

Université de Montréal

**Approximation de la distribution a posteriori d'un modèle
Gamma-Poisson hiérarchique à effets mixtes**

par

Annick Joëlle Nembot Simo

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistiques

janvier 2011

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Approximation de la distribution a posteriori d'un modèle
Gamma-Poisson hiérarchique à effets mixtes**

présenté par

Annick Joëlle Nembot Simo

a été évalué par un jury composé des personnes suivantes :

Mylène Bédard

(président-rapporteur)

Alejandro Murua

(directeur de recherche)

Pierre Duchesne

(membre du jury)

Mémoire accepté le:

10 Janvier 2011

SOMMAIRE

La méthode que nous présentons pour modéliser des données dites de "comptage" ou données de Poisson est basée sur la procédure nommée Modélisation multi-niveau et interactive de la régression de Poisson (PRIMM) développée par Christiansen et Morris (1997). Dans la méthode PRIMM, la régression de Poisson ne comprend que des effets fixes tandis que notre modèle intègre en plus des effets aléatoires. De même que Christiansen et Morris (1997), le modèle étudié consiste à faire de l'inférence basée sur des approximations analytiques des distributions a posteriori des paramètres, évitant ainsi d'utiliser des méthodes computationnelles comme les méthodes de Monte Carlo par chaînes de Markov (MCMC). Les approximations sont basées sur la méthode de Laplace et la théorie asymptotique liée à l'approximation normale pour les lois a posteriori. L'estimation des paramètres de la régression de Poisson est faite par la maximisation de leur densité a posteriori via l'algorithme de Newton-Raphson. Cette étude détermine également les deux premiers moments a posteriori des paramètres de la loi de Poisson dont la distribution a posteriori de chacun d'eux est approximativement une loi gamma. Des applications sur deux exemples de données ont permis de vérifier que ce modèle peut être considéré dans une certaine mesure comme une généralisation de la méthode PRIMM. En effet, le modèle s'applique aussi bien aux données de Poisson non stratifiées qu'aux données stratifiées ; et dans ce dernier cas, il comporte non seulement des effets fixes mais aussi des effets aléatoires liés aux strates. Enfin, le modèle est appliqué aux données relatives à plusieurs types d'effets indésirables observés chez les participants d'un essai clinique impliquant un vaccin quadrivalent contre la rougeole, les oreillons, la rubéole et la varicelle. La régression de Poisson comprend l'effet fixe correspondant à la variable traitement/contrôle, ainsi que des effets aléatoires liés aux systèmes biologiques du corps humain auxquels sont attribués les effets indésirables considérés.

Mots clés : Méthodes bayésiennes approximatives, Méthode de Laplace, Modèles à effets mixtes, Modèle hiérarchique, Régression de Poisson.

SUMMARY

We propose a method for analysing count or Poisson data based on the procedure called Poisson Regression Interactive Multilevel Modeling (PRIMM) introduced by Christiansen and Morris (1997). The Poisson regression in the PRIMM method has fixed effects only, whereas our model incorporates random effects. As well as Christiansen and Morris (1997), the model studied aims at doing inference based on adequate analytical approximations of posterior distributions of the parameters. This avoids the use of computationally expensive methods such as Markov chain Monte Carlo (MCMC) methods. The approximations are based on the Laplace's method and asymptotic theory. Estimates of Poisson mixed effects regression parameters are obtained through the maximization of their joint posterior density via the Newton-Raphson algorithm. This study also provides the first two posterior moments of the Poisson parameters involved. The posterior distribution of these parameters is approximated by a gamma distribution. Applications to two datasets show that our model can be somehow considered as a generalization of the PRIMM method since it also allows clustered count data. Finally, the model is applied to data involving many types of adverse events recorded by the participants of a drug clinical trial which involved a quadrivalent vaccine containing measles, mumps, rubella and varicella. The Poisson regression incorporates the fixed effect corresponding to the covariate treatment/control as well as a random effect associated with the biological system of the body affected by the adverse events.

Keys words : Approximate bayesian methods, Hierarchical model, Laplace's method, Mixed effects models, Poisson regression.

TABLE DES MATIÈRES

Sommaire	iii
Summary	v
Liste des figures	viii
Liste des tableaux	ix
Remerciements	1
Introduction	2
Chapitre 1. Description du modèle original de Christiansen et Morris (1997) ..	6
1.1. Modèle descriptif	6
1.2. Modèle inférentiel	7
1.3. Distributions des hyperparamètres (β, ξ)	8
1.4. Distributions des paramètres λ_i	10
Chapitre 2. Présentation du modèle	13
2.1. Modèle descriptif	13
2.2. Modèle inférentiel	15
2.3. Fonction de vraisemblance	16
2.4. Distribution des hyperparamètres	19
2.4.1. Dérivées de $\ell(\theta)$ par rapport aux hyperparamètres	23
2.4.2. Dérivées de $\ell_R(\theta)$ par rapport aux hyperparamètres	27
2.5. Algorithme pour l'estimation des hyperparamètres	29

2.6. Distribution des paramètres individuels $\{\lambda_{ij}\}$	31
Chapitre 3. Comparaison de résultats avec les données de Christiansen et Morris (1997)	36
3.1. Exemple 1 : Taux de panne des pompes	36
3.2. Exemple 2 : Données de transplantation du coeur	37
Chapitre 4. Application aux données sur les effets indésirables d'un essai clinique	39
4.1. Présentation des données	39
4.2. Résultats de la modélisation.....	41
Conclusion	45
Bibliographie	47
Annexe A. Code R du programme	1

LISTE DES FIGURES

2.1	Représentation graphique de la fonction g	21
-----	---	----

LISTE DES TABLEAUX

3.1	Résultats de l'analyse des données sur les pannes des pompes par l'utilisation de deux modèles.	37
3.2	Résultats de l'analyse des données de transplantation du coeur par l'utilisation de deux modèles.	38
4.1	Résultats d'un essai clinique sur les effets indésirables(EIs).	40
4.2	Résultats d'un essai clinique sur les effets indésirables(EIs). (<i>Suite</i>).....	41
4.3	Estimateurs des paramètres $\beta = (\beta_0, \beta_1)'$ et $\xi_j(j = 1, \dots, p)$ résultants de l'analyse des données sur les EIs.	42
4.4	Estimateurs des B_{ij} , λ_{ij} et $\sigma_{\lambda_{ij}}^2$ selon le groupe (T=traitement, C=contrôle) résultant de l'analyse des données sur les EIs.	43
4.5	Estimateurs des B_{ij} , λ_{ij} et $\sigma_{\lambda_{ij}}^2$ selon le groupe (T=traitement, C=contrôle) résultant de l'analyse des données sur les EIs. (<i>Suite</i>).....	44

REMERCIEMENTS

Je commencerais par adresser mes sincères remerciements à mon directeur de recherche, monsieur Alejandro Murua, qui a bien voulu superviser ce mémoire. Je le remercie pour son appui scientifique, ses commentaires pertinents et sa grande disponibilité tout au long de l'élaboration de ce travail.

Je remercie le département de mathématiques et statistique et la Faculté des Études Supérieures et Postdoctorales de l'Université de Montréal à travers lesquels des bourses m'ont été accordées au cours de ma formation.

Je tiens tout particulièrement à remercier mon époux Évariste pour son immense support, ses encouragements et dont la présence à mes côtés m'apporte bonheur et réconfort quotidiennement.

Je suis infiniment reconnaissante à l'endroit de mes parents, ma famille et mes amis qui m'ont soutenu durant ma formation et ont toujours cru en mes capacités de mener à bien ce que j'entreprends.

Et par-dessus tout, je rends gloire à L'Éternel Dieu pour son Amour infini et les grâces dont Il me comble tous les jours de ma vie.

INTRODUCTION

Le travail rapporté dans ce document présente une méthode d'analyse des données dites de "comptage" ou données de Poisson. Les variables de comptage décrivent le nombre de fois où un événement s'est produit pendant une certaine période, par exemple le nombre de patients décédés au cours d'une année dans un hôpital, et sont présentes dans plusieurs domaines tels que les sciences sociales ou encore la médecine. Les techniques de modélisation de telles données sont donc d'un grand intérêt dans les études statistiques.

Les modèles de Poisson ont largement été utilisés dans les analyses de régression des données de comptage. Holford (1980) utilise des modèles log-linéaires pour l'analyse de proportions (ou données de Poisson) et de taux de survie dont l'estimation des paramètres est faite à travers des algorithmes tel que l'ajustement proportionnel itératif (IPF). Frome (1983) considère des modèles de régression de Poisson incluant des modèles log-linéaires, quasi-linéaires et non linéaires. Les estimateurs des paramètres peuvent être obtenus par une méthode des moindres carrés pondérés itérative (IRLS). Il fait une application à travers un modèle non linéaire pour analyser la mortalité due au cancer du poumon parmi les médecins anglais fumant régulièrement de la cigarette.

Il est cependant reconnu que les données de comptage présentent souvent une variation extra-poissonnienne (ou surdispersion) considérable relative à un modèle de Poisson. Ainsi, de nombreux travaux ont développé des méthodes et applications statistiques pour le traitement des données de Poisson surdispersées. Albert (1992) analyse des données concernant le nombre de coups de circuit frappés par des joueurs de baseball à travers deux modèles hiérarchiques de Poisson. Le deuxième modèle est considéré comme étant le mieux approprié étant donné qu'il prend en compte le problème de surdispersion des données ; ceci est fait en utilisant le mélange Gamma-Poisson. Les moments a posteriori

d'intérêt sont calculés par l'application de l'algorithme d'échantillonnage de Gibbs présenté par Gelfand et Smith (1990), ou encore Casella et George (1992).

Parmi ces études, il y en a qui intègrent des effets aléatoires dans la régression de Poisson. Vonesh (1990) présente une régression de Poisson à effets mixtes pour analyser les facteurs de risque associés à la péritonite, une infection bactérienne du péritoine. Le modèle qu'il propose comprend des effets fixes correspondant à l'information existante collectée à travers les individus, ainsi que des effets aléatoires liés aux individus. La méthode du maximum de vraisemblance est utilisée pour estimer les paramètres du modèle. Il montre ensuite que le modèle à effets mixtes s'ajuste mieux aux données comparé à la régression de Poisson qui ignore les effets aléatoires liés aux individus. Tempelman R.J. et Gianola D.(1996) développent des modèles qui s'appliquent à la reproduction animale et qui prennent en compte l'extra-variabilité des données que le modèle de Poisson mixte usuel ne prend pas en compte. Une distribution gamma est assignée aux paramètres de Poisson conduisant à un modèle binomial négatif. Le logarithme de l'espérance du paramètre de Poisson est exprimé comme fonction linéaire d'effets polygéniques fixes et aléatoires. L'estimation de ces effets est faite par la méthode du maximum a posteriori (MAP) via l'algorithme de Newton-Raphson. Puis ils effectuent des simulations pour évaluer la qualité des estimateurs des composantes de la variance. Une application à l'analyse du nombre d'inséminations artificielles jusqu'à la conception chez des génisses laitières est ensuite présentée pour illustrer les différences potentielles dans l'inférence sur les composantes de la variance sous des modèles de Poisson intégrant et n'intégrant pas la surdispersion.

Le modèle que nous présentons dans cette étude est une extension de la procédure nommée Modélisation multi-niveau et interactive de la régression de Poisson (PRIMM) développée par Christiansen et Morris (1997). Ces auteurs ont démontré les avantages de leur approche par rapport à d'autres méthodes telles que les méthodes de modèle linéaire généralisé en présence de surdispersion ou encore les méthodes de Monte Carlo par chaînes de Markov (MCMC). En effet, contrairement à d'autres modèles de Poisson, la méthode PRIMM fournit les estimateurs et intervalles pour les paramètres de la loi de Poisson et les coefficients de la régression log-linéaire. De plus, bien que l'utilisation des méthodes MCMC fournit un certain nombre d'inférences, la rapidité d'implémentation

de la procédure PRIMM permet son évaluation à travers un usage répété de simulations, vérifiant ainsi qu'elle a de bonnes propriétés fréquentistes.

Dans la procédure PRIMM, la régression de Poisson ne comprend que des effets fixes. Nous avons pensé qu'il serait intéressant d'étendre leur modèle en y ajoutant des effets aléatoires (ou effets de groupes). En effet, les données dans plusieurs études apparaissent très souvent par groupes ou strates. Par exemple, considérons une étude réalisée sur les petits d'une portée pour un échantillon de souris. Les observations relevées sur les petits d'une même souris vont donc constituer une strate. Dans le cas d'une étude faite sur les patients d'un échantillon de cliniques, les observations provenant d'une clinique donnée sont considérées comme une strate. Étant donné que les observations à l'intérieur d'une strate tendent à être plus similaires que les observations provenant de différentes strates, les analyses qui ignorent ce fait pourraient conduire à des résultats erronés (Agresti (2002), page 491). Nous aurons donc dans notre analyse une régression de Poisson à effets mixtes comme dans certains travaux mentionnés précédemment, à la différence que la méthode d'estimation est basée sur la procédure PRIMM.

De même que Christiansen et Morris (1997), la méthode que nous développons consiste à faire de l'inférence basée sur des approximations analytiques des distributions a posteriori des paramètres, évitant ainsi d'utiliser des méthodes computationnelles comme les méthodes MCMC. Les approximations sont basées sur la méthode de Laplace et la théorie asymptotique. La procédure d'ajustement du modèle fournit les estimateurs des coefficients de la régression log-linéaire et des paramètres de la loi de Poisson. Cependant, nous utilisons la méthode de Laplace plus que Christiansen et Morris (1997) car l'ajout d'effets aléatoires implique plus de paramètres dans le modèle.

Avant de présenter le modèle proprement dit, nous en définissons les termes de base et concepts. Nous considérons des données de Poisson par strates ou groupes. Nous avons un nombre total de p strates et dans chaque strate j , il y a n_j observations de Poisson. À chaque observation ij ($i = 1, \dots, n_j$ et $j = 1, \dots, p$), est associée un paramètre de Poisson λ_{ij} . Les taux $\{\lambda_{ij}\}$ sont appelés les "paramètres individuels". Ces paramètres individuels sont prédits par $(r - 1) + (q - 1)$ variables explicatives, dont $q - 1$ sont des variables spécifiques aux strates, à travers un modèle log-linéaire ayant comme coefficients de régression $\beta = (\beta_0, \dots, \beta_{r-1})'$ et $\{b_j = (b_{j0}, \dots, b_{j,q-1})', j = 1, \dots, p\}$. Nous appelons ces

coefficients de régression les "paramètres structurels". Nous considérons des paramètres structurels supplémentaires que nous notons $\{\xi_j, j = 1, \dots, p\}$. Le paramètre ξ_j représente la variabilité "inter-individuelle" dans le groupe j . Lorsque les données ne sont pas stratifiées, ce paramètre est simplement noté ξ et représente la variabilité "inter-individuelle" dans la population considérée.

Le reste du travail est organisé comme suit. Dans le chapitre 1 nous décrivons le modèle original de Christiansen et Morris (1997). Nous présentons notre extension avec effets aléatoires dans le chapitre 2. Nous faisons une comparaison de résultats avec les données utilisées par Christiansen et Morris (1997) dans le chapitre 3. Les résultats de l'application du modèle aux données sur les effets indésirables d'un essai clinique sont reportés dans le chapitre 4. Finalement, nous présentons la conclusion de notre étude.

Chapitre 1

DESCRIPTION DU MODÈLE ORIGINAL DE CHRISTIANSEN ET MORRIS (1997)

Dans ce chapitre nous faisons une description du modèle développé par Christiansen et Morris (1997) sur lequel s'est basée notre étude. Dans ce modèle, le paramètre ξ représente la variabilité "inter-individuelle" dans la population considérée et les coefficients de la régression du modèle log-linéaire sont composés des effets fixes $\beta = (\beta_0, \dots, \beta_{r-1})'$.

1.1. MODÈLE DESCRIPTIF

Il y a deux niveaux dans le modèle hiérarchique de Christiansen et Morris. Le niveau 1 définit la distribution des données observées, $\{z_i\}$ ($i = 1, \dots, n$), sachant les paramètres individuels $\{\lambda_i\}$. Le niveau 2 spécifie la distribution des $\{\lambda_i\}$ sachant les hyperparamètres $\alpha \equiv (\beta, \xi)$.

Niveau 1 : les données observées (z_1, \dots, z_n) ont des distributions de Poisson indépendantes avec espérances $E(z_i) = e_i \lambda_i$, pour des expositions connues $e_i > 0$. C'est-à-dire $z_i | \lambda_i \sim \text{Pois}(e_i \lambda_i)$, et les taux observés $\{y_i = z_i / e_i\}$ ont pour espérances $E(y_i) = \lambda_i$. Les expositions ici représentent le nombre d'unités exposées à l'évènement considéré. Par exemple, dans une étude du taux de mortalité des patients dans n hopitaux, on a l'observation z_i ($i = 1, \dots, n$), représentant le nombre de décès dans l'hôpital i et dans ce cas, l'exposition e_i est le nombre total de patients de cet hôpital.

Niveau 2 : les paramètres individuels $\{\lambda_i\}$ suivent indépendamment des distributions gamma sachant les hyperparamètres $\alpha = (\beta_0, \dots, \beta_{r-1}, \xi)$ et ont pour espérances

$E(\lambda_i|\alpha) = \mu_i$, $\lambda_i|\alpha \sim \text{Gam}(\xi, \xi/\mu_i)$. Ensuite, la régression log-linéaire est définie par $\log(\mu_i) = X_i'\beta$, où $X_i = (1, x_{i1}, \dots, x_{i,r-1})'$ représente le vecteur des $(r - 1)$ variables explicatives pour les effets fixes $\beta = (\beta_0, \dots, \beta_{r-1})'$.

Distribution a priori des hyperparamètres : la densité a priori conjointe des hyperparamètres est définie par $h(\beta, \xi) = z_0/(\xi + z_0)^2$. Ce qui équivaut à $B_0 \equiv \xi/(\xi + z_0) \sim \text{Uniforme}(0, 1)$ (voir Christiansen et Morris (1997)). La constante z_0 est la médiane de ξ . Ce choix permet d'assurer une distribution a priori moins informative lorsque z_0 prend de petites valeurs. Christiansen et Morris (1997) propose ici une distribution a priori impropre pour β , tandis que celle de ξ est propre. Ils choisissent une telle distribution a priori propre pour ξ afin de s'assurer que l'estimateur du maximum de vraisemblance de ξ ne soit pas infini. En effet, lorsque cela arrive, les "facteurs de retrait" $B_i = \xi/(\xi + e_i\mu_i)$ tendent vers 1, annulant ainsi l'effet du terme $e_i\mu_i$ qui dans B_i apporte l'information sur les données.

1.2. MODÈLE INFÉRENTIEL

Le niveau 1 dans cette section spécifie les distributions marginales des données observées $\{z_i\}$ qui sont des binomiales négatives indépendantes, résultant du mélange de distributions Gamma-Poisson. Le niveau 2 définit les distributions conditionnelles des paramètres de Poisson $\{\lambda_i\}$ sachant les hyperparamètres (β, ξ) et les données.

Niveau 1 : les distributions marginales des observations sont des binomiales négatives,

$$z_i|\beta, \xi \sim NB(\xi, 1 - B_i), \quad (1.1)$$

où $B_i = \xi/(\xi + e_i\mu_i)$.

