BABACAR SECK

Estimation pour les modèles linéaires généralisés : Approche marginale, approche conditionnelle et application

Essai présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en statistique
pour l'obtention du grade de Maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

Septembre 2006

©Babacar Seck, 2006

Avant-propos

Tout d'abord c'est à mon directeur de recherche, Thierry Duchesne, professeur au Département de mathématiques et de statistique de l'Université Laval, que je tiens à adresser un grand merci. Ses conseils avisés, son dynamisme et sa disponibilité m'ont permis de mener à bien ce travail.

Ensuite, je voudrais, à travers la personne de M. Duchesne, rendre un hommage déférent ainsi qu'un témoignage de gratitude à tous les professeurs et professionnels du Département de mathématiques et de statistique de l'Université Laval, pour leur appui et soutien permanent et surtout leur effort d'octroyer des bourses de recherche qui ont fait de mon séjour dans cette université une expérience des plus agréables et des plus enrichissantes.

Je tiens aussi à remercier Daniel Fortin, professeur au Département de biologie de l'Université Laval, pour m'avoir fourni les données sur les bisons pour l'exemple d'application du chapitre 4.

Toute ma profonde gratitude à mes amis du Canada, du Sénégal, de la Suisse et partout dans le monde, pour leur support continu.

Enfin, c'est un merci du fond du coeur que je lance à mes frères et soeurs plus particulièrement à Adama Seck qui, en plus d'être un frère, est l'un de mes meilleurs amis. Je réserve le plus grand des mercis à mon défunt père, à ma mère et à ma femme Ramatoulaye Seye. Ils ont été, sont et resteront les piliers qui proposent un appui sans limite.

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Introduction

Lorsqu'il s'agit de modéliser des phénomènes naturels, les modèles linéaires généralement utilisés ne permettent que des variables endogènes continues de distribution normale. Malgré la prédominance bien justifiée de cette fameuse loi, il est néanmoins possible d'imaginer une foule de situations où supposer une distribution normale ne sera pas approprié. 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, les modèles linéaires classiques ont été étendus à la classe plus large de modèles que sont les modèles linéaires généralisés (GLM), permettant de modéliser la distribution d'une variable endogène Y en fonction de variables exogènes $\boldsymbol{x}$, en autant que cette distribution fasse partie de la famille exponentielle. En général, l'estimation des paramètres du modèle se fait sur la base d'observations indépendantes.

De plus, il existe un grand nombre de situations impliquant des données groupées pour lesquelles les réponses sont corrélées :

• Par exemple, supposons un ensemble de données constitué de patients enregistrés dans plusieurs hôpitaux au sein d'un état ou d'une province. Supposons aussi que les données qui nous intéressent se rapportent à un certain type de procédure médicale. Il est probable que chaque hôpital ait son propre protocole de traitement de sorte qu'il existe une corrélation des effets de traitement au sein des hôpitaux qui n'existe pas entre les hôpitaux.

• En sciences appliquées, il est fréquent de prendre des mesures sur un individu à plusieurs moments dans le temps, ce qui fait qu'il est cependant raisonnable de prendre

en considération la corrélation des mesures pour un individu donné dans les analyses statistiques. Par exemple, dans les études sur les traitements contre l'épilepsie, pour chaque individu i, on mesure $Y_{i,1}$,....,Y_{i,n_i}, le nombre de crises dans plusieurs périodes de temps consécutives. Bien que l'hypothèse d'indépendance soit raisonnable entre $Y_{i,j}$ et $Y_{i'j'}$ si $i \neq i'$ (individus différents), le nombre de crises dans deux périodes différentes pour un même individu, $Y_{i,j}$ et $Y_{i,j'}$, sont fort probablement corrélés. De façon similaire, si $Y_{i,j}$ est une variable binaire dénotant la présence ou l'absence d'une tumeur chez le j^{eme} rat de la i^{eme} portée (famille) après injection d'un cancérogène, il est probable que les rats d'une même famille réagiront de façon similaire au cancérogène et, donc, que les variables $Y_{i,j}$ et $Y_{i,j'}$ soient corrélées.

Les GLM s'appliquent particulièrement aux exemples de données cités précédemment. On les appelle des données longitudinales ; elles sont également appelées des données en panels.

Sur un autre plan, la modélisation de la corrélation pouvant intervenir dans l'explication du phénomène étudié s'avère parfois nécessaire. Ainsi, des effets aléatoires ont été introduits dans le modèle linéaire généralisé pour donner naissance au modèle linéaire généralisé mixte permettant la modélisation de cette corrélation (GLMM).

Dans ce travail, nous nous intéressons essentiellement à l'estimation des paramètres des modèles linéaires généralisés pour données corrélées.

Dans un deuxième chapitre, nous adopterons l'approche marginale basée sur les équations d'estimation généralisées (GEE) pour l'estimation des paramètres du modèle de régression. Ici, on ne cherche pas à modéliser la corrélation mais plutôt à corriger nos estimateurs et leurs estimateurs de variances et covariances pour tenir compte de la corrélation dans nos inférences sur les coefficients de régression. Hardin et Hilbe (2002) ont fait des GEE le sujet principal d'un livre. La procédure GENMOD de SAS utilise l'approche GEE pour l'estimation des paramètres du modèle linéaire généralisé. Cependant, des inférences sur les coefficients de régression peuvent être faites, mais on ne peut pas effectuer de prévisions pour un individu donné, car ceci nécessite la modélisation des effets aléatoires qui causent la corrélation.

Ainsi, dans le but de modéliser la corrélation entre les mesures, nous abordons dans le troisième chapitre l'approche conditionnelle basée sur les modèles linéaires généralisés mixtes. Dans ce chapitre, nous allons, d'une part, présenter succinctement la théorie des modèles linéaires généralisés mixtes, et d'autre part, développer des méthodes qui serviront à l'estimation des paramètres d'un tel modèle. Les procédures GLIMMIX et NLMIXED de SAS permettent d'ajuster ces genres de modèles.

Dans un quatrième chapitre, nous exposons en détail une application concrète de ces modèles à des données sur les bisons. Dans ce chapitre, nous présentons le milieu et la manière dont s'est effectuée la collecte de données. On expliquera également les différentes variables. Les méthodes développées aux chapitres précédents seront utilisées pour faire l'analyse.

Finalement, nous concluons cet essai avec une courte discussion des éléments importants abordés et en donnant quelques idées de recherche future au chapitre 5.

Chapitre 2

Approche marginale basée sur les équations d'estimation généralisées (GEE)

2.1 Introduction

Dans l'introduction, nous avons expliqué par des exemples concrets qu'il arrive fréquemment que, sachant les valeurs des variables exogènes $\boldsymbol{x}_1$,...,$\boldsymbol{x}_n$, les variables endogènes $\boldsymbol{Y}_1$,...,$\boldsymbol{Y}_n$ soient corrélées. Ainsi, dans ce chapitre, nous verrons comment il est possible de tenir compte de ce type de corrélation dans nos inférences sur les coefficients de régression $\boldsymbol{\beta}$ d'un modèle linéaire généralisé. L'approche que nous adapterons utilisera le concept d'équations d'estimation généralisées (GEE). Cette approche ne spécifie pas entièrement la distribution conjointe des $\boldsymbol{Y}_i$, mais plutôt une modélisation de la moyenne et une spécification de la structure de corrélation de *travail*. Dans le contexte longitudinal, différentes formes de la structure de corrélation de *travail* sont utilisées et les estimateurs sont solution des GEE. Un élément attrayant de cette approche est que les estimateurs des paramètres du modèle sont convergents même dans l'éventualité où la structure de corrélation de *travail* serait mal spécifiée. Nous terminons ce chapitre par une illustration basée sur des modèles de régression de Poisson et logistique.

2.2 Quelques hypothèses et définitions

Supposons que pour chaque individu (ou groupe) i, nous avons un vecteur $\boldsymbol{Y}_i = (Y_{i1},....,Y_{in_i})'$ de plusieurs observations de la variable endogène ainsi que la matrice de dimension $n_i \times p$ $\boldsymbol{X}_i = (\boldsymbol{x}_{i1},....,\boldsymbol{x}_{in_i})'$, où $\boldsymbol{x_{ij}} = (x_{ij1},....,x_{ijp})'$ représente la valeur des variables exogènes pour l'observation j de l'individu (du groupe) i. On suppose que la distribution marginale de chaque variable endogène Y_{ij} étant donné $\boldsymbol{x}_{ij}$ est un membre de la famille exponentielle. Plus précisément, on suppose que la fonction de densité de Y_{ij} étant donné $\boldsymbol{x}_{ij}$ s'écrit

$$f(y_{ij}|\boldsymbol{x}_{ij}) = \exp\left\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij},\phi)\right\}, \quad i = 1,...,n; \qquad j = 1,...,n_i, \tag{2.1}$$

où $\mathrm{E}(Y_{ij}|\boldsymbol{x}_{ij}) = \mu_{ij} = g^{-1}(\eta_{ij}) = g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta})$, pour une fonction de lien $g(\cdot)$ connue. Le paramètre θ_{ij} est un paramètre canonique et ϕ un paramètre de dispersion. $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ est le prédicteur linéaire. Le lien entre la i^{eme} composante de ce prédicteur linéaire et l'espérance de $\boldsymbol{Y}_i$ s'établit par l'intermédiare de la fonction de lien $g(\cdot)$. Parmi toutes les fonctions de lien, celle qui permet d'égaler le prédicteur linéaire et le paramètre canonique est appelée fonction de lien canonique. Les fonctions $b(\cdot)$ et $c(\cdot)$ sont spécifiques à chaque distribution et la fonction $a(\phi)$ est généralement de la forme $\frac{\phi}{\omega_{ij}}$, où la valeur de ϕ reste constante pour toutes les observations tandis que ω_{ij} est une valeur connue qui peut varier d'observation en observation.

Cette famille de lois regroupe un certain nombre de lois dont les lois classiques : binomiale, Poisson, normale, gamma, etc. Dans le tableau TAB.2.1 ci-après, extrait de la thèse de Trottier (1998) (on peut également le retrouver dans l'aide de SAS sur GLIMMIX), on décrit pour chacune de ces lois, l'expression du paramètre canonique θ en fonction des paramètres naturels de la loi, le paramètre ϕ et les fonctions $a(\cdot)$ et $b(\cdot)$ associées. Pour simplifier la lecture du tableau, on a omis l'indice ij.

TAB. 2.1 – *Paramètre canonique et fonctions $a(\cdot)$ et $b(\cdot)$ caractérisant les lois usuelles de la famille exponentielle.*

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

Pour toute distribution de la famille exponentielle donnée en (2.1), l'espérance et la

variance de la variable associée s'expriment à l'aide des fontions $a(\cdot)$ et $b(\cdot)$. Ainsi, on a

$$E(Y_{ij}|\boldsymbol{x}_{ij}) = \mu_{ij} = b'(\theta_{ij}) \tag{2.2}$$
$$Var(Y_{ij}|\boldsymbol{x}_{ij}) = a(\phi)b''(\theta_{ij}). \tag{2.3}$$

Les égalités (2.2) et (2.3) sont démontrées dans l'annexe B.
Il est donc important de souligner qu'il existe une relation directe entre l'espérance de $Y_{ij}|\boldsymbol{x}_{ij}$ et sa variance :

$$Var(Y_{ij}|\boldsymbol{x}_{ij}) = a(\phi)b''(b'^{-1}(\mu_{ij})).$$

On désignera par la suite par $V = b''o\ b'^{-1}$ cette fonction de variance. D'où

$$Var(Y_{ij}|\boldsymbol{x}_{ij}) = a(\phi)V(\mu_{ij}).$$

Nous donnons, dans le tableau TAB.2.2, extrait de la thèse de Trottier (1998), l'expression de l'espérance en fonction du (des) paramètre(s) naturel(s), du paramètre canonique ainsi que de la fonction de variance des lois usuelles de la famille exponentielle.

TAB. 2.2 – *Expression de l'espérance et de la variance des lois usuelles de la famille exponentielle.*

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

Étant donné les variables exogènes, on suppose que les variables endogènes d'un vecteur $\boldsymbol{Y}_i$ sont indépendantes de celles d'un vecteur $\boldsymbol{Y}_{i'}$, pour $i \neq i'$.

2.3 Équations d'estimation sous l'indépendance

2.3.1 Estimation des paramètres de régression

Nous allons tout d'abord commencer par estimer le vecteur des coefficients de régression $\boldsymbol{\beta}$ en supposant que les variables aléatoires à l'intérieur d'un vecteur $\boldsymbol{Y}_i$ sont indépendantes. Cette hypothèse est généralement fausse, mais elle nous permet

d'amorcer la procédure d'estimation. Dans ce cas, la matrice de variance de $\boldsymbol{Y}_i$ est donnée par

$$\boldsymbol{V}_i = \mathrm{Var}(\boldsymbol{Y}_i) = \phi \boldsymbol{A}_i{}^{\frac{1}{2}} \boldsymbol{R}_i \boldsymbol{A}_i{}^{\frac{1}{2}},$$

avec $\boldsymbol{A}_i = \mathrm{Diag}\big\{Var(Y_{ij}|\boldsymbol{x}_{ij}),\ j = 1, ..., n_i\big\}$, et où $\boldsymbol{R}_i = \frac{(\boldsymbol{A}_i{}^{\frac{1}{2}})^{-1}\boldsymbol{V}_i(\boldsymbol{A}_i{}^{\frac{1}{2}})^{-1}}{\phi} = \mathbf{I}_{n_i \times n_i}$ est la matrice des corrélations des éléments du vecteur $\boldsymbol{Y}_i$, qui est dans ce cas égale à la matrice identité de dimension $n_i \times n_i$. Nous appellerons cette matrice la structure de corrélation de "travail" pour $\boldsymbol{Y}_i$.

