m)-1}(y - T^{(m)} U E_c^{(m)})\,.
\end{aligned}$$

• Cas d'erreurs à variances hétérogènes :

On reprend uniquement les équations qui subissent une modification, c'est-à-dire celles concernant les variances des erreurs. Ainsi, l'annulation des dérivées donne :

$$\forall i \in \{1,\ldots,I\}\,, \ \sigma_{0_i}^{2(m+1)} = \frac{E\left((y_i - X_i\beta - \sigma_i U_i\xi)' R_i^{-1}(y_i - X_i\beta - \sigma_i U_i\xi) | y, \Phi^{(m)}\right)}{n_i}$$

Et on change donc dans le schéma itératif précédent la réactualisation de σ_0^2 par :
$\forall i \in \{1,\ldots,I\}$,

$$\begin{aligned}
n_i\sigma_{0_i}^{2(m+1)} &= (y_i - X_i\beta^{(m)})' R_i^{-1}(y_i - X_i\beta^{(m)}) - 2\sigma_i^{(m)}(y_i - X_i\beta^{(m)})' R_i^{-1} U E_c^{(m)} \\
&\quad + \sigma_i^{2(m)} E_c^{(m)\prime} U_i' R_i^{-1} U_i E_c^{(m)} + \sigma_i^{2(m)} \mathrm{tr}[U_i' R_i^{-1} U_i V_c^{(m)}]\,.
\end{aligned}$$

4.4.4.2 Algorithme LINHE_REML

Ce deuxième algorithme adopte une optique EM pour REML. Il a été proposé dans le cadre des L2M hétérogènes par Foulley et Quaas (1995). Il combine à la fois la démarche ayant conduit à l'algorithme 2 et l'écriture du modèle hétérogène (avec σ sorti). Ce qui correspond à l'algorithme 4 évoqué précédemment. Nous le décrivons dans ses grandes lignes, en traitant directement le cas d'erreurs à variances hétérogènes, le cas homogène s'en déduisant très facilement.

À la log-vraisemblance du paramètre Φ pour les données complètes de la section précédente, se rajoute l'a priori sur β avec une variance fixée à l'infini qui n'apporte pas d'information sur $\Phi = (\sigma'^2, \sigma_0'^2)'$. Ainsi, les dérivées restent identiques et l'on a :

$\forall i \in \{1,\ldots,I\}$,

$$\begin{aligned}
\frac{\partial l(\Phi; y, \xi, \beta)}{\partial \sigma_i^2} &= \frac{\frac{1}{\sigma_i}\xi' U_i' R_i^{-1}(y_i - X_i\beta) - \xi' U_i' R_i^{-1} U_i \xi}{2\sigma_{0_i}^2} \\
\frac{\partial l(\Phi; y, \xi, \beta)}{\partial \sigma_{0_i}^2} &= -\frac{n_i}{2\sigma_{0_i}^2} + \frac{(y_i - X_i\beta - \sigma_i U_i\xi)' R_i^{-1}(y_i - X_i\beta - \sigma_i U_i\xi)}{2\sigma_{0_i}^4}\,.
\end{aligned}$$

D'où les réactualisations des paramètres :
$\forall i \in \{1, \ldots, I\}$,

$$\sigma_i^{(m+1)} = \frac{E\left(\xi' U_i' R_i^{-1}(y_i - X_i\beta)|y, \Phi^{(m)}\right)}{E\left(\xi' U_i' R_i^{-1} U_i \xi|y, \Phi^{(m)}\right)} \tag{4.6}$$

$$\sigma_{0_i}^{2(m+1)} = \frac{E\left((y_i - X_i\beta - \sigma_i U_i \xi)' R_i^{-1}(y_i - X_i\beta - \sigma_i U_i \xi)|y, \Phi^{(m)}\right)}{n_i} \tag{4.7}$$

À ce stade, la méthodologie préconise de réaliser le calcul des espérances à l'aide des éléments du système de Henderson, qui s'écrit ici :

$$\begin{pmatrix} X'W^{-1}X & X'W^{-1}TU \\ U'TW^{-1}X & U'TW^{-1}TU + I_q \end{pmatrix} \begin{pmatrix} \beta \\ \xi \end{pmatrix} = \begin{pmatrix} X'W^{-1}y \\ Z'W^{-1}y \end{pmatrix} .$$

Alors, en notant comme précédemment $\hat{\beta}$ et $\hat{\xi}$ les solutions de ce système, et C l'inverse de la matrice des coefficients : $C = \begin{pmatrix} C_{\beta\beta} & C_{\beta\xi} \\ C_{\xi\beta} & C_{\xi\xi} \end{pmatrix}$, on a :

$$E(\xi' U_i' R_i^{-1}(y_i - X_i\beta)|y, \Phi^{(m)}) = \hat{\xi}' U_i' R_i^{-1}(y_i - X_i\hat{\beta}) - \mathrm{tr}(U_i' R_i^{-1} X_i C_{\beta\xi})$$

$$E(\xi' U_i' R_i^{-1} U_i \xi|y, \Phi^{(m)}) = \hat{\xi}' U_i' R_i^{-1} U_i \hat{\xi} + \mathrm{tr}(U_i' R_i^{-1} U_i C_{\xi\xi})$$

$$\begin{aligned} &E((y_i - X_i\beta - \sigma_i U_i \xi)' R_i^{-1}(y_i - X_i\beta - \sigma_i U_i \xi)|y, \Phi^{(m)}) = \\ &\qquad (y_i - X_i\hat{\beta} - \sigma_i^{(m)} U_i \hat{\xi})' R_i^{-1}(y_i - X_i\hat{\beta} - \sigma_i^{(m)} U_i \hat{\xi}) \\ &\qquad + \mathrm{tr}(X_i' R_i^{-1} X_i C_{\beta\beta}) + 2\sigma_i^{(m)} \mathrm{tr}(U_i' R_i^{-1} X_i C_{\beta\xi}) + \sigma_i^{2(m)} \mathrm{tr}(U_i' R_i^{-1} U_i C_{\xi\xi}) . \end{aligned}$$

En replaçant ces expressions dans (4.6) et (4.7), on obtient alors le schéma itératif de l'algorithme.

4.4.4.3 Résultats sur l'exemple

Nous reprenons le petit exemple présenté en début de cette section. Nous donnons ci-dessous les résultats obtenus par chacun des deux algorithmes dans les deux types de modélisation avec ou sans erreurs à variances hétérogènes.

• Cas d'erreurs à variances homogènes :

		LINHE_ML	LINHE_REML
$\hat{\sigma}^2_{\text{err}}$		16365.22	17447.40
$\hat{\sigma}^2_{\text{eff}}$	1	679.73	987.60
	2	3744.46	5452.92
	3	5516.45	8895.20
$\hat{\beta}$	1	398.54	398.58
	2	521.82	522.19
	3	583.59	587.80
niter		39	46

• Cas d'erreurs à variances hétérogènes :

		LINHE_ML	LINHE_REML
$\hat{\sigma}^2_{\text{err}}$	1	3615.31	3793.80
	2	17410.67	18703.50
	3	34052.87	36972.49
$\hat{\sigma}^2_{\text{eff}}$	1	789.35	1145.29
	2	3833.50	5523.34
	3	5772.37	9246.40
$\hat{\beta}$	1	398.78	398.85
	2	519.54	520.00
	3	589.47	593.96
niter		39	47

4.4.5 Simulations

Nous venons de présenter deux algorithmes pour l'estimation des composantes de la variance dans un modèle linéaire mixte à variances hétérogènes, c'est-à-dire pour lequel la variance des effets aléatoires dépend de l'environnement. Nous mettons maintenant ces algorithmes à l'epreuve sur des jeux de données simulées, de façon à analyser numériquement la qualité des estimations obtenues. Une fois de plus, nous utilisons le plan d'expérience du même exemple et nous simulons des données à partir du modèle (4.2), ceci dans les deux cas d'erreurs à variances homogènes et hétérogènes. Dans les tableaux ci-dessous, nous faisons figurer à chaque fois :

- les valeurs des paramètres du modèle en simulation,
- la moyenne et l'écart-type, sur 200 jeux de données simulées, des estimations obtenues par chacun des deux algorithmes.