Niveau 2 : Christiansen et Morris (1997) montrent que les distributions conditionnelles des paramètres de Poisson $\{\lambda_i\}$ sachant (β, ξ) et les données sont des distributions gamma. C'est-à-dire

$$\lambda_i|(\text{data}, \beta, \xi) \sim \text{Gam}(z_i + \xi, e_i + \xi/\mu_i), \quad (1.2)$$

où $\text{data} \equiv \{z_i\} (i = 1, \dots, n)$. Les paramètres de cette loi gamma sont : $a_i = z_i + \xi$ et $b_i = e_i + \xi/\mu_i$. On a $1/b_i = \mu_i/(e_i\mu_i + \xi) = (1 - B_i)/e_i$. Les moyennes et variances des $\lambda_i | (\text{data}, \beta, \xi)$ sont alors données par :

$$\begin{aligned} \lambda_i^* \equiv E(\lambda_i | \text{data}, \beta, \xi) &= a_i/b_i = (z_i + \xi)(1 - B_i)/e_i \\ &= (1 - B_i)y_i + B_i\mu_i, \end{aligned} \quad (1.3)$$

et

$$(\sigma_i^*)^2 \equiv \text{var}(\lambda_i | \text{data}, \beta, \xi) = a_i/b_i^2 = \lambda_i^*(1 - B_i)/e_i. \quad (1.4)$$

Ayant précédemment défini les distributions pour les données $\{z_i\}$ et la distribution a priori des hyperparamètres (β, ξ) , il sera question dans la section suivante de déterminer la distribution a posteriori de ces hyperparamètres.

1.3. DISTRIBUTIONS DES HYPERPARAMÈTRES (β, ξ)

L'estimation des hyperparamètres (β, ξ) est faite par la maximisation du logarithme de leur densité a posteriori en utilisant ses dérivées de premier et second ordre par rapport aux hyperparamètres via l'algorithme de Newton-Raphson. Etant donné que la densité a posteriori des hyperparamètres dépend de la fonction de vraisemblance des données, Christiansen et Morris (1997) dérivent au préalable l'expression de cette fonction de vraisemblance et ses dérivées de premier et second ordre par rapport aux hyperparamètres.

La fonction de vraisemblance des hyperparamètres sachant les données est basée sur la distribution marginale des observations qui est la loi binomiale négative obtenue en (1.1). Elle se présente comme suit :

$$L(\beta, \xi) = \prod_{i=1}^n \frac{\Gamma(\xi + z_i)}{z_i! \Gamma(\xi)} (1 - B_i)^{z_i} B_i^\xi. \quad (1.5)$$

La log-vraisemblance est notée $\mathcal{L}(\beta, \xi) \equiv \log(L(\beta, \xi))$. On définit $M \equiv \max[z_1, \dots, z_n]$, $n_s \equiv$ nombre de $\{i : z_i = s\}$ pour $s = 0, \dots, M$, $N_s \equiv$ nombre de $\{i : z_i \geq s\}$, $\bar{z} = \sum_i z_i/n$ et $X = (X_1, \dots, X_n)'$ la matrice $n \times r$ des variables explicatives. Alors,

$$\begin{aligned} \mathcal{L}(\beta, \xi) &= \sum_{s=1}^M N_s \log(\xi + s - 1) - n\bar{z} \log(\xi) \\ &\quad + z'X\beta - \sum_{i=1}^n (\xi + z_i) \log(1 + e_i\mu_i/\xi). \end{aligned} \quad (1.6)$$

Afin d'améliorer les approximations de la vraisemblance, $\tau \equiv \log \xi$ est utilisé dans la suite au lieu de ξ .

Les dérivées première et seconde de $\mathcal{L}(\beta, \xi)$ par rapport à β sont données par

$$\frac{\partial \mathcal{L}(\beta, \xi)}{\partial \beta} = \sum_{i=1}^n (z_i - e_i \mu_i) B_i X_i, \quad (1.7)$$

$$-\frac{\partial^2 \mathcal{L}(\beta, \xi)}{\partial(\beta\beta')} = \sum_{i=1}^n X_i X_i' e_i B_i \lambda_i^* = X' D X, \quad (1.8)$$

où D est la matrice diagonale $n \times n$ dont l'élément i de la diagonale est $e_i B_i \lambda_i^* > 0$ et λ_i^* est défini en (1.3).

La dérivée croisée de $\mathcal{L}(\beta, \xi)$ par rapport à β et τ est

$$\frac{\partial^2 \mathcal{L}(\beta, \xi)}{\partial \beta \partial \tau} = \sum_{i=1}^n (z_i - e_i \mu_i) B_i (1 - B_i) X_i. \quad (1.9)$$

Les dérivées première et seconde de $\mathcal{L}(\beta, \xi)$ par rapport à $\tau = \log \xi$ sont

$$\frac{\partial \mathcal{L}(\beta, \xi)}{\partial \tau} = \sum_{s=1}^M N_s \frac{\xi}{\xi + s - 1} + \xi \sum_{i=1}^n \log(B_i) - \sum_{i=1}^n (z_i - e_i \mu_i) B_i, \quad (1.10)$$

$$-\frac{\partial^2 \mathcal{L}(\beta, \xi)}{\partial \tau^2} = \sum_{s=1}^M N_s \left(\frac{\xi}{\xi + s - 1} \right)^2 - \sum_{i=1}^n z_i B_i + \sum_{i=1}^n (z_i - e_i \mu_i) B_i (1 - B_i) - \frac{\partial \mathcal{L}(\beta, \xi)}{\partial \tau}. \quad (1.11)$$

La densité a posteriori des hyperparamètres (β, τ) est déduite de leur densité a priori et de la fonction de vraisemblance obtenue précédemment et s'écrit

$$p(\beta, \tau) \equiv c_0 L(\beta, \xi | \text{data}) \xi / (\xi + z_0)^2, \text{ où } c_0 \text{ est une constante.}$$

Christiansen et Morris (1997) considèrent ensuite l'approche du maximum de vraisemblance restreint (REML) pour faire un ajustement de cette densité a posteriori afin d'obtenir de meilleurs estimations des composantes de la variance (voir Liang et Zeger 1995). Cette approche estime τ en considérant β comme paramètre nuisible; ce qui conduit à l'approximation suivante pour la densité a posteriori

$$p(\tau) = c_2 \left| \hat{H}_\tau \right|^{-1/2} L(\hat{\beta}_\tau, \xi | \text{data}) \xi / (\xi + z_0)^2, \quad (1.12)$$

où $\hat{\beta}_\tau$ représente la racine de la fonction en (1.7) et dépend de τ , \hat{H}_τ se définit comme $\hat{H}_\tau \equiv -\frac{\partial^2 \mathcal{L}(\beta, \xi)}{\partial(\beta\beta')} \Big|_{\hat{\beta}_\tau}$ et dépend aussi de τ . Cette approximation de la densité a posteriori est aussi appelée la vraisemblance restreinte et son logarithme, en ignorant le terme constant

c_2 , est

$$\ell_R(\beta, \tau) \equiv \log \left\{ L(\beta, \tau | \text{data}) \frac{\xi}{(\xi + z_0)^2} |H_\tau|^{-1/2} \right\}. \quad (1.13)$$

Les estimateurs $\hat{\beta}$ et $\hat{\tau}$ qui maximisent $\ell_R(\beta, \tau)$ sont obtenus en utilisant les dérivées de $\ell_R(\beta, \tau)$ via l'algorithme de Newton-Raphson. Une approximation du terme $|H_\tau|^{-1/2}$ est utilisée afin de simplifier ces dérivées. On note que les dérivées de $\ell_R(\beta, \tau)$ par rapport à β sont les mêmes qu'avec $\mathcal{L}(\beta, \xi)$ obtenues en (1.7), (1.8) et (1.9). Les dérivées par rapport à τ quant à elles dépendent des résultats en (1.10) et (1.11) mais impliquent également d'autres termes. La distribution a posteriori des hyperparamètres (β, τ) est ensuite déterminée, selon la théorie asymptotique, comme étant approximativement une distribution conjointe normale de moyenne $(\hat{\beta}, \hat{\tau})$ et avec une matrice de variance-covariance égale à l'inverse de la matrice d'information de Fisher observée. Le négatif des dérivées de second ordre de la log-densité $\ell_R(\beta, \tau)$ définit cette matrice d'information de Fisher.

Il reste à présent à déterminer les distributions des paramètres individuels λ_i . Ceci qui fait l'objet de la section suivante.

1.4. DISTRIBUTIONS DES PARAMÈTRES λ_i

La distribution approximative des hyperparamètres (β, τ) déterminée dans la section précédente est utilisée ici pour calculer la distribution de chaque λ_i . La moyenne et la variance a posteriori $\hat{\lambda}_i$ et $\hat{\sigma}_{\lambda_i}^2$ de λ_i vont être obtenues au préalable, ensuite sa distribution sera déterminée.

Pour commencer, il est utile de déterminer la moyenne et la variance de $\mu_i = \exp(X_i' \beta)$, car elles interviennent dans le calcul de $\hat{\lambda}_i$ et $\hat{\sigma}_{\lambda_i}^2$. La distribution de μ_i est une distribution lognormale. En effet, $\log(\mu_i) = X_i' \beta$ suit une loi normale telle que $E(\log \mu_i) = X_i' E(\beta)$ et $\text{var}(\log \mu_i) = X_i' \text{var}(\beta) X_i$. Ainsi,

$$\hat{\mu}_i \equiv E(\mu_i | \text{data}) = \exp[X_i' E(\beta) + X_i' \text{var}(\beta) X_i / 2],$$

$$\hat{\sigma}_{\mu_i}^2 \equiv \text{var}(\mu_i | \text{data}) = \hat{\mu}_i^2 (\exp[X_i' \text{var}(\beta) X_i] - 1),$$

où $E(\beta) = \hat{\beta}$ et $\text{var}(\beta)$ est le premier bloc de la diagonale de la matrice de variance-covariance de (β, τ) .

La moyenne et la variance de λ_i sont :

$$E(\lambda_i|\text{data}) = E\lambda_i^* = (1 - EB_i)y_i + E(\mu_i B_i), \quad (1.14)$$

$$\begin{aligned} \text{var}(\lambda_i|\text{data}) &= E(\sigma_i^*)^2 + \text{var}(\lambda_i^*) \\ &= E\lambda_i^*(1 - EB_i)/e_i + \text{var}(\mu_i B_i - B_i y_i), \end{aligned} \quad (1.15)$$

où λ_i^* et $(\sigma_i^*)^2$ ont été obtenues en (1.3) et (1.4).

On remarque que le calcul de ces deux moments requiert les espérances de la forme $E(\mu_i^s B_i^t)$ ($s, t = 0, 1, 2$) qui apparaissent dans leur formulation. On définit $b_i \equiv \text{cov}(X_i' \beta, \tau - X_i' \beta) / \text{var}(\tau - X_i' \beta)$, et $E_s B_i^t \equiv E(\mu_i^s B_i^t) / E\mu_i^s$. On montre ensuite que $E_s B_i^t$ s'exprime comme

$$E_s B_i^t = \frac{EB_i^t (B_i / (1 - B_i))^{sb_i}}{E(B_i / (1 - B_i))^{sb_i}}. \quad (1.16)$$

Ainsi, une fois qu'on a $E_s B_i^t$, on peut déterminer $E(\mu_i^s B_i^t) = E_s B_i^t E\mu_i^s$. Cependant, le calcul des espérances dans l'expression de $E_s B_i^t$ nécessite d'avoir la distribution de B_i . Pour ce faire, la méthode dite "méthode de la densité ajustée" (ADM) (Christiansen et Morris (1997)) est utilisée afin de trouver les paramètres de la distribution de B_i . L'ADM est une méthode qui permet de faire une approximation de la distribution a posteriori d'un paramètre afin d'obtenir ses moments. On rappelle que $B_i = e^\tau / (e^\tau + e_i \mu_i)$, et donc $0 < B_i < 1$. La distribution beta est alors choisie pour B_i . Dans ce cas, l'ADM va ajuster la densité de B_i afin d'obtenir une distribution beta pour B_i et par suite ses moments.

Si on définit $u_i \equiv \text{logit}(B_i)$, ou de façon équivalente, $B_i \equiv e^{u_i} / (1 + e^{u_i})$, tel que $u_i \sim N(m_i, w_i^2)$, alors la densité de B_i n'est pas celle d'une distribution beta. En effet, la fonction de repartition de B_i s'écrit

$$P(B_i < b) = P(\log(B_i / (1 - B_i)) < \log(b / (1 - b))) = \Phi((\log(b / (1 - b)) - m_i) / w_i).$$

Ainsi, la densité de B_i prend la forme

$$p(b) = \frac{dP(B_i < b)}{db} = \frac{1}{w_i} \phi((\log(b / (1 - b)) - m_i) / w_i) \frac{1}{b(1-b)}, \text{ où } \phi \sim N(0, 1).$$

Alors l'ADM fait une approximation de cette densité avec la densité beta afin de faciliter le calcul des moments. L'application de la méthode permet, pour la distribution de B_i telle que $B_i \sim \text{Beta}(a_{i1}, a_{i2})$, d'avoir les paramètres $a_{i1} = 1 / (w_i^2 (1 - \hat{B}_i))$ et $a_{i2} = 1 / (w_i^2 \hat{B}_i)$ avec $\hat{B}_i \equiv e^{m_i} / (1 + e^{m_i})$. Ainsi, pour $t \geq 0$ et $-a_1 < k < a_2$, on a l'approximation ADM

suivante

$$\frac{EB_i^t(B_i/(1-B_i))^k}{E(B_i/(1-B_i))^k} \approx \frac{(a_{i1}+k)^{(t)}}{(a_{i1}+a_{i2})^{(t)}}, \quad (1.17)$$

où on définit $a^{(t)} \equiv \Gamma(a+t)/\Gamma(a) = a(a+1)\dots(a+t-1)$.

Ainsi, pour calculer les expressions $E_s B_i^t$ ($s, t = 0, 1, 2$) en (1.16), il suffit d'utiliser cette approximation en remplaçant k par sb_i . De plus, pour évaluer m_i et w_i^2 , on sait que

$$\begin{aligned} u_i &= \text{logit}(B_i) \\ &= \log B_i - \log(1 - B_i) \\ &= \tau - X_i' \beta - \log(e_i). \end{aligned}$$

Il s'en suit que $m_i \equiv E(u_i) = \hat{\tau} - X_i' \hat{\beta} - \log(e_i)$ et $w_i^2 \equiv \text{var}(u_i) = \text{var}(\tau_j - X_i' \beta)$.

On a ainsi tous les éléments pour calculer les $E(\mu_i^s B_i^t)$ ($s, t = 0, 1, 2$) qu'on utilisera pour obtenir les expressions finales de la moyenne et la variance a posteriori $\hat{\lambda}_i \equiv E(\lambda_i | \text{data})$ et $\hat{\sigma}_{\lambda_i}^2 \equiv \text{var}(\lambda_i | \text{data})$ des paramètres λ_i . Finalement, on conclut que conditionnellement aux données, ces paramètres ont chacun approximativement une distribution gamma,

$$\lambda_i | \text{data} \sim \text{Gam}(\hat{\lambda}_i^2 / \hat{\sigma}_{\lambda_i}^2, \hat{\lambda}_i / \hat{\sigma}_{\lambda_i}^2).$$

Ce résultat se justifie bien par le fait que la distribution de λ_i en (1.2) conditionnellement aux données et (β, τ) est une gamma.

Chapitre 2

PRÉSENTATION DU MODÈLE

Dans cette partie, nous présentons le modèle que nous avons développé à partir de celui de Christiansen et Morris (1997). Contrairement à eux, notre modèle considère des données de Poisson par strates. Ainsi, pour chaque strate j ($j = 1, \dots, p$), il y a une variabilité inter-individuelle représentée par le paramètre ξ_j . De plus, les coefficients de la régression log-linéaire sont composés non seulement des effets fixes $\beta = (\beta_0, \dots, \beta_{r-1})'$ mais aussi des effets aléatoires $\{b_j = (b_{j0}, \dots, b_{j,q-1})', j = 1, \dots, p\}$ liés aux strates.

2.1. MODÈLE DESCRIPTIF

Nous avons un modèle hiérarchique à deux niveaux. Dans cette section, le niveau 1 spécifie la distribution des données observées, $\{z_{ij}\}$ ($i = 1, \dots, n_j$ et $j = 1, \dots, p$), sachant les paramètres individuels $\{\lambda_{ij}\}$. Le niveau 2 définit la distribution des $\{\lambda_{ij}\}$ sachant les hyperparamètres $\theta \equiv (\beta, \xi_1, \dots, \xi_p, b_1, \dots, b_p)$.

Niveau 1 : les données observées $\{z_{ij}\}$ ont des distributions de Poisson indépendantes avec espérances $E(z_{ij}) = e_{ij}\lambda_{ij}$, pour des expositions connues $e_{ij} > 0$, c'est-à-dire $z_{ij}|\lambda_{ij} \sim \text{Pois}(e_{ij}\lambda_{ij})$. Les taux observés $\{y_{ij} = z_{ij}/e_{ij}\}$ ont donc pour espérances $E(y_{ij}) = \lambda_{ij}$.

Niveau 2 : les paramètres individuels $\{\lambda_{ij}\}$ suivent indépendamment des distributions gamma sachant les hyperparamètres $\theta = (\beta, \xi_1, \dots, \xi_p, b_1, \dots, b_p)$ et avec espérances $E(\lambda_{ij}|\theta) = \mu_{ij}$, c'est-à-dire $\lambda_{ij}|\theta \sim \text{Gam}(\xi_j, \xi_j/\mu_{ij})$.

On définit ensuite la régression de Poisson par $\log(\mu_{ij}) = X'_{ij}\beta + W'_{ij}b_j$, où $X_{ij} =$

$(1, x_{ij,1}, \dots, x_{ij,r-1})'$ représente le vecteur des valeurs des $(r - 1)$ variables explicatives pour les effets fixes $\beta = (\beta_0, \dots, \beta_{r-1})'$. Le vecteur $b_j = (1, b_{j1}, \dots, b_{j,q-1})'$ représente le vecteur des effets aléatoires pour la strate j et $W_{ij} = (1, w_{ij,1}, \dots, w_{ij,q-1})'$ le vecteur de ses variables explicatives. En effet, pour certaines données, on pourrait disposer de variables explicatives caractérisant les strates. Prenons l'exemple d'une étude qui collecte des données sur les patients provenant de différents hôpitaux ; les données d'un même hôpital étant considéré comme une strate. Des variables décrivant les strates (les hôpitaux) pourraient être le secteur (public/privé), l'équipement, la taille du personnel, la qualité des soins ou encore le statut socio-économique du quartier où se trouve l'hôpital.

Distribution a priori des hyperparamètres :

On pose $\tau_j = \log \xi_j$ et $\tau = \log \xi$. On suppose que pour chaque τ_j la distribution est $\tau_j \sim N(\tau, \sigma_\tau^2)$, où $\sigma_\tau^2 \sim \text{Inv-}\chi^2(\nu_0, s_0^2)$. De façon analogue à Christiansen et Morris (1997), nous considérons que $\xi/(\xi + z_0) \sim \text{Uniforme}(0, 1)$. Cette distribution uniforme pour $\xi/(\xi + z_0)$ est en fait équivalente à avoir pour les hyperparamètres (β, ξ) la densité a priori conjointe $p(\beta, \xi) = z_0/(\xi + z_0)^2$, où la constante z_0 est la médiane de ξ . Ceci suppose une distribution a priori constante pour β , donc impropre. Le choix de densités a priori propres pour ξ et les τ_j est fait pour éviter des problèmes de convergence dans l'estimation du maximum de vraisemblance de ces paramètres et nous assurer d'avoir des distributions a posteriori propres. En effet, Hobert et Casella (1996) montrent que l'utilisation de distributions a priori impropres pour des composantes de la variance dans les modèles linéaires hiérarchiques mixtes peut souvent conduire à des distributions a posteriori impropres. En ce qui concerne la distribution des b_j , nous supposons que $b_j \sim N_q(0, \Psi)$, où $\Psi \sim W^{-1}(m_0, \Lambda_0)$ et W^{-1} est la distribution Wishart-inversée. Comme dans tout modèle bayésien hiérarchique mixte typique, nous supposons que le vecteur des effets aléatoires b_j a une distribution multivariée normale. La loi Wishart-inversée est utilisée comme distribution a priori de la matrice de variance-covariance Ψ et dont le paramètre m_0 est le nombre de degrés de liberté et Λ_0 est la matrice désignant le paramètre d'échelle.