Nous obtenons l'estimateur de $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, en maximisant la fonction de vraisemblance du paramètre $\boldsymbol{\beta}$ sous l'indépendance. Puisque

$$f(y_{ij}|\boldsymbol{x}_{ij}) = \exp\left\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right\} i = 1, ..., n; \qquad j = 1, ..., n_i,$$

et que l'on suppose que les $\boldsymbol{Y}_i$ sont indépendantes, la fonction de vraisemblance est donnée par

$$L(\boldsymbol{\beta}, \phi; y_{ij}) = \exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + \sum_{i=1}^{n}\sum_{j=1}^{n_i} c(y_{ij}, \phi)\right\}. \tag{2.4}$$

Quant à elle, la fonction de log-vraisemblance est donnée par

$$\ell(\boldsymbol{\beta}, \phi; y_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + \sum_{i=1}^{n}\sum_{j=1}^{n_i} c(y_{ij}, \phi). \tag{2.5}$$

Comme dans le cas de la famille exponentielle la fonction de vraisemblance est "régulière" (propriétés de la famille exponentielle), on trouve la valeur de $\boldsymbol{\beta}$ qui la maximise en résolvant le sytème d'équations

$$\begin{aligned}
\left\{\frac{\partial \ell(\boldsymbol{\beta}, \phi; y_{ij})}{\partial \beta_k}\right\}_{k=1,...,p} &= \left(\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{a(\phi)}(y_{ij}\frac{\partial \theta_{ij}}{\partial \beta_k} - b'(\theta_{ij})\frac{\partial \theta_{ij}}{\partial \beta_k})\right\}_{k=1,...,p}\right)_{p\times 1} = [0]_{p\times 1} \\
&= \left(\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{a(\phi)}(y_{ij}\frac{\partial \theta_{ij}}{\partial \eta_{ij}}\frac{\partial \eta_{ij}}{\partial \beta_k} - b'(\theta_{ij})\frac{\partial \theta_{ij}}{\partial \eta_{ij}}\frac{\partial \eta_{ij}}{\partial \beta_k})\right\}_{k=1,...,p}\right)_{p\times 1} \\
&= [0]_{p\times 1} \\
&= \left(\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{a(\phi)}\frac{\partial \theta_{ij}}{\partial \eta_{ij}}x_{ijk}(y_{ij} - \mu_{ij})\right\}_{k=1,...,p}\right)_{p\times 1} = [0]_{p\times 1},
\end{aligned}$$

où p est le nombre de colonnes de la matrice $\boldsymbol{X}$. En ré-écrivant sous forme matricielle, on a

$$U_{indep}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\frac{1}{a(\phi)}\boldsymbol{X}_i'\boldsymbol{\Delta}_i(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = \mathbf{0}, \tag{2.6}$$

où $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i})'$ avec $\boldsymbol{x}_{ij} = (x_{ij1}, ..., x_{ijp})'$ et $\boldsymbol{\mu_i}(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), ..., \mu_{in_i}(\boldsymbol{\beta}))'$. $\boldsymbol{\Delta_i}$ est une matrice diagonale de dimension $n_i \times n_i$ dont l'élément en position (j, j) est $\frac{\partial \theta_{ij}}{\partial \eta_{ij}}$, où θ_{ij} est le paramètre canonique de la famille exponentielle. (Notez que si le lien $g(\cdot)$ est le lien canonique, alors $\theta_{ij} = \mu_{ij}$ pour la majorité des distributions de la famille exponentielle, (McCullagh et Nelder, 1989)). Illustrons les équations d'estimation sous l'indépendance (2.6) avec des modèles de Poisson et logistique :

- **Modèle de Poisson**

Considérons le cas où les variables $Y_{ij}|\boldsymbol{x}_{ij} \sim$ Poisson(μ_{ij}), $Y_{ij}|\boldsymbol{x}_{ij}$ indépendantes, $i = 1, ..., n$; $j = 1, ..., n_i$ avec

$$P(Y_{ij} = y_{ij}|\boldsymbol{x}_{ij}) = e^{-\mu_{ij}} \frac{\mu_{ij}^{y_{ij}}}{y_{ij}!}, \qquad y_{ij} = 0, 1, 2, ... \tag{2.7}$$

Dans le cas du lien log, on a

$$\log \mu_{ij} = \boldsymbol{\beta}'\boldsymbol{x}_{ij} = \eta_{ij} \Longleftrightarrow \mu_{ij} = \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij}) = \exp(\eta_{ij}).$$

On a donc

$$P(Y_{ij} = y_{ij}|\boldsymbol{x}_{ij}) = \frac{e^{-\exp(\eta_{ij})}[\exp(\eta_{ij})]^{y_{ij}}}{y_{ij}!}. \tag{2.8}$$

Puisque les observations dans un panel sont indépendantes, nous écrivons la probabilité d'un vecteur de résultats pour le panel i comme

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{x}_i) = \prod_{j=1}^{n_i} \frac{e^{-\exp(\eta_{ij})}[\exp(\eta_{ij})]^{y_{ij}}}{y_{ij}!}. \tag{2.9}$$

La vraisemblance est ainsi donnée par le produit des probabilités d'un vecteur de résultats pour le panel i :

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{x}_i) && (2.10)\\
&= \prod_{i=1}^{n}\prod_{j=1}^{n_i} \frac{e^{-\exp(\eta_{ij})}[\exp(\eta_{ij})]^{y_{ij}}}{y_{ij}!} && (2.11)\\
&= \prod_{i=1}^{n}\prod_{j=1}^{n_i} \frac{e^{-\exp(\eta_{ij})} \times e^{(y_{ij}\eta_{ij})}}{e^{\log y_{ij}!}} && (2.12)\\
&= \prod_{i=1}^{n}\prod_{j=1}^{n_i} e^{-\exp(\eta_{ij}) + y_{ij}\eta_{ij} - \log y_{ij}!} && (2.13)\\
&= \exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}(y_{ij}\eta_{ij} - \exp(\eta_{ij}) - \log y_{ij}!)\right\}. && (2.14)
\end{aligned}$$

La log-vraisemblance est

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}(y_{ij}\eta_{ij} - \exp(\eta_{ij}) - \log y_{ij}!). \tag{2.15}$$

Pour $t = 1, ..., p$, p étant le nombre de colonnes de la matrice $\mathbf{X}$, on a :

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_t} = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \eta}\frac{\partial \eta}{\partial \beta_t} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}(y_{ij} - \exp(\eta_{ij}))x_{ijt} \tag{2.16}$$

ou, sous forme matricielle,

$$U_{indep}(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_t} = \sum_{i=1}^{n} \boldsymbol{X}_i'\boldsymbol{\Delta}_i(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})),$$

où $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i})'$ avec $\boldsymbol{x_{ij}} = (x_{ij1}, ..., x_{ijp})'$, $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), ..., \mu_{in_i}(\boldsymbol{\beta}))'$ et où $\boldsymbol{\Delta_i}$ est une matrice diagonale de dimension $n_i \times n_i$ dont l'élément en position (j, j) est $\exp(\eta_{ij}) = \frac{\partial \mu_{ij}}{\partial \eta_{ij}}$.

• **Modèle logistique**

Considérons maintenant le cas où les variables $Y_{ij}|\boldsymbol{x}_{ij} \sim$ Bernoulli(μ_{ij}), où $Y_{ij}|\boldsymbol{x}_{ij}$ indépendantes, $i = 1, ..., n$; $j = 1, ..., n_i$ avec

$$P(Y_{ij} = y_{ij}|\boldsymbol{x}_{ij}) = \mu_{ij}^{y_{ij}}(1 - \mu_{ij})^{1-y_{ij}}, \qquad y_{ij} = 0, 1. \tag{2.17}$$

La fonction de lien canonique est le lien logit, c'est-à-dire

$$\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \boldsymbol{\beta}'\boldsymbol{x}_{ij} = \eta_{ij} \Longleftrightarrow \mu_{ij} = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})}{1 + \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij})} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}.$$

Et donc,

$$\begin{aligned}
P(Y_{ij} = y_{ij}|\boldsymbol{x}_{ij}) &= \left(\frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}\right)^{y_{ij}}\left(\frac{1}{1 + \exp(\eta_{ij})}\right)^{1-y_{ij}} && (2.18)\\
&= \exp\left\{y_{ij}\log\frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}\right\}\\
&\times \exp\left\{(1 - y_{ij})\log\frac{1}{1 + \exp(\eta_{ij})}\right\} && (2.19)\\
&= \exp\left(y_{ij}\log\frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} + (1 - y_{ij})\log\frac{1}{1 + \exp(\eta_{ij})}\right) && (2.20)\\
&= \exp\left[y_{ij}\eta_{ij} - \log\left\{1 + \exp(\eta_{ij})\right\}\right]. && (2.21)
\end{aligned}$$

Puis que les observations dans un panel sont indépendantes, nous écrivons la probabilité d'un vecteur de résultats pour le panel i comme

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{x}_i) = \prod_{j=1}^{n_i} \exp\left[y_{ij}\eta_{ij} - \log\left\{1+\exp(\eta_{ij})\right\}\right]. \tag{2.22}$$

La vraisemblance est ainsi donnée par le produit des probabilités d'un vecteur de résultats pour le panel i :

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{x}_i) && (2.23)\\ &= \prod_{i=1}^{n}\prod_{j=1}^{n_i} \exp\left[y_{ij}\eta_{ij} - \log\left\{1+\exp(\eta_{ij})\right\}\right] && (2.24)\\ &= \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[y_{ij}\eta_{ij} - \log\left\{1+\exp(\eta_{ij})\right\}\right]\right). && (2.25) \end{aligned}$$

La log-vraisemblance est

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[y_{ij}\eta_{ij} - \log\left\{1+\exp(\eta_{ij})\right\}\right]. \tag{2.26}$$

Pour $t = 1, ..., p$, on a

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_t} = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \eta}\frac{\partial \eta}{\partial \beta_t} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(y_{ij} - \frac{\exp\eta_{ij}}{1+\exp\eta_{ij}}\right)x_{ijt} \tag{2.27}$$

ou, sous forme matricielle,

$$U_{indep}(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_t} = \sum_{i=1}^{n} \boldsymbol{X}_i'\boldsymbol{\Delta}_i\left\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\right\},$$

où $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i})'$ avec $\boldsymbol{x}_{ij} = (x_{ij1}, ..., x_{ijp})'$, $\boldsymbol{\mu_i}(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), ..., \mu_{in_i}(\boldsymbol{\beta}))'$ et où $\boldsymbol{\Delta}_i$ est une matrice diagonale de dimension $n_i \times n_i$ dont l'élément en position (j,j) est $\frac{\exp(\eta_{ij})}{\left\{1+\exp(\eta_{ij})\right\}^2} = \frac{\partial \mu_{ij}}{\partial \eta_{ij}}$.

Si on définit la matrice $\boldsymbol{A}_i$ comme étant la matrice diagonale dont l'élément en position (j,j) est $b''(\theta_{ij})$, alors on a que $\hat{\boldsymbol{\beta}}$ qui résoud (2.6) sera, sous l'hypothèse d'indépendance, approximativement de distribution normale multivariée de moyenne $\boldsymbol{\beta}$ et de variance estimée par

$$\hat{\mathbf{V}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i'\boldsymbol{\Delta}_i\boldsymbol{A}_i\boldsymbol{\Delta}_i\boldsymbol{X}_i\right)^{-1}\Bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}. \tag{2.28}$$

Mathématiquement,

$$\hat{\boldsymbol{\beta}} \approx \mathcal{N}(\boldsymbol{\beta}, \hat{\mathbf{V}}). \tag{2.29}$$

2.3.2 Intervalles de confiance et tests d'hypothèse pour β_j

On utilise le fait que $\hat{\boldsymbol{\beta}} \approx \mathcal{N}(\boldsymbol{\beta}, \hat{\mathbf{V}})$. Ainsi pour un paramètre individuel β_j, soit V_{jj}, l'élément de $\hat{\mathbf{V}}$ correspondant à la variance de $\hat{\beta}_j$. Alors de l'équation (2.29), on a que

$$P\left[-z_{\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{\sqrt{V_{jj}}} \leq z_{\alpha/2} \right] \approx 1 - \alpha.$$

Ceci suggère donc l'intervalle de confiance de niveau $(1-\alpha)100\%$ suivant pour β_j :

$$\hat{\beta}_j \pm z_{\alpha/2}\sqrt{V_{jj}}.$$

Pour tester une hypothèse de la forme $H_0 : \beta_j = \beta_{j0}$, la procédure est simple. On calcule tout d'abord, sous H_0, la statistique du test $Z_0 = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{V_{jj}}}$. Si la contre hypothèse est $H_1 : \beta_j \neq \beta_{j0}$, le seuil du test est $2\mathrm{P}[\mathcal{N}(0,1) \geq |Z_0|]$, si la contre hypothèse est $H_1 : \beta_j > \beta_{j0}$, le seuil du test est $\mathrm{P}[\mathcal{N}(0,1) \geq Z_0]$ et si la contre hypothèse est $H_1 : \beta_j < \beta_{j0}$, le seuil du test est $\mathrm{P}[\mathcal{N}(0,1) \leq -Z_0]$.

2.4 Équations d'estimation généralisées (GEE)

Les équations d'estimation généralisées sont en fait une généralisation des équations d'estimation (2.6) où l'on peut supposer une structure de corrélation de *travail* autre que l'indépendance pour $\boldsymbol{R}_i$. Pour compenser le fait que la structure de corrélation de *travail* puisse ne pas être la vraie structure de corrélation, la variance de l'estimateur $\hat{\boldsymbol{\beta}}$ sera estimée par un estimateur de variance robuste.

Soit $\boldsymbol{V}_i = \phi(\boldsymbol{A}_i)^{\frac{1}{2}}\boldsymbol{R}_i(\boldsymbol{\alpha})(\boldsymbol{A}_i)^{\frac{1}{2}}$, la matrice de variance de *travail* pour $\boldsymbol{Y}_i$, où $\boldsymbol{R}_i(\boldsymbol{\alpha})$ est une structure de corrélation de *travail* pour $\boldsymbol{Y}_i$. Ici on suppose que cette matrice contient certains paramètres inconnus que l'on représente par le vecteur $\boldsymbol{\alpha}$. L'idée est d'essayer de "deviner" la vraie structure de corrélation de $\boldsymbol{Y}_i$. Si on spécifie une mauvaise structure, les inférences sur $\boldsymbol{\beta}$ seront quand même valides, mais si on spécifie la structure correctement, on aura des inférences plus efficaces (variance des estimateurs plus faible). Les équations d'estimation sont ensuite données par

$$\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} = 0, \tag{2.30}$$

où $\boldsymbol{D}_i = \boldsymbol{A}_i \boldsymbol{\Delta}_i \boldsymbol{X}_i$. Si on pose $\boldsymbol{R}_i(\boldsymbol{\alpha}) = \boldsymbol{I}_{n_i \times n_i}$, alors (2.30) se simplifie à (2.6). Nous donnons plus bas un algorithme permettant de trouver la valeur de l'estimateur $\hat{\boldsymbol{\beta}}$ qui résoud (2.30).

Afin d'estimer les paramètres de la matrice de variance et de vérifier l'ajustement du modèle, on peut définir les résidus $e_{ij} = \frac{Y_{ij}-\mu_{ij}}{\sqrt{b"(\theta_{ij})}}$, que l'on évalue à $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$. Pour estimer le paramètre de dispersion ϕ, on pose

$$\hat{\phi} = \frac{1}{N-p}\sum_{i=1}^{n}\sum_{j=1}^{n_i} e_{ij}^2,$$

où $N = \sum_{i=1}^{n} n_i$, est le nombre total d'observations dans l'échantillon et $p = \dim(\boldsymbol{\beta})$.

On estime ensuite $\boldsymbol{\beta}$ en utilisant l'algorithme suivant :
1. Estimer $\boldsymbol{\beta}$ sous l'hypothèse d'indépendance et dénoter l'estimateur obtenu $\hat{\boldsymbol{\beta}}_0$.
2. Estimer $\boldsymbol{\alpha}$ et ϕ à partir de $\hat{\boldsymbol{\beta}}$ et des e_{ij}.
3. Poser $\boldsymbol{V}_i = \hat{\phi}(\boldsymbol{A}_i)^{\frac{1}{2}}\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})(\boldsymbol{A}_i)^{\frac{1}{2}}$.
4. Mettre la valeur de $\hat{\boldsymbol{\beta}}$ à jour :

$$\hat{\boldsymbol{\beta}}_{m+1} = \hat{\boldsymbol{\beta}}_m + \left(\sum_{i=1}^{n} \boldsymbol{D}_i'\boldsymbol{V}_i{}^{-1}\boldsymbol{D}_i\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{D}_i'\boldsymbol{V}_i{}^{-1}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}_m)\}\right).$$

5. Itérer les étapes 2 à 4 jusqu'à convergence (différence entre $\hat{\boldsymbol{\beta}}_{m+1}$ et $\hat{\boldsymbol{\beta}}_m$ plus petite qu'une tolérance spécifiée).

Si $\boldsymbol{R}_i(\boldsymbol{\alpha})$ était la vraie structure de corrélation pour $\boldsymbol{Y}_i$, alors la variance de $\hat{\boldsymbol{\beta}}$ serait estimée par

$$\boldsymbol{V}_T = \left(\sum_{i=1}^{n} \boldsymbol{D}_i'\boldsymbol{V}_i{}^{-1}\boldsymbol{D}_i\right)^{-1}\Bigg|_{\substack{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}\\ \phi=\hat{\phi}\\ \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}}.$$

Mais comme $\boldsymbol{R}_i(\boldsymbol{\alpha})$ n'est qu'une matrice de corrélation de *travail* et est possiblement fausse, alors on estime la variance de $\hat{\boldsymbol{\beta}}$ par un estimateur de matrice de variance sandwich robuste :

$$\boldsymbol{V}_S = \boldsymbol{V}_T\left(\sum_{i=1}^{n} \boldsymbol{D}_i'\boldsymbol{V}_i{}^{-1}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}'\boldsymbol{V}_i{}^{-1}\boldsymbol{D}_i\right)\Bigg|_{\substack{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}\\ \phi=\hat{\phi}\\ \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}}\boldsymbol{V}_T. \qquad (2.31)$$

Le terme "sandwich" vient du fait que dans l'expression (2.31), une correction empirique est prise en "sandwich" entre deux estimateurs de variance basés sur le modèle de travail (Fay & Graubard (2001)).