4.4.5.1 Simulations avec erreurs à variances homogènes

		Valeurs simulées	Valeurs estimées			
			LINHE_ML		LINHE_REML	
			moy.	e.t.	moy.	e.t.
σ^2_{err}		1.5	1.30	0.35	1.38	0.38
σ^2_{eff}	1	0.5	0.41	0.46	0.56	0.62
	2	3	2.40	2.39	3.30	3.20
	3	2	1.52	1.71	2.22	2.39
β	1	2	2.04	0.45	2.04	0.45
	2	0	0.03	0.92	0.04	0.93
	3	-1.5	-1.43	0.84	-1.43	0.84

De façon générale, nous constatons que les moyennes des estimations obtenues sont relativement fidèles aux valeurs en simulation. C'est particulièrement le cas pour les estimations des effets fixes β, qui, par ailleurs, ne diffèrent pas entre les deux algorithmes. En ce qui concerne l'estimation des variances des effets aléatoires, les deux procédures ont réussi à déceler l'hétérogénéité. On retrouve les différences entre les procédures ML et REML, avec une légère sous-estimation pour la première et une légère surestimation pour la deuxième. Notons enfin que les écart-types pour ces mêmes estimations sont du même ordre que les moyennes.

4.4.5.2 Simulations avec erreurs à variances hétérogènes

• Variances des erreurs inférieures à celles des effets :

		Valeurs simulées	Valeurs estimées			
			LINHE_ML		LINHE_REML	
			moy.	e.t.	moy.	e.t.
σ^2_{err}	1	0.5	0.45	0.18	0.46	0.19
	2	1	0.84	0.42	0.91	0.46
	3	1.5	1.29	0.70	1.40	0.78
σ^2_{eff}	1	3	2.13	1.84	2.84	2.44
	2	2.5	1.84	1.75	2.46	2.32
	3	4	2.89	2.80	3.93	3.75
β	1	2	1.89	0.78	1.89	0.78
	2	0	-0.09	0.79	-0.09	0.79
	3	-1.5	-1.62	0.96	-1.62	0.96

• Variances des erreurs de même grandeur que celles des effets :

		Valeurs simulées	Valeurs estimées			
			LINHE_ML		LINHE_REML	
			moy.	e.t.	moy.	e.t.
σ^2_{err}	1	0.5	0.47	0.18	0.49	0.19
	2	1.5	1.18	0.58	1.29	0.64
	3	1	0.82	0.40	0.88	0.43
σ^2_{eff}	1	1	0.78	0.72	1.05	0.95
	2	0.5	0.54	0.66	0.74	0.89
	3	1.5	1.14	1.12	1.64	1.57
β	1	2	2.08	0.54	2.08	0.54
	2	0	0.03	0.47	0.03	0.47
	3	-1.5	-1.39	0.73	-1.39	0.73

• Variances des erreurs supérieures à celles des effets :

		Valeurs simulées	Valeurs estimées			
			LINHE_ML		LINHE_REML	
			moy.	e.t.	moy.	e.t.
σ^2_{err}	1	3	2.55	1.01	2.69	1.08
	2	2.5	2.24	1.09	2.39	1.20
	3	4	3.64	1.99	3.92	2.17
σ^2_{eff}	1	0.5	0.57	0.70	0.81	0.95
	2	1	0.97	1.07	1.42	1.51
	3	1.5	1.50	2.28	2.40	3.62
β	1	2	1.98	0.55	1.98	0.55
	2	0	0.01	0.73	0.01	0.74
	3	-1.5	-1.50	0.89	-1.50	0.90

Au vu de ces trois tableaux de résultats, on constate que l'introduction de l'hétérogénéité sur les erreurs n'altère pas la qualité de l'estimation de β et que globalement on arrive à retrouver les différences entre les variances estimées selon les environnements. Il n'est pas surprenant de s'apercevoir que les grandes valeurs des paramètres de variance ont plus de mal à être atteintes par l'estimation. En ce qui concerne les composantes de l'effet, on remarque comme précédemment que les écart-types sont du même ordre de grandeur que les moyennes. D'autre part, les deux

estimations ML et REML, selon les valeurs des variances résiduelles, encadrent la vraie valeur du paramètre ou sont décallées à droite ou à gauche de celle-ci. Plus les erreurs sont grandes et plus les estimations des variances de l'effet s'éloignent de la vraie valeur. Les composantes résiduelles sont quant à elles toujours légèrement sous-estimées.
Notons enfin que les écart-types des estimations des composantes de β sont rangés dans le même ordre que la somme des composantes de la variance pour chaque environnement, ce qui a du sens étant donné le modèle choisi (où chaque composante de β est affectée à un environnement).

4.5 Estimation des composantes de la variance dans un GL2M hétérogène

Après avoir traité le cas gaussien, nous en arrivons naturellement à l'estimation des composantes de la variance dans les modèles linéaires généralisés mixtes à variances hétérogènes. Nous avons défini précisément, en section 4.2, cette classe de modèles et en particulier la modélisation de l'hétérogénéité que nous avons adoptée et que nous conservons dans tout ce chapitre. Dans cette problématique, nous avons pu constater, en section 4.3, que l'estimation par utilisation directe de l'algorithme EM (avec approximation intégrale par quadrature gaussienne pour le calcul des espérances) était peu envisageable (même si souvent adoptée dans le cas de la surdispersion). Pourtant, selon les résultats de la section 4.4, la démarche EM semble être un outil très intéressant pour faire face au problème de l'hétérogénéité. Dans cette dernière section, nous proposons donc une méthode d'estimation dans les GL2M hétérogènes mettant à profit cet atout.

Cette méthode est une méthode itérative dont chaque itération se présente en deux étapes :

- *la linéarisation* réalisée par l'introduction d'une variable dépendante en suivant la procédure d'estimation classique dans les modèles linéaires généralisés,
- *l'estimation* ayant alors lieu dans le modèle linéarisé obtenu à l'étape précédente, sur lequel nous formulons des hypothèses de modèle linéaire mixte à variances hétérogènes.

C'est une démarche à laquelle nous sommes accoutumés puisqu'elle sous-tendait déjà le travail présenté au chapitre 2. La première étape de linéarisation suit le même schéma et se justifie de façon identique. La deuxième étape s'appuie essentiellement sur les travaux présentés à la section 4.4 dans le cadre des modèles linéaires mixtes à variances hétérogènes.

Nous nous limitons pour présenter cette méthode à un modèle avec un seul effet aléatoire. Nous évoquerons ensuite son extension possible à plusieurs effets aléatoires.

Reprenons les notations du modèle au travers des trois hypothèses suivantes.

- Le vecteur d'observations $y = (y'_1, ..., y'_i, ..., y'_I)'$ rassemble les observations des I environnements (n_i observations pour l'environnement i) et le vecteur d'effet aléatoire $\xi = (\xi_1, ..., \xi_q)'$ possède q niveaux. Le vecteur y est réalisation du vecteur réponse Y dont les composantes, conditionnellement à ξ, sont indépendantes et distribuées selon une loi de la famille exponentielle :

$$\forall i \in \{1, ..., I\}\ , \forall j \in \{1, ..., n_i\}\ ,\ f(y_{ij}|\xi) = \exp\left(\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\Phi)} + c(y_{ij}, \Phi)\right).$$

  Dans cette expression, la fonction a est soit une fonction à valeurs connues, soit elle s'écrit à l'aide d'un paramètre inconnu à estimer (comme pour la loi gamma par exemple). On conserve donc la notation $a(\Phi) = \Phi$, et l'on ne tiendra pas compte le cas échéant des équations concernant l'estimation de Φ, que l'on aura pris soin de remplacer alors par la valeur connue.

- Le prédicteur linéaire inclut l'hétérogénéité :

$$\forall i \in \{1, ..., I\}\ , \qquad \eta_i = X_i\beta + \sigma_i U_i \xi$$

  où X_i et U_i ($n_i \times p$ et $n_i \times q$) sont des matrices fixées,
  β et σ_i des paramètres à estimer.

  On suppose $\xi \sim \mathcal{N}(0, I_q)$ (on pourrait aussi remplacer I_q par une matrice symétrique définie positive A).