2.2. MODÈLE INFÉRENTIEL

Les distributions marginales des données observées $\{z_{ij}\}$ sont des binomiales négatives indépendantes. Ceci résulte des combinaisons de lois gamma avec des distributions de Poisson. En effet, si on a des variables aléatoires telles que $z|\lambda \sim f(z|\lambda) = \text{Pois}(\lambda)$ et $\lambda \sim \pi(\lambda) = \text{Gam}(a, b)$, alors la distribution marginale de z est donnée par $m(z|a, b) = \int_{\lambda} f(z|\lambda)\pi(\lambda)d\lambda$. Ainsi,

$$\begin{aligned} m(z|a, b) &= \int_{\lambda} \frac{\lambda^z e^{-\lambda}}{z!} \frac{b^a \lambda^{a-1} e^{-\lambda b}}{\Gamma(a)} d\lambda, \\ &= \frac{b^a \Gamma(z+a)}{z! \Gamma(a) (1+b)^{z+a}} \int_{\lambda} \frac{(1+b)^{z+a} \lambda^{z+a-1} e^{-\lambda(1+b)}}{\Gamma(z+a)} d\lambda, \\ &= \frac{\Gamma(z+a)}{z! \Gamma(a)} b^a \left(\frac{1}{1+b} \right)^{z+a}, \\ &= \frac{\Gamma(z+a)}{z! \Gamma(a)} \left(\frac{b}{1+b} \right)^a \left(\frac{1}{1+b} \right)^z. \end{aligned}$$

D'où $z|a, b \sim NB(a, \frac{1}{1+b})$.

Dans notre cas, on a $z_{ij}|\lambda_{ij} \sim \text{Pois}(e_{ij}\lambda_{ij})$ et $e_{ij}\lambda_{ij}|\theta \sim \text{Gam}(\xi_j, \xi_j/(e_{ij}\mu_{ij}))$. La distribution marginale de z_{ij} est alors une binomiale négative que nous spécifions au niveau 1 suivant.

Niveau 1 : la distribution marginale de chaque z_{ij} est

$$\begin{aligned} m(z_{ij}|\theta) &= \frac{\Gamma(\xi_j + z_{ij})}{z_{ij}! \Gamma(\xi_j)} \left(\frac{\xi_j}{\xi_j + e_{ij}\mu_{ij}} \right)^{\xi_j} \left(\frac{e_{ij}\mu_{ij}}{\xi_j + e_{ij}\mu_{ij}} \right)^{z_{ij}} \\ &= \frac{\Gamma(\xi_j + z_{ij})}{z_{ij}! \Gamma(\xi_j)} B_{ij}^{\xi_j} (1 - B_{ij})^{z_{ij}}, \end{aligned}$$

où $B_{ij} = \xi_j / (\xi_j + e_{ij}\mu_{ij})$.

En d'autres termes,

$$z_{ij}|\beta, \xi_j, b_j \sim NB(\xi_j, 1 - B_{ij}). \quad (2.1)$$

On a, $0 < B_{ij} < 1$. Pour des ξ_j très grands, ce qui implique des "facteurs de retrait" B_{ij} proches de l'unité, la distribution binomiale négative ci-dessus est approximativement une distribution de Poisson (voir Gelman et al.(1995), pages 350-351). Le paramètre ξ_j permet de mesurer la variation extra-poissonnienne pour les observations de la strate j .

Niveau 2 : les distributions conditionnelles des paramètres $\{\lambda_{ij}\}$ sachant θ et les données sont des distributions gamma. En effet, puisqu'on a

$z_{ij}|\lambda_{ij} \sim f(z_{ij}|\lambda_{ij}) = \text{Pois}(e_{ij}\lambda_{ij})$ et $\lambda_{ij}|\theta \sim f(\lambda_{ij}|\theta) = \text{Gam}(\xi_j, \xi_j/\mu_{ij})$ alors, $f(\lambda_{ij}|z_{ij}, \theta) \propto f(z_{ij}|\lambda_{ij})f(\lambda_{ij}|\theta)$. Or, $f(z_{ij}|\lambda_{ij}) \propto \lambda_{ij}^{z_{ij}} e^{-e_{ij}\lambda_{ij}}$ et $f(\lambda_{ij}|\theta) \propto \lambda_{ij}^{\xi_j-1} e^{-\frac{\xi_j\lambda_{ij}}{\mu_{ij}}}$ donc, $f(\lambda_{ij}|z_{ij}, \theta) \propto \lambda_{ij}^{z_{ij}+\xi_j-1} e^{-(e_{ij}+\frac{\xi_j}{\mu_{ij}})\lambda_{ij}}$. Il s'en suit que :

$$\lambda_{ij}|\text{data}, \theta \sim \text{Gam}(z_{ij} + \xi_j, e_{ij} + \xi_j/\mu_{ij}), \quad (2.2)$$

avec $\text{data} \equiv \{z_{ij}\} (i = 1, \dots, n_j \text{ et } j = 1, \dots, p)$.

Les moyennes et variances, de façon analogue au modèle de Christiansen et Morris (1997) (section 1.2), sont données par :

$$\lambda_{ij}^* \equiv E(\lambda_{ij}|\text{data}, \theta) = (1 - B_{ij})y_{ij} + B_{ij}\mu_{ij} \quad (2.3)$$

$$(\sigma_{ij}^*)^2 \equiv \text{var}(\lambda_{ij}|\text{data}, \theta) = \lambda_{ij}^*(1 - B_{ij})/e_{ij}. \quad (2.4)$$

Les distributions marginales des données $\{z_{ij}\}$ et les distributions a priori des hyperparamètres $\theta \equiv (\beta, \xi_1, \dots, \xi_p, b_1, \dots, b_p)$ ayant été définies, nous allons procéder au calcul de la fonction vraisemblance dans la section 2.3. Nous allons ensuite utiliser cette fonction dans l'évaluation de la distribution a posteriori des hyperparamètres dans la section 2.4.

2.3. FONCTION DE VRAISEMBLANCE

La fonction de vraisemblance des données est basée sur la loi binomiale négative obtenue en (2.1) $NB(\xi_j, B_{ij})$ et s'écrit

$$L(\beta, \xi_1, \dots, \xi_p, b_1, \dots, b_p) = \prod_{j=1}^p \prod_{i=1}^{n_j} \frac{\Gamma(\xi_j + z_{ij})}{z_{ij}! \Gamma(\xi_j)} (1 - B_{ij})^{z_{ij}} B_{ij}^{\xi_j}, \quad (2.5)$$

son logarithme est noté $\mathcal{L}(\theta) \equiv \log(L(\theta))$.

Proposition 1. *On définit $M \equiv \max [z_{ij}]$, $n_s \equiv$ nombre de $\{(i, j) : z_{ij} = s\}$ pour $s = 0, \dots, M$, $N_s \equiv$ nombre de $\{(i, j) : z_{ij} \geq s\}$, $\bar{z}_j = \sum_i z_{ij}/n_j$, $\bar{z} = \sum_j \bar{z}_j/p$, $X_j = (X_{1j}, \dots, X_{n_jj})'$ la matrice $n_j \times r$ des variables explicatives, $W_j = (W_{1j}, \dots, W_{n_jj})'$*

la matrice $n_j \times q$ des variables explicatives liées à la strate j . Alors,

$$\begin{aligned} \mathcal{L}(\theta) = & \sum_{s=1}^M N_s \log(\xi_j + s - 1) - \sum_{j=1}^p \sum_{i=1}^{n_j} \log(z_{ij}!) + \sum_{j=1}^p \sum_{i=1}^{n_j} z_{ij} \log(e_{ij}) - \sum_{j=1}^p n_j \bar{z}_j \log(\xi_j) \\ & + \sum_{j=1}^p z'_j X_j \beta + \sum_{j=1}^p z'_j W_j b_j - \sum_{j=1}^p \sum_{i=1}^{n_j} (\xi_j + z_{ij}) \log(1 + e_{ij} \mu_{ij} / \xi_j). \end{aligned}$$

Preuve :

À partir de (2.5) on obtient

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{j=1}^p \sum_{i=1}^{n_j} \left\{ \log \frac{\Gamma(\xi_j + z_{ij})}{z_{ij}! \Gamma(\xi_j)} - \log z_{ij}! + z_{ij} \log(1 - B_{ij}) + \xi_j \log B_{ij} \right\} \\ &= S_1 + S_2 + S_3 - \sum_{j=1}^p \sum_{i=1}^{n_j} \log(z_{ij}!). \end{aligned} \quad (2.6)$$

Les sommes S_1 , S_2 , et S_3 sont définies par

$$\begin{aligned} S_1 &\equiv \sum_{j=1}^p \sum_{i=1}^{n_j} \log \frac{\Gamma(\xi_j + z_{ij})}{z_{ij}! \Gamma(\xi_j)} = \sum_{j=1}^p \sum_{i=1}^{n_j} \log \{ \xi_j (\xi_j + 1) \dots (\xi_j + z_{ij} - 1) \} \\ &= \sum_{j=1}^p \sum_{s=1}^M N_s \log(\xi_j + s - 1), \end{aligned}$$

$$\begin{aligned} S_2 &\equiv \sum_{j=1}^p \sum_{i=1}^{n_j} z_{ij} \log(1 - B_{ij}) \\ &= \sum_{j=1}^p \sum_{i=1}^{n_j} \{ z_{ij} \log e_{ij} + z_{ij} \log \mu_{ij} - z_{ij} \log \xi_j (1 + e_{ij} \mu_{ij} / \xi_j) \} \\ &= \sum_{j=1}^p \sum_{i=1}^{n_j} z_{ij} \log e_{ij} + \sum_{j=1}^p z'_j X_j \beta + \sum_{j=1}^p z'_j W_j b_j - \sum_{j=1}^p n_j \bar{z}_j \log \xi_j \\ &\quad - \sum_{j=1}^p \sum_{i=1}^{n_j} z_{ij} \log(1 + e_{ij} \mu_{ij} / \xi_j), \end{aligned}$$

et

$$\begin{aligned} S_3 &\equiv \sum_{j=1}^p \sum_{i=1}^{n_j} \xi_j \log B_{ij} = \sum_{j=1}^p \sum_{i=1}^{n_j} \xi_j (\log \xi_j - \log \xi_j - \log(1 + e_{ij} \mu_{ij} / \xi_j)) \\ &= - \sum_{j=1}^p \sum_{i=1}^{n_j} \xi_j \log(1 + e_{ij} \mu_{ij} / \xi_j). \end{aligned}$$

On obtient donc le résultat de la proposition 1 en remplaçant les expressions de S_1 , S_2 et S_3 dans l'équation (2.6).

Notre objectif est de déterminer les estimateurs des hyperparamètres $\beta, b_1, \dots, b_p, \tau_1, \dots, \tau_p$, en maximisant leur densité à posteriori. En fait nous considérons un ajustement de cette densité à posteriori en utilisant l'approche du maximum de vraisemblance restreint (REML). On obtient ainsi une densité ajustée, appelée vraisemblance restreinte, dont on va calculer les dérivées de premier et second ordre par rapport aux hyperparamètres. Ces dérivées sont en effet nécessaires pour l'implémentation de l'algorithme de Newton-Raphson utilisé dans notre programme de maximisation. Etant donné que le logarithme de la vraisemblance restreinte dépend de la log-vraisemblance $\mathcal{L}(\theta)$ obtenue dans la proposition 1, il est utile dans un premier temps de calculer les dérivées de $\mathcal{L}(\theta)$ par rapport aux hyperparamètres.

Les dérivées première et seconde de $\mathcal{L}(\theta)$ par rapport à β sont données par

$$\frac{\partial \mathcal{L}(\theta)}{\partial \beta} = \sum_{j=1}^p \sum_{i=1}^{n_j} (z_{ij} - e_{ij} \mu_{ij}) B_{ij} X_{ij}, \quad (2.7)$$

$$H_\beta \equiv -\frac{\partial^2 \mathcal{L}(\theta)}{\partial(\beta\beta')} = \sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij} X'_{ij} e_{ij} B_{ij} \lambda_{ij}^* = \sum_{j=1}^p X'_j D_j X_j, \quad (2.8)$$

où D_j est la matrice diagonale $n_j \times n_j$ dont l'élément i de la diagonale est $e_{ij} B_{ij} \lambda_{ij}^* > 0$ et λ_{ij}^* est défini en (2.3).

Les dérivées croisées de $\mathcal{L}(\theta)$ par rapport à β et b_j puis par rapport à β et τ_j sont

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial \beta \partial b_j} = \sum_{i=1}^{n_j} X_{ij} W'_{ij} [-z_{ij} - B_{ij} (e_{ij} \mu_{ij} - z_{ij})] B_{ij}, \quad (2.9)$$

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial \beta \partial \tau_j} = \sum_{i=1}^{n_j} (z_{ij} - e_{ij} \mu_{ij}) B_{ij} (1 - B_{ij}) X_{ij}. \quad (2.10)$$

Nous avons ensuite les dérivées première, et seconde de $\mathcal{L}(\theta)$ par rapport à b_j suivantes

$$\frac{\partial \mathcal{L}(\theta)}{\partial b_j} = \sum_{i=1}^{n_j} (z_{ij} - e_{ij} \mu_{ij}) B_{ij} W_{ij} \quad (2.11)$$

$$-\frac{\partial^2 \mathcal{L}(\theta)}{\partial(b_j b'_j)} = \sum_{i=1}^{n_j} W_{ij} W'_{ij} e_{ij} B_{ij} \lambda_{ij}^* = W'_j D_j W_j \quad (2.12)$$

La dérivée croisée de $\mathcal{L}(\theta)$ par rapport à b_j et τ_j est

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial b_j \partial \tau_j} = \sum_{i=1}^{n_j} (z_{ij} - e_{ij} \mu_{ij}) B_{ij} (1 - B_{ij}) W_{ij}. \quad (2.13)$$

Enfin, les dérivées première et seconde de $\mathcal{L}(\theta)$ par rapport à τ_j sont données par

$$\begin{aligned} \frac{\partial \mathcal{L}(\theta)}{\partial \tau_j} &= \sum_{s=1}^M N_s \frac{\xi_j}{\xi_j + s - 1} + \xi_j \sum_{i=1}^{n_j} \log(B_{ij}) - \sum_{i=1}^{n_j} (z_{ij} - e_{ij}\mu_{ij})B_{ij}, \\ -\frac{\partial^2 \mathcal{L}(\theta)}{\partial \tau_j^2} &= \sum_{s=1}^M N_s \left(\frac{\xi_j}{\xi_j + s - 1}\right)^2 - \sum_{i=1}^{n_j} z_{ij} B_{ij} + \sum_{i=1}^{n_j} (z_{ij} - e_{ij}\mu_{ij})B_{ij}(1 - B_{ij}) - \frac{\partial \mathcal{L}(\theta)}{\partial \tau_j}. \end{aligned}$$

La log-vraisemblance $\mathcal{L}(\theta)$ et ses dérivées ayant ainsi été obtenues, nous allons dans la section suivante déterminer la densité à posteriori des hyperparamètres $\beta, b_1, \dots, b_p, \tau_1, \dots, \tau_p$ et ensuite la vraisemblance restreinte et ses dérivées.

2.4. DISTRIBUTION DES HYPERPARAMÈTRES

La densité a posteriori de $\theta = (\beta, b_1, \dots, b_p, \tau_1, \dots, \tau_p)$ peut s'écrire comme

$$\begin{aligned} p(\theta \mid \text{data}) &\equiv p(\beta, b_1, \dots, b_p, \tau_1, \dots, \tau_p, \tau, \sigma_\tau^2, \Psi \mid \text{data}) \\ &= c_0 L(\theta \mid \text{data}) f(\tau) f(\sigma_\tau^2) f(\Psi) \prod_{j=1}^p f(b_j) \prod_{j=1}^p f(\tau_j), \end{aligned}$$

où c_0 est une constante.

Dans cette section, contrairement à Christiansen et Morris (1997)(voir page 622), nous n'allons pas directement utiliser la densité a posteriori $p(\theta \mid \text{data})$ ainsi définie pour construire la vraisemblance restreinte. En effet, la présence des effets aléatoires $\{b_i\}$ dans notre modèle a rendu l'expression de $p(\theta \mid \text{data})$ un peu plus compliquée, car en plus de nos paramètres d'intérêts que sont $\beta, b_1, \dots, b_p, \tau_1, \dots, \tau_p$, nous avons aussi les paramètres τ, σ_τ^2 et Ψ . Nous allons donc considérer ces derniers comme paramètres nuisibles et intégrer $p(\theta \mid \text{data})$ successivement par rapport à Ψ, σ_τ^2 et τ .

Proposition 2. *Le logarithme de la densité a posteriori de θ après avoir enlevé les paramètres nuisibles s'écrit*

$$\ell(\theta) \equiv \log \left\{ L(\theta \mid \text{data}) \left| \Lambda_0 + \sum_{j=1}^p b_j b'_j \right|^{-(m_0+p)/2} A^{-(\nu_0+p)/2} e^{g(a)} \left(-\frac{d^2 g(\tau)}{d\tau^2} \Big|_{\tau=a} \right)^{-1/2} \right\},$$

où $g(\tau) = -b \log(1 + \frac{(\tau-a)^2}{A}) + \tau - 2 \log(e^\tau + z_0)$, $A = \frac{\nu_0 s_0^2}{p} + \frac{1}{p} \sum_{j=1}^p (\tau_j - a)^2$, $a = \frac{1}{p} \sum_{j=1}^p \tau_j$ et $b = (\nu_0 + p)/2$.

Preuve :

En intégrant $p(\theta \mid \text{data})$ par rapport à Ψ , on a

$$p_1(\theta \mid \text{data}) = \int_{\Psi} p(\theta \mid \text{data}) d\Psi = C_1 \mathcal{I}_1,$$

où C_1 est une constante par rapport à Ψ et

$$\mathcal{I}_1 = \int_{\Psi} f(\Psi) \prod_{j=1}^p f(b_j) d\Psi.$$

Rappelons que $b_j \sim N_q(0, \Psi)$ et $\Psi \sim W^{-1}(m_0, \Lambda_0)$, alors

$$f(b_j) = (2\pi)^{-q/2} |\Psi|^{-1/2} \exp[-b_j' \Psi^{-1} b_j / 2] \text{ et}$$

$$f(\Psi) = \frac{|\Lambda_0|^{m_0/2} |\Psi|^{-(m_0+q+1)/2} \exp[-\text{trace}(\Lambda_0 \Psi^{-1})/2]}{2^{m_0 q/2} \Gamma_q(m_0/2)}.$$

On obtient

$$\mathcal{I}_1 = \frac{\pi^{-pq/2} |\Lambda_0|^{m_0} \Gamma_q((m_0 + p)/2)}{\left| \Lambda_0 + \sum_{j=1}^p b_j b_j' \right|^{(m_0+p)/2} \Gamma_q(m_0/2)}.$$

Ensuite, l'intégrale de $p_1(\theta \mid \text{data})$ par rapport à σ_τ^2 est

$$p_2(\theta \mid \text{data}) = \int_{\sigma_\tau^2} p_1(\theta \mid \text{data}) d\sigma_\tau^2 = C_2 \mathcal{I}_2,$$

où C_2 ne dépend pas de σ_τ^2 et

$$\mathcal{I}_2 = \int_{\sigma_\tau^2} f(\sigma_\tau^2) \prod_{j=1}^p f(\tau_j) d\sigma_\tau^2.$$

Rappelons également que $\tau_j \sim N(\tau, \sigma_\tau^2)$ et $\sigma_\tau^2 \sim \text{Inv-}\chi^2(\nu_0, s_0^2)$, alors

$$f(\tau_j) = (2\pi\sigma_\tau^2)^{-1/2} \exp[-(\tau_j - \tau)^2 / 2\sigma_\tau^2] \text{ et}$$

$$f(\sigma_\tau^2) = \frac{(s_0^2 \nu_0 / 2)^{\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma_\tau^2)^{-(\nu_0/2+1)} \exp(-\nu_0 s_0^2 / 2\sigma_\tau^2).$$

Le calcul de l'intégrale \mathcal{I}_2 donne

$$\mathcal{I}_2 = \frac{(s_0^2 \nu_0 / 2)^{\nu_0/2} \Gamma((\nu_0 + p)/2) 2^{\nu_0+p/2}}{(2\pi)^{p/2} \Gamma(\nu_0/2)} [pA(1 + (\tau - a)^2/A)]^{-(\nu_0+p)/2},$$

où $A = \frac{\nu_0 s_0^2}{p} + \frac{1}{p} \sum_{j=1}^p (\tau_j - a)^2$ et $a = \frac{1}{p} \sum_{j=1}^p \tau_j$.