2.5 Spécification et estimation de $\boldsymbol{R}(\boldsymbol{\alpha})$

2.5.1 Introduction

Il y'a plusieurs manières dont nous pouvons spécifier la structure de corrélation de *travail*. Dans cette section, nous donnons une liste des formes les plus communes pour la structure de corrélation de *travail* $\boldsymbol{R}(\boldsymbol{\alpha})$ et déterminons son estimateur pour chacune de ces formes. Une liste plus exhaustive des formes possibles de $\boldsymbol{R}(\boldsymbol{\alpha})$ est donnée dans le livre de Hardin & Hilbe (2002).

2.5.2 Corrélation échangeable

La forme la plus simple de la matrice de corrélation de *travail* est la matrice identité ($\boldsymbol{R}(\alpha) = \mathbf{I}_{n_i \times n_i}$), où l'on assume que la corrélation entre Y_{ij} et $Y_{ij'}$ est nulle pour $j \neq j'$. Dans une extension simple de cette structure, on fait l'hypothèse que la corrélation entre Y_{ij} et $Y_{ij'}$ est α pour $j \neq j'$. Ce type de corrélation est la corrélation échangeable. On l'appelle également la corrélation commune, la corrélation égale, ou la corrélation composée. Dans ce cas, α est un scalaire et la matrice de corrélation de *travail* a la structure suivante :

$$\boldsymbol{R}(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha & \dots & \alpha \\ \alpha & 1 & \alpha & \dots & \alpha \\ \alpha & \alpha & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \dots & 1 \end{pmatrix}.$$

Cette hypothèse est généralement appropriée pour des ensembles de données dans lesquels les mesures répétées n'ont aucune dépendance de temps. S'il y a peu d'individus et beaucoup d'observations par individu, une matrice de corrélation de *travail* échangeable est un bon choix. La corrélation échangeable assume un seul facteur de corrélation entre deux mesures répétées quelconques et la même variance pour chaque mesure répétée. Par exemple, une étude de santé dans laquelle les panels représentent les cliniques et les mesures répétées les patients dans les cliniques est un bon exemple de ce type de données.

Des GEE avec une structure de corrélation échangeable utilisent les estimés des résidus de Pearson,

$$\hat{r}_{ij} = \frac{(y_{ij} - \hat{\mu}_{ij})}{\sqrt{V(\hat{\mu}_{ij})}},$$

de l'ajustement du modèle pour estimer le paramètre commun de corrélation (Hardin & Hilbe (2002)). L'estimateur de α utilisant ces résidus est

$$\hat{\alpha} = \frac{1}{\hat{\phi}} \sum_{i=1}^{n} \left(\frac{\sum_{u=1}^{n_i} \sum_{v=1}^{n_i} \hat{r}_{iu} \hat{r}_{iv} - \sum_{u=1}^{n_i} \hat{r}_{iu}^2}{n_i(n_i - 1)} \right).$$

2.5.3 Corrélation auto-régressive d'ordre 1 (AR(1))

Dans ce type de structure, on suppose que la corrélation entre Y_{ij} et $Y_{ij'}$ est $\alpha^{|j'-j|}$ pour $j \neq j'$. Dans ce cas, α est toujours un scalaire et la matrice de corrélation de travail a la structure suivante :

$$\boldsymbol{R}(\alpha) = \begin{pmatrix} 1 & \alpha & \dots & \alpha^{n_i-1} \\ \alpha & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha \\ \alpha^{n_i-1} & \dots & \alpha & 1 \end{pmatrix}.$$

Comme cette structure implique que la corrélation diminue à mesure que l'écart entre j et j' augmente, ce type de corrélation est indiqué pour des ensembles de données dans lesquels les mesures répétées ont une dépendance temporelle, l'indice j dénotant l'ordre (chronologique) dans lequel les mesures ont été prises. En effet, il est souvent raisonnable que les corrélations entre les mesures répétées prises ensemble dans le temps soient plus fortes que celles prises après de longs intervalles de temps. Dans ce type de structure, on exploite la corrélation temporelle des mesures répétées. Bref, une matrice de corrélation de *travail* autorégressive permet de tenir compte de l'effet de l'auto-corrélation temporelle. Une étude de santé dans laquelle les panels sont représentés par les patients avec plusieurs mesures sur chaque patient dans le temps est un bon exemple pour ce type de données.

Comme dans la structure de corrélation échangeable, on utilise les estimés des résidus de Pearson $\hat{r}_{ij}$ de l'ajustement du modèle pour estimer les corrélations (SAS Institute Inc. (2004)). L'estimateur de α utilisant ces résidus est

$$\hat{\alpha} = \frac{1}{(K_1 - p)\hat{\phi}} \sum_{i=1}^{n} \sum_{j=1}^{n_i - 1} \hat{r}_{i,j} \hat{r}_{i,j+1},$$

où $K_1 = \sum_{i=1}^{n} (n_i - 1)$.

2.5.4 Corrélation non structurée

Ce type de structure suppose que la corrélation entre Y_{ij} et $Y_{ij'}$ est $\alpha_{jj'}$ pour $j \neq j'$. La matrice de corrélation de *travail* non structurée est donc la plus générale des structures de corrélations discutées. Elle a la structure suivante :

$$\boldsymbol{R}(\alpha) = \begin{pmatrix} 1 & \alpha_{1,2} & \dots & \alpha_{1,n_i} \\ \alpha_{1,2} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{n_i-1,n_i} \\ \alpha_{1,n_i} & \dots & \alpha_{n_i-1,n_i} & 1 \end{pmatrix}.$$

Cette hypothèse n'impose aucune structure particulière à la matrice de corrélation de *travail*. En d'autres termes, aucune structure particulière n'est assumée sur les covariances entre Y_{ij} et $Y_{ij'}$ pour $j \neq j'$. Ceci implique que chaque paire d'observations a sa propre corrélation. S'il y a peu d'observations par individu et plusieurs individus, une matrice de corrélation de travail non structurée est sans doute un bon choix. Un bon exemple de données adaptées à ce type de structure est donné par Stokes, Davis, et Koch (1995), qui ajustent un modèle GEE pour données binaires, où l'on considère des données d'essais cliniques comparant deux traitements pour un désordre respiratoire. Ils disposent de deux centres et les patients à chacun des deux centres sont aléatoirement affectés aux groupes recevant le traitement ou un placebo. Pendant le traitement, le statut respiratoire est codé (0 = mauvais, 1 = bon) pour chacune de quatre visites. Ici les panels sont représentés par les patients (qui sont au nombre de 111) avec quatre observations par patient (une observation par visite).

Comme dans les sous-sections précédentes, on utilise les estimés des résidus de Pearson $\hat{r}_{ij}$ de l'ajustement du modèle pour estimer les corrélations (Hardin & Hilbe (2002)). L'estimateur de $\boldsymbol{R}(\alpha)$ utilisant ces résidus est

$$\hat{\boldsymbol{R}}(\alpha) = \frac{\sum_{i=1}^{n} n_i}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} \frac{\hat{r}_{ij}^2}{n_i}} G,$$

où

$$\mathbf{G} = \begin{pmatrix} g_{1,1}\hat{r}_{i,1}^2 & g_{1,2}\hat{r}_{i,1}\hat{r}_{i,2} & \dots & g_{1,n_i}\hat{r}_{i,1}\hat{r}_{i,n_i} \\ g_{2,1}\hat{r}_{i,2}\hat{r}_{i,1} & g_{2,2}\hat{r}_{i,2}^2 & \ddots & g_{2,n_i}\hat{r}_{i,2}\hat{r}_{i,n_i} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n_i,1}\hat{r}_{i,n_i}\hat{r}_{i,1} & g_{n_i,2}\hat{r}_{i,n_i}\hat{r}_{i,2} & \dots & g_{n_i,n_i}\hat{r}_{i,n_i}^2 \end{pmatrix},$$

$g_{uv} = (\sum_{i=1}^{n} I(i,u,v)^{-1})$,

$$I(i,u,v) = \begin{cases} 1 & \text{si le panel } i \text{ à des observations aux indices } u \text{ et } v \\ 0 & \text{sinon.} \end{cases}$$

Malheureusement, la matrice de corrélation ainsi estimée n'est pas nécessairement inversible et des problèmes numériques peuvent survenir, particulièrement pour des ensembles de données non équilibrés, c'est-à-dire le cas où l'on n'a pas le même nombre d'observations par panel ou des données manquantes. Dans de tels cas, il est recommandé d'utiliser une structure de corrélation de *travail* plus simple, comme celles présentées aux sous-sections précédentes.

Par l'approche GEE, on ne modélise pas directement la corrélation entre les mesures. De plus les inférences sont marginales (inférences sur les effets moyens des variables exogènes dans la population, et non pas sur les effets sur les groupes). L'ajout d'effets aléatoires dans le modèle linéaire généralisé au chapitre suivant permettra la modélisation de cette corrélation et de faire des inférences conditionnelles (inférences au niveau des groupes), et on parlera ainsi de modèles linéaires généralisés mixtes.

Chapitre 3

Approche conditionnelle (GLMM)

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. Les modèles à effets aléatoires constituent un moyen plus élaboré d'étudier cette variabilité. Ainsi, l'introduction d'effets aléatoires permet, d'une part, de séparer la variabilité totale en deux parties : la variabilité due aux effets aléatoires et celle que l'on affecte aux erreurs. D'autre part, elle permet de modéliser la corrélation entre les variables endogènes.

Mais qu'est-ce qu'un effet aléatoire ? Tentons de répondre à cette question à l'aide d'une illustration basée sur un exemple purement fictif dans lequel on oppose les deux natures possibles des effets : effet fixe/effet aléatoire. Imaginons que l'on s'intéresse à l'effet de trois 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 chacun un type de médicament de façon à ce que chaque type soit administré à quatre personnes différentes. Pour chaque personne, on relève, après chacune des quatre prises du médicament (en quatre occasions différentes), le temps de disparition des maux de tête. On a donc mentionné deux facteurs pouvant avoir effet : le médicament administré et la personne concernée. 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 effet fixe. Cependant, les 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 variabilité totale. Le facteur personne est donc considéré comme facteur à effet aléatoire.

Grâce à cette notion d'effet aléatoire, les modèles linéaires classiques ne contenant que des effets fixes ont pu être enrichis et élargis en modèles linéaires mixtes en y introduisant des effets aléatoires. De même que les effets aléatoires ont été introduits dans les modèles linéaires, ils peuvent l'être tout naturellement au sein des modèles linéaires généralisés (qui sont des modèles de régression ordinaire, où la variable endogène peut suivre une distribution autre que la normale) pour donner naissance aux modèles linéaires généralisés mixtes (GLMM), l'objet de ce chapitre, qui permettent de modéliser la corrélation entre les mesures, ce qui n'était pas le cas avec l'approche GEE du chapitre précédent. De plus, on suppose que les effets aléatoires dans un modèles linéaire généralisé mixte suivent, comme pour le modèle linéaire mixte, une distribution normale. Conditionnellement aux effets aléatoires, la variable endogène suit une loi faisant partie de la famille exponentielle.

Dans ce chapitre, nous allons parler dans une première partie de la structure du modèle linéaire généralisé mixte et de ses propriétés, ainsi que de l'interprétation des coefficients de régression $\boldsymbol{\beta}$. Une deuxième partie sera entièrement consacrée à l'estimation des paramètres du modèle et des inférences.

3.2 Structure du modèle

Soient $\mathbf{Y}, \boldsymbol{\gamma}$ et $\boldsymbol{\varepsilon}$, le vecteur $n \times 1$ des valeurs de la variable endogène, le vecteur $r \times 1$ de variables aléatoires (nommées effets aléatoires, qui ne sont pas observés) et le vecteur $n \times 1$ des termes d'erreur. Soient $\mathbf{X}$ et $\mathbf{Z}$, la matrice de schéma $n \times p$ et la matrice de schéma $n \times r$ connues. Le modèle linéaire généralisé mixte suppose que sachant $\mathbf{X}$, les Y_{ij} sont indépendants des $Y_{i'j'}$ pour tout $i \neq i'$. De plus, on suppose que $(Y_{i1}|\boldsymbol{x}_{i1}, \boldsymbol{\gamma}_i), (Y_{i2}|\boldsymbol{x}_{i2}, \boldsymbol{\gamma}_i), ..., (Y_{in_i}|\boldsymbol{x}_{in_i}, \boldsymbol{\gamma}_i)$ sont indépendants pour tout i. (Notez que ceci signifie, en général, que $(Y_{i1}|\boldsymbol{x}_{i1}), (Y_{i2}|\boldsymbol{x}_{i2}), ..., (Y_{in_i}|\boldsymbol{x}_{in_i})$ sont corrélés.)

Comme son nom l'indique, le modèle linéaire généralisé mixte est un hybride des modèles linéaires mixte et linéaire généralisé. On supposera donc que $(Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i) \sim f(y|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i)$, où

$$f(y|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i) = \exp\left\{\frac{y\theta_{ij} - b(\theta_{ij})}{a(\phi)} - c(y, \phi)\right\}, \quad i = 1, ..., n; \qquad j = 1, ..., n_i \tag{3.1}$$

avec $E(Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i) = \mu_{ij} = b'(\theta_{ij})$ (égalité démontrée à l'Annexe B) et $Var(Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i) = a(\phi)b''(\theta_{ij}) = a(\phi)V(\mu_{ij})$ (égalité démontrée à l'Annexe B).

Le modèle postule que l'effet des variables exogènes et des effets aléatoires sur la

distribution de Y_{ij} consiste à modifier la valeur de μ_{ij} ainsi :

$$g(\mu_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i, \tag{3.2}$$

où $g(\cdot)$ est une fonction de lien connue, $\boldsymbol{\beta}$ est le vecteur des paramètres ($\dim\beta = p$) et $\boldsymbol{z}_{ij}$ est la rangée de la matrice $\mathbf{Z}$ qui correspond à la j^{eme} observation du groupe i. Par exemple, si

$$g(\mu_{ij}) = (\beta_0 + \gamma_{0i}) + (\beta_1 + \gamma_{1i})x_{ij1} + (\beta_2 + \gamma_{2i})x_{ij2} + \cdots + (\beta_p + \gamma_{pi})x_{ijp},$$

alors $\boldsymbol{z}_{ij} = (1, x_{ij1}, x_{ij2}, \cdots, x_{ijp})'$. Afin de compléter la spécification du modèle, nous devons supposer une distribution pour les effets aléatoires, puisque ces derniers ne sont pas observés. Pour l'instant, nous nous contenterons de supposer que les $\boldsymbol{\gamma}_i$ sont iid selon une distribution avec une densité connue $f_{\boldsymbol{\gamma}}$,

$$\boldsymbol{\gamma}_i \sim f_{\boldsymbol{\gamma}}. \tag{3.3}$$

Nous verrons qu'en fait les logiciels assument que $\boldsymbol{\gamma}_i$ suit une loi normale multivariée de moyenne $\mathbf{0}$ et avec une matrice de covariance $\mathbf{D}$ de structure connue (par exemple non-structurée, symétrie composée, etc...). De plus, la variance de Y sachant $\boldsymbol{\gamma}$ peut s'écrire

$$\mathrm{Var}(Y|\boldsymbol{\gamma}) = a(\phi)\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{R}\boldsymbol{A}^{\frac{1}{2}},$$

avec $\boldsymbol{A}^{\frac{1}{2}}$ une matrice diagonale comme au chapitre 2, mais contenant $\sqrt{\frac{\mathrm{Var}(\boldsymbol{Y}_i)}{a(\phi)}}$ sur la diagonale principale, et $\boldsymbol{R}$ est une matrice de corrélation dont la forme générale doit être définie par l'utilisateur.