- Le lien entre μ_{ij} et η_{ij} est :

$$\forall i \in \{1, ..., I\}\ ,\ \forall j \in \{1, ..., n_i\} \quad : \quad \eta_{ij} = g(\mu_{ij})\ .$$

4.5.1 Proposition d'une méthode d'estimation

4.5.1.1 Étape de linéarisation

Conditionnellement à ξ, le modèle considéré est tout simplement un GLM. Ainsi, comme nous l'avons déjà réalisé au chapitre 2, en suivant le schéma classique des moindres carrés pondérés itérés dans les GLM, nous introduisons le vecteur z dont chaque composante est définie par :

$$\forall i \in \{1, ..., I\}\ ,\ \forall j \in \{1, ..., n_i\}\ , \qquad z_{ij} = x'_{ij}\beta + \sigma_i u'_{ij}\xi + g'(\mu_{ij})(y_{ij} - \mu_{ij})\ .$$

En écrivant le modèle initial conditionnel à ξ sous la forme :

$$\begin{array}{rcccc} Y_{ij} & = & \mu_{ij} & + & e_{ij} \\ & = & g^{-1}(\eta_{ij}) & + & e_{ij} \ , \end{array}$$

où
$$\begin{array}{rcl} E(e_{ij}|\xi) & = & 0 \\ \mathrm{var}(e_{ij}|\xi) & = & \Phi \ b''(\theta_{ij}) \ , \end{array}$$

cette expression de z_{ij} peut être vue comme un développement limité au premier ordre de $g(y_{ij})$ en μ_{ij}.

Notons $\forall i \in \{1, ..., I\}$, $\forall j \in \{1, ..., n_i\}$, $\varepsilon_{ij} = g'(\mu_{ij})(Y_{ij} - \mu_{ij}) = g'(\mu_{ij})e_{ij}$. Nous considérons alors la modélisation suivante pour le vecteur aléatoire Z (dont z est issu) :

$$\forall i \in \{1, ..., I\} \ , \ \forall j \in \{1, ..., n_i\} \ , \qquad Z_{ij} = x'_{ij}\beta + \sigma_i u'_{ij}\xi + \varepsilon_{ij} \ , \tag{4.8}$$

écriture à laquelle sont associées des hypothèses de lois normales pour les distributions de ξ et ε et donc de Z. La loi de ξ reste $\mathcal{N}(0, I_q)$. Quant à celle de ε, elle respecte :

$$\forall i \in \{1, \ldots, I\} \ , \ \forall j \in \{1, \ldots, n_i\} \ , \qquad \begin{array}{rcl} E(\varepsilon_{ij}|\xi) & = & 0 \\ \mathrm{var}(\varepsilon_{ij}|\xi) & = & \Phi \ g'(\mu_{ij})^2 \ b''(\theta_{ij}) \ . \end{array}$$

Conditionnellement à ξ, les ε_{ij} étant indépendants, la matrice diagonale, de taille $N = \sum_{i=1}^{I} n_i$, définie par :

$$W_\xi = \{_d \ \Phi \ g'(\mu_{ij})^2 \ b''(\theta_{ij}) \ \}_{i=1,\ldots,I, j=1,\ldots,n_i} \ ,$$

est la matrice de variance résiduelle, conditionnelle à ξ, du modèle (4.8). Dans le cas d'un lien canonique, cette matrice s'écrit :

$$W_\xi = \{_d \ \Phi \ g'(\mu_{ij}) \ \}_{i=1,\ldots,I, j=1,\ldots,n_i} \ .$$

On peut aussi la découper en blocs de matrices diagonales et on note $W_{\xi,i}$ la matrice résiduelle correspondant au $i^{\text{ème}}$ environnement :

$$W_{\xi,i} = \{_d \ \Phi \ g'(\mu_{ij})^2 \ b''(\theta_{ij}) \ \}_{j=1,\ldots,n_i}$$

Enfin, on notera W et W_i les espérances de ces matrices :

$$\begin{array}{rcccl} \mathrm{Var}(\varepsilon) & = & W & = & E(W_\xi) \\ \mathrm{Var}(\varepsilon_i) & = & W_i & = & E(W_{\xi,i}) \ . \end{array}$$

Et on désignera par W_ξ^* et W^* les matrices mises à l'échelle : $W_\xi = \Phi W_\xi^*$ et $W = \Phi W^*$, de même pour $W_{\xi,i}^*$ et W_i^*.

Ce modèle, défini par (4.8) et les hypothèses qui suivent, sera appelé *modèle linéarisé*. De même qu'au deuxième chapitre, toutes les hypothèses sont donc réunies pour reconnaître dans ce modèle linéarisé un modèle linéaire mixte mais cette fois-ci à variances hétérogènes pour l'effet aléatoire. Pourtant, cette hypothèse gaussienne va en contradiction avec le fait que, de façon générale, les matrices W et W_ξ sont différentes. Cette distinction a d'ailleurs abouti aux deux méthodes (celle proposée par Schall et la notre) dont nous avons discuté au chapitre 2. Pour ne pas reprendre ici ce débat, nous adopterons la notation $\mathcal{W}$ pour désigner indifféremment W et W_ξ. De même, nous noterons $\mathcal{W}_i$ le bloc de $\mathcal{W}$ correspondant à l'environnement i et $\mathcal{W}^*$, $\mathcal{W}_i^*$ ces mêmes matrices mises à l'échelle.

D'autre part, d'un point de vue de l'estimation et comme dans le cas classique des GLM, les paramètres inconnus du modèle linéarisé interviennent à la fois dans la partie explicative du modèle, dans la matrice de variance résiduelle $\mathcal{W}$ ainsi que dans la définition de z. C'est pourquoi nous envisageons une procédure itérative pour l'estimation. À chaque itération, z et $\mathcal{W}$ sont évaluées aux valeurs courantes des paramètres. Ainsi, à la $t^{\text{ème}}$ itération, pour les valeurs $\beta^{[t]}$ et $\sigma_i^{[t]}$ $(i = 1, ..., I)$ des paramètres et pour $\xi^{[t]}$, la valeur de z_{ij} sera calculée par :

$$z_{ij}^{[t]} = x_{ij}'\beta^{[t]} + \sigma_i^{[t]} u_{ij}'\xi^{[t]} + g'(\mu_{ij}^{[t]})(y_{ij} - \mu_{ij}^{[t]}) \ .$$

Le vecteur z prendra donc une nouvelle valeur à chaque fois.

Pour cet algorithme itératif, au pas t de l'itération, on considèrera donc le modèle linéarisé dans lequel :
- les valeurs observées des Z_{ij} sont les $z_{ij}^{[t]}$,
- la matrice de variance résiduelle (conditionnelle ou non) est $\mathcal{W}^{[t]}$.
On obtient alors les valeurs $[t+1]$ en procédant à l'estimation au sein de ce modèle, ce que nous décrivons ci-dessous.

4.5.1.2 Étape d'estimation

Dans toute cette section, nous nous plaçons donc dans le cadre du modèle linéarisé obtenu après la $t^{\text{ème}}$ linéarisation et décrit dans les lignes précédentes. Nous le regardons comme un *modèle linéaire mixte à variances hétérogènes*, dans lequel nous pouvons utiliser les algorithmes LINHE_ML et LINHE_REML établi en section 4.4. Nous reprenons ici plus spécialement la démarche ayant conduit à l'algorithme LINHE_REML et basée sur les travaux de Foulley et Quaas (1995). Comme eux, nous envisageons un traitement bayésien du vecteur β. Pour cela, un a priori non informatif (avec une variance infinie) est supposé pour ce paramètre. La procédure proposée permet alors d'estimer, dans une démarche EM, les paramètres $\sigma^2 = (\sigma_1^2, ..., \sigma_I^2)'$ et Φ le cas échéant.

Notations :
• Pour alléger l'écriture dans toute la suite, nous supprimons les exposants des vecteurs z et Z ainsi que des matrices $\mathcal{W}$ et $\mathcal{W}_i$ et leurs correspondantes rééchelonnées (autrement dit tous les élément dont la valeur a été fixée par la $t^{\text{ème}}$ itération) mais c'est bien $z^{[t]}$, $Z^{[t]}$, $\mathcal{W}^{[t]}$, $\mathcal{W}_i^{[t]}$, $\mathcal{W}^{*[t]}$ et $\mathcal{W}_i^{*[t]}$ que nous désignons.
• On note toujours ξ à la fois le vecteur aléatoire et sa réalisation.
• La procédure mise en place dans ce $t^{\text{ème}}$ modèle linéarisé est une procédure itérative dont m représentera l'indice d'itération.