Enfin, en intégrant $p_2(\theta \mid \text{data})$ par rapport à τ , on a

$$p_3(\theta \mid \text{data}) = \int_{\tau} p_2(\theta \mid \text{data}) d\tau = C_3 \mathcal{I}_3,$$

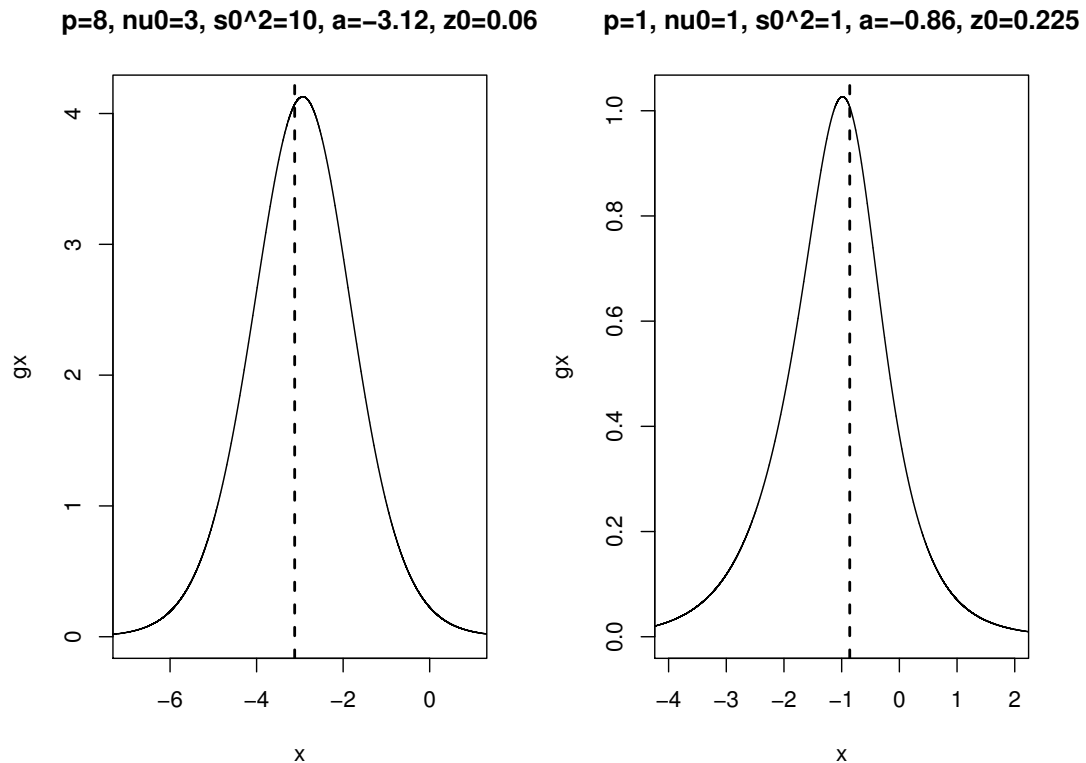
où C_3 est une constante par rapport à τ et

$$\mathcal{I}_3 = \int_{\tau} \left(1 + \frac{(\tau - a)^2}{A}\right)^{-b} \frac{e^{\tau}}{(e^{\tau} + z_0)^2} d\tau,$$

avec $b = (\nu_0 + p)/2$. Notons que la densité de $\xi = e^{\tau}$ est $z_0/(\xi + z_0)^2$, alors celle de τ est $f(\tau) = z_0 e^{\tau}/(e^{\tau} + z_0)^2$.

On considère $g(\tau) = -b \log\left(1 + \frac{(\tau - a)^2}{A}\right) + \tau - 2 \log(e^{\tau} + z_0)$, le logarithme de la fonction dans l'intégrale \mathcal{I}_3 . Le maximum de cette fonction ne pouvant pas être obtenu analytiquement, nous allons l'estimer en effectuant une approximation. La figure 2.1 ci-dessous montre la représentation graphique de la fonction g pour certaines valeurs des paramètres de la fonction g basées sur des données réelles, et la droite verticale pointillée marque le point a . Nous constatons que le point où la fonction g atteint son maximum est proche a ; nous allons donc approximer le maximum de g par $g(a)$ et utiliser ensuite la méthode de Laplace pour approximer l'intégrale \mathcal{I}_3 . Avant d'appliquer cette méthode à notre cas, nous allons au préalable décrire en quoi elle consiste.

FIGURE 2.1. Représentation graphique de la fonction g



La méthode de Laplace est l'approche habituellement utilisée pour faire une approximation analytique des intégrales souvent rencontrées dans l'inférence bayésienne (Gelman et al.(1995), page 304). Afin de décrire comment marche l'application standard de cette méthode, considérons $f(x)$ une fonction non-négative avec $x = (x_1, \dots, x_n)'$. Soient x_0 le mode de $\log f(x)$ et $H(x)$ sa matrice hessienne $n \times n$ dont l'élément (i, j) est $\partial^2 \log f(x) / \partial x_i \partial x_j$. Le développement en série de Taylor du second ordre de $\log f(x)$ autour de x_0 est le suivant :

$$\log f(x) \approx \log f(x_0) + (x - x_0) \frac{\partial \log f(x_0)}{\partial x} + \frac{1}{2} (x - x_0)' H(x_0) (x - x_0).$$

Étant donné que x_0 est le mode, $\partial \log f(x_0) / \partial x = 0$ et on a :

$$f(x) \approx f(x_0) \exp\left(\frac{1}{2} (x - x_0)' H(x_0) (x - x_0)\right).$$

Cette approximation de $f(x)$ est proportionnelle à une densité normale multivariée. Son intégrale donne alors

$$\begin{aligned} \int f(x) dx &\approx f(x_0) \int \exp\left(\frac{1}{2} (x - x_0)' H(x_0) (x - x_0)\right) dx \\ &= (2\pi)^{n/2} | -H(x_0) |^{-1/2} f(x_0). \end{aligned}$$

Ainsi, en appliquant l'approximation de Laplace à notre cas, nous avons

$$\mathcal{I}_3 \approx e^{g(a)} (2\pi)^{1/2} (-g''(a))^{-1/2},$$

$$\text{où } g(a) = a - 2 \log(e^a + z_0) \text{ et } g''(a) \equiv \left. \frac{d^2 g(\tau)}{d\tau^2} \right|_{\tau=a} = -\frac{2b}{A} - 2 \frac{z_0 e^a}{(e^a + z_0)^2}.$$

Finalement, l'approximation pour la densité a posteriori de θ est

$$p_3(\theta | \text{data}) \approx c_0 L(\theta | \text{data}) \mathcal{I}_1 K \mathcal{I}_3 \quad (2.14)$$

$$\text{où } K = \frac{(s_0^2 \nu_0 / 2)^{\nu_0 / 2} \Gamma((\nu_0 + p) / 2) 2^{\nu_0 + p} / 2}{(2\pi)^{p/2} \Gamma(\nu_0 / 2)} (pA)^{-(\nu_0 + p) / 2}.$$

Le logarithme de $p_3(\theta | \text{data})$ donne le résultat de la proposition 2, en ignorant les termes constants.

Avant de construire la vraisemblance restreinte et de calculer les dérivées de son logarithme par rapport à nos paramètres d'intérêt, il est utile au préalable de déterminer les dérivées de l'approximation du logarithme de la densité a posteriori $\ell(\theta)$ obtenue précédemment. En effet, les dérivées du logarithme de la vraisemblance restreinte dépendent de celles de $\ell(\theta)$.

2.4.1. Dérivées de $\ell(\theta)$ par rapport aux hyperparamètres

Les dérivées de $\ell(\theta)$ relatives à β sont les mêmes qu'avec $\mathcal{L}(\theta)$ obtenues en (2.7), (2.8), (2.9) et (2.10).

Afin de simplifier les dérivées de $\ell(\theta)$ par rapport à b_j , nous faisons une approximation pour le terme $\log \left| \Lambda_0 + \sum_{j=1}^p b_j b_j' \right|$ dans $\ell(\theta)$. Considérons A , une matrice carré d'ordre r et I la matrice identité d'ordre r . Soient $(\gamma_i)_{i=1, \dots, r}$ les valeurs propres de A , le logarithme du déterminant de A peut s'écrire comme

$$\begin{aligned} \log |A| &= \log \left(\prod_{i=1}^r \gamma_i \right) = \sum_{i=1}^r \log(\gamma_i) \\ &\approx \sum_{i=1}^r \log \left(\frac{1}{r} \text{trace}(A) \right) \\ &= r \log \left(\frac{1}{r} \text{trace}(A) \right). \end{aligned} \quad (2.15)$$

L'idée ici est d'approximer chaque valeur propre γ_i par leur moyenne.

Ainsi, dans notre cas on a

$$\log \left| \Lambda_0 + \sum_{j=1}^p b_j b_j' \right| \approx \log |\Lambda_0| + q \log \left(1 + \frac{1}{q} \sum_{j=1}^p b_j' \Lambda_0^{-1} b_j \right).$$

En effet, $\left| \Lambda_0 + \sum_{j=1}^p b_j b_j' \right| = |\Lambda_0| \left| I + \Lambda_0^{-1} \sum_{j=1}^p b_j b_j' \right|$ et

$\log \left| I + \Lambda_0^{-1} \sum_{j=1}^p b_j b_j' \right| \approx q \log \left(\frac{1}{q} \text{trace} \left(I + \Lambda_0^{-1} \sum_{j=1}^p b_j b_j' \right) \right)$, en utilisant le résultat en (2.15).

Or,

$$\begin{aligned} \text{trace} \left(I + \Lambda_0^{-1} \sum_{j=1}^p b_j b_j' \right) &= q + \sum_{j=1}^p \text{trace}(\Lambda_0^{-1} b_j b_j') \\ &= q + \sum_{j=1}^p \text{trace}(b_j' \Lambda_0^{-1} b_j) \\ &= q + \sum_{j=1}^p b_j' \Lambda_0^{-1} b_j. \end{aligned}$$

D'où $\log \left| I + \Lambda_0^{-1} \sum_{j=1}^p b_j b_j' \right| \approx q \log \left(1 + \frac{1}{q} \sum_{j=1}^p b_j' \Lambda_0^{-1} b_j \right)$.

Finalement, $\log \left| \Lambda_0 + \sum_{j=1}^p b_j b_j' \right| \approx \log |\Lambda_0| + q \log \left(1 + \frac{1}{q} \sum_{j=1}^p b_j' \Lambda_0^{-1} b_j \right)$.

Les dérivées de premier et second ordre de $\ell(\theta)$ par rapport à b_j sont

$$\begin{aligned}\frac{\partial \ell(\theta)}{\partial b_j} &= \frac{\partial \mathcal{L}(\theta)}{\partial b_j} - q(m_0 + p)/2 \frac{\partial \log(1 + \frac{1}{q} \sum_{j=1}^p b'_j \Lambda_0^{-1} b_j)}{\partial b_j} \\ &= \frac{\partial \mathcal{L}(\theta)}{\partial b_j} - q(m_0 + p) \frac{\Lambda_0^{-1} b_j}{q + \sum_{j=1}^p b'_j \Lambda_0^{-1} b_j}\end{aligned}\quad (2.16)$$

$$\begin{aligned}-\frac{\partial^2 \ell(\theta)}{\partial (b_j b'_j)} &= -\frac{\partial^2 \mathcal{L}(\theta)}{\partial (b_j b'_j)} + q(m_0 + p) \frac{[\Lambda_0^{-1}(q + \sum_{j=1}^p b'_j \Lambda_0^{-1} b_j) - 2\Lambda_0^{-1} b_j b'_j \Lambda_0^{-1}]}{(q + \sum_{j=1}^p b'_j \Lambda_0^{-1} b_j)^2} \\ &= M_{b_j} + N_{b_j} \equiv H_{b_j}\end{aligned}\quad (2.17)$$

Pour $i \neq k$, la dérivée croisée de $\ell(\theta)$ par rapport à b_j et b_k est

$$\frac{\partial^2 \ell(\theta)}{\partial b_j \partial b_k} = 2q(m_0 + p) \frac{\Lambda_0^{-1} b_k b'_j \Lambda_0^{-1}}{q + \sum_{j=1}^p b'_j \Lambda_0^{-1} b_j)^2}.\quad (2.18)$$

La dérivée croisée de $\ell(\theta)$ par rapport à b_j et τ_j est la même qu'avec $\mathcal{L}(\theta)$ telle qu'obtenue en (2.13).

Les dérivées première, seconde et croisée de $\ell(\theta)$ relatives à τ_j sont

$$\frac{\partial \ell(\theta)}{\partial \tau_j} = \frac{\partial \mathcal{L}(\theta)}{\partial \tau_j} - b \frac{\partial \log A}{\partial \tau_j} + \frac{\partial g(a)}{\partial \tau_j} - \frac{1}{2} \frac{\partial \log(-g''(a))}{\partial \tau_j}\quad (2.19)$$

$$\frac{\partial^2 \ell(\theta)}{\partial \tau_j^2} = \frac{\partial^2 \mathcal{L}(\theta)}{\partial \tau_j^2} - b \frac{\partial^2 \log A}{\partial \tau_j^2} + \frac{\partial^2 g(a)}{\partial \tau_j^2} - \frac{1}{2} \frac{\partial^2 \log(-g''(a))}{\partial \tau_j^2}\quad (2.20)$$

$$\frac{\partial^2 \ell(\theta)}{\partial \tau_j \partial \tau_k} = \frac{\partial^2 \mathcal{L}(\theta)}{\partial \tau_j \partial \tau_k} - b \frac{\partial^2 \log A}{\partial \tau_j \partial \tau_k} + \frac{\partial^2 g(a)}{\partial \tau_j \partial \tau_k} - \frac{1}{2} \frac{\partial^2 \log(-g''(a))}{\partial \tau_j \partial \tau_k}.\quad (2.21)$$

Ces dérivées incluent les termes $\partial \log A / \partial \tau_j$, $\partial^2 \log A / \partial \tau_j^2$ et $\partial^2 \log A / \partial \tau_j \partial \tau_k$ qui sont calculés comme suit

$$\begin{aligned}\frac{\partial \log A}{\partial \tau_j} &= \frac{2(\tau_j - a)}{pA}, \\ \frac{\partial^2 \log A}{\partial \tau_j^2} &= \frac{2[(p-1)A - 2(\tau_j - a)^2]}{p^2 A^2}, \\ \frac{\partial^2 \log A}{\partial \tau_j \partial \tau_k} &= \frac{2[(2(\tau_j - a)(\tau_k - a) - A)]}{p^2 A^2}.\end{aligned}$$

Elles contiennent également les termes $\partial g(a)/\partial\tau_j$, $\partial^2 g(a)/\partial\tau_j^2$ et $\partial^2 g(a)/\partial\tau_j\partial\tau_k$ qui sont évalués par

$$\begin{aligned}\frac{\partial g(a)}{\partial\tau_j} &= 1/p - \frac{2e^a}{p(e^a + z_0)}, \\ \frac{\partial^2 g(a)}{\partial\tau_j^2} &= -\frac{2z_0e^a}{p^2(e^a + z_0)^2}, \\ \frac{\partial^2 g(a)}{\partial\tau_j\partial\tau_k} &= -\frac{2z_0e^a}{p^2(e^a + z_0)^2}.\end{aligned}$$

L'expression $\frac{1}{2}\frac{\partial\log(-g''(a))}{\partial\tau_j}$ dans la dérivée d'ordre 1 en (2.19) donne

$$\frac{1}{2}\frac{\partial\log(-g''(a))}{\partial\tau_j} = \frac{\partial\log B}{\partial\tau_j} = \frac{B'}{B},$$

où

$$\begin{aligned}B &\equiv \frac{1}{2}(-g''(a)) = \frac{b}{A} + \frac{e^a z_0}{(e^a + z_0)^2}, \\ B' &\equiv \frac{\partial B}{\partial\tau_j} = -\frac{2b(\tau_j - a)}{pA^2} + \frac{z_0(z_0e^a - e^{2a})}{p(e^a + z_0)^3}.\end{aligned}$$

Nous avons ensuite le terme

$$\frac{1}{2}\frac{\partial^2\log(-g''(a))}{\partial\tau_j^2} = \frac{B''B - (B')^2}{B^2}$$

dans la dérivée d'ordre 2 en (2.20), où

$$B'' \equiv \frac{\partial^2 B}{\partial\tau_j^2} = -\frac{2b[(p-1)A - 4(\tau_j - a)^2]}{p^2A^3} + \frac{z_0e^a(e^{2a} - 4z_0e^a + z_0^2)}{p^2(e^a + z_0)^4}.$$

Enfin, l'expression $\frac{1}{2}\frac{\partial^2\log(-g''(a))}{\partial\tau_j\partial\tau_k}$ dans la dérivée croisée en (2.21) est donnée par

$$\frac{1}{2}\frac{\partial^2\log(-g''(a))}{\partial\tau_j\partial\tau_k} = \frac{\frac{\partial B'}{\partial\tau_k}B - \frac{\partial B}{\partial\tau_k}B'}{B^2},$$

où

$$\begin{aligned}\frac{\partial B'}{\partial\tau_k} &= \frac{2b[A + 4(\tau_k - a)(\tau_j - a)]}{p^2A^3} + \frac{z_0e^a(e^{2a} - 4z_0e^a + z_0^2)}{p^2(e^a + z_0)^4}, \\ \frac{\partial B}{\partial\tau_k} &= -\frac{2b(\tau_k - a)}{pA^2} + \frac{z_0(z_0e^a - e^{2a})}{p(e^a + z_0)^3}.\end{aligned}$$

À présent, nous allons faire un ajustement de la densité a posteriori des hyperparamètres $p_3(\theta | \text{data})$ obtenue en (2.14), en considérant l'approche du maximum de vraisemblance restreint (REML). Ceci produit en effet une meilleure estimation des paramètres $\{\tau_j\}$ que ceux obtenus en maximisant $p_3(\theta | \text{data})$ (Liang et Zeger 1995).

Nous rappelons que $p_3(\theta | \text{data}) = c_0 L(\theta | \text{data}) \mathcal{I}_1(b^*) K(\tau^*) \mathcal{I}_3(\tau^*)$, avec $b^* = (b_1, \dots, b_p)'$ et $\tau^* = (\tau_1, \dots, \tau_p)'$.

L'approche REML estime les $\{\tau_j\}$ en considérant β et b^* comme des paramètres nuisibles. Ainsi, la vraisemblance restreinte est obtenue en intégrant $p_3(\theta | \text{data})$ par rapport aux paramètres $\phi \equiv (\beta, b^*)$. Nous allons calculer cette intégrale en utilisant l'approximation de Laplace. Pour cela, considérons le développement en série de Taylor du second ordre de $\log p_3(\theta | \text{data})$ suivant :

$$\begin{aligned} \log p_3(\theta | \text{data}) &\approx \log p_3(\hat{\phi}, \tau^*) + \frac{\partial \log p_3(\hat{\phi}, \tau^*)}{\partial \phi} (\phi - \hat{\phi}) \\ &\quad + \frac{1}{2} (\phi - \hat{\phi})' \frac{\partial^2 \log p_3(\hat{\phi}, \tau^*)}{\partial (\phi \phi')} (\phi - \hat{\phi}). \end{aligned}$$

Or, $\frac{\partial \log p_3(\hat{\phi}, \tau^*)}{\partial \phi} = 0$.