3.3 Propriétés du modèle et interprétation

3.3.1 Introduction

Le modèle tel que décrit à la section 3.2 est un modèle conditionnel, c'est-à-dire que les moyennes μ_{ij} sont sachant la valeur des effets aléatoires $\boldsymbol{\gamma}_i$. Elles correspondent donc aux attributs des observations du groupe i, et non aux attributs de la population générale. Si nous désirons faire des inférences sur la population générale, il nous faut les moyennes, variances et covariances **marginales**.

3.3.2 Moyenne marginale

$$\begin{aligned} E(Y_{ij}|\boldsymbol{x}_{ij}) &= E_{\boldsymbol{\gamma}_i}\Big\{E(Y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\gamma}_i)\Big\} && (3.4)\\ &= E_{\boldsymbol{\gamma}_i}(\mu_{ij}) && (3.5)\\ &= E_{\boldsymbol{\gamma}_i}\Big\{g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\} && (3.6)\end{aligned}$$

À moins que la fonction de lien $g(\cdot)$ ne soit spécifiée, cette expression ne peut se simplifier. Pour l'illustration avec une fonction de lien $g(\cdot)$ particulière, supposons qu'on a un lien log. Ainsi $\mathrm{g}(\mu_{ij}) = \log(\mu_{ij})$ et $g^{-1}(x) = \exp(x)$. Alors, on a

$$\begin{aligned} E(Y_{ij}|\boldsymbol{x}_{ij}) &= E_{\boldsymbol{\gamma}_i}\Big\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\} && (3.7)\\ &= \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})E_{\boldsymbol{\gamma}_i}\Big\{\exp(\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\} && (3.8)\\ &= \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}'_{ij}), && (3.9)\end{aligned}$$

où $M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}_{ij})$ est la fonction génératrice des moments conjoints de $\boldsymbol{\gamma_i}$ évaluée en $\boldsymbol{z}_{ij}$. Supposons en outre que $\boldsymbol{\gamma}_i$ suit une loi normale multivariée de moyenne $\mathbf{0}$ et de variance-covariance $\mathbf{D}$. On a

$$E(Y_{ij}|\boldsymbol{x}_{ij}) = \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\exp(\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2)$$

ou

$$\log E(Y_{ij}|\boldsymbol{x}_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2. \quad (3.10)$$

3.3.3 Variance marginale

$$\begin{aligned} \mathrm{Var}(Y_{ij}|\boldsymbol{x}_{ij}) &= \mathrm{Var}_{\boldsymbol{\gamma}_i}\Big\{E(Y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\gamma}_i)\Big\} + E_{\boldsymbol{\gamma}_i}\Big\{\mathrm{Var}(Y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\gamma}_i)\Big\} && (3.11)\\ &= \mathrm{Var}_{\boldsymbol{\gamma}_i}(\mu_{ij}) + E_{\boldsymbol{\gamma}_i}\Big\{a(\phi)V(\mu_{ij})\Big\} && (3.12)\\ &= \mathrm{Var}_{\boldsymbol{\gamma}_i}\Big\{g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\} \\ &+ E_{\boldsymbol{\gamma}_i}\Big\{a(\phi)V(g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i))\Big\}, && (3.13)\end{aligned}$$

ce qui, encore une fois ne peut être simplifié sans faire des hypotèses spécifiques au sujet de la forme de la fonction de lien $g(\cdot)$ et/ou de la distribution conditionnelle de Y. Pour l'illustration, supposons qu'on a une fonction de lien log ($g(\mu_{ij}) = \log(\mu_{ij})$ et $g^{-1}(x) = \exp(x)$) et supposons que les Y_{ij} sachant $\boldsymbol{\gamma}_i$ sont indépendantes avec une distribution de Poisson ($a(\phi)V(\mu_{ij}) = \mu_{ij}$,$a(\phi) = 1$ et $V(\mu_{ij}) = \mu_{ij}$) et $\boldsymbol{\gamma}_i$ suit une loi normale multivariée de moyenne $\mathbf{0}$ et de variance-covariance $\mathbf{D}$. Alors

$$
\begin{aligned}
\mathrm{Var}(Y_{ij}|\boldsymbol{x}_{ij}) &= \mathrm{Var}_{\boldsymbol{\gamma}_i}\Big\{E(Y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\gamma}_i)\Big\} + E_{\boldsymbol{\gamma}_i}\Big\{\mathrm{Var}(Y_{ij}|\boldsymbol{x}_{ij},\boldsymbol{\gamma}_i)\Big\} \\
&= \mathrm{Var}_{\boldsymbol{\gamma}_i}(\mu_{ij}) + E_{\boldsymbol{\gamma}_i}(\mu_{ij}) \\
&= \mathrm{Var}_{\boldsymbol{\gamma}_i}\Big\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\} - E_{\boldsymbol{\gamma}_i}\Big\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\} \\
&= E_{\boldsymbol{\gamma}_i}\Big\{\exp(2(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i))\Big\} - \Big\{E_{\boldsymbol{\gamma}_i}\Big\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\}\Big\}^2 \\
&+ E_{\boldsymbol{\gamma}_i}\Big\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i)\Big\} \\
&= \exp(2(\boldsymbol{x}'_{ij}\boldsymbol{\beta}))\Big\{M_{\boldsymbol{\gamma}_i}(2\boldsymbol{z}_{ij}) - (M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}_{ij}))^2 + \exp(-\boldsymbol{x}'_{ij}\boldsymbol{\beta})M_{\boldsymbol{\gamma}_i}(2\boldsymbol{z}_{ij})\Big\} \\
&= \exp(2(\boldsymbol{x}'_{ij}\boldsymbol{\beta}))\Big\{\exp(2\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}) - \exp(\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}) \\
&+ \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\exp(\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2)\Big\} \\
&= \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2) \\
&\times \Big\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\big\{\exp(3\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2) - \exp(\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2)\big\} + 1\Big\} \\
&= E(Y_{ij}|\boldsymbol{x}_{ij})\Big[1 + \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\big\{\exp(3\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2) \\
&- \exp(\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2)\big\}\Big]. \qquad (3.14)
\end{aligned}
$$

Le terme entre crochets dans (3.14) est supérieur à 1, ce qui indique que la variance marginale est supérieure à la moyenne marginale, même dans le cas de la distribution de Poisson. Dans ce sens nous pouvons voir les effets aléatoires comme une manière de modéliser ou d'attribuer la surdispersion à une source particulière. Elle peut par exemple survenir lorsque la valeur moyenne de la variable endogène dépend d'une variable aléatoire latente (non observée), dans ce cas-ci $\boldsymbol{\gamma}_i$, que nous n'incluons pas dans le modèle de régression.

3.3.4 Covariance marginale

$$\begin{aligned}\mathrm{Cov}(Y_{ij}, Y_{ij'}|\boldsymbol{x}_{ij}, \boldsymbol{x}_{ij'}) &= \mathrm{Cov}\Big\{E(Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i), E(Y_{ij'}|\boldsymbol{x}_{ij'}, \boldsymbol{\gamma}_i)\Big\} \\ &+ E\Big\{\mathrm{Cov}(Y_{ij}, Y_{ij'}|\boldsymbol{x}_{ij}, \boldsymbol{x}_{ij'}, \boldsymbol{\gamma}_i)\Big\}.\end{aligned}$$

Comme nous avons supposé que sachant $\boldsymbol{\gamma}_i$ et les variables exogènes, les Y_{ij} sont indépendantes, on a que

$$\mathrm{Cov}(Y_{ij}, Y_{ij'}|\boldsymbol{x}_{ij}, \boldsymbol{x}_{ij'}, \boldsymbol{\gamma}_i) = 0.$$

D'où

$$\begin{aligned}\mathrm{Cov}(Y_{ij}, Y_{ij'}|\boldsymbol{x}_{ij}, \boldsymbol{x}_{ij'}) &= \mathrm{Cov}(\mu_{ij}, \mu_{ij'}) + E(0) \\ &= \mathrm{Cov}\Big\{g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i), g^{-1}(\boldsymbol{x}'_{ij'}\boldsymbol{\beta} + \boldsymbol{z}'_{ij'}\boldsymbol{\gamma}_i)\Big\}. \quad (3.15)\end{aligned}$$

Encore une fois, à moins que la fonction de lien $g(\cdot)$ ne soit spécifiée, cette expression ne peut se simplifier. Pour l'illustration pour une fonction de lien $g(\cdot)$ particulière, supposons qu'on a un lien log, ainsi $g(\mu_{ij}) = \log(\mu_{ij})$ et $g^{-1}(x) = \exp(x)$. Alors, on a

$$\begin{aligned}\mathrm{Cov}(Y_{ij}, Y_{ij'}|\boldsymbol{x}_{ij}, \boldsymbol{x}_{ij'}) &= \mathrm{Cov}\Big\{\exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i), \exp(\boldsymbol{x}'_{ij'}\boldsymbol{\beta} + \boldsymbol{z}'_{ij'}\boldsymbol{\gamma}_i)\Big\} \\ &= \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{x}'_{ij'}\boldsymbol{\beta})\mathrm{Cov}\Big\{\exp(\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i), \exp(\boldsymbol{z}'_{ij'}\boldsymbol{\gamma}_i)\Big\} \\ &= \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{x}'_{ij'}\boldsymbol{\beta})\Big\{M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}_{ij} + \boldsymbol{z}_{ij'}) - M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}_{ij})M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}_{ij'})\Big\}.\end{aligned}$$

Supposons, en outre, que $\boldsymbol{\gamma}_i$ suit une loi normale multivariée de moyenne $\mathbf{0}$ et de variance-covariance $\mathbf{D}$. On a

$$\begin{aligned}\mathrm{Cov}(Y_{ij}, Y_{ij'}|\boldsymbol{x}_{ij}, \boldsymbol{x}_{ij'}) = \exp\Big\{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{x}'_{ij'}\boldsymbol{\beta}\Big\} \Big[\exp\Big\{\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2 + \boldsymbol{z}'_{ij'}\mathbf{D}\boldsymbol{z}_{ij'}/2\Big\} \\ \times \Big(\exp\Big\{\boldsymbol{z}'_{ij}(\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}/2 + \boldsymbol{z}'_{ij'}\mathbf{D}\boldsymbol{z}_{ij'}/2)\boldsymbol{z}_{ij'}\Big\} - 1\Big) \Big]\end{aligned}$$

qui est égale à 0 si $\boldsymbol{z}'_{ij}\boldsymbol{z}_{ij'} = 0$ (c'est à dire si les deux observations n'ont pas d'effet aléatoire en commun) et est positive ailleurs (cas où $\boldsymbol{z}'_{ij}\boldsymbol{z}_{ij'} \neq 0$).

3.3.5 Interprétation des coefficients

Comment interpréter les $\boldsymbol{\beta}$ dans un modèle linéaire généralisé mixte ? Tout d'abord, l'interprétation est plus simple dans le cas d'inférences conditionnelles. Ainsi si le coefficient devant x_{ijk} est $(\beta_j + \gamma_{ki})$, alors une hausse d'une unité de x_{ijk}, toute autre variable exogène demeurant inchangée, augmentera $g^{-1}(\mu_{ij})$ de $(\beta_j + \gamma_{ki})$ unités. Bien sûr, s'il n'y a pas de pente aléatoire devant x_{ijk}, alors l'effet devient une hausse de β_j unités, ce qui est la même interprétation que dans le cadre des modèles linéaires généralisés, à la différence que l'effet en est un sur le prédicteur linéaire de chaque groupe, et non sur le prédicteur linéaire de la population entière.

Peut-on quand même avoir une idée de l'effet d'une hausse de la k^{ieme} variable exogène sur la distribution de Y dans la population globale, pas seulement sur la distribution conditionnelle de Y sachant les groupes ? La réponse est oui, dans certains cas particuliers. Par exemple, retournons à la sous-section 3.3.2 de la régression de Poisson avec lien log. Alors si nous n'avons pas de terme aléatoire devant la k^{ieme} variable exogène,

$$\frac{E(Y_{ij}|x_{ijk} = x + 1)}{E(Y_{ij}|x_{ijk} = x)} = \exp(\beta_k).$$

On a donc exactement le même effet que dans le cadre des modèles linéaires généralisés. Malheureusement, si nous avions eu un terme aléatoire devant x_{ijk}, alors nous n'aurions pas le même résultat car $\boldsymbol{z}_{ij}$ avec $x_{ijk} = x + 1$, disons $\boldsymbol{z}^*$, ne serait pas le même que $\boldsymbol{z}_{ij}$ avec $x_{ijk} = x$, disons $\boldsymbol{z}^{**}$. En effet, dans ce cas on obtient

$$\frac{E(Y_{ij}|x_{ijk} = x + 1)}{E(Y_{ij}|x_{ijk} = x)} = \exp(\beta_k)\frac{M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}^*)}{M_{\boldsymbol{\gamma}_i}(\boldsymbol{z}^{**})},$$

où $\boldsymbol{z}^* = (1, x_{ij1}, \cdots, x_{ijk-1}, x + 1, x_{ijk+1}, \cdots, x_{ijp})'$ et
$z^{**} = (1, x_{ij1}, \cdots, x_{ijk-1}, x, x_{ijk+1}, \cdots, x_{ijp}), p = dim(\boldsymbol{\beta})'$.

En général, pour les inférences marginales, il est plus simple de passer par l'approche basée sur les équation d'estimation généralisées (GEE) décrite au chapitre précédent. En fait, ceci nous mène à une question intéressante : Doit-on s'attendre à une estimation différente de la valeur de $\boldsymbol{\beta}$ entre une approche marginale basée sur les équation d'estimation généralisées (GEE) et une approche basée sur un modèle linéaire généralisé mixte ? La réponse est oui, ce qui ne devrait pas être surprenant puis que, tel que vu ci-dessus, l'effet des variables exogènes sur les moyennes conditionnelles de chaque groupe et l'effet sur la moyenne marginale de la population ne sont pas toujours égaux. Pour l'illustrer, supposons le cas particulier

$$Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i \sim \text{binomiale}(1, \pi_{ij}),$$

où $\Phi^{-1}(\pi_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i$ et $\boldsymbol{\gamma}_i$ suit une loi normale multivariée de moyenne $\mathbf{0}$ et de variance-covariance $\mathbf{D}$. Alors, on a

$$\begin{aligned} E(Y_{ij}|\boldsymbol{x}_{ij}) = \pi_{ij} &= \Phi(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i) \\ &= P(Z \leq \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i) \end{aligned}$$

où $Z \sim$ normale(0,1). On a donc

$$E(Y_{ij}|\boldsymbol{x}_{ij}) = P(Z - \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i \leq \boldsymbol{x}'_{ij}\boldsymbol{\beta})$$

Posons $W = Z - \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i$. On obtient

$$\begin{aligned} E(W) &= E(Z) - E(\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i) \\ &= 0 - 0 \\ &= 0 \\ \mathrm{Var}(W) &= \mathrm{Var}(Z) + \mathrm{Var}(\boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i) \\ &= 1 + \boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij} \\ E(Y_{ij}|\boldsymbol{x}_{ij}) &= P(W \leq \boldsymbol{x}'_{ij}\boldsymbol{\beta}) \\ &= P\left(Z \leq \frac{\boldsymbol{x}'_{ij}\boldsymbol{\beta}}{\sqrt{1 + \boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}}}\right) \\ &= \Phi\left(\frac{\boldsymbol{x}'_{ij}\boldsymbol{\beta}}{\sqrt{1 + \boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}}}\right) \\ &= \Phi(\boldsymbol{x}'_{ij}\boldsymbol{\beta}^*), \end{aligned}$$

où $\boldsymbol{\beta}^* = \frac{\boldsymbol{\beta}}{\sqrt{1+\boldsymbol{z}'_{ij}\mathbf{D}\boldsymbol{z}_{ij}}}$. Comme le terme à l'interieur de la racine carrée est supérieur à 1, $\boldsymbol{\beta}^*$ sera toujours inférieur à $\boldsymbol{\beta}$ en valeur absolue, ce qui signifie que l'effet moyen des variables exogènes sur une population est une atténuation des effets moyens à l'intérieur des groupes.