Nous nous attachons dans un premier temps à écrire la vraisemblance des données complètes : $X = (Z', \xi', \beta')'$; z jouant le rôle, dans l'algorithme EM, des données observées, et ξ et β celui des données manquantes. Nous notons $\delta = (\sigma^{2\prime}, \Phi)'$ le vecteur des paramètres à estimer. Ainsi, la vraisemblance complète s'écrit :

$$\begin{aligned} L(\delta; x) &= \ln[f(z, \xi, \beta|\delta)] \\ &= \ln[f(z|\xi, \beta, \delta)] + \ln[f(\xi|\beta, \delta)] + \ln[f(\beta|\delta)] \; . \end{aligned}$$

Avec les hypothèses de distributions normales du modèle linéarisé, on a :

$$\begin{aligned} f(z|\xi, \beta, \delta) &= \prod_{i=1}^{I} \frac{1}{(2\pi)^{\frac{n_i}{2}} |\mathcal{W}_i|^{\frac{1}{2}}} \exp[-\frac{1}{2}(z_i - X_i\beta - \sigma_i U_i \xi)' \mathcal{W}_i^{-1} (z_i - X_i\beta - \sigma_i U_i \xi)] \\ &= \prod_{i=1}^{I} \frac{1}{(2\pi)^{\frac{n_i}{2}} |\mathcal{W}_i^*|^{\frac{1}{2}} \Phi^{\frac{n_i}{2}}} \exp[-\frac{1}{2\Phi} \varepsilon_i' \mathcal{W}_i^{*-1} \varepsilon_i] \; , \end{aligned}$$

d'où $\quad \ln[f(z|\xi, \beta, \delta)] = \text{const} - \sum_{i=1}^{I} \frac{n_i}{2} \ln(\Phi) - \sum_{i=1}^{I} \frac{1}{2\Phi} \varepsilon_i' \mathcal{W}_i^{*-1} \varepsilon_i.$

On a aussi : $\ln[f(\xi|\beta, \delta)] = -\frac{q}{2} \ln(2\pi) - \frac{1}{2} \xi' \xi.$

Ce n'est pas directement cette vraisemblance complète L que l'on maximise, mais son espérance conditionnelle aux observations, étant données les valeurs courantes des paramètres ; autrement dit la fonction $Q(\delta|\delta^{[m]}) = E(\ln[f(z, \xi|\delta)]|z, \delta^{[m]})$, à l'itération $[m]$ de l'algorithme. Dans l'objectif de la maximisation de cette fonction, nous conservons uniquement les éléments contenant une information sur les paramètres et qui ne disparaîtront pas à la dérivation. Ce qui donne :

$$Q(\delta|\delta^{[m]}) = \text{const} - \frac{N}{2}\ln(\Phi) - \frac{1}{2\Phi}\sum_{i=1}^{I} E(\ \varepsilon_i' \mathcal{W}_i^{*-1} \varepsilon_i\ |z, \delta^{[m]})\ .$$

Il s'agit maintenant de dériver cette fonction par rapport aux divers paramètres composant δ. Notons que les matrices de variance $\mathcal{W}$ et ses associées s'expriment à l'aide de ces paramètres. Cependant dans le modèle linéarisé, elles ont été fixées par les valeurs à l'étape t et sont donc considérées ici comme constantes. On obtient alors les dérivées suivantes :

$$\begin{aligned}
\forall i \in \{1,..,I\}\ , \quad \frac{\partial Q(\delta|\delta^{[m]})}{\partial \sigma_i^2} &= -\frac{1}{2\Phi} \quad \Big\{ \frac{1}{\sigma_i} E\Big[\ \xi' U_i' \mathcal{W}_i^{*-1}(z_i - X_i\beta)\ |z, \delta^{[m]}\Big] \\
&\qquad - E\Big[\ \xi' U_i' \mathcal{W}_i^{*-1} U_i \xi\ |z, \delta^{[m]}\Big]\Big\} \\
\text{et} \qquad \frac{\partial Q(\delta|\delta^{[m]})}{\partial \Phi} &= -\frac{1}{2\Phi} \quad \Big\{ N - \frac{1}{\Phi}\sum_{i=1}^{I} E\Big[\ \varepsilon_i' \mathcal{W}_i^{*-1} \varepsilon_i\ |z, \delta^{[m]}\Big]\Big\}\ .
\end{aligned}$$

Ce qui conduit, en annulant ces dérivées, au schéma itératif :

$$\begin{aligned}
\forall i \in \{1,..,I\}\ , \quad \sigma_i^{[m+1]} &= \frac{E\Big[\ \xi' U_i' \mathcal{W}_i^{*-1}(z_i - X_i\beta)\ |z, \delta^{[m]}\Big]}{E\Big[\ \xi' U_i' \mathcal{W}_i^{*-1} U_i \xi\ |z, \delta^{[m]}\Big]} \\
\text{et,} \qquad \Phi^{[m+1]} &= \frac{\sum_{i=1}^{I} E\Big[\ \varepsilon_i' \mathcal{W}_i^{*-1} \varepsilon_i\ |z, \delta^{[m]}\Big]}{N}\ .
\end{aligned}$$

Pour calculer ces espérances conditionnelles, nous utilisons les solutions du système des équations de Henderson du modèle linéaire mixte et la matrice des coefficients de ce système. Pour le modèle linéarisé (4.8), ce système d'équations est le suivant (en notant toujours sans exposants $z^{[t]}$ et $\mathcal{W}_i^{[t]}$) :

$$\begin{bmatrix} \sum_{i=1}^{I} X_i' \mathcal{W}_i^{-1} X_i & \sum_{i=1}^{I} \sigma_i^{[m]} X_i' \mathcal{W}_i^{-1} U_i \\ \sum_{i=1}^{I} \sigma_i^{[m]} U_i' \mathcal{W}_i^{-1} X_i & \sum_{i=1}^{I} \sigma_i^{2[m]} U_i' \mathcal{W}_i^{-1} U_i + I \end{bmatrix} \begin{bmatrix} \beta \\ \xi \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{I} X_i' \mathcal{W}_i^{-1} z_i \\ \sum_{i=1}^{I} \sigma_i^{[m]} U_i' \mathcal{W}_i^{-1} z_i \end{bmatrix}$$

On note toujours $\hat{\beta}$ et $\hat{\xi}$ les solutions obtenues à ce système d'équations et C l'inverse de la matrice des coefficients. Les espérances conditionnelles s'expriment alors sous la forme :

$$\begin{aligned}
E1_i^{[m]} &= E\left[\,\xi' U_i' \mathcal{W}_i^{*-1}(z_i - X_i\beta)\;|z, \delta^{[m]}\right] \\
&= \hat{\xi}' U_i' \mathcal{W}_i^{*-1}(z_i - X_i\hat{\beta}) - \mathrm{tr}(U_i' \mathcal{W}_i^{*-1} X_i C_{\beta\xi})
\end{aligned}$$

$$E2_i^{[m]} = E\left[\,\xi' U_i' \mathcal{W}_i^{*-1} U_i \xi\;|z, \delta^{[m]}\right] = \hat{\xi}' U_i' \mathcal{W}_i^{*-1} U_i \hat{\xi} + \mathrm{tr}(U_i' \mathcal{W}_i^{*-1} U_i C_{\xi\xi})$$

$$\begin{aligned}
E3_i^{[m]} &= E\left[\,\varepsilon_i' \mathcal{W}_i^{*-1} \varepsilon_i\;|z, \delta^{[m]}\right] \\
&= E\left[\,(z_i - X_i\beta - \sigma_i U_i \xi)' \mathcal{W}_i^{*-1}(z_i - X_i\beta - \sigma_i U_i \xi)\;|z, \delta^{[m]}\right] \\
&= (z_i - X_i\hat{\beta})' \mathcal{W}_i^{*-1}(z_i - X_i\hat{\beta}) + \mathrm{tr}(X_i' \mathcal{W}_i^{*-1} X_i C_{\beta\beta}) \\
&\quad -2\sigma_i^{[m]}\, E1_i^{[m]} + \sigma_i^{2[m]}\, E2_i^{[m]}\,.
\end{aligned}$$

Ceci permet d'affecter les nouvelles valeurs suivantes à σ_i et Φ :

$$\forall i \in \{1, \ldots, N\}\,, \qquad \sigma_i^{[m+1]} = \frac{E1_i^{[m]}}{E2_i^{[m]}}\,, \qquad \text{et} \qquad \Phi^{[m+1]} = \frac{\sum_{i=1}^{I} E3_i^{[m]}}{N}\,.$$

Si l'on veut procéder à l'estimation des paramètres du modèle linéarisé dans lequel nous nous sommes placés au début de cette section, il reste à itérer ce processus, c'est-à-dire former le nouveau système de Henderson à l'étape $[m+1]$ pour obtenir des estimations $[m+2]$ et ainsi de suite ... À la convergence de cette procédure à l'étape $[t]$, on obtient les valeurs $\dot{\sigma}_i$ et $\dot{\Phi}$, et l'on pose alors $\sigma_i^{[t+1]} = \dot{\sigma}_i$, $\Phi^{[t+1]} = \dot{\Phi}$, $\beta^{[t+1]} = \hat{\beta}$ et $\xi^{[t+1]} = \hat{\xi}$. À l'aide de ces nouvelles valeurs, il est maintenant possible de procéder à une nouvelle linéarisation, i.e. calculer les $z^{[t+1]}$ et $\mathcal{W}^{[t+1]}$ et donc définir le nouveau modèle linéarisé.