Alors en prenant l'exponentielle du développement précédent on a

$p_3(\theta | \text{data}) \approx p_3(\hat{\phi}, \tau^*) e^{-\frac{1}{2}(\phi - \hat{\phi})' \hat{H} \phi (\phi - \hat{\phi})}$, où $\hat{H} \phi = -\frac{\partial^2 \ell(\theta)}{\partial (\phi \phi')} \Big|_{\hat{\phi}}$ et on a

$$\frac{\partial^2 \ell(\theta)}{\partial (\phi \phi')} = \begin{pmatrix} \frac{\partial^2 \ell(\theta)}{\partial (\beta \beta')} & \frac{\partial^2 \ell(\theta)}{\partial (\beta b^*)} \\ \frac{\partial^2 \ell(\theta)}{\partial (\beta b^*)} & \frac{\partial^2 \ell(\theta)}{\partial (b^* b^*)} \end{pmatrix}.$$

En intégrant $p_3(\theta | \text{data})$ par rapport à ϕ , on a la vraisemblance restreinte suivante :

$$p_4(\tau^* | \text{data}) = c_4 \left| \hat{H} \phi \right|^{-1/2} L(\hat{\beta}, \hat{b}^*, \tau^* | \text{data}) \mathcal{I}_1(\hat{b}^*) K(\tau^*) \mathcal{I}_3(\tau^*).$$

En ignorant les termes constants de $p_4(\tau^* | \text{data})$ on définit son logarithme comme suit :

$$\begin{aligned} \ell_R(\theta) &\equiv \log \left\{ |H_\phi|^{-1/2} L(\theta | \text{data}) \Big|_{\Lambda_0 + \sum_{j=1}^p b_j b_j'} \right\}^{-(m_0+p)/2} \\ &\quad \times A^{-(\nu_0+p)/2} e^{g(a)} \left(-\frac{d^2 g(\tau)}{d\tau^2} \Big|_{\tau=a} \right)^{-1/2} \Big\}. \end{aligned}$$

Nous allons approximer le terme $\log(|H_\phi|)$ dans $\ell_R(\theta)$ afin de simplifier les dérivées relatives aux hyperparamètres θ . L'approche que nous utilisons ici pour faire cette approximation de la vraisemblance restreinte est différente de celle proposée par Christiansen et

Morris (1997)(voir page 622) car $|H_\phi|$ dépend non seulement de β mais aussi des effets aléatoires $\tau^* = (\tau_1, \dots, \tau_p)'$.

En nous référant au résultat obtenu en (2.15), on a $\log(|H_\phi|) \approx d \log(\frac{1}{d} \text{trace} H_\phi)$, où $d \equiv r + pq$. Mais $\text{trace} H_\phi = \text{trace} H_\beta + \sum_{j=1}^p \text{trace} H_{b_j}$, avec

$$\begin{aligned} \text{trace} H_\beta &= \text{trace} \left(\sum_{j=1}^p \sum_{i=1}^{n_j} X_{ij} X_{ij}^t e_{ij} B_{ij} \lambda_{ij}^* \right) = \sum_{j=1}^p \sum_{i=1}^{n_j} \sum_{t=0}^{r-1} e_{ij} B_{ij} \lambda_{ij}^* x_{ij,t}^2, \\ \text{trace} H_{b_j} &= \text{trace} \left(\sum_{i=1}^{n_j} W_{ij} W_{ij}^t e_{ij} B_{ij} \lambda_{ij}^* \right) = \sum_{i=1}^{n_j} \sum_{t=0}^{q-1} e_{ij} B_{ij} \lambda_{ij}^* w_{ij,t}^2. \end{aligned}$$

Les matrices H_β et H_{b_j} sont définies par les formules (2.8) et (2.17).

On considère l'approximation \tilde{B}_{ij} de B_{ij} telle que :

$$\begin{aligned} \tilde{B}_{ij} &= \frac{e^{\tau_j}}{e^{\tau_j} + e_{ij} m_{0j}}, \\ \text{où } m_{0j} &= E \left(\sum_{i=1}^{n_j} z_{ij} / \sum_{i=1}^{n_j} e_{ij} \right). \end{aligned}$$

De plus, le taux observé y_{ij} étant un estimateur de λ_{ij}^* , on simplifie les expressions de $\text{trace} H_\beta$ et $\text{trace} H_{b_j}$ en remplaçant donc λ_{ij}^* par y_{ij} qui lui ne dépend pas de B_{ij} ni de μ_{ij} . Ceci nous permettra de calculer plus aisément les dérivées de $\ell_R(\theta)$.

D'où,

$$\begin{aligned} \text{trace} H_\beta &\approx \sum_{j=1}^p \sum_{i=1}^{n_j} \sum_{t=0}^{r-1} e_{ij} \tilde{B}_{ij} y_{ij} x_{ij,t}^2, \\ \text{trace} H_{b_j} &\approx \sum_{i=1}^{n_j} \sum_{t=0}^{q-1} e_{ij} \tilde{B}_{ij} y_{ij} w_{ij,t}^2. \end{aligned}$$

2.4.2. Dérivées de $\ell_R(\theta)$ par rapport aux hyperparamètres

Les dérivées de $\ell_R(\theta)$ relatives à β sont les mêmes qu'avec $\ell(\theta)$ telles que définies dans les formules (2.7), (2.8), (2.9) et (2.10).

Les dérivées de $\ell_R(\theta)$ relatives à b_j sont les mêmes qu'avec $\ell(\theta)$ telles que définies dans les formules (2.16), (2.17), et (2.18).

Les dérivées première, seconde et croisée de $\ell_R(\theta)$ relatives à τ_j sont

$$\frac{\partial \ell_R(\theta)}{\partial \tau_j} = \frac{\partial \ell(\theta)}{\partial \tau_j} - \frac{1}{2} \frac{\partial \log(|H_\phi|)}{\partial \tau_j} \quad (2.22)$$

$$\frac{\partial^2 \ell_R(\theta)}{\partial \tau_j^2} = \frac{\partial^2 \ell(\theta)}{\partial \tau_j^2} - \frac{1}{2} \frac{\partial^2 \log(|H_\phi|)}{\partial \tau_j^2} \quad (2.23)$$

$$\frac{\partial^2 \ell_R(\theta)}{\partial \tau_j \partial \tau_k} = \frac{\partial^2 \ell(\theta)}{\partial \tau_j \partial \tau_k} - \frac{1}{2} \frac{\partial^2 \log(|H_\phi|)}{\partial \tau_j \partial \tau_k}. \quad (2.24)$$

Le terme $\frac{\partial \log(|H_\phi|)}{\partial \tau_j}$ dans la dérivée d'ordre 1 en (2.22) s'exprime comme

$$\frac{\partial \log(|H_\phi|)}{\partial \tau_j} = d \frac{(\text{trace} H_\phi)'}{\text{trace} H_\phi},$$

on définit

$$(\text{trace} H_\phi)' \equiv \frac{\partial (\text{trace} H_\phi)}{\partial \tau_j} = (\text{trace} H_\beta)' + (\text{trace} H_{b_j})',$$

où

$$(\text{trace} H_\beta)' \equiv \frac{\partial (\text{trace} H_\beta)}{\partial \tau_j} = \sum_{i=1}^{n_j} \sum_{t=0}^{r-1} e_{ij} y_{ij} x_{ij,t}^2 \tilde{B}_{ij} (1 - \tilde{B}_{ij}),$$

$$(\text{trace} H_{b_j})' \equiv \frac{\partial (\text{trace} H_{b_j})}{\partial \tau_j} = \sum_{i=1}^{n_j} \sum_{t=0}^{q-1} e_{ij} y_{ij} w_{ij,t}^2 \tilde{B}_{ij} (1 - \tilde{B}_{ij}).$$

Dans la dérivée d'ordre 2 en (2.23), le calcul de $\frac{\partial^2 \log(|H_\phi|)}{\partial \tau_j^2}$ donne

$$\frac{\partial^2 \log(|H_\phi|)}{\partial \tau_j^2} = d \frac{(\text{trace} H_\phi)'' (\text{trace} H_\phi) - (\text{trace} H_\phi)' (\text{trace} H_\phi)'}{(\text{trace} H_\phi)^2},$$

avec

$$(\text{trace} H_\phi)'' \equiv \frac{\partial^2 (\text{trace} H_\phi)}{\partial \tau_j^2} = (\text{trace} H_\beta)'' + (\text{trace} H_{b_j})'',$$

où

$$(\text{trace} H_\beta)'' \equiv \frac{\partial^2 (\text{trace} H_\beta)}{\partial \tau_j^2} = \sum_{i=1}^{n_j} \sum_{t=0}^{r-1} e_{ij} y_{ij} x_{ij,t}^2 \tilde{B}_{ij} (1 - \tilde{B}_{ij}) (1 - 2\tilde{B}_{ij}),$$

$$(\text{trace} H_{b_j})'' \equiv \frac{\partial^2 (\text{trace} H_{b_j})}{\partial \tau_j^2} = \sum_{i=1}^{n_j} \sum_{t=0}^{q-1} e_{ij} y_{ij} w_{ij,t}^2 \tilde{B}_{ij} (1 - \tilde{B}_{ij}) (1 - 2\tilde{B}_{ij}).$$

Nous avons enfin dans la dérivée croisée en (2.24), l'expression $\frac{\partial^2 \log(|H_\phi|)}{\partial \tau_j \partial \tau_k}$ qui est donnée

par

$$\frac{\partial^2 \log(|H_\phi|)}{\partial \tau_j \partial \tau_k} = d \frac{(\frac{\partial (\text{trace} H_\phi)'}{\partial \tau_k}) (\text{trace} H_\phi) - (\frac{\partial (\text{trace} H_\phi)'}{\partial \tau_j}) (\text{trace} H_\phi)'}{(\text{trace} H_\phi)^2},$$

où

$$\begin{aligned}\frac{\partial(\text{trace}H_\phi)'}{\partial\tau_k} &= 0, \\ \frac{\partial\text{trace}H_\phi}{\partial\tau_k} &= \sum_{i=1}^{n_j} \sum_{t=0}^{r-1} e_{ik}y_{ik} \cdot x_{ik,t}^2 \tilde{B}_{ik}(1 - \tilde{B}_{ik}).\end{aligned}$$

Les estimateurs $\hat{\theta} = (\hat{\beta}, \hat{\tau}_1, \dots, \hat{\tau}_p, \hat{b}_1, \dots, \hat{b}_p)$ sont obtenus par la maximisation de $\ell_R(\theta)$ en utilisant les dérivées calculées précédemment et ceci à travers la méthode de Newton-Raphson connue pour sa convergence rapide. Le programme de maximisation que nous avons effectué en utilisant cette méthode s'est fait en trois étapes. A partir de valeurs initiales données aux hyperparamètres $\theta = (\beta, \tau_1, \dots, \tau_p, b_1, \dots, b_p)$, nous avons d'abord déterminé les estimateurs $\hat{\tau}_1, \dots, \hat{\tau}_p$ qui maximisent $\ell_R(\theta)$; puis nous avons utilisé ces estimateurs des τ_j pour calculer $\hat{\beta}$ et enfin, nous avons déterminé les estimateurs $\hat{b}_1, \dots, \hat{b}_p$ en utilisant les valeurs $\hat{\tau}_1, \dots, \hat{\tau}_p$ et $\hat{\beta}$ obtenues précédemment.

En nous référant à la théorie asymptotique, la distribution a posteriori des hyperparamètres est approximativement une distribution conjointe normale de moyenne $\hat{\theta} = (\hat{\beta}, \hat{\tau}_1, \dots, \hat{\tau}_p, \hat{b}_1, \dots, \hat{b}_p)$ et avec une matrice de variance-covariance égale à l'inverse de la matrice d'information de Fisher observée (Gelman et al.(1995), pages 94-95, 100-101) évaluée à $\hat{\theta}$. Particulièrement, les éléments de la diagonale de la matrice de variance-covariance représentent les variances estimées des hyperparamètres.

2.5. ALGORITHME POUR L'ESTIMATION DES HYPERPARAMÈTRES

Dans cette section nous décrivons l'algorithme numérique utilisé pour le calcul des estimateurs $\hat{\theta} = (\hat{\beta}, \hat{\tau}_1, \dots, \hat{\tau}_p, \hat{b}_1, \dots, \hat{b}_p)'$. La fonction à maximiser est le logarithme de la vraisemblance restreinte que nous avons noté $\ell_R(\theta)$. Le programme de maximisation, basé sur la méthode de Newton-Raphson, s'est effectué en trois étapes.

Estimation des τ_j

Elle se fait selon les étapes suivantes :

1. On donne des valeurs initiales $\theta^{(0)} = (\beta^{(0)}, \tau_1^{(0)}, \dots, \tau_p^{(0)}, b_1^{(0)}, \dots, b_p^{(0)})'$ aux paramètres.
2. On définit $\tau^{(s)} = (\tau_1^{(s)}, \dots, \tau_p^{(s)})'$, l'approximation de $\hat{\tau} = (\hat{\tau}_1, \dots, \hat{\tau}_p)'$ à l'itération s

($s = 0, 1, 2, \dots$).

3. On définit $u_{\tau_j} = \partial \ell_R(\theta) / \partial \tau_j$ et $h_{\tau_j} = \partial^2 \ell_R(\theta) / \partial \tau_j^2$. Pour chaque τ_j de l'étape 2, on considère $u_{\tau_j}^{(t)}$ et $h_{\tau_j}^{(t)}$ les valeurs de u_{τ_j} et h_{τ_j} évaluées à $\tau_j^{(t)}$, la t^{ieme} approximation de $\hat{\tau}_j$ ($t = 0, 1, 2, \dots$). La valeur suivante à l'itération $t + 1$ est donnée par

$$\tau_j^{(t+1)} = \tau_j^{(t)} - (h_{\tau_j}^{(t)})^{-1} u_{\tau_j}^{(t)} \text{ (Agresti(2002), pages 143-144).}$$

On continue les itérations jusqu'à ce que la différence en valeur absolue entre $\tau_j^{(t+1)}$ et $\tau_j^{(t)}$ soit suffisamment proche de zéro. Après avoir fait ceci pour chaque τ_j on obtient $\tau^{(s+1)} = (\tau_1^{(s+1)}, \dots, \tau_p^{(s+1)})'$.

4. On considère ensuite le vecteur de différences $\tau^{(s+1)} - \tau^{(s)}$ dont on prend le plus grand élément en valeur absolue. Si cette valeur n'est pas suffisamment proche de zéro, on retourne à l'étape 2 en utilisant les valeurs de $\tau^{(s+1)}$.

Estimation de β

Le paramètre β est de dimension r , $\beta = (\beta_0, \dots, \beta_{r-1})'$.

1. On considère les valeurs initiales $\theta^{(0)} = (\beta^{(0)}, \tau_1^{(0)}, \dots, \tau_p^{(0)}, b_1^{(0)}, \dots, b_p^{(0)})'$, où les $\tau_j^{(0)}$ cette fois prennent les valeurs des estimateurs $\hat{\tau}$ obtenus précédemment.

2. On définit $u_\beta = (\partial \ell_R(\theta) / \partial \beta_0, \dots, \partial \ell_R(\theta) / \partial \beta_{r-1})'$ et H_β la matrice hessienne $r \times r$ dont l'élément (i, j) est $\partial^2 \ell_R(\theta) / \partial \beta_i \partial \beta_j$. Soient $u_\beta^{(t)}$ et $H_\beta^{(t)}$ les valeurs de u_β et H_β évaluées à $\beta^{(t)}$, la t^{ieme} approximation de $\hat{\beta}_j$ ($t = 0, 1, 2, \dots$). La valeur suivante à l'itération $t + 1$ est donnée par

$$\beta_j^{(t+1)} = \beta_j^{(t)} - (H_\beta^{(t)})^{-1} u_\beta^{(t)}.$$

3. En considérant le vecteur de différences $\beta^{(t+1)} - \beta^{(t)}$, si son plus grand élément en valeur absolue n'est pas suffisamment proche de zéro, on retourne à l'étape 2 en utilisant les valeurs de $\beta^{(t+1)}$.

Estimation des b_j

Le programme pour l'estimation des b_j est le même que celui des τ_j , mais le vecteur des valeurs initiales $\theta^{(0)}$ contient les estimateurs $\hat{\tau}$ et $\hat{\beta}$ déjà obtenus.

Nous avons rencontré des problèmes de convergence pour l'estimation des τ_j lorsque nous considérons le vecteur entier pour faire la maximisation. Pour résorber le problème,

nous avons adopté une estimation un par un à l'intérieur d'un cycle de Newton-Raphson comme décrit plus haut.

En ce qui concerne le choix des valeurs initiales, nous avons procédé comme suit. Pour le paramètre β , nous considérons la régression linéaire des $\log(y_{ij})$ en fonction des variables explicatives $X_{ij} = (x_{ij,0}, \dots, x_{ij,r-1})'$, les y_{ij} étant les taux observés. Les coefficients de cette régression sont pris comme valeurs initiales de β . Étant donné que dans le niveau 2 de notre modèle on a $\lambda_{ij}|\theta \sim \text{Gam}(\xi_j, \xi_j/\mu_{ij})$, on en tire que $E(\lambda_{ij}|\theta) = \mu_{ij}$ et $\xi_j = \mu_{ij}^2/\text{var}(\lambda_{ij}|\theta)$. Ainsi, nous prenons comme valeur initiale pour $\tau_j = \log(\xi_j)$, $\tau_j^{(0)} = \log(\text{moy}(y_j)^2/\text{var}(y_j))$, où $\text{moy}(y_j)$ et $\text{var}(y_j)$ sont la moyenne et la variance empiriques des taux observés dans la strate j .

En considérant la régression linéaire des $\log(y_{ij})$ en fonction des X_{ij} à l'intérieur de la strate j , $E(\log(y_{ij})) = \alpha_{0j} + \alpha_{1j}x_{ij,1} + \dots + \alpha_{r-1j}x_{ij,r-1}$, α_{0j} est pris comme valeur initiale de b_j .

Après l'estimation des hyperparamètres θ , il nous reste à présent à déterminer la distribution a posteriori et les deux premiers moments de chaque paramètre de la loi de Poisson λ_{ij} , c'est ce qui l'objet de la section qui suit.

2.6. DISTRIBUTION DES PARAMÈTRES INDIVIDUELS $\{\lambda_{ij}\}$

Ayant obtenu la distribution approximative des hyperparamètres $\theta = (\beta, \xi_1, \dots, \xi_p, b_1, \dots, b_p)$, nous allons l'utiliser pour calculer la distribution de chaque λ_{ij} . Il est aussi question de calculer sa moyenne et sa variance a posteriori $E(\lambda_{ij}|\text{data})$ et $\text{var}(\lambda_{ij}|\text{data})$.

Il est utile au préalable de déterminer la moyenne et la variance de $\mu_{ij} = e^{X'_{ij}\beta + W'_{ij}b_j}$ car ils interviennent dans le calcul de $E(\lambda_{ij}|\text{data})$ et $\text{var}(\lambda_{ij}|\text{data})$. La distribution de μ_{ij} est une distribution lognormale. En effet, $\log(\mu_{ij}) = X'_{ij}\beta + W'_{ij}b_j$ suit une loi normale telle que :

$$E(\log \mu_{ij}) = X'_{ij}E(\beta) + W'_{ij}E(b_j) \text{ et}$$

$$\text{var}(\log \mu_{ij}) = X'_{ij}\text{var}(\beta)X_{ij} + W'_{ij}\text{var}(b_j)W_{ij} + 2X'_{ij}\text{cov}(\beta, b_j)W_{ij},$$

où $E(\beta) = \hat{\beta}$, $E(b_j) = \hat{b}_j$ et $\text{var}(\beta)$, $\text{var}(b_j)$, $\text{cov}(\beta, b_j)$ sont des éléments de la matrice de variance-covariance de θ mentionnée à la fin de la section 2.4.