Notez que le phénomène d'atténuation observé dans l'illustration n'est pas spécifique à nos suppositions et se produit en général. En outre, plus les groupes sont hétérogènes (c'est à dire les $\boldsymbol{\gamma}_i$ de chaque groupe sont différents), plus l'atténuation sera importante. Pour plus de détails, voir McCullagh et Searle (2001, section 8.5).

3.4 Estimation par maximum de vraisemblance

3.4.1 La vraisemblance

Comme les $(Y_{ij}|x_{ij}, \gamma_i)$ sont supposées indépendantes pour $i = 1, ..., n$ et $j = 1, ..., n_i$, en combinant (3.1), (3.2) et (3.3), la fonction de vraisemblance basée sur les données observées s'écrit

$$L(\boldsymbol{\beta}) = \int \prod_{i=1}^{n} \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_i) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) d\boldsymbol{\gamma}. \tag{3.16}$$

3.4.2 Equations de vraisemblance

Dans cette sous-section, nous allons parler en premier lieu des équations de vraisemblance pour les paramètres à effet fixe et en second lieu de celles pour les paramètres dans la distribution de $f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})$. En effet, bien que les équations de vraisemblance soient numériquement complexes, nous pouvons les écrire sous une forme plus simple. De (3.16), on a

$$\ell(\boldsymbol{\beta}) = \log \int f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) d\boldsymbol{\gamma} = \log f_Y(y), \tag{3.17}$$

où $f_{Y|\boldsymbol{\gamma}}$ est la densité conditionnelle de Y par rapport à $\boldsymbol{\gamma}$.

$$\begin{aligned} \Rightarrow \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \int f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) d\boldsymbol{\gamma} / f_Y(y) \\ &= \int \left[\frac{\partial}{\partial \boldsymbol{\beta}} f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}) \right] f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) d\boldsymbol{\gamma} / f_Y(y). \end{aligned} \tag{3.18}$$

Or,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}) &= \left(\frac{1}{f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma})} \frac{\partial f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} \right) f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}) \\ &= \frac{\partial \log f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}). \end{aligned} \tag{3.19}$$

On peut donc ré-écrire (3.18) comme

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \int \frac{\partial \log f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) d\boldsymbol{\gamma} / f_Y(y) \\ &= \int \frac{\partial \log f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} f_{\boldsymbol{\gamma}|Y}(\boldsymbol{\gamma}|y) d\boldsymbol{\gamma} \end{aligned} \tag{3.20}$$

$$= \int X'W^*(Y-\mu) f_{\boldsymbol{\gamma}|Y}(\boldsymbol{\gamma}|y) d\boldsymbol{\gamma}$$

$$= X'E(W^*Y|Y) - X'E(W^*\mu|Y), \tag{3.21}$$

où l'espérance est calculée par rapport à $f_{\boldsymbol{\gamma}|Y}$ (la densité conditionnelle de $\boldsymbol{\gamma}$ étant donné Y), et où $W^* = \text{Diag}\left\{[a(\phi)V(\mu_{ij})g_\mu(\mu_{ij})]^{-1}\right\}$ avec $g_\mu(x) = \frac{dg(x)}{dx}$.
L'équation de vraisemblance pour le paramètre $\boldsymbol{\beta}$ est donc

$$X'E(W^*Y|Y) = X'E(W^*\mu|Y). \tag{3.22}$$

Un résultat similaire à (3.20) peut être dérivé pour les équations de maximum de vraisemblance pour les paramètres dans la distribution de $f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma})$. Soit φ ces paramètres. Alors

$$\begin{aligned} \frac{\partial \ell(\beta)}{\partial \varphi} &= \int \frac{\partial \log f_\gamma(\gamma)}{\partial \varphi} f_{\gamma|Y}(\gamma|y) d\gamma \\ &= E\left[\frac{\partial \log f_\gamma(\gamma)}{\partial \varphi}\middle| Y\right]. \end{aligned} \tag{3.23}$$

Il est impossible de simplifier cette expression, à moins de spécifier une forme particulière de la distribution des effets aléatoires.

Maintenant, la question fondamentale qu'on se pose est de savoir comment résoudre l'équation (3.22) en pratique. Des méthodes sont illustrées dans la section suivante afin de trouver une réponse à cette question.

3.5 Inférences : Implantation

3.5.1 Estimation des paramètres

Dans un modèle mixte, l'estimation des coefficients de régression et la prévision des effets aléatoires sont beaucoup plus complexes que dans le cas d'un modèle où tous les effets sont fixes. En effet, en plus d'estimer les p composantes du vecteur $\boldsymbol{\beta}$, il faut prévoir les r composantes des n vecteurs $\boldsymbol{\gamma}_i$, toutes les composantes de la structure de corrélation de travail $\boldsymbol{R}$ pour Y et toutes les composantes de la matrice de variance-covariance $\mathbf{D}$ de $\boldsymbol{\gamma}$. Une première approche pour estimer les coefficients de régression de même que les effets aléatoires consiste à se définir une pseudo-variable réponse t_{ij} ainsi :

$$t_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\gamma}_i + g_\mu(\mu_{ij})\{y_{ij} - \mu_{ij}(\boldsymbol{\beta})\}, \tag{3.24}$$

où $g_\mu(x) = \frac{dg(x)}{dx}$.

Si on pose $\boldsymbol{\Delta}_i = \text{Diag}(g_\mu(\mu_{ij}))$; $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i})'$; $\mathbf{Z_i} = (\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{in_i})'$; $\boldsymbol{\mu_i}(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), ..., \mu_{in_i}(\boldsymbol{\beta})'$; $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{in_i})'$ et $\boldsymbol{t}_i = (t_{i1}, ..., t_{in_i})'$, alors on peut ré-écrire cette expression sous forme matricielle :

$$\boldsymbol{t}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\Delta}_i\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}. \tag{3.25}$$

Comme $\boldsymbol{\Delta}_i$ dans (3.25) dépend de $\boldsymbol{\mu_i}$, les espérance et variance de $\boldsymbol{t}_i$ sont de forme plutôt complexe. Cependant, si on réduit $\boldsymbol{t}_i$ en remplaçant le $\boldsymbol{\gamma}_i$ dans $\boldsymbol{\Delta}_i$ par zéro, alors on a

$$\boldsymbol{t}_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\Delta}_i^*\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i\}, \tag{3.26}$$

où $\boldsymbol{\Delta}_i^* = \text{Diag}\{g_\mu[g^{-1}(\boldsymbol{x}_{ij}'\boldsymbol{\beta})]\}$. Sous cette simplification,

$$E(\boldsymbol{t}_i^*) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i$$

et

$$\text{Var}(\boldsymbol{t}_i^*) = \boldsymbol{Z}_i\mathbf{D}\boldsymbol{Z}_i' + \boldsymbol{\Delta}_i^*\text{Var}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)\boldsymbol{\Delta}_i^*.$$

Si on pose $\mathbf{V} = \boldsymbol{\Delta}_i^*\text{Var}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)\boldsymbol{\Delta}_i^*$, on reconnaît la structure de moyenne et de variance d'un modèle linéaire mixte. En effet, l'algorithme de Schall (1991) pour ajuster un modèle linéaire généralisé mixte sous cette approche consiste à itérer les étapes suivantes dans lesquelles un modèle linéaire mixte est ajusté pour obtenir les estimateurs de $\boldsymbol{\beta}$ et de $\boldsymbol{\gamma}$.

1. Ajuster un modèle linéaire mixte aux données en remplaçant les $\boldsymbol{Y}_i$ par les $\boldsymbol{t}_i^*$; plus précisément, le modèle suivant :

$$\boldsymbol{t}_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}^*,$$

où $E(\boldsymbol{\varepsilon}^*) = 0$ et $\text{Var}(\boldsymbol{\varepsilon}^*) = \mathbf{V} = \boldsymbol{\Delta}_i^*\text{Var}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)\boldsymbol{\Delta}_i^*$. Et donc $\text{Var}(\boldsymbol{t}_i^*) = \boldsymbol{Z}_i\mathbf{D}\boldsymbol{Z}_i' + \mathbf{V}$.
2. Estimer les paramètres $\boldsymbol{\beta}$ et $\boldsymbol{\gamma}$ en résolvant les équations de Henderson (les équations du modèle mixte) :

$$\begin{pmatrix} \boldsymbol{X}_i'\mathbf{V}^{-1}\boldsymbol{X}_i & \boldsymbol{X}_i'\mathbf{V}^{-1}\boldsymbol{Z}_i \\ \boldsymbol{Z}_i'\mathbf{V}^{-1}\boldsymbol{X}_i & \boldsymbol{Z}_i'\mathbf{V}^{-1}\boldsymbol{Z}_i + \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_i'\mathbf{V}^{-1}\boldsymbol{t}_i^* \\ \boldsymbol{Z}_i'\mathbf{V}^{-1}\boldsymbol{t}_i^* \end{pmatrix} \tag{3.27}$$

et dénoter les estimateurs de $\boldsymbol{\beta}$ et de $\boldsymbol{\gamma}$ obtenus $\hat{\boldsymbol{\beta}}_0$ et $\hat{\boldsymbol{\gamma}}_0$, respectivement.
3. Calculer $\hat{\boldsymbol{\mu}}_{0i} = g^{-1}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0 + \boldsymbol{Z}_i\hat{\boldsymbol{\gamma}}_0)$.
4. Définir un nouveau $\boldsymbol{t}_i^* = \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_0 + \boldsymbol{Z}_i\hat{\boldsymbol{\gamma}}_0 + \boldsymbol{\Delta}_i^*\{\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_{0i}\}$.
5. Comparer les anciens estimés de $\mathbf{D}$ et $\boldsymbol{R}$ avec les nouveaux estimés (où $\boldsymbol{R}$ est une matrice de corrélation telle que définie à la section 3.2).

6. Itérer les étapes 2 à 5 jusqu' à convergence (différence entre les anciens estimés de $\mathbf{D}$ et $\boldsymbol{R}$ avec les nouveaux estimés négligeable). Par défaut, l'algorithme de la procédure GLIMMIX de SAS s'arrête si la différence entre les anciens et les nouveaux estimés est inférieure ou égale à 10^{-8}.

Les estimateurs $\hat{\boldsymbol{\beta}}$ et $\hat{\boldsymbol{\gamma}}$ ainsi obtenus sont respectivement les meilleurs estimateurs linéaires sans biais (BLUE) et les meilleurs prédicteurs linéaires sans biais (BLUP) (McCulloch & Searle (2001)).

Une justification complètement différente de cette approche est via ce qu'on appelle la quasi-vraisemblance pénalisée (PQL). Se rappeler que la quasi-vraisemblance ne spécifie pas une distribution, mais seulement le rapport moyenne-variance. Ce n'est pas aussi une base exhaustive pour estimer la structure de variance-covariance. Une suggestion (Green et Silverman, 1994) pour remédier à ce problème est d'ajouter une fonction de pénalité de la forme $\frac{1}{2}\boldsymbol{\gamma}'\mathbf{D}^{-1}\boldsymbol{\gamma}$ à la quasi-vraisemblance, d'où

$$PQL = \sum Q_i - \frac{1}{2}\boldsymbol{\gamma}'\mathbf{D}^{-1}\boldsymbol{\gamma}, \tag{3.28}$$

où $Q_i = \int_{y_{ij}}^{\mu_{ij}} \frac{y_{ij}-t}{a(\phi)V(t)}dt$.

Les équations du maximum de quasi-vraisemblance s'obtiennent en dérivant respectivement par rapport à $\boldsymbol{\beta}$ et à $\boldsymbol{\gamma}$ l'équation (3.28) :

$$\begin{aligned}\frac{\partial PQL}{\partial \beta} &= \frac{\partial}{\partial \beta}\sum_i Q_i = 0 \\ &= \sum \frac{y_{ij}-\mu_{ij}}{a(\phi)V(\mu_{ij})}\frac{\partial \mu_{ij}}{\partial \beta} = 0.\end{aligned}$$

Or

$$\begin{aligned}\frac{\partial \mu_{ij}}{\partial \boldsymbol{\beta}} &= \frac{\partial \mu_{ij}}{\partial g(\mu_{ij})}\frac{\partial g(\mu_{ij})}{\partial \boldsymbol{\beta}} \\ &= \left(\frac{\partial g(\mu_{ij})}{\partial \mu_{ij}}\right)^{-1}\frac{\partial \boldsymbol{x}'_{ij}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} \\ &= \left(\frac{\partial g(\mu_{ij})}{\partial \mu_{ij}}\right)^{-1} \times \boldsymbol{x}'_{ij}.\end{aligned}$$

Donc, on a

$$\sum \frac{y_{ij}-\mu_{ij}}{a(\phi)V(\mu_{ij})g_\mu(\mu_{ij})}\boldsymbol{x}'_{ij} = 0$$

ou bien, sous forme matricielle,

$$\frac{1}{a(\phi)}\boldsymbol{X}'_i\boldsymbol{W}_i\boldsymbol{\Delta}_i(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = 0, \tag{3.29}$$

où $\boldsymbol{W}_i = \text{Diag}\{[V(\mu_{ij})g^2_\mu(\mu_{ij})]^{-1}\}$ et

$$\frac{\partial PQL}{\partial \boldsymbol{\gamma}} = \frac{1}{a(\phi)}\boldsymbol{Z}_i'\boldsymbol{W}_i\boldsymbol{\Delta}_i(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) - \mathbf{D}^{-1}\boldsymbol{\gamma}_i = 0. \tag{3.30}$$

Ceux-ci mènent Breslow et Clayton (1993) à nouveau à un algorithme similaire à celui de Schall (1991).

Encore une autre justification de cette approche est obtenue par l'intermédiaire des approximations de Laplace. Le principe de l'approximation de Laplace est basé sur le développement en séries de Taylor au second ordre pour évaluer l'intégrale de dimension élevée dans la vraisemblance et prend la forme

$$\log \int_{\Re^q} e^{h(\boldsymbol{\gamma})} d\boldsymbol{\gamma} = h(\boldsymbol{\gamma}_0) + \frac{q}{2}\log 2\pi - \frac{1}{2}\log\left| - \frac{\partial^2 h(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'}\right|, \tag{3.31}$$

où $\boldsymbol{\gamma}_0$ est la solution de l'équation

$$\left.\frac{\partial h(\gamma)}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0} = 0. \tag{3.32}$$

On utilise ce résultat pour approximer la log-vraisemblance du GLMM via

$$\begin{aligned}
\ell &= \log \int f_{Y|\boldsymbol{\gamma}} f_(\boldsymbol{\gamma}) d\boldsymbol{\gamma} \\
&= \log \int e^{\log f_{Y|\boldsymbol{\gamma}} + \log f_{\boldsymbol{\gamma}}} d\boldsymbol{\gamma} \\
&= \log \int e^{h(\boldsymbol{\gamma})} d\boldsymbol{\gamma},
\end{aligned} \tag{3.33}$$

avec $h(\boldsymbol{\gamma}) = \log f_{Y|\boldsymbol{\gamma}} + \log f_{\boldsymbol{\gamma}}$.