4.5.1.3 La procédure

La procédure que nous proposons alterne donc entre ces deux phases de *linéarisation* et d'*estimation*. D'un point de vue algorithmique, deux itérations sont ici imbriquées : la procédure globale de réactualisation du modèle linéarisé, qui, pour chaque modèle, fait appel à une procédure d'estimation, elle même itérative (comme la plupart des procédures d'estimation des composantes de la variance évoquées jusqu'ici). Pour l'implémentation, nous avons choisi de ne réaliser qu'un seul pas de la procédure d'estimation à l'intérieur de chaque phase de linéarisation. Nous avions déjà réalisé ce même choix au chapitre 2, et ceci en vue de simulations à réaliser. D'autres alternatives pourraient consister à en réaliser un nombre de pas fixé, voire de la laisser converger, avant une nouvelle linéarisation. Il faudra de toute façon rester prudent puisque rien ne nous assure qu'avec les critères d'arrêt choisis, la convergence soit atteinte. De plus, cette dernière solution présente le désavantage d'être plus longue. Pour pallier cela, une dernière alternative peut encore être envisagée, qui consiste à avancer rapidement en début de procédure (avec un seul pas de EM) pour ensuite laisser EM davantage travailler sur la fin. Il nous semble que ce n'est pas ici le lieu pour discuter de ces différences algorithmiques.

D'autre part, des deux algorithmes présentés dans le cas linéaire hétérogène, nous avons uniquement repris ici l'algorithme LINHE_REML. Selon la présentation que nous en avons faite, l'avantage de celui-ci réside dans l'utilisation du système de Henderson qui, ne se contentant pas de permettre le calcul des espérances, fournit aussi des valeurs pour ξ. Or, comme nous l'avions souligné au chapitre 2, ceci est important pour la réactualisation des valeurs de z et éventuellement de $\mathcal{W}$ (si on utilise W_ξ) lors de la phase de linéarisation. Cependant, à défaut, il existe d'autres façons d'obtenir des valeurs prédites pour ξ. Nous en avons présenté au chapitre 2. Nous utiliserons donc aussi l'algorithme LINHE_ML lors des simulations mais nous ne le détaillons pas ici : il se réécrit de façon similaire.

Revenons un instant sur la matrice $\mathcal{W}$. En effet, pour reprendre une remarque de la section 4.5.1.1, la différence entre les matrices W et W_ξ fait naître une ambiguité dans l'approximation "linéaire" (hypothèse de loi normale) du modèle linéarisé. Le choix entre l'une de ces deux matrices reste un sujet de débat. Nous verrons dans la section suivante que dans le cas homogène, ce choix débouche sur les deux alternatives dont nous avons largement discuté au deuxième chapitre et dont les arguments ont déja été exposés.
Ici, lors de la mise en place de la procédure d'estimation, la première fois où la matrice $\mathcal{W}$ intervient se situe dans l'expression de la loi conditionnelle de Z à ξ (ceci est dû au fait que l'algorithme EM se construit à partir de la loi conjointe de Z et ξ). Il semblerait alors plus naturel d'utiliser la matrice W_ξ. Cependant, si tel était le cas, il serait alors nécessaire lors du calcul des espérances conditionnelles de tenir compte du fait que cette matrice est aléatoire. Ainsi, on devrait en calculer aussi son espérance ou sa covariance avec les autres variables aléatoires selon la distribution de ξ sachant z. Pour ces raisons, ainsi que celles évoquées au deuxième chapitre, nous préférons choisir pour $\mathcal{W}$ la matrice W et non W_ξ.

Il est important de rappeler que si les estimations de σ_i n'avaient pas été construites à l'aide de la démarche EM, nous n'aurions pas su à partir du système de Henderson obtenir des estimations de ces composantes de la variance. En effet, écrit sous la forme précédente, aucune équation de ce système ne nous conduit à cette estimation. Et même si nous avions laissé les σ_i^2 comme paramètres de variance

de la distribution des ξ, l'obtention des valeurs de ξ ne nous permettait pas non plus cette estimation, puisqu'une même composante peut intervenir pour des σ_i différents. Nous avons déjà évoqué ce problème lié à l'hétérogénéité en section 4.4.2. Ainsi, l'algorithme EM, écrit en sortant σ_i de la distribution de ξ, trouve ici une grande utilité.
De plus, le fait d'être passé par une étape de linéarisation présente au moins deux avantages :

- le problème du calcul intégral (et de son approximation) ne se pose plus. En effet, dans le cas gaussien, les règles de conditionnement répondent au problème.
- traiter l'hétérogénéité dans le cas linéaire est plus simple et nous en avons discuté des solutions précédemment.

Pour finir, nous avons présenté notre procédure uniquement dans le cas d'un modèle avec un seul effet aléatoire. Elle peut être étendue sans aucun problème au cas de plusieurs effets aléatoires. Le modèle linéarisé s'écrira pour K effets :

$$\forall i \in \{1, \ldots, I\} \ , \qquad Z_i = X_i\beta + \sum_{j=1}^{K} \sigma_{ij} U_{ij} \xi_j + \varepsilon_i \quad \text{où} \quad \mathrm{Var}(\varepsilon_i) = \mathcal{W}_i \ .$$

Il en découle ensuite l'écriture habituelle du système de Henderson.

4.5.2 Le cas particulier de l'homogénéité

Dans ce paragraphe, nous cherchons à savoir ce que devient la procédure d'estimation que nous venons de mettre en place, dans le cas particulier de l'homogénéité, c'est-à-dire dans le cas où $\forall i \in \{1, \ldots, I\}$, $\sigma_i^2 = \sigma^2$. Nous considérons à nouveau un modèle avec un seul effet aléatoire. Il n'est plus nécessaire, avec cette hypothèse d'homogénéité, de faire la distinction entre les différents environnements. On rassemble les N individus, que l'on indice dorénavant par i. Les matrices ne sont plus redécoupées en sous-blocs, x_i' et u_i' concernent le $i^{\text{ème}}$ individu et désignent les $i^{\text{èmes}}$ lignes de X et U. Le paramètre σ est toujours sorti de la distribution du vecteur ξ. Et enfin, le prédicteur linéaire peut s'écrire vectoriellement (η est alors de taille N) sous la forme simple :

$$\eta = X\beta + \sigma U\xi \ .$$

Dans l'étape de linéarisation, aucune modification de fond n'est nécessaire. Le modèle linéarisé s'écrit alors :

$$Z = X\beta + \sigma U\xi + \varepsilon$$

$$\begin{array}{llcll} \text{où} & \mathrm{Var}(\varepsilon|\xi) & = & W_\xi & = \{_d \ \Phi g'(\mu_i)^2 b''(\theta_i)\ \}_{i=1,\dots,N} \\ & \mathrm{Var}(\varepsilon) & = & W & = E(W_\xi) \\ & & = & \Phi\, W^* \,. & \end{array}$$

De même que précédemment, les paramètres à estimer étant présents à la fois dans la partie explicative et dans la matrice résiduelle, il est naturel d'envisager un algorithme d'estimation itératif. Ainsi, à l'itération t, la matrice de variance du modèle linéarisé est $\mathcal{W}^{[t]}$ ($\mathcal{W}$ désigne toujours indifféremment les matrices W ou W_ξ) et la réalisation de Z est $z^{[t]}$.
Dans toute la suite, on omettra les exposants de ces éléments et on écrira $\mathcal{W}$, $\mathcal{W}^*$ et z pour désigner $\mathcal{W}^{[t]}$, $\mathcal{W}^{*[t]}$ et $z^{[t]}$.