On a donc :

$$\begin{aligned}\hat{\mu}_{ij} &\equiv E(\mu_{ij}|\text{data}) = \exp[X'_{ij}E(\beta) + W'_{ij}E(b_j) + \text{var}(\log \mu_{ij})/2], \\ \hat{\sigma}_{\mu_{ij}}^2 &\equiv \text{var}(\mu_{ij}|\text{data}) = \hat{\mu}_{ij}^2 \exp[\text{var}(\log \mu_{ij}) - 1].\end{aligned}$$

La moyenne et la variance de λ_{ij} sont

$$\begin{aligned}\hat{\lambda}_{ij} &\equiv E(\lambda_{ij}|\text{data}) = E\lambda_{ij}^* \\ &= (1 - EB_{ij})y_{ij} + E(\mu_{ij}B_{ij}),\end{aligned}\quad (2.25)$$

$$\begin{aligned}\hat{\sigma}_{\lambda_{ij}}^2 &\equiv \text{var}(\lambda_{ij}|\text{data}) = E(\sigma_{ij}^*)^2 + \text{var}E(\lambda_{ij}^*), \\ &= E\lambda_{ij}^*(1 - EB_{ij})/e_{ij} + \text{var}(\mu_{ij}B_{ij} - B_{ij}y_{ij}),\end{aligned}\quad (2.26)$$

où λ_{ij}^* et $(\sigma_{ij}^*)^2$ ont été obtenues en (2.3) et (2.4).

On remarque que des espérances de la forme $E(\mu_{ij}^s B_{ij}^t)$ pour $s, t = 0, 1, 2$ apparaissent dans ces formules de $\hat{\lambda}_{ij}$ et $\hat{\sigma}_{\lambda_{ij}}^2$. Alors pour calculer ces dernières, nous devons au préalable évaluer les expressions $E(\mu_{ij}^s B_{ij}^t)$ ($s, t = 0, 1, 2$).

Proposition 3. *Pour $\beta, \tau_1, \dots, \tau_p, b_1, \dots, b_p$ suivant une distribution conjointe normale, on définit $h_{ij} \equiv \text{cov}[X'_{ij}\beta + W'_{ij}b_j, \tau_j - (X'_{ij}\beta + W'_{ij}b_j)]/\text{var}[\tau_j - (X'_{ij}\beta + W'_{ij}b_j)]$. Alors pour $t \geq 0$,*

$$E_s B_{ij}^t \equiv \frac{E(\mu_{ij}^s B_{ij}^t)}{E\mu_{ij}^s} = \frac{EB_{ij}^t (B_{ij}/(1 - B_{ij}))^{sh_{ij}}}{E(B_{ij}/(1 - B_{ij}))^{sh_{ij}}}.\quad (2.27)$$

Preuve : Posons $u_1 = X'_{ij}\beta + W'_{ij}b_j$ et $u_2 = \tau_j - (X'_{ij}\beta + W'_{ij}b_j)$, alors $h_{ij} = \text{cov}(u_1, u_2)/\text{var}(u_2)$. On rappelle que $B_{ij} = e^{\tau_j}/(e^{\tau_j} + e_{ij}\mu_{ij}) = e^{u_2}/(e^{u_2} + e_{ij})$. Alors $u_1 - h_{ij}u_2$ est indépendant de u_2 et par suite, de B_{ij} . En effet, les distributions de $u_1 - h_{ij}u_2$ et u_2 sont des lois normales (car fonctions linéaires de β, b_j et τ_j) et on a aussi

$$\begin{aligned}\text{cov}(u_1 - h_{ij}u_2, u_2) &= \text{cov}(u_1, u_2)\text{var}(u_2)/\text{var}(u_2) - h_{ij}\text{cov}(u_2, u_2) \\ &= h_{ij}\text{var}(u_2) - h_{ij}\text{var}(u_2) \\ &= 0.\end{aligned}$$

Par définition, $\mu_{ij} \equiv e^{X'_{ij}\beta + W'_{ij}b_j}$, d'où $\mu_{ij} = e^{u_1}$. On a également

$$\begin{aligned}\frac{B_{ij}}{1 - B_{ij}} &= \frac{e^{\tau_j}}{e_{ij}\mu_{ij}} = \frac{e^{u_1+u_2}}{e_{ij}e^{u_1}} \\ &= \frac{e^{u_2}}{e_{ij}}.\end{aligned}$$

Ainsi,

$$\begin{aligned}
\frac{E(\mu_{ij}^s B_{ij}^t)}{E\mu_{ij}^s} &= \frac{E(e^{s(u_1 - h_{ij}u_2)} e^{sh_{ij}u_2} B_{ij}^t)}{E(e^{s(u_1 - h_{ij}u_2)} e^{sh_{ij}u_2})} \\
&= \frac{E(e^{sh_{ij}u_2} B_{ij}^t)}{E(e^{sh_{ij}u_2})} \text{ (car } u_1 - h_{ij}u_2 \text{ indépendant de } u_2 \text{ et } B_{ij}) \\
&= \frac{EB_{ij}^t (B_{ij}/(1 - B_{ij}))^{sh_{ij}}}{E(B_{ij}/(1 - B_{ij}))^{sh_{ij}}}.
\end{aligned}$$

Ayant donc $E_s B_{ij}^t$, on peut déterminer $E(\mu_{ij}^s B_{ij}^t) = E_s B_{ij}^t E\mu_{ij}^s$. Cependant, le calcul des espérances dans l'expression de $E_s B_{ij}^t$ en (2.27) nécessite d'avoir la distribution de B_{ij} . La méthode ADM (méthode de la densité ajustée) présentée par Christiansen et Morris (1997) que nous allons utiliser ici permet de trouver les paramètres de la distribution de B_{ij} . L'ADM est une méthode qui permet de faire une approximation de la distribution a posteriori d'un paramètre afin d'obtenir ses moments. On rappelle que $B_{ij} = e^{\tau_j} / (e^{\tau_j} + e_{ij}\mu_{ij})$ et donc $0 < B_{ij} < 1$, la distribution beta est alors choisie pour B_{ij} . Dans ce cas, l'ADM va ajuster la densité de B_{ij} afin d'obtenir une distribution beta pour B_{ij} et par suite ses moments.

Pour montrer comment marche l'ADM pour la distribution beta, considérons $0 \leq B \leq 1$ avec une fonction de densité unimodale $p(B)$. Soit $u \equiv \text{logit}(B) = \log(B/(1 - B))$, on suppose que $u \sim N(m, w^2)$. Alors la densité de B , $p(B)$, n'est pas celle d'une distribution beta. En effet, la fonction de repartion de B s'écrit $P(B < b) = P(\log(B/(1 - B)) < \log(b/(1 - b))) = \Phi((\log(b/(1 - b)) - m)/w)$. Ainsi, la densité de B prend la forme

$$p(B) = \frac{1}{w} \phi\left(\frac{u - m}{w}\right) \frac{du}{dB} = \frac{1}{w} \phi\left(\frac{u - m}{w}\right) \frac{1}{B(1 - B)}$$

où $\phi \sim N(0, 1)$. Dans cette écriture, l'utilisation de B au lieu de b est faite juste pour des raisons de forme. On définit $l(B) = \log(B(1 - B)p(B)) = -\frac{1}{2w^2}(u - m)^2 - \frac{1}{2}\log(2\pi w^2)$. On ajuste ainsi la densité $p(B)$ avec le facteur $B(1 - B)$ nécessaire pour la distribution beta. On a $l'(B) = -\frac{1}{w^2}(u - m) \frac{du}{dB} = 0$ pour $\text{logit}(\hat{B}) = m$, c'est-à-dire $\hat{B} = e^m / (1 + e^m)$. $l''(B) = -\frac{1}{w^2}(u - m) \frac{d^2u}{dB^2} - \frac{1}{w^2} \left(\frac{du}{dB}\right)^2$. On définit $i_0 \equiv -l''(\hat{B}) = \frac{1}{w^2 \hat{B}^2 (1 - \hat{B}^2)}$. Si $p(B)$ est la densité de la distribution Beta(a_1, a_2), alors elle est proportionnelle à $B^{a_1-1}(1 - B)^{a_2-1}$ et dans ce cas $l(B) \propto \log(B^{a_1}(1 - B^{a_2}))$. De ceci on obtient $\hat{B} = a_1 / (a_1 + a_2)$, la racine de $l'(B) = 0$ et $i_0 \equiv -l''(\hat{B}) = (a_1 + a_2) / (\hat{B}(1 - \hat{B}))$. Alors, $a_1 = \hat{B}^2(1 - \hat{B})i_0$ et

$a_2 = \hat{B}(1 - \hat{B})^2 i_0$, on a donc l'approximation ADM

$$B \sim \text{Beta}(\hat{B}^2(1 - \hat{B})i_0, \hat{B}(1 - \hat{B})^2 i_0).$$

Ayant ainsi la densité approximative $p(B) \equiv \text{Beta}(a_1, a_2)$ de B , on peut donc estimer les intégrales $EB^t(B/(1 - B))^k \approx \int_0^1 b^{t+k}(1 - b)^{-k} p(b) db$ et $E(B/(1 - B))^k \approx \int_0^1 b^k(1 - b)^{-k} p(b) db$, dont le rapport donne

$$\frac{EB^t(B/(1 - B))^k}{E(B/(1 - B))^k} \approx \frac{(a_1 + k)^{(t)}}{(a_1 + a_2)^{(t)}}, \quad (2.28)$$

pour $t \geq 0$ et $-a_1 < k < a_2$. On définit $a^{(t)} \equiv \Gamma(a + t)/\Gamma(a) = a(a + 1)\dots(a + t - 1)$.

En appliquant la méthode à notre cas, on considère $u_{ij} = \text{logit}(B_{ij})$ et de façon équivalente $B_{ij} = e^{u_{ij}}/(1 + e^{u_{ij}})$. On suppose $u_{ij} \sim N(m_{ij}, w_{ij}^2)$. On a,

$$\begin{aligned} u_{ij} &= \text{logit}(B_{ij}), \\ &= \log B_{ij} - \log(1 - B_{ij}), \\ &= \tau_j - (X'_{ij}\beta + W'_{ij}b_j) - \log(e_{ij}). \end{aligned}$$

Alors, $m_{ij} \equiv E(u_{ij}) = \hat{\tau}_j - (X'_{ij}\hat{\beta} + W'_{ij}\hat{b}_j) - \log(e_{ij})$,

$$w_{ij}^2 \equiv \text{var}(u_{ij}) = \text{var}(\tau_j - (X'_{ij}\beta + W'_{ij}b_j)).$$

La distribution de B_{ij} par l'approximation ADM est

$$B_{ij} \sim \text{Beta}(a_{ij,1}, a_{ij,2}),$$

où $a_{ij,1} = \hat{B}_{ij}^2(1 - \hat{B}_{ij})i_0$, $a_{ij,2} = \hat{B}_{ij}(1 - \hat{B}_{ij})^2 i_0$. Or $\hat{B}_{ij} = e^{m_{ij}}/(1 + e^{m_{ij}})$ et $i_0 = \frac{1}{w_{ij}^2 \hat{B}_{ij}^2(1 - \hat{B}_{ij})}$, on obtient donc $a_{ij,1} = 1/(w_{ij}^2(1 - \hat{B}_{ij}))$ et $a_{ij,2} = 1/(w_{ij}^2 \hat{B}_{ij})$. On note aussi que \hat{B}_{ij} étant la moyenne de B_{ij} , on a

$$EB_{ij} \equiv \hat{B}_{ij} = \frac{a_{ij,1}}{a_{ij,1} + a_{ij,2}}. \quad (2.29)$$

Ainsi, pour avoir les expressions $E_s B_{ij}^t$ ($s, t = 0, 1, 2$) définies en (2.27), il suffit de remplacer dans l'approximation (2.28) les éléments k, B, a_1 et a_2 respectivement par $sh_{ij}, B_{ij}, a_{ij,1}$ et $a_{ij,2}$.

A partir de ces résultats, on dérive les expressions finales suivantes pour les deux premiers moments de la loi a posteriori de chaque λ_{ij} dont la distribution sachant les données est une loi Gamma.

La moyenne a posteriori de λ_{ij} est donnée par

$$\begin{aligned}\hat{\lambda}_{ij} \equiv E(\lambda_{ij}|\text{data}) &= (1 - E_0 B_{ij})y_{ij} + E(B_{ij}\mu_{ij}) \\ &= (1 - \hat{B}_{ij})y_{ij} + (E\mu_{ij})(E_1 B_{ij}),\end{aligned}$$

$$\text{où } E_1 B_{ij} = (a_{ij,1} + h_{ij}) / (a_{ij,1} + a_{ij,2}).$$

La variance quant à elle s'exprime comme

$$\hat{\sigma}_{\lambda_{ij}}^2 \equiv \text{var}(\lambda_{ij}|\text{data}) = \frac{1}{e_{ij}} y_{ij} E(1 - B_{ij})^2 + \frac{1}{e_{ij}} \hat{\mu}_{ij} E_1 B_{ij} (1 - B_{ij}) + \text{var}(B_{ij}\mu_{ij} - B_{ij}y_{ij}),$$

$$\text{où } E(1 - B_{ij})^2 = 1 - 2\hat{B}_{ij} + E_0 B_{ij}^2 \text{ et}$$

$$\begin{aligned}\text{var}(B_{ij}\mu_{ij} - B_{ij}y_{ij}) &= (E\mu_{ij}^2)(E_2 B_{ij}^2) + y_{ij}^2 E_0 B_{ij}^2 - 2y_{ij} \hat{\mu}_{ij} E_1 B_{ij}^2 \\ &\quad - (y_{ij} \hat{B}_{ij} - \hat{\mu}_{ij} E_1 B_{ij})^2.\end{aligned}$$

$$\text{Pour } s = 0, 1, 2, \text{ on a } E_s B_{ij}^2 = \frac{(a_{ij,1} + sh_{ij})(a_{ij,1} + sh_{ij} + 1)}{(a_{ij,1} + a_{ij,2})(a_{ij,1} + a_{ij,2} + 1)}.$$

Chapitre 3

COMPARAISON DE RÉSULTATS AVEC LES DONNÉES DE CHRISTIANSEN ET MORRIS (1997)

Comme nous l'avons déjà mentionné, le modèle que nous avons développé est une extension, avec effets aléatoires liés aux strates, de la procédure nommée Modélisation multi-niveau et interactive de la régression de Poisson (PRIMM) développée par Christiansen et Morris (1997). En effet, notre modèle convient également aux données de Poisson stratifiées et en présence d'une seule strate ($p = 1$), notre modèle et celui de Christiansen et Morris (1997) sont en principe identiques ; la seule différence se situe au niveau des approximations faites pour simplifier les dérivées de la vraisemblance restreinte. Afin de vérifier qu'il peut donc constituer une certaine généralisation de la méthode PRIMM, nous allons comparer les résultats obtenus par notre programme à ceux de Christiansen et Morris (1997) pour les deux exemples de données qu'ils ont utilisés.

3.1. EXEMPLE 1 : TAUX DE PANNE DES POMPES

Les données concernent les pannes de dix pompes dans une centrale nucléaire équipée d'un réacteur à eau pressurisée. Chaque donnée observée z_i ($i = 1, \dots, 10$) représente le nombre de pannes de la i^{ieme} pompe. Il y a une seule variable explicative qui prend les valeurs $x_{i1} = 1$ pour les quatre pompes qui opèrent de manière continue et $x_{i1} = 0$ pour les six pompes qui fonctionnent par intermittence. L'exposition e_i de la i^{ieme} pompe représente le nombre d'heures de fonctionnement. Ainsi, $y_i = z_i/e_i$ est le taux observé de panne de la pompe i . Il y a une seule strate pour ces données alors, $p = 1$.

La table 3.1 présente les résultats de l'analyse des données sur les pannes des pompes utilisant la méthode PRIMM (modèle 1) et ceux obtenus par notre modèle (modèle 2).

Les facteurs de retrait B_i et les deux premiers moments $\hat{\lambda}_i$ et $\hat{\sigma}_{\lambda_i}^2$, des distributions du taux de panne sont contenus dans la table 3.1 . Par ailleurs, l'estimateur du coefficient de la régression de Poisson $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ est de $(-0.54, -0.67)'$ pour le modèle 1 avec écart type de $(0.36, 0.29)'$ et pour le modèle 2 de $(-0.525, -0.658)'$ avec écart type de $(0.36, 0.29)'$. L'estimateur de la variabilité ξ prend la valeur 0.96 (écart type=0.47) pour le modèle 1 et 0.945 (écart type=0.50) pour le modèle 2.

Les valeurs de ces estimateurs pour le modèle 1 sont tirées de Christiansen et Morris (1997). En ce qui concerne le modèle 2, les estimateurs de β et ξ sont obtenus par l'implémentation de l'algorithme décrit dans la section 2.5. Les valeurs de \hat{B}_i , $\hat{\lambda}_i$ et $\hat{\sigma}_{\lambda_i}^2$ sont données par les formules (2.29), (2.25) et (2.26) de la section 2.6.

TABLE 3.1. Résultats de l'analyse des données sur les pannes des pompes par l'utilisation de deux modèles.

obs i	\hat{B}_i		$\hat{\lambda}_i$		$\hat{\sigma}_{\lambda_i}^2$	
	modèle1	modèle2	modèle1	modèle2	modèle1	modèle2
1	0.43	0.424	1.08	1.121	0.83	0.804
2	0.43	0.424	1.08	1.121	0.83	0.804
3	0.27	0.269	1.77	1.808	0.83	0.813
4	0.13	0.128	0.67	0.688	0.35	0.346
5	0.07	0.068	2.13	2.137	0.45	0.446
6	0.05	0.047	0.12	0.124	0.09	0.089
7	0.11	0.110	0.59	0.592	0.14	0.133
8	0.06	0.058	0.09	0.093	0.04	0.038
9	0.04	0.039	0.06	0.063	0.03	0.026
10	0.03	0.030	0.12	0.121	0.03	0.031

Dans la section suivante, nous présentons les résultats pour le second exemple, ensuite nous donnerons les conclusions que nous tirons de cette comparaison de résultats.

3.2. EXEMPLE 2 : DONNÉES DE TRANSPLANTATION DU COEUR

Ce second exemple concerne des taux de mortalité des patients dans 15 centres de transplantation du coeur aux Etats-Unis. Pour chaque centre i ($i = 1, \dots, 15$), z_i représente le nombre de décès dans le mois suivant la transplantation et l'exposition e_i est le nombre de patients traités dans le centre. L'index de sévérité x_{i1} utilisé comme variable explicative, reflète le nombre attendu de décès dans le centre i et est basé sur sept variables

démographiques et de santé considérées comme facteurs à risque pour les patients traités dans le centre.

Les résultats de l'analyse de ces données par les deux modèles cités précédemment sont présentés dans la table 3.2. L'estimateur du coefficient de la régression de Poisson $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ est pour le modèle 1 de $(-2.95, 1.19)'$ avec écart type de $(0.30, 1.85)'$ et pour le modèle 2 de $(-2.97, 1.50)'$ avec écart type de $(0.23, 1.76)'$. L'estimateur de la variabilité ξ prend la valeur 3.07 (écart type=2.20) pour le modèle 1 et 3.59 (écart type=2.47) pour le modèle 2.

TABLE 3.2. Résultats de l'analyse des données de transplantation du coeur par l'utilisation de deux modèles.

obs i	\hat{B}_i		$\hat{\lambda}_i$		$\hat{\sigma}_{\lambda_i}^2$	
	modèle1	modèle2	modèle1	modèle2	modèle1	modèle2
1	0.70	0.701	0.117	0.126	0.057	0.046
2	0.70	0.743	0.049	0.049	0.028	0.023
3	0.61	0.623	0.083	0.089	0.038	0.035
4	0.60	0.638	0.043	0.045	0.023	0.021
5	0.59	0.662	0.047	0.045	0.027	0.019
6	0.50	0.513	0.042	0.047	0.022	0.022
7	0.41	0.404	0.051	0.060	0.026	0.025
8	0.43	0.440	0.057	0.062	0.024	0.024
9	0.39	0.405	0.079	0.084	0.028	0.027
10	0.41	0.442	0.024	0.027	0.016	0.014
11	0.43	0.472	0.083	0.082	0.029	0.024
12	0.40	0.435	0.045	0.047	0.019	0.018
13	0.36	0.384	0.042	0.045	0.018	0.017
14	0.35	0.377	0.079	0.079	0.024	0.022
15	0.30	0.308	0.067	0.071	0.022	0.022

Bien que l'algorithme numérique utilisé et les approximations faites pour les dérivées relatives à la vraisemblance restreinte ne soient pas les mêmes pour les deux modèles, nous pouvons constater une similarité dans leurs résultats pour les deux exemples de données. Ceci montre donc que nos approximations semblent adéquates et fiables.