Pour construire l'approximation de Laplace, (3.32) doit être résolue et une expression pour $\frac{\partial^2 h(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'}$ est nécessaire. Si on suppose que $\boldsymbol{\gamma}$ suit une loi normale multivariée de moyenne $\mathbf{0}$ et de variance-covariance $\mathbf{D}$, alors

$$\log f_{\boldsymbol{\gamma}} = -\frac{1}{2}\boldsymbol{\gamma}'\mathbf{D}^{-1}\boldsymbol{\gamma} - \frac{q}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{D}|$$

et $h(\boldsymbol{\gamma})$ devient

$$\log f_{Y|\boldsymbol{\gamma}} + \log f_{\boldsymbol{\gamma}} = \log f_{Y|\boldsymbol{\gamma}} - \frac{1}{2}\boldsymbol{\gamma}'\mathbf{D}^{-1}\boldsymbol{\gamma} - \frac{q}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{D}|.$$

En dérivant $h(\boldsymbol{\gamma})$ par rapport à $\boldsymbol{\gamma}$, on obtient

$$\frac{\partial h(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \frac{\partial \log f_{Y|\boldsymbol{\gamma}}}{\partial \boldsymbol{\gamma}} - \mathbf{D}^{-1}\boldsymbol{\gamma}$$

$$= \frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{\Delta}_i (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) - \mathbf{D}^{-1} \boldsymbol{\gamma}_i. \tag{3.34}$$

Pour trouver $\boldsymbol{\gamma}_0$, il est nécessaire de résoudre l'équation

$$\frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{\Delta}_i (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{D}^{-1} \boldsymbol{\gamma}_i, \tag{3.35}$$

ce qui n'est pas aussi simple qu'il ne le paraît, puisque $\boldsymbol{W}_i$, $\boldsymbol{\Delta}_i$ et $\boldsymbol{\mu}_i$ dans la partie gauche de l'équation (3.35) sont toutes des fonctions de $\boldsymbol{\gamma}$. On aura aussi besoin de la dérivée seconde :

$$\frac{\partial^2 h(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = \frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{\Delta}_i \frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\gamma}'} + \frac{1}{a(\phi)} \boldsymbol{Z}_i' \frac{\partial \boldsymbol{W}_i \boldsymbol{\Delta}_i}{\partial \boldsymbol{\gamma}'} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) - \mathbf{D}^{-1}. \tag{3.36}$$

Pour certains modèles (exemple binomial ou Poisson) $\boldsymbol{W}_i \boldsymbol{\Delta}_i = \mathbf{I}$, d'où le second terme vaut zéro. En général, le second terme tend en moyenne vers zéro en ce qui concerne la distribution conditionnelle de Y sachant $\boldsymbol{\gamma}$. Ainsi, il est peut-être raisonnable de le considérer négligeable. Si c'est le cas, (3.36) devient

$$\begin{aligned} -\frac{\partial^2 h(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} &= \frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{\Delta}_i \boldsymbol{\Delta}_i^{-1} \boldsymbol{Z}_i + 0 + \boldsymbol{D}^{-1} \\ &= \frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{Z}_i + \mathbf{D}^{-1} \\ &= \left(\frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{Z}_i \mathbf{D} + \mathbf{I} \right) \mathbf{D}^{-1}. \end{aligned} \tag{3.37}$$

Utilisant (3.37) dans (3.31), on obtient

$$\begin{aligned} \ell &\simeq \log f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}_0) - \frac{1}{2} \boldsymbol{\gamma}_0' \mathbf{D}^{-1} \boldsymbol{\gamma}_0 - \frac{q}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{D}| + \frac{q}{2} \log 2\pi \\ &- \frac{1}{2} \log \left| \left(\frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{Z}_i \mathbf{D} + \mathbf{I} \right) \mathbf{D}^{-1} \right| \\ &= \log f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}_0) - \frac{1}{2} \boldsymbol{\gamma}_0' \mathbf{D}^{-1} \boldsymbol{\gamma}_0 + \frac{1}{2} \log \left| \frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{Z}_i \mathbf{D} + \mathbf{I} \right|. \end{aligned} \tag{3.38}$$

Ceci doit encore être maximisé par rapport à $\boldsymbol{\beta}$ pour trouver une estimation du maximum de vraisemblance. Une dérivation par rapport à $\boldsymbol{\beta}$ donne une équation de score approximative :

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \frac{\partial \log f_{Y|\boldsymbol{\gamma}}(y|\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\beta}} + \frac{\partial}{\partial \boldsymbol{\beta}} \frac{1}{2} \log \left| \frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{Z}_i \mathbf{D} + \mathbf{I} \right| \tag{3.39}$$

$$= \frac{1}{a(\phi)} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{\Delta}_i (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + \frac{\partial}{\partial \boldsymbol{\beta}} \frac{1}{2} \log \left| \frac{1}{a(\phi)} \boldsymbol{Z}_i' \boldsymbol{W}_i \boldsymbol{Z}_i \mathbf{D} + \mathbf{I} \right| \tag{3.40}$$

$$= \frac{1}{a(\phi)} \boldsymbol{X}_i' \boldsymbol{W}_i \boldsymbol{\Delta}_i (\boldsymbol{Y}_i - \boldsymbol{\mu}_i). \tag{3.41}$$

On résoud ainsi conjointement les équations

$$\frac{1}{a(\phi)}\boldsymbol{X}_i'\boldsymbol{W}_i\boldsymbol{\Delta}_i(\boldsymbol{Y}_i-\boldsymbol{\mu}_i) = 0 \tag{3.42}$$

$$\frac{1}{a(\phi)}\boldsymbol{Z}_i'\boldsymbol{W}_i\boldsymbol{\Delta}_i(\boldsymbol{Y}_i-\boldsymbol{\mu}_i) = \mathbf{D}^{-1}\boldsymbol{\gamma} \tag{3.43}$$

pour $\boldsymbol{\beta}$ et $\boldsymbol{\gamma}$.

Des méthodes pour résoudre ces équations s'appellent ainsi les méthodes PQL. Green (1990), Schall (1991) et Wolfinger (1993) en ont tous discuté. Malheureusement, ces méthodes ne fonctionnent pas très bien dans la pratique. Elles conduisent à des estimateurs asymptotiquement biaisés et par conséquent non convergents (Lin et Breslow (1996)). Mais néanmoins, contrairement à l'approche GEE du chapitre précédent, l'approche conditionnelle permet non seulement de faire des inférences sur les coefficients de régression mais aussi d'effectuer des prévisions pour un individu donné et ceci grâce à la modélisation de la corrélation (voir section 3.5.3).

3.5.2 Tests d'hypothèses

Considérons le test d'hypothèse de la forme

$$H_0 : \mathbf{L}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix} = \mathbf{d},$$

où $\mathbf{L}$ est une matrice de dimension $k \times (p+r)$, avec k le nombre d'hypothèses à tester et où $\mathbf{d}$ est un vecteur de dimension $k \times 1$. La statistique utilisée pour faire ce test est

$$F = \frac{\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{pmatrix}' \mathbf{L}'(\mathbf{L}\mathbf{C}^{-}\mathbf{L}')^{-1}\mathbf{L}\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{pmatrix} - \mathbf{d}}{\text{rang}(\mathbf{L})},$$

où

$$\mathbf{C} = \begin{pmatrix}\boldsymbol{X}_i'\mathbf{V}^{-1}\boldsymbol{X}_i & \boldsymbol{X}_i'\mathbf{V}^{-1}\boldsymbol{Z}_i\\ \boldsymbol{Z}_i'\mathbf{V}^{-1}\boldsymbol{X}_i & \boldsymbol{Z}_i'\mathbf{V}^{-1}\boldsymbol{Z}_i + \mathbf{D}^{-1}\end{pmatrix}$$

et pour la forme de la matrice $\mathbf{V}$, se référer à la sous-section précédente (3.5.1), algorithme de Schall(1991). $\mathbf{C}^{-}$ représente un inverse généralisé de $\mathbf{C}$.
Sous H_0, F suit une loi de Fisher avec ν_1 degrés de liberté au numérateur et ν_2 degrés de liberté au dénominateur. On a que ν_1 est le rang de la matrice $\mathbf{L}$, alors que ν_2 est obtenu par une approximation comme celle de Satterthwaite. Tout d'abord, on diagonalise la

matrice $\mathbf{LC}^{-}\mathbf{L}' = \mathbf{U}'\mathbf{QU}$, avec $\mathbf{U}$ une matrice orthogonale des vecteurs propres de la matrice $\mathbf{LC}^{-}\mathbf{L}'$, et $\mathbf{Q}$ une matrice diagonale contenant les ν_1 valeurs propres de la matrice $\mathbf{LC}^{-}\mathbf{L}'$. Les matrices $\mathbf{U}$ et $\mathbf{Q}$ sont toutes deux de dimensions $\nu_1 \times \nu_1$. Définissons maintenant $\mathbf{b_j}$ comme étant la j^{eme} ligne de $\mathbf{UL}$. Posons

$$\nu_j = \frac{2(Q_j)^2}{\boldsymbol{g'}_j \mathbf{A} \boldsymbol{g'}_j},$$

où Q_j est la j^{eme} élément sur la diagonale de $\mathbf{Q}$ et $\boldsymbol{g}_j$ est le gradient de $\mathbf{b_j C^{-} b'_j}$ par rapport à $\boldsymbol{\theta}$ et évalué en $\hat{\boldsymbol{\theta}}$ (avec $\boldsymbol{\theta}$ un vecteur de dimension $q \times 1$ contenant les paramètres inconnus des matrices $\mathbf{D}$ et $\boldsymbol{R}$). $\mathbf{A}$ est la matrice de variance-covariance de $\hat{\boldsymbol{\theta}}$. Posons finalement

$$E = \sum_{j=1}^{\text{rang}(\mathbf{L})} \frac{\nu_j}{\nu_j - 2} I(\nu_j > 2),$$

où

$$I(\nu_j > 2) = \begin{cases} 1 & \text{si } \nu_j > 2 \\ 0 & \text{sinon.} \end{cases}$$

Les degrés de liberté ν_2 sont donc

$$\nu_2 = \begin{cases} 2E/(E - \text{rang}(\mathbf{L})) & \text{si } E > \text{rang}(\mathbf{L}) \\ 0 & \text{sinon.} \end{cases}$$

3.5.3 Prévisions

Quand on parle d'une prévision pour la valeur moyenne d'un individu, on parle en fait d'une prévision de $E(Y|\boldsymbol{x}, \boldsymbol{\gamma})$ pour une valeur donnée de $\boldsymbol{x}$ et $\boldsymbol{z}$ étant donné les valeurs des Y_{ij} pour cet individu. Ceci est donné par $g^{-1}(\boldsymbol{x}'\hat{\boldsymbol{\beta}} + \boldsymbol{z}'\hat{\boldsymbol{\gamma}})$, où $g(\cdot)$ est une fonction de lien connue. Par exemple, dans un modèle linéaire généralisé mixte logistique ($Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\gamma}_{0i}, \boldsymbol{\gamma}_{1i}$; binomiale, lien logit), alors si nous désirons estimer la probabilité qu'un individu du groupe i avec variable exogène $\boldsymbol{x}$ ait $Y = 1$, on calcule tout simplement

$$\frac{\exp\left\{\hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\gamma}}_{0i} + (\hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\gamma}}_{1i})\boldsymbol{x}\right\}}{1 + \exp\left\{\hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\gamma}}_{0i} + (\hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\gamma}}_{1i})\boldsymbol{x}\right\}}.$$

Les intervalles de confiances pour un paramètre individuel ou une combinaison de paramètres peuvent être faits à l'aide de la méthode de Wald. On utilise le fait que les estimateurs sont approximativemnet de distribution normale et sans biais et avec

variance donnée en fonction de l'inverse de la matrice d'information. Ainsi pour un paramètre individuel β_j, on a

$$\frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{\text{Var}(\hat{\beta}_j)}} \approx \mathcal{N}(0,1).$$

Ou bien, pour une combinaison linéaire de paramètres,

$$K'\hat{\boldsymbol{\beta}} - K'\boldsymbol{\beta}_0 \approx \mathcal{N}(\mathbf{0}, K'\mathcal{I}^{-1}K),$$

où $\mathcal{I}$ est la matrice d'information.

Chapitre 4

Exemple d'application : Analyse des données sur les bisons

4.1 Introduction

L'objectif de cette analyse est d'étudier la distribution de la présence en des points précis de la population de bisons des plaines du Parc National de Prince Albert (53°44′N, 106°40′W), Saskatchewan (Canada), en fontion des attributs spatiaux du paysage. Pour ce faire, les méthodes décrites aux chapitres précédents seront appliquées à une base de données réelle.

Dans ce chapitre, nous allons tout d'abord présenter le milieu d'étude et parler de la collecte de données. Nous expliquerons les différentes variables utilisées, ainsi que les méthodes de l'analyse pour enfin terminer par la présentation et l'interprétation des résultats.

4.2 Présentation du milieu et collecte de données

Dans le parc, le paysage est couvert à 85% de forêts, et cette matrice de forêts est entremêlée de prés (10%), de lacs et de rivières (5%) (Fortin et al., 2003). Les régions agricoles se trouvant à côté du parc sont potentiellement accessibles aux bisons. Cette étude de choix d'habitat se base essentiellement sur la localisation de neuf femelles bisons équipées de radio-colliers avec système de positionnement global (GPS). Les

bisons sont localisés chaque heure, deux jours consécutifs par semaine pendant trois mois, du 2 septembre au 2 décembre. Pour modéliser la probabilité de Présence/Absence des bisons en un endroit donné, une variable binomiale Y a été créée. Cette variable prend la valeur 1 si le bison a visité l'endroit et 0 sinon. Ainsi, un modèle linéaire généralisé mixte logistique de la forme suivante est utilisé afin de faire cette analyse :

$$\text{logit}\left(\mathrm{P}[Y_{ij} = 1|\boldsymbol{x}_{ij}, \gamma_{0i}]\right) = (\beta_0 + \gamma_{0i}) + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_p x_{ijp},$$

où i représente le numéro du bison et j celui de l'observation, $\gamma_{0i} \sim \mathcal{N}(0, \sigma^2)$ représente l'effet aléatoire dû au bison i que l'on ne peut observer, permettant de modéliser la corrélation des observations prises pour ce bison.

On dispose de deux types de fichiers de données avec les mêmes covariables. Un premier, dénommé "Observed" (des sites visités par les bisons), contenant évidemment une colonne où Y vaut partout 1 et une colonne "Animal" signalant le code d'identification du bison. Un deuxième fichier, dénommé "Random" (des sites non visités par les bisons), avec une colonne où Y vaut partout 0 signalant l'absence de bisons. À la recommandation du chercheur, pour définir le sujet i, chaque endroit visité par le bison i a été apparié au fichier Random (endroits non visités). Ceci a été fait autant de fois qu'on a de bisons, donc neuf fois (voir figure FIG.4.1 et tableau TAB.4.1 ci-dessous).