L'étape suivante d'estimation reprend la démarche itérative EM, au cours de laquelle les expressions des paramètres à l'étape $[m+1]$ sont :

$$\begin{aligned} \sigma^{[m+1]} &= \frac{E\left[\ \xi' U' \mathcal{W}^{*-1}(Z - X\beta)\ |z, \delta^{[m]}\right]}{E\left[\ \xi' U' \mathcal{W}^{*-1} U \xi\ |z, \delta^{[m]}\right]} \\ \text{et,}\quad \Phi^{[m+1]} &= \frac{E\left[\ \varepsilon' \mathcal{W}^{*-1} \varepsilon\ |z, \delta^{[m]}\right]}{N}\,. \end{aligned}$$

Ils seront de même calculés à l'aide des coefficients et des solutions du système de Henderson, associé au modèle linéarisé.
À la convergence de cette procédure d'estimation, on obtient (on l'a vu en section 4.4.3) les estimations REML des paramètres du modèle linéarisé à l'étape $[t]$. Ces estimations sont donc identiques à celles auxquelles on aboutit dans le même modèle réécrit avec σ rentré dans la distribution de ξ, à l'aide de la résolution direct du système REML ou par le système de Henderson. Et c'est cette dernière démarche qu'adoptent les algorithmes présentés au chapitre 2.

Tout ceci reste vrai à chaque étape $[t]$ de linéarisation. C'est pourquoi de façon globale l'algorithme obtenu ici est équivalent à ceux du deuxième chapitre. Et le même propos peut être tenu pour les estimations ML avec l'algorithme LINHE_ML transposé au cas homogène. Cependant, ici, comme au chapitre 2, nous avons choisi de n'effectuer qu'un seul pas de la procédure d'estimation à l'intérieur de chaque phase de linéarisation. Ceci serait donc susceptible de perturber l'équivalence entre les algorithmes ; puisqu'après un seul pas, les estimations obtenues par les divers algorithmes sont cette fois-ci différentes et le modèle linéarisé à l'étape $[t+1]$ le sera donc aussi. Néanmoins, nous avons observé que, quel que soit l'algorithme d'estimation du modèle linéaire utilisé (EM, système REML, Henderson), cette implémentation n'affectait pas les estimations obtenues en fin de procédure.
Nous en concluons donc que la procédure proposée dans ce chapitre aboutit, dans

le cas homogène, aux mêmes estimations que :

- la procédure SCHALL lorsque $\mathcal{W} = W_\xi$,

- notre procédure du premier chapitre lorsque $\mathcal{W} = W$.

4.5.3 Simulations

Après avoir présenté une méthode d'estimation des composantes de la variance dans les GL2M hétérogènes, et avoir vu qu'elle n'était pas sans lien avec les méthodes discutées au chapitre 2, il est maintenant indispensable d'observer son comportement numérique. C'est ce que nous faisons sur des simulations. Nous considérons pour cela différentes lois de la famille exponentielle et différentes fonctions de lien. Bien entendu nous ne reprenons pas le cas de la loi normale, auquel la section 4.4 a été consacrée. Nous envisageons ici les cas :

- loi binomiale - lien logit,

- loi de Poisson - lien log,

- loi exponentielle - lien log.

En ce qui concerne le prédicteur, nous conservons toujours le plan d'expérience de l'exemple (cf. 4.4.1). Nous simulons 200 jeux de données, et nous donnons dans les tableaux ci-dessous les résultats des moyennes et écart-types des 200 estimations obtenues par les différents algorithmes : approche REML, approche ML soit avec W_ξ, soit avec W (dans le cas où ces matrices sont différentes). Evidemment, lorsque nous désignons les colonnes de ces tableaux par REML ou ML, nous ne faisons en aucun cas référence à des démarches de vraisemblance dans le GL2M d'origine, mais uniquement aux algorithmes choisis dans les modèles linéarisés successifs.
Ces simulations sont réalisées pour deux valeurs différentes du vecteur de paramètres σ^2 : $\sigma^2 = (1.5; 0.5; 1)'$ et $\sigma^2 = (0.05; 4; 1)'$. Pour β, nous conservons le vecteur nul de façon à éviter des problèmes numériques notamment dans le cas du lien logarithme où l'espérance des lois simulées pourrait alors atteindre de très grandes valeurs.

Enfin, avant de présenter ces tableaux, rappelons que, dans la configuration de référence, le vecteur des données est de taille $n = 36$ et que le vecteur des effets aléatoires possède $q = 4$ réalisations, deux chiffres relativement faibles.

• Cas binomial - lien logit

		Valeurs estimées							
		REML				ML			
	Valeurs	W_ξ		W		W_ξ		W	
	simulées	moy.	e.t.	moy.	e.t.	moy.	e.t.	moy.	e.t.
	1.5	1.32	1.03	1.33	1.04	1.00	0.79	1.00	0.79
σ^2	0.5	0.44	0.34	0.45	0.34	0.33	0.26	0.33	0.26
	1	0.87	0.69	0.87	0.69	0.65	0.53	0.65	0.53
	0	0.08	0.64	0.08	0.65	0.08	0.64	0.08	0.64
β	0	0.05	0.39	0.05	0.39	0.05	0.39	0.05	0.39
	0	0.08	0.53	0.08	0.53	0.08	0.52	0.08	0.53
	0.05	0.05	0.04	0.05	0.05	0.04	0.03	0.04	0.03
σ^2	4	3.42	2.79	3.42	2.74	2.56	2.10	2.59	2.14
	1	0.84	0.69	0.84	0.73	0.63	0.52	0.62	0.53
	0	0.00	0.12	0.00	0.12	0.00	0.12	0.00	0.12
β	0	0.05	0.94	0.05	0.95	0.05	0.93	0.05	0.95
	0	0.03	0.49	0.03	0.49	0.03	0.48	0.03	0.49

Manifestement, quelles que soient les valeurs de σ^2, les différents algorithmes ont très bien décelé l'hétérogénéité. Et même si les estimations sont plus ou moins proches des valeurs simulées selon les cas REML/ML, les ordres de grandeur des différentes valeurs des composantes de σ^2 sont respectés. De même, les résultats sur β sont très bons dans tous les cas.
En ce qui concerne les dispersions de ces estimations, notons que :

- pour σ^2, les écart-types restent de même grandeur que les moyennes obtenues donc suivent grosso modo les grandeurs simulées,
- l'écart-type de β_i est proche de $\dfrac{\sigma_i}{2}$!

Enfin, au moins dans ce cas, nous ne lisons pas de différence entre les procédures utilisant W et celles utilisant W_ξ.

Dans le tableau ci-dessous, nous observons l'évolution de ces estimations lorsque :

- nous doublons le nombre de données, c'est-à-dire que nous doublons le nombre d'observations par environnement et nous répétons une deuxième fois le plan d'expérience,
- nous doublons le nombre de données et nous doublons le nombre de pères, donc le nombre de réalisations de l'effet aléatoire.

Nous donnons uniquement les résultats de la procédure REML / W.

	Valeurs simulées	Valeurs estimées $n = 36$ $q = 4$ moy.	e.t.	$n = 72$ $q = 4$ moy.	e.t.	$n = 72$ $q = 8$ moy.	e.t.
σ^2	1.5	1.33	1.04	1.20	0.78	1.20	0.72
	0.5	0.45	0.34	0.40	0.27	0.40	0.25
	1	0.87	0.69	0.80	0.53	0.79	0.48
β	0	0.08	0.65	-0.01	0.58	-0.00	0.86
	0	0.05	0.39	-0.01	0.34	-0.01	0.50
	0	0.08	0.53	-0.01	0.48	-0.00	0.70

Il n'est pas surprenant de constater que le fait de doubler n entraîne une diminution de la dispersion des estimations. Pour σ^2, cette diminution s'accentue encore davantage en augmentant q le nombre de réalisations de l'effet aléatoire. Mais inversement, la dispersion des estimations de β augmente alors. À n fixé, la variance de $\hat{\beta}$ augmente avec le nombre de réalisations de l'effet aléatoire.