Chapitre 4

APPLICATION AUX DONNÉES SUR LES EFFETS INDÉSIRABLES D'UN ESSAI CLINIQUE

Avant de présenter les résultats de l'application du modèle, nous allons faire une description des données utilisées.

4.1. PRÉSENTATION DES DONNÉES

Berry et Berry (2004) présentent les résultats d'un essai clinique impliquant un vaccin quadrivalent contre la rougeole, les oreillons, la rubéole et la varicelle. Les participants étaient 296 jeunes enfants en bonne santé âgés de 12 à 18 mois et ont été aléatoirement assignés à deux groupes. Le groupe de traitement reçu le traitement rougeole-oreillons-rubéole-varicelle au jour 0 et le groupe contrôle reçu au jour 0 le traitement rougeole-oreillons-rubéole suivi de celui de la varicelle au jour 42. La question principale était d'évaluer les réactions locales et systémiques pour la composante varicelle. La comparaison des effets indésirables (EIs) a été faite entre le groupe de traitement pendant les jours 0-42 et le groupe contrôle pendant les jours 42-84. La taille de l'échantillon était plus petite pour le groupe contrôle ($N_C = 132$) que pour le traitement ($N_T = 148$) parce que seuls les participants ayant reçu l'injection pour la varicelle ont été considérés. Pour cet essai, 40 types de EIs ont été considérés et attribués à 8 systèmes biologiques du corps humain. Il existe en effet plusieurs systèmes qui composent le corps humain (le système nerveux, le système respiratoire, le système digestif,...etc). On note que les types de EIs dans différents systèmes biologiques du corps humain ne sont pas les mêmes. L'attribution des EIs aux systèmes biologiques du corps humain est faite sur des bases biologiques et non sur une observation empirique. Les nombres de EIs pour tous les 40 types et selon le groupe

sont donnés dans la table 4.1 ; z_{ij} représente, pour un état donné (traitement/contrôle), le nombre de patients expérimentant le type de EIs indiqué par i attribué au système biologique du corps humain j .

TABLE 4.1. Résultats d'un essai clinique sur les effets indésirables(EIs).

j	i	Type de EIs	Traitement ($N_T = 148$)		Contrôle ($N_T = 132$)	
			z_{ij}	Taux	z_{ij}	Taux
1	1	Asthénie/fatigue	57	0.385	40	0.303
1	2	Fièvre	34	0.230	26	0.197
1	3	Infection fongique	2	0.014	0	0.000
1	4	Infection virale	3	0.020	1	0.008
1	5	Malaise	27	0.182	20	0.152
3	1	Anorexie	7	0.047	2	0.015
3	2	Candidiasis oral	2	0.014	0	0.000
3	3	Constipation	2	0.014	0	0.000
3	4	Diarrhée	24	0.162	10	0.076
3	5	Gastro-entérite	3	0.020	1	0.008
3	6	Nausée	2	0.014	7	0.053
3	7	Vomissement	19	0.128	19	0.144
5	1	Lymphadénopathie	3	0.020	2	0.015
6	1	Déshydratation	0	0.000	2	0.015
8	1	Pleurs	2	0.014	0	0.000
8	2	Insomnie	2	0.014	2	0.015
8	3	Irritabilité	75	0.507	43	0.326
9	1	Bronchite	4	0.027	1	0.008
9	2	Congestion nasale	4	0.027	2	0.015
9	3	Congestion respiratoire	1	0.007	2	0.015
9	4	Toux	13	0.088	8	0.061
9	5	Infection respiratoire	28	0.189	20	0.152
9	6	Laryngotrachéobronchite	2	0.014	1	0.008
9	7	Pharyngite	13	0.088	8	0.061
9	8	Rhinorrhée	15	0.101	14	0.106
9	9	Sinusite	3	0.020	1	0.008
9	10	Angine	2	0.014	1	0.008
9	11	Toux asthmatiforme	3	0.020	1	0.008

TABLE 4.2. Résultats d'un essai clinique sur les effets indésirables(EIs). (Suite)

j	i	Type de EIs	Traitement ($N_T = 148$)		Contrôle ($N_T = 132$)	
			z_{ij}	Taux	z_{ij}	Taux
10	1	Piqûre	4	0.027	0	0.000
10	2	Eczéma	2	0.014	0	0.000
10	3	Prurit	2	0.014	1	0.008
10	4	Eruption	13	0.088	3	0.023
10	5	Erythème fessier	6	0.041	2	0.015
10	6	Eruption de rougeole ou rubéole	8	0.054	1	0.008
10	7	Eruption de varicelle	4	0.027	2	0.015
10	8	Urticatoire	0	0.000	2	0.015
10	9	Exanthème viral	1	0.007	2	0.015
11	1	Conjonctivite	0	0.000	2	0.015
11	2	Otite moyenne	18	0.122	14	0.106
11	3	Otorrhée	2	0.014	1	0.008

4.2. RÉSULTATS DE LA MODÉLISATION

Dans le cas des données sur les EIs, les systèmes biologiques du corps humain représentent les strates dont nous avons fait référence dans le modèle ; nous avons donc $p = 8$. Chaque observation z_{ij} ($i = 1, \dots, n_j$ et $j = 1, \dots, p$) représente le nombre de EIs pour le i^{ieme} type de EIs dans le j^{ieme} système biologique du corps humain. Le vecteur X_{ij} comporte une seule variable explicative (donc $r = 2$) qui prend les valeurs $x_{ij,1} = 0$ pour les observations du groupe contrôle et $x_{ij,1} = 1$ pour celles du groupe de traitement. Le vecteur W_{ij} des variables explicatives dans la strate j est juste égale à 1 (donc $q = 1$), car dans cette étude il n'existe pas de variables caractérisant les strates. Ainsi, l'effet aléatoire b_j lié au système biologique du corps humain j se réduit à un scalaire. L'exposition est $e_{ij} = 132$ pour les observations de traitement et $e_{ij} = 148$ pour celles de contrôle. Ainsi, y_{ij} représente le taux observé de EIs pour le type i dans le système biologique j , avec une moyenne de taux de 0.07 pour toutes les observations de traitement et 0.05 pour celles de contrôle.

Les résultats de notre analyse sont donnés dans les tables 4.3 et 4.4. Les estimateurs de $\beta = (\beta_0, \beta_1)'$, des $\xi_j \equiv \log(\tau_j)$ ($j = 1, \dots, p$) apparaissent dans la table 4.3 tandis que la table 4.4 contient les estimateurs des facteurs de retrait B_{ij} , les moyennes et variances

$\hat{\lambda}_{ij}$ et $\hat{\sigma}_{\lambda_{ij}}^2$, des distributions du taux de EIs. La dernière colonne de la table 4.4 donne pour chaque observation ij une estimation de la probabilité que la différence de taux λ_{ij} entre le groupe de traitement et le groupe contrôle soit négative.

TABLE 4.3. Estimateurs des paramètres $\beta = (\beta_0, \beta_1)'$ et $\xi_j (j = 1, \dots, p)$ résultants de l'analyse des données sur les EIs.

Estimateur	Valeur	Écart type
$\hat{\beta}_0$	1.018	1.389
$\hat{\beta}_1$	0.091	0.040
$\hat{\xi}_1$	0.125	0.006
$\hat{\xi}_2$	0.043	0.001
$\hat{\xi}_3$	0.000	0.000
$\hat{\xi}_4$	0.500	0.133
$\hat{\xi}_5$	0.450	0.082
$\hat{\xi}_6$	0.823	0.193
$\hat{\xi}_7$	0.875	0.142
$\hat{\xi}_8$	0.649	0.137

L'estimateur du coefficient de la régression de Poisson associé à la variable explicative est $\hat{\beta}_1 = 0.091$ avec un écart type de 0.040, il est donc statistiquement différent de zéro. Ce qui signifie que le taux de EIs est significativement plus important dans le groupe de traitement. On remarque que la variabilité ξ_j est plus élevée dans les 3 derniers systèmes biologiques, indiquant une surdispersion plus grande pour les observations de Poisson liées à ces systèmes biologiques que pour les autres. Les écarts types dans la table 3 sont obtenus à partir de la diagonale de la matrice de variance-covariance des hyperparamètres $\theta = (\beta, \tau_1, \dots, \tau_p, b_1, \dots, b_p)$. La moyenne générale des estimateurs des retraits B_{ij} est de 0.78, indiquant que les différences dans les taux de EIs sont en bonne partie expliquées par la variable traitement/contrôle. On remarque aussi que les estimateurs des retraits sont moins élevés pour les observations de traitement car ces dernières ont la plus grande exposition.

Dans la dernière colonne de la table 4.4, les probabilités ont été estimées en faisant des simulations. Pour chaque observation ij et pour chaque groupe (traitement et contrôle),

nous avons simulé un échantillon de taille 1000 suivant une loi Gamma de moyenne $\hat{\lambda}_{ij}$ et de variance $\hat{\sigma}_{\lambda_{ij}}^2$. Ensuite, nous avons calculé les différences entre les données des échantillons du groupe de traitement et ceux du groupe contrôle. Ainsi, nous avons considéré la proportion de différences négatives pour estimer la probabilité que la différence de taux λ_{ij} entre les deux groupes soit négative ; c'est-à-dire la probabilité que le traitement ait un taux de EIs moins élevé. Les résultats obtenus nous montrent que cette probabilité est petite (<10%) pour le type 4 du système biologique 3 (diarrhée) et pour le type 4 du système biologique 10 (éruption). L'effet du traitement serait donc considérable pour ces deux types de EIs.

TABLE 4.4. Estimateurs des B_{ij} , λ_{ij} et $\sigma_{\lambda_{ij}}^2$ selon le groupe (T=traitement, C=contrôle) résultant de l'analyse des données sur les EIs.

j	i	\hat{B}_{ij}		$\hat{\lambda}_{ij}$		$\hat{\sigma}_{\lambda_{ij}}^2$		Probabilité de diff<=0
		T	C	T	C	T	C	
1	1	0.219	0.393	0.302	0.185	0.051	0.048	0.317
1	2	0.219	0.393	0.180	0.120	0.039	0.038	0.337
1	3	0.219	0.393	0.011	0.001	0.010	0.003	0.546
1	4	0.219	0.393	0.017	0.005	0.012	0.008	0.263
1	5	0.219	0.393	0.143	0.093	0.035	0.034	0.357
3	1	0.815	0.911	0.010	0.002	0.018	0.012	0.296
3	2	0.815	0.911	0.003	0.001	0.011	0.006	0.916
3	3	0.815	0.911	0.003	0.001	0.011	0.006	0.911
3	4	0.815	0.911	0.031	0.007	0.032	0.023	0.075
3	5	0.815	0.911	0.005	0.001	0.013	0.009	0.752
3	6	0.815	0.911	0.003	0.005	0.011	0.020	0.911
3	7	0.815	0.911	0.025	0.013	0.029	0.031	0.221
5	1	0.990	0.996	0.002	0.001	0.034	0.036	1.000
6	1	0.947	0.976	0.000	0.000	0.004	0.010	1.000
8	1	0.674	0.827	0.005	0.001	0.010	0.005	0.883
8	2	0.674	0.827	0.005	0.003	0.010	0.012	0.860
8	3	0.674	0.827	0.166	0.057	0.058	0.049	0.156
9	1	0.704	0.846	0.010	0.002	0.014	0.009	0.142
9	2	0.704	0.846	0.010	0.003	0.014	0.011	0.173
9	3	0.704	0.846	0.004	0.003	0.008	0.011	0.750
9	4	0.704	0.846	0.028	0.010	0.023	0.019	0.169
9	5	0.704	0.846	0.058	0.024	0.033	0.030	0.198

TABLE 4.5. Estimateurs des B_{ij} , λ_{ij} et $\sigma_{\lambda_{ij}}^2$ selon le groupe (T=traitement, C=contrôle) résultant de l'analyse des données sur les EIs. (Suite)

j	i	\hat{B}_{ij}		$\hat{\lambda}_{ij}$		$\hat{\sigma}_{\lambda_{ij}}^2$		Probabilité de diff<=0
		T	C	T	C	T	C	
9	6	0.704	0.846	0.006	0.002	0.010	0.009	0.473
9	7	0.704	0.846	0.028	0.010	0.023	0.019	0.153
9	8	0.704	0.846	0.032	0.017	0.025	0.025	0.257
9	9	0.704	0.846	0.008	0.002	0.012	0.009	0.270
9	10	0.704	0.846	0.006	0.002	0.010	0.009	0.490
9	11	0.704	0.846	0.008	0.002	0.012	0.009	0.269
10	1	0.901	0.955	0.003	0.000	0.012	0.005	1.000
10	2	0.901	0.955	0.002	0.000	0.009	0.005	1.000
10	3	0.901	0.955	0.002	0.001	0.009	0.007	0.622
10	4	0.901	0.955	0.009	0.001	0.020	0.011	0.087
10	5	0.901	0.955	0.005	0.001	0.014	0.009	0.160
10	6	0.901	0.955	0.006	0.001	0.016	0.007	0.170
10	7	0.901	0.955	0.003	0.001	0.012	0.009	0.373
10	8	0.901	0.955	0.001	0.001	0.005	0.009	0.893
10	9	0.901	0.955	0.001	0.001	0.007	0.009	0.883
11	1	0.785	0.894	0.001	0.002	0.005	0.011	0.964
11	2	0.785	0.894	0.027	0.012	0.027	0.026	0.212
11	3	0.785	0.894	0.004	0.001	0.010	0.009	0.556

Berry et Berry(2004) pour ces données sur les effets indésirables, ont présenté une méthode pour calculer la probabilité que le groupe traitement ait un taux de EIs plus grand que celui du groupe contrôle. Ils trouvent que le seul type de EIs pour lequel cette probabilité est élevée (= 0.780) est le type 3 du système biologique 8 (irritabilité). Dans notre modèle, pour cet effet indésirable, nous obtenons que la probabilité que le groupe traitement ait un taux de EIs moins élevée est de 0.156. Ce qui signifie une probabilité de 0.844 que le groupe traitement ait un taux de EIs plus grand que celui du groupe contrôle. Cette valeur est assez proche de ce qu'ils ont trouvé ; mais nos résultats montrent un effet du traitement encore plus considérable pour deux autres types de EIs (diarrhée et éruption).

CONCLUSION

Dans cette étude, nous avons développé une méthode pour modéliser des données de comptage à travers un modèle de Poisson hiérarchique à effets mixtes. Le modèle étudié est une extension avec effets aléatoires de la procédure nommée Modélisation multi-niveau et interactive de la régression de Poisson (PRIMM) développé par Christiansen et Morris (1997). Ces auteurs ont démontré les avantages de leur approche par rapport à d'autres méthodes telles que les méthodes de modèle linéaire généralisé en présence de surdispersion ou encore les méthodes de Monte Carlo par chaînes de Markov (MCMC). Dans la procédure PRIMM, la régression de Poisson ne comprend que des effets fixes tandis que notre modèle intègre en plus des effets aléatoires.

L'objectif principal de notre travail était d'estimer les coefficients de la régression log-linéaire ainsi que les paramètres de la loi de Poisson. Ceci a été fait en effectuant diverses approximations analytiques telles que la méthode de Laplace et la théorie asymptotique liée à l'approximation normale pour la loi a posteriori, évitant ainsi d'utiliser des méthodes computationnelles comme les méthodes MCMC.

L'estimation des paramètres structurels du modèle a été faite par la maximisation de leur densité a posteriori via l'algorithme de Newton-Raphson ; cette densité ayant été ajustée en considérant l'approche du maximum de vraisemblance restreint (REML). La distribution approximative conjointe de ces paramètres structurels sachant les données est une loi normale. Ce résultat contribue à déterminer les deux premiers moments a posteriori des paramètres individuels du modèle. Ces paramètres de la loi de Poisson ont chacun approximativement une distribution gamma.

Ensuite, nous avons comparé les résultats que nous obtenons avec notre modèle aux résultats du modèle de Christiansen et Morris (1997) pour les deux exemples de données qu'ils ont utilisés. Dans le cas de ces données où il n'y a qu'une seule strate et pas d'effets

aléatoires, les deux modèles sont en principe identiques et ne diffèrent en fait que par l'algorithme numérique utilisé et les approximations faites pour simplifier les dérivées relatives à la vraisemblance restreinte. Nous avons constaté que les résultats obtenus sont assez similaires ; permettant ainsi de valider nos approximations.

Enfin, le modèle est appliqué aux données relatives au nombre d'effets indésirables observés chez les participants d'un essai clinique impliquant un vaccin quadrivalent contre la rougeole, les oreillons, la rubéole et la varicelle. Pour ces données, la régression de Poisson comprend l'effet fixe correspondant à la variable traitement/contrôle, ainsi que des effets aléatoires liés aux systèmes biologiques du corps humain auxquels sont attribués les effets indésirables considérés. L'analyse des résultats a montré un effet du traitement pour deux types d'effets indésirables, la diarrhée qui est dans le système 3 et l'éruption dans le système 10.

BIBLIOGRAPHIE

Agresti, A. (2002). *Categorical Data Analysis*. Seconde édition, Wiley.

Albert, J. (1992). A Bayesian Analysis of a Poisson Random Effects Model for Home Run Hitters. *The American Statistician*, Vol. 46, No. 4, pp. 246-253.

Berry, S. M. et Berry, D. A. (2004). Accounting for Multiplicities in Assessing Drug Safety : A Three-Level Hierarchical Mixture Model. *Biometrics* 60, pp. 418-426.

Casella, G. et Berger, R. L. (2002). *Statistical Inference*. Seconde édition, Duxbury Press.

Casella, G. et George, E. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, pp. 167-174.

Christiansen, C. L. et Morris, C. N. (1997). Hierarchical Poisson Regression Modeling. *Journal of the American Statistical Association*, Vol. 92, No. 438, pp. 618- 632.

Frome, E. L. (1983). The Analysis of Rates using Poisson Regression Models. *Biometrics*, 39, pp. 665-674.

Gelfand, A. E. et Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85, pp. 398-409.

Gelman, A., Carlin, J. B., Stern, H. S. et Rubin, D. B. (1995). *Bayesian Data Analysis*.

Chapman and Hall, London.

Hobert J.P. et Casella G.(1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *Journal of the American Statistical Association*, 91, pp. 1461- 1473.

Holford, T. R. (1980). The Analysis of Rates and of Survivorship using Log-linear Models. *Biometrics*, 36, pp. 299-305.

John M.Chambers(2008). *Software for Data Analysis : Programming with R*. Springer, NewYork.

Liang, K. Y. et Zeger, S. L. (1995). Inference Based on Estimating Functions in the Presence of Nuisance Parameters. *Statistical Science*, 10, 158-173.

Nott, D. J., Fielding M. et Leonte, D. (2009). On a Generalization of the Laplace Approximation. *Statistics and Probability Letters* 79, pp. 1397-1403.

Tempelman R.J. et Gianola D.(1996). A Mixed Effects Model for Overdispersed Count Data in Animal Breeding. *Biometrics*, 52, pp. 265-279.

Venables, W.N et Ripley, B.D. (2002). *Modern Applied Statistics with S*. Springer, NewYork.

Vonesh, E.F. (1990). Modelling Peritonitis Rates and Associated Risk Factors for Individuals on Continuous Ambulatory Peritoneal Dialysis. *Statistics in Medecine*, 9, pp. 263-271.