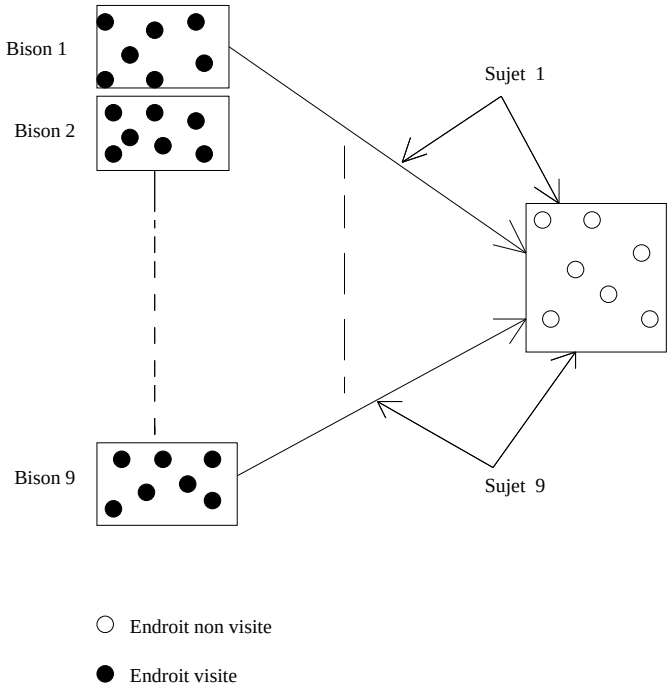

FIG. 4.1 – *Organisation schématique de la base de données*

TAB. 4.1 – *Représentation de la base de données*

Bisons	Y	Observations
1	1	$x_{111}, x_{112}, \cdots, x_{11p}$
1	1	$x_{121}, x_{122}, \cdots, x_{12p}$
⋮	⋮	⋮
1	1	$x_{1n_i1}, x_{1n_i2}, \cdots, x_{1n_ip}$
1	0	Ici, on a les lignes du fichier "Random"(Y=0)
⋮	⋮	avec les mêmes covariables. Ces
1	0	lignes sont les mêmes pour tous les bisons.
2	1	$x_{211}, x_{212}, \cdots, x_{21p}$
2	1	$x_{221}, x_{222}, \cdots, x_{22p}$
⋮	⋮	⋮
2	1	$x_{2n_i1}, x_{2n_i2}, \cdots, x_{2n_ip}$
2	0	Ici, on a les lignes du fichier "Random"(Y=0)
⋮	⋮	avec les mêmes covariables. Ces
2	0	lignes sont les mêmes pour tous les bisons.
3	1	$x_{311}, x_{312}, \cdots, x_{31p}$
3	1	$x_{321}, x_{322}, \cdots, x_{32p}$
⋮	⋮	⋮
3	1	$x_{3n_i1}, x_{3n_i2}, \cdots, x_{3n_ip}$
3	0	Ici, on a les lignes du fichier "Random"(Y=0)
⋮	⋮	avec les mêmes covariables. Ces
3	0	lignes sont les mêmes pour tous les bisons.
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
9	1	$x_{911}, x_{912}, \cdots, x_{91p}$
9	1	$x_{921}, x_{922}, \cdots, x_{92p}$
⋮	⋮	⋮
9	1	$x_{9n_i1}, x_{9n_i2}, \cdots, x_{9n_ip}$
9	0	Ici, on a les lignes du fichier "Random"(Y=0)
⋮	⋮	avec les mêmes covariables. Ces
9	0	lignes sont les mêmes pour tous les bisons.

4.3 Explication des variables

La variable réponse Y a été bien décrite dans la section précédente. Cependant, pour faire cette étude, nous nous sommes servis de pas mal de variables explicatives, généralement liées aux aspects spatiaux du paysage et de ses multiples attributs pour chacun des endroits des fichiers "Observed" et "Random". Dans la description de certaines des variables, on utilise le concept de zone tampon. Ainsi, la j^{eme} zone tampon du i^{eme} bison est le rayon de 300m autour du point où le bison i se trouve lors de l'observation j. Voici une liste des variables considérées et leur signification :

DSTMD : Distance à un pré.

dstrd : Distance à une route.

DSTWAT : Distance à l'eau.

HABVAR300 : Le nombre de types d'habitats différents dans la zone tampon.

PROPAG300 : Proportion de la zone tampon d'agriculture.

PROPMD300 : Proportion de la zone tampon de pré.

PROPWAT300 : Proportion de la zone tampon d'eau.

EDGED300 : Longueur des arêtes entre forêt/non forêt/eau.

sumrd : Densité d'une route.

Basé sur une image satellite de Landsat TM dans le système d'information géographique, une variable TYPE D'HABITATION a été classifiée en sept modalités qui sont

conif : Stands de conifères.

decid : Stands à feuilles caduques.

road : Routes.

agric : Régions agricoles.

meadow : Prés.

riparian : Secteurs ripicoles.

water : Eau

4.4 Méthodes

Hormis la variable de base water, les six autres variables caractérisant la variable à sept modalités TYPE D'HABITATION seront forcées à demeurer dans le modèle, alors que les autres variables seront sélectionnées par la méthode d'élimination selon deux seuils différents : 5% et 1%. Dans cette étude la procédure GLIMMIX de SAS sera utilisée. La procédure GLIMMIX utilise la quasi-vraisemblance pénalisée (PQL) pour l'estimation des paramètres. Cette procédure offre beaucoup de possibilités avec ses nombreux énoncés et options (voir l'aide de SAS sur GLIMMIX), entre autres l'énoncé RANDOM sera utilisé afin d'obtenir les effets aléatoires sur l'ordonnée à l'origine. Dans cette présente analyse, la matrice des covariances pour le bison i est de dimension 1. Cette matrice contient un seul élément σ^2 qui est la variance des γ_{0i}.

L'idée dans un premier temps est de faire la sélection de modèle à l'aide de la procédure GLIMMIX de SAS basée sur les critères d'Akaike (AIC) et de Schwartz (BIC), aussi bien que sur les p-values liées aux statistiques de Fisher. Les AIC des différents sous-modèles sont comparés afin de choisir le meilleur modèle qui sera ajusté aux données et ensuite de confirmer ce choix par la comparaison des BIC. Les expressions des critères de décision énoncés ci-dessus sont les suivantes :

$$AIC = -2\ell + 2d$$

$$BIC = -2\ell + d\ln(n),$$

où ℓ est la valeur maximale du logarithme de la vraisemblance ou de la quasi-vraisemblance, d est le nombre de paramètres du modèle et n est le nombre d'observations. Plus petite est la valeur du critère, meilleur est le modèle.

Malheureusement, ceci n'a pas été fait dans la présente analyse , parce que les sorties de la procédure GLIMMIX ne donnent ni AIC, ni BIC pour notre cas car ils sont invalides. Pourquoi ? GLIMMIX utilise la méthode PQL et comme on a un effet aléatoire, la linéarisation de la log-vraisemblance ne serait pas la même dans tous les sous-modèles et par conséquent les AIC ne seraient pas comparables. Ainsi, on a fait

une sélection de modèles basée uniquement sur les p-values liées aux statistiques de Fisher selon deux seuils d'élimination différents : 5% et 1%. En plus des variables caractérisant la variable TYPE D'HABITATION qui sont forcées à demeurer dans le modèle, les autres variables sélectionnées dans le modèle final restent les mêmes selon les deux seuils d'élimination et sont les suivantes :
DSTMD, DSTWAT, HABVAR300, PROPWAT300, et EDGED300.
Cependant, cette méthode reste peu efficace car elle n'est pas à mesure de comparer tous les sous-modèles possibles.

Pour contourner ce problème, on a essayé de faire l'ajustement avec la procédure NLMIXED de SAS qui calcule la vraisemblance par approximation de Laplace pour l'estimation des paramètres. Les sorties montrent bien les AIC et les BIC des modèles mais les résultats obtenus sont erronés (paramètres mal estimés, des points à la place de certains estimateurs de paramètres et de certaines p-values, etc.), et donc qui suggère que l'algorithme n'a pas convergé. Ce résultat était cependant prévisible parce que dans notre base de données on a peu d'individus et beaucoup d'observations par individu alors que NLMIXED est meilleure si on a le schéma inverse, c'est-à-dire le cas où l'on a beaucoup d'individus et peu d'observations par individu.

Une dernière tentative est de faire la sélection de modèle à l'aide de la procédure GENMOD de SAS avec l'énoncé REPEATED. La procédure GENMOD utilise la méthode GEE développée au chapitre 2 pour l'estimation des paramètres. La sélection de modèles a été basée sur deux seuils d'élimination différents : 5% et 1%. Ainsi, en plus des variables caractérisant la variable TYPE D'HABITATION qui sont forcées à demeurer dans le modèle, les variables sélectionnées dans le modèle final restent les mêmes selon les deux seuils d'élimination et sont les suivantes :
DSTMD, DSTWAT, HABVAR300, PROPWAT300, et EDGED300.

On remarque qu'on tombe sur le même modèle final obtenu avec la sélection faite par GLIMMIX basée uniquement sur les p-values liées aux statistiques de Fisher selon deux seuils d'élimination de 5% et de 1%. Et finalement, comme on veut modéliser la corrélation, le modèle final qui sera choisi est celui ajusté à l'aide de GLIMMIX. Les résultats sont présentés et interprétés dans la section suivante.

4.5 Présentation et interprétation des résultats

Dans cette section nous allons d'une part présenter les résultats de l'analyse, et d'autre part interpréter ces résultats selon un seuil de signification $\alpha = 5\%$. Les résultats

de l'analyse sont présentés dans le tableau suivant :

TAB. 4.2 – *sortie GLIMMIX*

Solutions for Fixed Effects			
Effect	Estimate	Standard Error	Seuils observés
Intercept	2.0480	0.1543	<.0001
DSTMD	-0.00751	0.000257	<.0001
DSTWAT	-0.00175	0.000073	<.0001
HABVAR300	-0.3255	0.008384	<.0001
PROPWAT300	-4.0749	0.2704	<.0001
EDGED300	91.8079	3.0475	<.0001
conif	-2.0349	0.1121	<.0001
decid	0.2389	0.1030	0.0203
road	0.1823	0.1502	0.2249
agric	-2.2918	0.1498	<.0001
meadow	0.2129	0.1061	0.0447
riparian	-2.0690	0.4134	<.0001

Les résultats dans le tableau ci-dessus indiquent que la distribution des bisons est liée aux multiples attributs du paysage. La probabilité de l'occurence du bison a augmenté dans les secteurs à proximité de prés (coefficient estimé à -0.00751, p< .0001) et d'eau (coefficient estimé à -0.00175, p< .0001). Les bisons ont également choisi les secteurs où le nombre de types d'habitat différents est faible (coefficient estimé à -0.3255, p< .0001), les endroits entourés par une petite proportion de la zone tampon en eau (coefficient estimé à -4.0749, p< .0001), ainsi que les secteurs où la longueur des arêtes entre forêt/non forêt/eau est grande (coefficient estimé à 91.8079, p< .0001). Relativement à l'eau, ils semblent préférer les stands à feuilles caduques (coefficient estimé à 0.2389, p=0.0203) et les prés (coefficient estimé à 0.2129, p=0.0447). Par contre, les bisons semblent éviter les stands de conifères (coefficient estimé à -2.0349, p< .0001), les régions agricoles (coefficient estimé à -2.2918, p< .0001), ainsi que les secteurs ripicoles (coefficient estimé à -2.0690, p< .0001) par rapport à l'eau. Cependant, il semble qu'ils aiment autant les routes (p=0.2249) que l'eau. La variance de la distribution des γ_{0i} est estimée dans notre présente étude à 0.05660.

Bien que les interprétations des effets données ci-dessus soient valides, il est à noter que les coefficients de régression représentent l'effet d'une hausse d'un facteur lorsque tous les autres facteurs demeurent inchangés. Comme en pratique les variables mesurées pour décrire les types d'habitats sont corrélées, il est plutôt rare que la valeur d'un facteur change et que celle des autres demeure inchangée lorsque nous passons d'un point dans l'habitat à un autre. Il faut donc interpréter les coefficients obtenus avec une certaine précaution.

Le code décrivant les différentes procédures SAS qui ont servi à faire cette analyse, ainsi que les manipulations faites dans les fichiers de base, est présenté dans l'Annexe A. Les différentes étapes de la procédure d'élimination à l'aide de GENMOD ont été aussi présentées dans l'Annexe A.

Chapitre 5

Conclusion

Dans cet essai, nous avons parcouru d'une manière globale les modèles linéaires généralisés pour données corrélées tant sur le point de vue marginal que conditionnel. Ainsi, nous nous sommes rendu compte durant ce travail de l'importance de ce type de modélisation, surtout dans le domaine de la santé (l'exemple des traitements contre l'épilepsie vu dans l'introduction) mais aussi dans bien d'autres situations encore comme en biologie animale (l'exemple d'application de l'essai).

Dans le chapitre 2, sous le modèle marginal, nous avons utilisé l'approche par équations d'estimation généralisées (GEE) pour estimer les paramètres du modèle. Les GEE utilisent des estimateurs robustes pour l'estimation des paramètres du modèle et de leur matrice de variance-covariance du fait que la matrice des corrélations de *travail* peut être possiblement fausse. Et comme dans cette matrice on trouve certains paramètres inconnus, nous avons spécifié une liste de ses formes les plus communes et donné l'estimateur pour chacune des formes. Cependant, dans le livre de Hardin et Hilbe (2002), beaucoup d'autres formes de matrices des corrélations de *travail* ont été spécifiées, ainsi que leurs méthodes d'estimation respectives.

Au chapitre 3, nous avons introduit des effets aléatoires dans le modèle linéaire généralisé de façon à modéliser la corrélation dans le but de pouvoir faire des inférences sur un individu dans un panel donné. Nous avons de plus développé dans ce chapitre des méthodes de résolution des équations de vraisemblance afin de trouver une estimation des paramètres du modèle.

Dans le chapitre 4, nous avons appliqué les méthodes développées aux chapitres 2 et 3 à des données sur les bisons. Cependant, l'analyse reste encore à améliorer. En effet, au lieu d'utiliser un seul facteur aléatoire dans le modèle, on aurait pu essayer d'en

introduire plusieurs. On aurait également désirer obtenir les critères AIC et BIC avec GLIMMIX et par conséquent la sélection de modèle serait beaucoup plus simple et peut-être une meilleure prévision des variables étudiées. Pour de plus amples informations sur cette étude, consulter l'article de Craiu, Duchesne, et Fortin (2006).

Dans cet essai, nous n'avons pas fait d'étude théorique sur les méthodes de validation des hypothèses et de la qualité de l'ajustement d'un modèle linéaire généralisé mixte (GLMM). Ces méthodes peu développées dans les travaux concernant les GLMM, pourraient être le sujet principal d'une étude à elles seules. Cependant, certains auteurs ont commencé à s'y intéresser (Ritz(2004) et Waagep(2005)). Une poursuite intéressante pourrait donc être d'étudier de façon plus appronfondie ces méthodes dans notre cadre précis.

Bibliographie

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

[2] Craiu R.V., Duchesne T., & Fortin D. (2006). *Generalized Estimating Equations and Model Selection for Longitudinal Conditional Logistic Regression*. Article non publié.

[3] Diggle P.J., Liang K.Y., & Zeger S.L. (1994). *Analysis of Longitudinal Data*. Oxford University Press, Oxford, UK.

[4] Dobson A.J. (1990). *An Introduction to Generalized Linear Models*. Chapman & Hall, London.

[5] Fay M.P., & Graubard B.I. (2001). *Small-Sample Adjustments for Wald-Type Tests Using Sandwich Estimators*. Biometrics, 57, 1198-1206.

[6] Hardin, J.W., & Hilbe J.M. (2002). *Generaralized Estimating Equations*. Chapman & Hall/CRC, Boca Raton, Florida 33431.

[7] Liang K.Y., & Zeger S.L. (1986). *Longitudinal Data Analysis Using Generalized linear Models*. Biometrika, 73(1), 13-22.

[8] Lin X., & Breslow N.E. (1996). *Bias correction in Generalized Linear Mixed Models with Multiple Components of Dispersion*. Journal of the American Statistical Association, Vol. 91.

[9] McCullagh P., & Nelder J.A. (1989). *Generalized linear Models*. Chapman & Hall, New York.

[10] McCulloch C.E., & Searle S.R. (2001). *Generalized, Linear, and Mixed Models*. Wiley, New York.

[11] Ritz C. (2004). *Goodness-of-fit tests for Mixed Models*. Scandinavian Journal of Statistics, 31, 443-458.

[12] SAS Institute Inc. (2004). *Help and Documentation*. Cary, NC, USA.

[13] Trottier C. (1998). *Estimation dans les Modèles Linéaires Généralisés à Effets Aléatoires*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, Grenoble.

[14] Waagep.etersen, R. (2005). *A Simulation-based Goodness-of-fit test for Random Effects in Generalized linear Models*. Document non publié.