• Cas Poisson - lien log

	Valeurs simulées	Valeurs estimées REML W_ξ moy.	e.t.	REML W moy.	e.t.	ML W_ξ moy.	e.t.	ML W moy.	e.t.
σ^2	1.5	1.49	1.62	1.70	1.70	0.98	1.03	1.34	1.46
	0.5	0.83	1.26	1.08	1.84	0.51	0.65	0.78	1.37
	1	1.11	1.35	1.53	2.45	0.66	0.75	1.03	1.71
β	0	0.09	0.64	-0.05	0.70	0.13	0.62	-0.06	0.70
	0	-0.10	0.56	-0.20	0.62	-0.07	0.53	0.19	0.62
	0	0.05	0.63	-0.07	0.70	0.08	0.62	-0.07	0.70
σ^2	0.05	0.19	0.28	0.33	0.72	0.12	0.18	0.25	0.60
	4	2.54	1.71	2.58	1.79	1.65	1.20	1.96	1.49
	1	1.12	1.47	1.60	2.48	0.67	0.75	1.06	1.70
β	0	-0.07	0.30	-0.13	0.34	-0.07	0.29	-0.12	0.34
	0	0.22	0.90	0.06	0.95	0.30	0.87	0.06	0.92
	0	-0.02	0.61	-0.15	0.75	0.01	0.59	-0.13	0.74

Les estimations obtenues ont su rendre l'idée de l'hétérogénéité simulée, même si les grandes valeurs de σ^2 restent difficiles à atteindre. On note une plus grande disparité entre les algorithmes utilisant W et ceux utilisant W_ξ, particulièrement pour l'estimation des composantes de la variance. Cette différence tend sensiblement vers une plus grande dispersion des estimations dans le cas de l'utilisation de W.

• **Cas exponentiel - lien log**

	Valeurs simulées	Valeurs estimées REML moy.	REML e.t.	ML moy.	ML e.t.
σ^2	1.5	1.62	1.49	1.17	1.12
	0.5	0.68	0.80	0.49	0.58
	1	1.29	1.59	0.85	1.02
β	0	-0.16	0.57	-0.16	0.57
	0	-0.11	0.45	-0.11	0.45
	0	-0.18	0.59	-0.16	0.59
σ^2	0.05	0.14	0.20	0.10	0.14
	4	3.34	3.10	2.37	2.30
	1	1.14	2.21	0.67	1.24
β	0	-0.05	0.32	-0.05	0.32
	0	-0.12	1.19	-0.10	1.20
	0	-0.06	0.72	-0.06	0.69

Dans ce cas, les matrices W et W_ξ sont les mêmes puisqu'elles sont égales à la matrice identité. Les résultats sont relativement similaires à ceux du cas binomial pour l'estimation de σ^2 : une fois de plus l'hétérogénéité est bien détectée avec toujours la même différence sensible entre les estimations REML et ML ! Notons, en ce qui concerne β, que la qualité de l'estimation n'est pas sans lien avec l'importance des valeurs des composantes de l'effet.

Conclusion

La modélisation avec effets aléatoires a ouvert de nouveaux horizons sur les façons d'envisager, au sein de modèles, l'explication d'une variable observée lors d'une expérience. La génétique animale en est actuellement un domaine d'application privilégié, mais ce type de modélisation peut s'avérer utile en bien d'autres situations encore. Ainsi, la pertinence de la classe des modèles linéaires généralisés à effets aléatoires n'est plus à démontrer.

Dans ce travail, nous nous sommes intéressés à la question de l'estimation des paramètres d'effets fixes et des composantes de la variance de tels modèles.

Nous l'avons souligné à plusieurs reprises : ces GL2M sont définis conditionnellement aux effets aléatoires, qui ne sont pas directement observés. Toute la difficulté réside donc dans la levée du conditionnement. C'est pourquoi, pour faire face à ce problème, les méthodes d'estimation construites s'appuient sur diverses approximations. Tout en gardant un point de vue global sur ces modèles, nous avons tenté de dégager les différents types d'approximations mis en jeu et les différents niveaux de déconditionnement du modèle réalisé.

Au cours des chapitres 2 et 3, nous avons pu explorer deux voies principales de raisonnement :

- soit en se plaçant au niveau du modèle conditionnel, la linéarisation est alors effectuée au tout début de la démarche (chapitre 2),

- soit en ce plaçant au niveau du modèle marginal, la linéarisation est alors réalisée en cours de route (chapitre 3).

Nous avons essayé d'éclairer et de comparer ces différents points de vue, en notant aussi les cas où ils coïncidaient.

Dans le chapitre 2, nous avons proposé une méthode simple d'estimation, combinant les techniques GLM et L2M, par une relecture de la procédure GLM à l'aide du modèle linéarisé. Cette méthode s'est avérée proche de celle proposée par Schall (1991), les différences étant principalement résumées par l'utilisation respective des matrices W et W_ξ. Si cette dernière méthode a reçu les éclairages de diverses démarches permettant d'y aboutir, il serait intéressant d'étudier comment ces mêmes raisonnements peuvent être appliquées pour justifier, de façon similaire, l'utilisation de la matrice W.

La méthode que nous proposons permet en outre de tirer partie d'une information, apportée dans certains cas par les GL2M, sur la structure de variance. Mais cela ne nous prive pas pour autant, si nécessaire, d'une éventuelle estimation du paramètre de dispersion.

Les simulations que nous avons réalisées ne montrent pas de différence entre l'utilisation de W et celle de W_ξ, autrement dit entre notre méthode et celle de Schall. Il est donc nécessaire, dans un premier temps, de compléter ces simulations, et, le cas échéant, à défaut d'exhiber des différences, de déterminer les raisons théoriques qui rendraient semblables ces deux méthodes.

Dans le chapitre 3, nous avons étendu la méthodologie développée par Gilmour, Anderson, et Rae (1985). Nous espérons avoir donné un cadre plus général à son utilisation, tout en soulignant le domaine de validité restreint de cette méthode aux faibles valeurs des composantes.

Au chapitre 4, nous avons introduit une notion d'hétérogénéité au sein des modèles linéaires généralisés à effets aléatoires. C'est en nous appuyant sur les travaux réalisés au chapitre 2, que nous avons pu mettre en place une méthode d'estimation dans ce nouveau cadre. L'outil principal de ce chapitre est l'algorithme EM. Nous n'avons traité que de sa mise en place pour répondre à notre problématique, sans pour autant faire d'étude théorique de cet algorithme. Celui-ci a par ailleurs déjà fait l'objet de nombreux travaux. Une poursuite intéressante pourrait donc être d'étudier de façon plus approfondie, à l'aide de ces travaux, les propriétés d'un tel algorithme dans notre cadre précis.

De plus, à la lumière de ce chapitre, différentes autres modélisations de l'hétérogénéité mériteraient aussi d'être étudiées.

De façon générale, à la question de l'estimation des paramètres d'un modèle succèdent bien entendu celles sur la possibilité de construire des tests sur ces paramètres, ou encore d'établir des méthodes de choix de modèle. Ce sujet est d'autant plus délicat que les méthodes d'estimation sont, elles mêmes, approchées et que ces approximations sont plus ou moins bien maitrisées. Cependant, certains auteurs ont commencé à s'y intéresser (cf. Lin (1997)). Enfin, l'article fondamental de Lee et Nelder (1996) ouvre aussi des perspectives à ce sujet dans le cadre plus large des HGLM.

Références

AITKIN, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Applied Statistics 36*, 332–339.

AITKIN, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing 6*(3), 251–262.

ALBERT, J. H. et S. CHIB (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association 88*(422), 669–679.

ANDERSON, D. A. et M. AITKIN (1985). Variance Components Models with Binary Response : Interviewer Variability. *Journal of the Royal Statistical Society, B 47*(2), 203–210.

ANDERSON, D. A. et J. P. HINDE (1988). Random effects in generalized linear models and the EM algorithm. *Communications in Statistics - Theory and Methods 17*(11), 3847–3856.

BOCK, D. et M. AITKIN (1981). Marginal maximum likelihood estimation of item parameters : Application of an EM algorithm. *Psychometrika 46*, 443–459.

BRESLOW, N. E. et D. G. CLAYTON (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association 88*(421), 9–25.