[15] Wolfinger R., & O'Connell R. (1993). *Generalized Linear Mixed Models : A Pseudo-likelihood Approach*. Journal of Statistical Computation and Simulation, 48, 233-243.

[16] Zeger S.L., & Liang K.Y. (1986). *Longitudinal Data Analysis for Dicrete and Continuous Outcomes*. Biometrics, 42(1), 121-130.

[17] Zheng B. (2000). *Summarizing the Goodness of Fit of Generalized Linear Models for Longitudinal Data*. Statistics in Medicine, 19(10), 1265-1275.

[18] Zorn C.J.W. (2001). *Generaralized Estimating Equation Models for Correlated Data : A Review with Applications*. American Journal of Political Science, 45(2), 470-490.

Annexe A

Annexe A

A.1 Sélection de modèle à l'aide de la procédure GENMOD de SAS.

Hormis la variable de base water, les six autres variables caractérisant le type d'habitat sont forcées à demeurer dans le modèle, alors que les autres variables seront sélectionnées par la méthode d'élimination selon deux seuils différents : 5% et 1%.

Étape 1 : Modèle complet.

Le modèle converge mais on constate les erreurs de compilation suivantes dans le log de SAS :

```
ERROR: Error in computing the variance function.
ERROR: Error in parameter estimate covariance computation.
ERROR: Error in estimation routine.
```

J'ai essayé d'enlever pas mal de variables dans le modèle mais on constate toujours des erreurs de compilation et même des fois on n'avait pas la convergence de certains modèles. Par chance, je suis arrivé à enlever la variable PROPAG300 du modèle et tout devient normal et les résultats obtenus sont présentés dans l'étape suivante :

Étape 2 : Modèle sans la variable PROPAG300.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	47E3	21653.6337	0.4631
Scaled Deviance	47E3	21653.6337	0.4631
Pearson Chi-Square	47E3	53014.3856	1.1338
Scaled Pearson X2	47E3	53014.3856	1.1338
Log Likelihood		-10826.8169	

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	$Pr > \|Z\|$
Intercept	2.1092	0.2450	1.6291	2.5894	8.61	<.0001
DSTMD	-0.0078	0.0005	-0.0088	-0.0069	-15.62	<.0001
dstrd	-0.0001	0.0000	-0.0002	0.0000	-1.59	0.1108
DSTWAT	-0.0018	0.0003	-0.0023	-0.0013	-6.96	<.0001
HABVAR300	-0.3188	0.0233	-0.3645	-0.2731	-13.68	<.0001
PROPMD300	0.1445	0.3671	-0.5751	0.8640	0.39	0.6940
PROPWAT300	-4.1432	0.4438	-5.0130	-3.2733	-9.34	<.0001
EDGED300	89.0097	7.2446	74.8104	103.2089	12.29	<.0001
sumrd	-0.0020	0.0063	-0.0144	0.0104	-0.32	0.7525
conif	-2.1152	0.2068	-2.5206	-1.7098	-10.23	<.0001
decid	0.2371	0.1535	-0.0637	0.5379	1.54	0.1224
road	0.1272	0.1935	-0.2521	0.5064	0.66	0.5111
agric	-2.4460	0.9559	-4.3195	-0.5725	-2.56	0.0105
meadow	0.1580	0.1722	-0.1796	0.4955	0.92	0.3590
riparian	-2.1171	0.5456	-3.1865	-1.0477	-3.88	0.0001

Étape 3 : Modèle sans les variables PROPAG300 et sumrd.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	47E3	21655.0961	0.4631
Scaled Deviance	47E3	21655.0961	0.4631
Pearson Chi-Square	47E3	53164.2396	1.1370
Scaled Pearson X2	47E3	53164.2396	1.1370
Log Likelihood		-10827.5481	

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	$Pr > \|Z\|$
Intercept	2.1130	0.2484	1.6262	2.5998	8.51	<.0001
DSTMD	-0.0079	0.0005	-0.0088	-0.0070	-16.81	<.0001
dstrd	-0.0001	0.0000	-0.0002	0.0000	-1.39	0.1659
DSTWAT	-0.0018	0.0003	-0.0023	-0.0013	-7.09	<.0001
HABVAR300	-0.3193	0.0226	-0.3636	-0.2750	-14.12	<.0001
PROPMD300	0.1608	0.3440	-0.5135	0.8351	0.47	0.6403
PROPWAT300	-4.1257	0.4628	-5.0328	-3.2185	-8.91	<.0001
EDGED300	88.1827	8.6957	71.1394	105.2260	10.14	<.0001
conif	-2.1184	0.2078	-2.5257	-1.7111	-10.20	<.0001
decid	0.2375	0.1532	-0.0629	0.5378	1.55	0.1212
road	0.0900	0.2525	-0.4050	0.5849	0.36	0.7217
agric	-2.4502	0.9619	-4.3354	-0.5649	-2.55	0.0109
meadow	0.1559	0.1744	-0.1859	0.4978	0.89	0.3713
riparian	-2.1171	0.5471	-3.1894	-1.0449	-3.87	0.0001

Étape 4 : Modèle sans les variables PROPAG300, sumrd et PROPMD300.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	47E3	21656.6705	0.4631
Scaled Deviance	47E3	21656.6705	0.4631
Pearson Chi-Square	47E3	53538.8223	1.1449
Scaled Pearson X2	47E3	53538.8223	1.1449
Log Likelihood		-10828.3353	

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	$Pr > \|Z\|$
Intercept	2.1624	0.2330	1.7057	2.6191	9.28	<.0001
DSTMD	-0.0080	0.0005	-0.0089	-0.0070	-16.63	<.0001
dstrd	-0.0001	0.0001	-0.0002	0.0000	-1.38	0.1678
DSTWAT	-0.0018	0.0003	-0.0023	-0.0013	-7.10	<.0001
HABVAR300	-0.3206	0.0240	-0.3676	-0.2736	-13.37	<.0001
PROPWAT300	-4.1645	0.4670	-5.0798	-3.2492	-8.92	<.0001
EDGED300	88.9120	9.3210	70.6432	107.1807	9.54	<.0001
conif	-2.1432	0.1949	-2.5253	-1.7612	-11.00	<.0001
decid	0.2187	0.1325	-0.0409	0.4784	1.65	0.0987
road	0.0654	0.2151	-0.3561	0.4869	0.30	0.7610
agric	-2.4771	0.9604	-4.3594	-0.5948	-2.58	0.0099
meadow	0.1545	0.1728	-0.1842	0.4932	0.89	0.3714
riparian	-2.1459	0.5094	-3.1443	-1.1475	-4.21	<.0001

Étape 5 : Modèle sans les variables PROPAG300, sumrd, PROPMD300 et dstrd.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	47E3	21685.4125	0.4637
Scaled Deviance	47E3	21685.4125	0.4637
Pearson Chi-Square	47E3	54067.1197	1.1562
Scaled Pearson X2	47E3	54067.1197	1.1562
Log Likelihood		-10842.7062	

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	$Pr > \|Z\|$
Intercept	2.0527	0.2719	1.5198	2.5856	7.55	<.0001
DSTMD	-0.0079	0.0005	-0.0089	-0.0069	-15.39	<.0001
DSTWAT	-0.0018	0.0003	-0.0023	-0.0013	-6.80	<.0001
HABVAR300	-0.3305	0.0283	-0.3861	-0.2750	-11.66	<.0001
PROPWAT300	-4.1528	0.4642	-5.0626	-3.2430	-8.95	<.0001
EDGED300	93.1214	11.0727	71.4192	114.8236	8.41	<.0001
conif	-2.1740	0.1917	-2.5498	-1.7983	-11.34	<.0001
decid	0.2396	0.1448	-0.0441	0.5233	1.66	0.0979
road	0.1758	0.2768	-0.3667	0.7183	0.64	0.5253
agric	-2.4041	0.9893	-4.3432	-0.4650	-2.43	0.0151
meadow	0.1741	0.1889	-0.1960	0.5443	0.92	0.3565
riparian	-2.1167	0.5345	-3.1643	-1.0690	-3.96	<.0001

Ainsi termine la procédure d'élimination et nous concluons avec ce dernier modèle.

A.2 Programme SAS

```
/*****************************************
  Procédures d'importation des données
  d'EXCEL vers SAS.
*****************************************/

PROC IMPORT OUT= WORK.presence
            DATAFILE= "C:\Documents and
Settings\Proprietaire\Bureau\don nee1.xls"
            DBMS=EXCEL REPLACE;
     SHEET="Feuil1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;
PROC IMPORT OUT= WORK.PRESENCE2
            DATAFILE= "C:\Documents and
```

```
Settings\Proprietaire\Bureau\don nee1.xls"
            DBMS=EXCEL REPLACE;
     SHEET="Feuil2$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

/*************************************
Manipulation des fichiers de données
Observed et Random.
*************************************/

/*Création d'une colonne où la variable réponse
prend partout 1 dans le fichier Observed.*/

data observe;set presence;
 rep=1;
run;

/*On crée une colonne où la variable réponse
prend partout 0 dans le fichier Random et
une colonne Animal qui prend le numéro de
l'animal. Et ceci est répété neuf fois
 (autant de fois qu'on a de bisons).*/

data random1;set presence2;
 rep=0;
 Animal=511;
run;

data random2;set presence2;
 rep=0;
 Animal=512;
run;

data random3;set presence2;
 rep=0;
```

```
 Animal=513;
run;

data random4;set presence2;
 rep=0;
 Animal=514;
run;

data random5;set presence2;
 rep=0;
 Animal=515;
run;

data random6;set presence2;
 rep=0;
 Animal=516;
run;

data random7;set presence2;
 rep=0;
 Animal=517;
run;

data random8;set presence2;
 rep=0;
 Animal=518;
run;

data random9;set presence2;
 rep=0;
 Animal=519;
run;

/*****************************************
 Conception de la base de données qui a
 servi à faire l'analyse.
*****************************************/
/*On définit certaines variables comme dstrd,
 sumrd, road et type d'habitat selon certains
 critères recommandés par le chercheur.*/
```

```
data global; set observe random1 random2 random3
random4 random5 random6 random7 random8 random9;
dstrd = min(dstrd1,dstrd2);
sumrd = sum1rd300 + sum2rd300;
if panprlc < 112 then decid = 1;
if panprlc >= 112 then decid = 0;
if panprlc >= 112 and panprlc < 128 then conif = 1;
if panprlc < 112 or panprlc > 128 then conif = 0;
if panprlc >= 202 and panprlc < 214 then meadow = 1;
if panprlc < 202 or panprlc > 300 then meadow = 0;
if panprlc = 301 then water1 = 1;
if panprlc = 303 then water1 = 1;
if panprlc NE 301 and panprlc NE 303 then water1=0;
if panprlc = 320 then riparian=1;
if panprlc NE 320 then riparian=0;
if panprlc = 401 then agric=1;
if panprlc = 402 then agric=1;
if panprlc NE 401 and panprlc NE 402 then agric=0;
if panprlc = 411 then road1=1;
if panprlc NE 411 then road1=0;
if panprlc = 412 then road2=1;
if panprlc NE 412 then road2=0;
road = road1 + road2;
 drop;
run;

/*****************************
    procédures SAS utilisées.
*****************************/
/*Création d'une sortie graphique de type html.
 Sortie beaucoup plus lisible que la sortie
 standard de SAS.*/

ods html;
ods graphics on;

/*Procédure GLIMMIX*/

 proc glimmix data=global;
class Animal;
```

```
model rep = dstmd  dstwat  habvar300  propwat300
edged300  conif  decid road  agric  meadow
riparian/ solution dist=binomial link=logit;
random intercept/subject=Animal type=cs;
run;

/*Procédure NLMIXED: Les valeurs initiales
 des paramètres sont les estimés obtenus
 par GENMOD*/

proc nlmixed data=global;
parms B0=2.0527  B1=-0.0079  B2=-0.0018
B3=-0.3305 B4=-4.1528  B5=93.1214 B6=-2.1740
B7= 0.2396 B8=0.1758 B9=-2.4041  B10=0.1741
B11=-2.1167  SD=1;
eta = B0 + B1*dstmd + B2*dstwat + B3*habvar300
+ B4*propwat300 + B5*edged300 + B6*conif + B7*decid
+ B8*road + B9*agric + B10*meadow + B11*riparian + U;
p = exp(eta)/(1+exp(eta));
model rep ~ binomial(1,p);
random U ~ NORMAL(0,SD*SD) subject=Animal;
run;

/*Procédure GENMOD*/

proc genmod descending;
class Animal;
model rep = dstmd  dstwat habvar300  propwat300
edged300  conif  decid  road agric  meadow
riparian/dist=binomial link=logit;
repeated subject=Animal/ type=cs;
run;

ods graphics off;
ods html close;
```

Annexe B

Annexe B

B.1 Paramètre canonique et fonctions $a(\cdot)$ et $b(\cdot)$ caractérisant les lois usuelles de la famille exponentielle.

Le tableau suivant est extrait de la thèse de Trottier (1998). On peut également le retrouver dans l'aide de SAS sur GLIMMIX.

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

Pour chacune de ces lois, l'espérance et la variance de la variable associée s'expriment à l'aide des fonctions $a(\cdot)$ et $b(\cdot)$. Ces résultats sont démontrés dans la section suivante.

B.2 Démonstration des formules pour l'espérance et la variance d'une loi faisant partie de la famille exponentielle.

La fonction génératrice des moments d'une loi faisant partie de la famille exponentielle est

$$\begin{aligned}
M_Y(t) = E(e^{ty}) &= \int_{\Re} e^{ty+\frac{y\theta-b(\theta)}{a(\phi)}-c(y,\phi)}dy \\
&= e^{\frac{-b(\theta)}{a(\phi)}} \int_{\Re} e^{ty+\frac{y\theta}{a(\phi)}-c(y,\phi)}dy \\
&= \frac{e^{\frac{-b(\theta)}{a(\phi)}}}{e^{\frac{-b(a(\phi)t+\theta)}{a(\phi)}}} \int_{\Re} e^{\frac{a(\phi)ty+y\theta}{a(\phi)}-c(y,\phi)-\frac{b(a(\phi)t+\theta)}{a(\phi)}}dy \\
&= \frac{e^{\frac{-b(\theta)}{a(\phi)}}}{e^{\frac{-b(a(\phi)t+\theta)}{a(\phi)}}} \underbrace{\int_{\Re} e^{\frac{y(a(\phi)t+\theta)-b(a(\phi)t+\theta)}{a(\phi)}-c(y,\phi)}dy}_{=1} \\
&= e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}}.
\end{aligned}$$

L'espérance de $\mathbf{Y}$ est obtenue comme suit :

$$\begin{aligned}
E(Y) \;=\; M'_Y(t)|_{t=0} \qquad &= \left(e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}}\right)' \Bigg|_{t=0} \\
&= e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}} \left(\frac{b'(a(\phi)t+\theta)a(\phi)}{a(\phi)}\right) \Bigg|_{t=0} \\
&= e^{\frac{-b(\theta)+b(\theta)}{a(\phi)}} b'(\theta) \\
&= b'(\theta).
\end{aligned}$$

La variance de $\mathbf{Y}$ est obtenue ainsi :

$$Var(Y) \;= (M''_Y(t) - (M'_Y(t))^2)|_{t=0}.$$

Or,

$$\begin{aligned}
M''_Y(t) &= \left(e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}}\right)'' \Bigg|_{t=0} \\
&= e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}} \left\{(b'(a(\phi)t+\theta))^2+b''(a(\phi)t+\theta)a(\phi)\right\}\Big|_{t=0} \\
&= e^{\frac{-b(\theta)+b(\theta)}{a(\phi)}} \left\{(b'(\theta))^2 + b''(\theta)a(\phi)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= (b'(\theta))^2 + b''(\theta)a(\phi) \\
Var(Y) &= (b'(\theta))^2 + b''(\theta)a(\phi) - (b'(\theta))^2 \\
&= b''(\theta)a(\phi).
\end{aligned}
$$