COOK, R. D. et S. WEISBERG (1983). Diagnostics for heteroscedasticity in regression. *Biometrika 70*, 1–10.

CRESSIE, N. et S. N. LAHIRI (1993). The Asymptotic Distribution of REML Estimators. *Journal of multivariate analysis 45*(2), 217–233.

DAVIDIAN, M. et R. J. CARROLL (1988). A note on extended quasi-likelihood. *Journal of the Royal Statistical Society, B 50*(1), 74–82.

DEMPSTER, A. P., N. M. LAIRD, et D. B. RUBIN (1977). Maximum likelihood for incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B 39*, 1–38.

DE STEFANO, A. L. (1983). *Identifying and quantifying sources heterogeneous residual and sire variances in dairy production data.* Ph. D. thesis, Cornell University.

DIETZ, E. (1992). Estimation of heterogeneity - a GLM approach. In L. Fahrmeir and al. (Eds.), *Advances in GLIM and statistical modelling. Proceedings of the GLIM' 92 conference and the 7th international workshop on statistical modelling - Munich*, Number 78 in Lecture Notes in Statistics, pp. 66–71. Springer-Verlag.

DIETZ, E. et D. BÖHNING (1995). Statistical inference based on a general model of unobserved heterogeneity. In G. Seeber, B. Francis, R. Hatzinger, and G. Steckel-Berger (Eds.), *Statistical modelling. Proceedings of the 10th international workshop on statistical modelling - Innsbruck*, Number 104 in Lecture Notes in Statistics, pp. 75–82. Springer-Verlag.

DUCROCQ, V. (1990). Estimation of genetic parameters arising in non linear models. In *4th World Congr. Genet. Appl. Livestock Prod.*, Volume 13, pp. 419–428. W.G. Hill and R. Thompson and J.A. Wooliams, Edinburgh.

ENGEL, B. et A. KEEN (1994). A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica 48*(1), 1–22.

FAHRMEIR, L. et H. KAUFMANN (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics 13*(1), 342–368.

FAHRMEIR, L. et G. TUTZ (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Series in statistics. Springer-Verlag.

FOULLEY, J. L., D. GIANOLA, et S. IM (1987). Genetic evaluation of traits distributed as poisson-binomial with reference to reproductive characters. *Theoretical Applied Genetics 73*, 870–877.

FOULLEY, J. L., D. GIANOLA, et S. IM (1990). Genetic Evaluation for Discrete Polygenic Traits in Animal Breeding. In D. Gianola and K. Hammond (Eds.), *Advances in Statistical Methods for Genetic Improvement of Livestock*, pp. 361–409. Springer-Verlag.

FOULLEY, J. L. et S. IM (1993). A marginal quasi-likelihood approach to the analysis of poisson variables with generalized linear mixed models. *Genetics, Selection, Evolution 25*(1), 101–107.

FOULLEY, J. L. et E. MANFREDI (1991). Approches statistiques de l'évaluation génétique des reproducteurs pour des caractères binaires à seuils. *Genetics, Selection, Evolution 23*, 309–338.

FOULLEY, J. L. et R. L. QUAAS (1995). Heterogeneous variances in gaussian linear mixed models. *Genetics, Selection and Evolution 27*(3), 211–228.

Foulley, J. L., M. San Cristobal, D. Gianola, et S. Im (1992). Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Computational Statistics and Data Analysis 13*(3), 291–305.

Garrick, D. J., E. J. Pollak, R. L. Quaas, et L. Van Vleck (1989). Variance heterogeneity in direct and maternal weight traits by sex and percent purebred Simmental sired calves. *Journal of Animal Science 67*(10), 2515–2528.

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

Gianola, D. (1980). Genetic evaluation of animals for traits with categorical responses. *Journal of Animal Science 51*, 1272–1276.

Gianola, D. (1986). On selection criteria and estimation of parameters when the variance is heterogeneous. *Theory of Applied Genetics 72*, 671–677.

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

Gianola, D., J. L. Foulley, R. L. Fernando, C. Henderson, et K. Weigel (1992). Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. *Journal of Dairy Science 75*(10), 2805–2923.

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

Gilmour, A. R., R. D. Anderson, et A. L. Rae (1987). Variance components on an underlying scale for ordered multiple threshold categorical data using a generalized linear mixed model. *Journal of Animal Breeding and Genetics 104*, 149–155.

Godambe, V. P. et M. E. Thompson (1989). An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference 22*(2), 137–152.

Green, P. J. (1987). Penalized Likelihood for General Semi-Parametric Regression Models. *International Statistical Review 55*(422), 245–259.

Harville, D. A. (1977). Maximum-likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association 72*, 320–340.

Harville, D. A. et R. W. Mee (1984). A Mixed-Models Procedure for Analyzing Oredered Categorical Data. *Biometrics 40*, 393–408.

Henderson, C. R., O. Kempthorne, S. R. Searle, et C. VonKrosig (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*, 192–218.

Hill, W. G. (1984). On selection among groups with heterogeneous variance. *Animal Production 39*, 473–477.

Hillis, S. L. et C. S. Davis (1994). A Simple Justification of the Iterative Fitting Procedure for Generalized Linear Models. *The American Statistician 48*(4), 288–289.

Hinde, J. (1982). Compound poisson regression model. In R. Gilchrist (Ed.), *GLIM 82: Proceedings of the International Conference on Generalized Linear Models*, Number 14 in Lecture Notes in Statistics, pp. 109–121. Springer-Verlag.

Im, S. et D. Gianola (1988). Mixed models for binomial data with an application to lamb mortality. *Applied Statistics 37*(2), 196–204.

Jacqmin-Gadda, H. et D. Commenges (1995). Tests of homogeneity for generalized linear models. *Journal of the American Statistical Association 90*(432), 1237–1246.

Jansen, J. et J. A. Hoekstra (1993). The analysis of proportions in agricultural experiments by a generalized linear mixed model. *Statistica Neerlandica 47*(3), 161–174.

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

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika 84*(2), 309–326.

McCullagh, P. et J. Nelder (1989). *Generalized Linear Models.* (seconde ed.). Chapman and Hall.

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

McLachlan, G. J. et T. Krishnan (1997). *The EM Algorithm and Extensions.* John Wiley & Sons.

Nelder, J. A. et Y. Lee (1992). Likelihood, Quasi-likelihood and Pseudolikelihood: Some Comparisons. *Journal of the Royal Statistical Society, B 54*(1), 273–284.

Nelder, J. A. et D. Pregibon (1987). An extended quasi-likelihood function. *Biometrika 74*, 221–231.

Nelder, J. A. et R. W. M. Wedderburn (1972). Generalized Linear Models. *Journal of the Royal Statistical Society, A 135*, 370–384.

Rao, C. R. et J. Kleffe (1988). *Estimation of variance components and applications.* Number 3 in Series in Statistics and Probability. North Holland.

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

SEARLE, S. R., G. CASELLA, et C. E. MC CULLOCH (1992). *Variance components.* John Wiley & Sons.

SHAEFFER, L. R. et J. W. WILTON (1976). Methods of sire evaluation for calving ease. *Journal of Dairy Science 59*, 544–551.

SOLOMON, P. J. et D. R. COX (1992). Non linear component of variance models. *Biometrika 79*(1), 1–11.

STEELE, B. M. (1996). A modified EM algorithm for estimation in generalized mixed models. *Biometrics 52*(4), 1295–1310.

STIRATELLI, R., N. LAIRD, et J. H. WARE (1984). Random-Effects Models for Serial Observations with Binary Response. *Biometrics 40*, 961–971.

STROUD, A. H. et D. SECREST (1966). *Gaussian quadrature formulas.* Prentice Hall.

SWEETING, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Annals of statistics 8*, 1375–1381.

THOMPSON, R. et K. MEYER (1986). Estimation of variance components : what is missing in the EM algorithm ? *Journal of Statistical Computation and Simulation 24*, 215–230.

WALKER, S. (1996). An EM algorithm for non linear random effects models. *Biometrics 52*(3), 934–944.

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

WILLIAMS, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics 31*(2), 144–148.

WU, C. J. J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics 11*, 95–103.

ZEGER, S. L., K. Y. LIANG, et P. S. ALBERT (1988). Models for longitudinal data : a generalized estimating equation approach. *Biometrics 44*, 1049–1060.