Annexe A

CODE R DU PROGRAMME

Nous présentons ici les lignes du programme que nous avons développé dans le logiciel R pour la mise en application de notre modèle. Pour illustrer la structure des données à entrer, nous considérons l'exemple des données sur les effets indésirables d'un essai clinique que nous avons décrit dans la section 4.1. Ainsi, nous avons le vecteur z qui contient les données de comptage (le nombre d'EIs observé pour chaque type de EIs), le vecteur x qui est la variable explicative traitement/contrôle prend la valeur 0 pour les observations liées au groupe contrôle et 1 pour celles liées au traitement. Il peut y avoir plusieurs variables explicatives, dans ce cas on aurait un vecteur pour chacune d'elles. Les expositions sont contenues dans le vecteur e . Pour désigner la strate qui dans cet exemple est le système biologique du corps humain, il y a le vecteur $group$ qui contient les numéros des différents systèmes biologiques. Pour ces données, il n'y a pas de variable explicative spécifique à l'effet aléatoire alors le vecteur $x.b$ ne contient que des 1. Lorsque nous entrons ces données dans le programme, il en ressort les résultats que nous avons dans les tables 4.3 et 4.4.

```

#p= nombre de groupes (strates)
#e= vecteur des expositions
#z= vecteur des données de comptage
#x= variable explicative associé à l'effet fixe
#x.b=variable explicative associé à l'effet aléatoire

#####
data0=read.table(bodyAEs2.txt,h=T) #lecture des données
e=data0$e
z=data0$z
x=data0$x
group=data0$group
l=length(z)

####creation de e,z,x,y par groupe
if (p==1){group=c(rep(1,l))}
y=z/e
group=as.factor(group)
z.s=split(z,group)
e.s=split(e,group)
y.s=split(y,group)
x.s=split(x,group)
p=length(z.s)
r=2

####creation de la matrice x et du vecteur x.b par groupe
n=rep(0,p)
for (j in 1:p){
  n[j]=length(e.s[[j]])
x.s[[j]] <- as.matrix(cbind(rep(1, n[j]), x.s[[j]] ))
}
x.b=split(rep(1,l),group)
if (p==1){x.b=split(rep(0,l),group)}

####recherche de valeurs initiales
mu0=rep(0,p)
xi=rep(0,p)
tau=rep(0,p)
b.prev=rep(0,p)
z0=sum(z)/sum(e)
for (j in 1:p){
mu0[j]=sum(z.s[[j]])/sum(e.s[[j]])
xi[j]=(mean(y.s[[j]])^2/var(y.s[[j]]))
if (j==1){xi[j]=sum(z.s[[j]])/sum(e.s[[j]])}
tau[j]=log(xi[j])
}

for (j in 1:p){
reg.b=lm(log(y.s[[j]]+0.0001) ~ x.s[[j]][,2])
b.prev[j]=reg.b$coef[1]
}
if (p==1){b.prev=0}

reg1=lm(log(y+0.0001) ~ x)
beta.prev=c(log(mean(y)),reg1$coef[2])

#####
### CREATION DE LA FONCTION QUI CALCULE LA VRAISEMBLANCE
#####
L.theta<- fonction(theta)
{
  beta.prev=theta[1:r]
  tau=theta[(r+1):(p+r)]
  b.prev=theta[(p+r+1):(r+2*p)]
  xi=exp(tau)
}

```

```

eta=split(rep(0,l),group)
mu=split(rep(0,l),group)
B=split(rep(0,l),group)
lam=split(rep(0,l),group)
sigm=split(rep(0,l),group)

for (j in 1:p){
  eta[[j]]=x.s[[j]]%% beta.prev + x.b[[j]]*b.prev[j]
  mu[[j]]=exp(eta[[j]])
  B[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu[[j]])
  lam[[j]]=(1-B[[j]])*y.s[[j]]+B[[j]]*mu[[j]]
  sigm[[j]]=sqrt(lam[[j]]*(1-B[[j]])/e.s[[j]])
}
#####
##calcul derivees de beta
sum1.1=matrix(0,r,p)
for (j in 1:p){
  mu[[j]]=as.vector(mu[[j]])
  B[[j]]=as.vector(B[[j]])
  for(i in 1:n[j]){
    sum1.1[,j]=(z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*x.s[[j]][i,]+sum1.1[,j]
  }
}
d1.beta=c(rep(0,r))
for (k in 1:r){
d1.beta[k]=sum(sum1.1[k,])
}
liste1=list(matrix(0,r,r))
sum1.2=list(matrix(0,r,r))
while(length(sum1.2)<p){
sum1.2=c(sum1.2,liste1)
}
d2.beta=matrix(0,r,r)
for (j in 1:p){
  for(i in 1:n[j]){
    sum1.2[[j]]=-e.s[[j]][i]*B[[j]][i]*lam[[j]][i]*x.s[[j]][i,]%*(x.s[[j]][i,]+
    sum1.2[[j]])
  }
}
d2.beta=sum1.2[[j]]+d2.beta
}
#####
## calcul des derivees des b_j
d1.b=c(rep(0,p))
d2.b=c(rep(0,p))
trNb.0=c(rep(0,p))
trNb.1=c(rep(0,p))
trNb.2=c(rep(0,p))
Nb=c(rep(0,p))
q=1
m0=2
Lam0=1

for (j in 1:p){
  Nb[j]=q*(m0+p)*(Lam0^(-1))*(q+sum(b.prev*Lam0^(-1)*b.prev))-2*Lam0^(-1)*
  b.prev[j]*b.prev[j]*Lam0^(-1)/(q+sum(b.prev*Lam0^(-1)*b.prev))^2
  trNb.0[j]=Nb[j]
  trNb.1[j]=-q*(m0+p)*2*b.prev[j]*(2+q+sum(b.prev^2))/(q+sum(b.prev^2))^3
  trNb.2[j]=-2*q*(m0+p)*((2+q+sum(b.prev^2))+2*b.prev[j]^2)*(q+sum(b.prev^2))-
  6*b.prev[j]^2*(2+q+sum(b.prev^2))/(q+sum(b.prev^2))^4
  d1.b[j]=sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]]*x.b[[j]]-q*(m0+p)*(Lam0^(-1)*
  b.prev[j])/(q+sum(b.prev*Lam0^(-1)*b.prev))
  d2.b[j]=-sum(e.s[[j]]*B[[j]]*lam[[j]]*x.b[[j]]*x.b[[j]])-Nb[j]
}

```

```
#####
##calcul des derivees pour la matrice H.beta.b
d2.beta.b=matrix(0,r,p)
d2.beta.tau=matrix(0,r,p)
for (j in 1:p){
  for(i in 1:n[j]){
    d2.beta.b[,j]=(-z.s[[j]][i]-B[[j]][i]*(e.s[[j]][i]*mu[[j]][i]-
    z.s[[j]][i]))* B[[j]][i]* x.s[[j]][i]*x.b[[j]][i]+d2.beta.b[,j]
    d2.beta.tau[,j]=(z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*
    (1-B[[j]][i])*x.s[[j]][i]+ d2.beta.tau[,j]
  }
}
d2.b.b=matrix(0,p,p)
trNb.b=matrix(0,p,p)
d2.b.tau=matrix(0,p,p)
for (j in 1:p){
  for (k in 1:p){
    trNb.b[j,k]=-2*q*(m0+p)*((2*b.prev[j]*b.prev[k])*(q+sum(b.prev^2))-
    6*b.prev[j]*b.prev[k]*(2+q+sum(b.prev^2)))/(q+sum(b.prev^2))^4
    d2.b.b[j,k]=2*q*(m0+p)*(Lam0^(-1)*b.prev[j]*b.prev[k]*Lam0^(-1))/
    (q+sum(b.prev*Lam0^(-1)*b.prev))^2
    d2.b.tau[j,j]=sum((z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*(1-B[[j]][i])*x.b[[j]][i])
  }
}
#####
##vraisemblance
m.z=rep(0,p)
for (j in 1:p){
  m.z[j]=max(z.s[[j]])
}
M=max(m.z)
N <- c(rep(0,M))
for (s in 0:M){
  for (j in 1:p){
    for (i in 1:(n[j])){
      if (z.s[[j]][i]>=s) {N[s] <- 1+N[s]}
    }
  }
}
nu0=3
s02=10
a=mean(tau)
A=nu0*s02/p+(1/p)*sum((tau-a)^2)
b=(nu0+p)/2
Sm=c(1:M)
L1.1=rep(0,p)
for (j in 1:p){
  L1.1[j]=sum(N*log(xi[j]+Sm-1))-sum(log(z.s[[j]]+0.0001))+sum(z.s[[j]]*
  log(e.s[[j]]))- n[j]*mean(z.s[[j]])*log(xi[j])+t(z.s[[j]])%*%x.s[[j]]%*%
  beta.prev+t(z.s[[j]])%*%x.b[[j]]%*%
  b.prev[j]-sum((xi[j]+z.s[[j]])*log(1+e.s[[j]]*mu[[j]]/xi[j]))
}
L1=sum(L1.1)
L2=log(abs(Lam0+sum(b.prev*b.prev))^(-(m0+p)/2)*A^(-(nu0+p)/2)*exp(a-2*
log(exp(a)+z0))*(2*b/A+2*exp(a)*z0/(exp(a)+z0)^2)^(-1/2))
H.beta.b=matrix(0,r+p,r+p)
H.beta.b[c(1,r),c(1,r)]=-d2.beta
for (j in (r+1):(r+p)){
  H.beta.b[j,j]=-d2.b[j-r]
  H.beta.b[c(1,r),j]=-d2.beta.b[,j-r]
  H.beta.b[j,c(1,r)]=-d2.beta.b[,j-r]}
}

```

```

    if ((r+1)<(r+p-1)){
      for (j in (r+1):(r+p-1)){
        for (k in (j+1):(r+p)){
          H.beta.b[j,k]=-d2.b.b[j-r,k-r]
          H.beta.b[k,j]=-d2.b.b[j-r,k-r]}
        }
      }
    L3=log(abs(det(H.beta.b))^-1/2)
    L.R=-(L1+L2+L3)
    if(p==1){L.R=L1+det(-d2.beta)}
    L.R
  }
#####
##CREATION DE LA FONCTION QUI CALCULE LE GRADIENT
#####
grad.theta<- function(theta)
{
  beta.prev=theta[1:r]
  tau=theta[(r+1):(r+p)]
  b.prev=theta[(p+r+1):(r+2*p)]
  xi=exp(tau)
  eta=split(rep(0,l),group)
  mu=split(rep(0,l),group)
  B=split(rep(0,l),group)
  lam=split(rep(0,l),group)
  for (j in 1:p){
    eta[[j]]=x.s[[j]]%% beta.prev + x.b[[j]]*b.prev[j]
    mu[[j]]=exp(eta[[j]])
    B[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu[[j]])
    lam[[j]]=(1-B[[j]])*y.s[[j]]+B[[j]]*mu[[j]]
  }
#####
##calcul dérivées des tau_j
  m.z=rep(0,p)
  for (j in 1:p){
    m.z[j]=max(z.s[[j]])
  }
  M=max(m.z)
  N <- c(rep(0,M))
  for (s in 0:M){
    for (j in 1:p){
      for (i in 1:(n[j])){
        if (z.s[[j]][i]>=s) {N[s] <- 1+N[s]}
      }
    }
  }
  r=2
  q=1
  S1=c(rep(0,p))
  part1.1=c(rep(0,p))
  part1.2=c(rep(0,p))
  part1.3=c(rep(0,p))
  eta=split(rep(0,l),group)
  mu=split(rep(0,l),group)
  B=split(rep(0,l),group)
  nu0=3
  s02=10
  a=mean(tau)
  A=nu0*s02/p+(1/p)*sum((tau-a)^2)
  b=(nu0+p)/2
  B.0=c(rep(0,p))
  B.1=c(rep(0,p))
  for (j in 1:p){
    S1[j]=0
    for (s in 1:M){
      S1[j]=N[s]*(xi[j]/(xi[j]+s-1))+S1[j]}
  }
}

```

```

eta[[j]]=x.s[[j]]%% beta.prev + x.b[[j]]*b.prev[j]
mu[[j]]=exp(eta[[j]])
B[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu[[j]])
part1.1[j]=S1[j]+xi[j]*sum(log(B[[j]]))-sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]])
B.0[j]=b/A+exp(a)*z0/(exp(a)+z0)^2
B.1[j]=-2*b*(tau[j]-a)/p*A^2+z0*(z0*exp(a)-exp(2*a))/(p*(exp(a)+z0)^3)
part1.2[j]=-2*b*(tau[j]-a)/p*A + 1/p-2*exp(a)/(p*(exp(a)+ z0))-B.1[j]/(2*B.0[j])
}
m0=2
Lam0=1
Nb=c(rep(0,p))
B.tilde=split(rep(0,l),group)
for (j in 1:p){
  B.tilde[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu0[j])
Nb[j]=q*(m0+p)*(Lam0^(-1))*(q+sum(b.prev*Lam0^(-1)*b.prev))-2*Lam0^(-1)*
b.prev[j]*b.prev[j]* Lam0^(-1)/(q+sum(b.prev*Lam0^(-1)*b.prev))^2
}
sum1=split(rep(0,l),group)
sum2=c(rep(0,p))
trHbeta.1=c(rep(0,p))
trHbeta.2=c(rep(0,p))
trHb.0=c(rep(0,p))
trHb.1=c(rep(0,p))
trHb.2=c(rep(0,p))
for (j in 1:p){
  for (i in 1:n[j]){
    sum1[[j]][i]=sum(x.s[[j]][i,]^2)
    sum2[j]=sum(e.s[[j]]*B.tilde[[j]]*y.s[[j]]*sum1[[j]])
    trHbeta.1[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*y.s[[j]]*sum1[[j]])
    trHb.0[j]=sum(e.s[[j]]*B.tilde[[j]]*y.s[[j]]*x.b[[j]]^2)+Nb[j]
trHb.1[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*y.s[[j]]*x.b[[j]]^2)
    trH.1[j]=trHbeta.1[j]+trHb.1[j]
    trHbeta.2[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*(1-2*B.tilde[[j]])*
y.s[[j]]*sum1[[j]])
    trHb.2[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*(1-2*B.tilde[[j]])*
y.s[[j]]*x.b[[j]]^2)
    trH.2[j]=trHbeta.2[j]+trHb.2[j]
  }
trHbeta.0=sum(sum2)
trH.0=trHbeta.0+sum(trHb.0)
d1.tau=c(rep(0,p))
for (j in 1:p){
  part1.3[j]=-(r+p*q)*trH.1[j]/(2*trH.0)
  d1.tau[j]=part1.1[j]+part1.2[j]+part1.3[j]
}
#####
#calcul des dérivées de beta
sum1.1=matrix(0,r,p)
for (j in 1:p){
  mu[[j]]=as.vector(mu[[j]])
  B[[j]]=as.vector(B[[j]])
  for(i in 1:n[j]){
    sum1.1[,j]=(z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*x.s[[j]][i,]+sum1.1[,j]
  }
d1.beta=c(rep(0,r))
for (k in 1:r){
d1.beta[k]=sum(sum1.1[k,])
}
#####
##calcul des dérivées des b_j
d1.b=c(rep(0,p))
trNb.0=c(rep(0,p))
trNb.1=c(rep(0,p))
trNb.2=c(rep(0,p))
Nb=c(rep(0,p))
q=1

```

```

m0=2
Lam0=1
for (j in 1:p){
Nb[j]=q*(m0+p)*(Lam0^(-1))*(q+sum(b.prev*Lam0^(-1)*b.prev))-2*Lam0^(-1)*
p.prev[j]*b.prev[j]*Lam0^(-1))/(q+sum(b.prev*Lam0^(-1)*b.prev))^2
trNb.0[j]=Nb[j]
trNb.1[j]=-q*(m0+p)*2*b.prev[j]*(2+q+sum(b.prev^2))/(q+sum(b.prev^2))^3
trNb.2[j]=-2*q*(m0+p)*((2+q+sum(b.prev^2)+2*b.prev[j]^2)*(q+sum(b.prev^2))-
6*b.prev[j]^2*(2+q+sum(b.prev^2)))/(q+sum(b.prev^2))^4
d1.b[j]=sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]]*x.b[[j]]-q*(m0+p)*(Lam0^(-1)*
b.prev[j])/(q+sum(b.prev*Lam0^(-1)*b.prev))
}
Vect=c(d1.beta,d1.tau,d1.b)
if(p==1){Vect=c(d1.beta,d1.tau,0)}
Vect
}
#####
###CREATION DE LA FONCTION QUI CALCULE LA MATRICE HESSIENNE
#####
mat.hess<- function(theta0)
{
beta.prev=theta0[1:r]
tau=theta0[(r+1):(p+r)]
b.prev=theta0[(p+r+1):(r+2*p)]
xi=exp(tau)
eta=split(rep(0,l),group)
mu=split(rep(0,l),group)
B=split(rep(0,l),group)
lam=split(rep(0,l),group)
for (j in 1:p){
eta[[j]]=x.s[[j]]*beta.prev + x.b[[j]]*b.prev[j]
mu[[j]]=exp(eta[[j]])
B[[j]]=xi[[j]]/(xi[[j]]+e.s[[j]]*mu[[j]])
lam[[j]]=(1-B[[j]])*y.s[[j]]+B[[j]]*mu[[j]]
}
#####
##calcul dérivées des tau_j
m.z=rep(0,p)
for (j in 1:p){
m.z[j]=max(z.s[[j]])
M=max(m.z)
N <- c(rep(0,M))
for (s in 0:M){
for (j in 1:p){
for (i in 1:(n[j])){
if (z.s[[j]][i]>=s) {N[s] <- 1+N[s]}}
}
r=2
q=1
S1=c(rep(0,p))
part1.1=c(rep(0,p))
part1.2=c(rep(0,p))
part1.3=c(rep(0,p))
eta=split(rep(0,l),group)
mu=split(rep(0,l),group)
B=split(rep(0,l),group)
nu0=3
s02=10
a=mean(tau)
A=nu0*s02/p+(1/p)*sum((tau-a)^2)
b=(nu0+p)/2
B.0=c(rep(0,p))
B.1=c(rep(0,p))
for (j in 1:p){
S1[j]=0
for (s in 1:M){S1[j]=N[s]*(xi[[j]]/(xi[[j]]+s-1))+S1[j]}
}
}
}

```



```

eta[[j]]=x.s[[j]]%% beta.prev + x.b[[j]]*b.prev[j]
mu[[j]]=exp(eta[[j]])
B[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu[[j]])
part1.1[j]=S1[j]+xi[j]*sum(log(B[[j]]))-sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]])
B.0[j]=b/A+exp(a)*z0/(exp(a)+z0)^2
B.1[j]=-2*b*(tau[j]-a)/p*A^2+z0*(z0*exp(a)-exp(2*a))/(p*(exp(a)+z0)^3)
part1.2[j]=-2*b*(tau[j]-a)/p*A + 1/p-2*exp(a)/(p*(exp(a)+z0))-B.1[j]/(2*B.0[j])
}
m0=2
Lam0=1
Nb=c(rep(0,p))
B.tilde=split(rep(0,l),group)
for (j in 1:p){
  B.tilde[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu0[j])
  Nb[j]=q*(m0+p)*(Lam0^(-1)*q+sum(b.prev*Lam0^(-1)*b.prev))-2*Lam0^(-1)*
  b.prev[j]*b.prev[j]*Lam0^(-1))/(q+sum(b.prev*Lam0^(-1)*b.prev))^2
}
sum1=split(rep(0,l),group)
sum2=c(rep(0,p))
trHbeta.1=c(rep(0,p))
trHbeta.2=c(rep(0,p))
trHb.0=c(rep(0,p))
trHb.1=c(rep(0,p))
trHb.2=c(rep(0,p))
trH.1=c(rep(0,p))
trH.2=c(rep(0,p))
for (j in 1:p){
  for (i in 1:n[j]){sum1[[j]][i]=sum(x.s[[j]][i,]^2)}
  sum2[j]=sum(e.s[[j]]*B.tilde[[j]]*y.s[[j]]*sum1[[j]])
  trHbeta.1[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*y.s[[j]]*sum1[[j]])
  trHb.0[j]=sum(e.s[[j]]*B.tilde[[j]]*y.s[[j]]*x.b[[j]]^2)+Nb[j]
  trHb.1[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*y.s[[j]]*x.b[[j]]^2)
  trH.1[j]=trHbeta.1[j]+trHb.1[j]
  trHbeta.2[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*(1-2*B.tilde[[j]])*
  y.s[[j]]*sum1[[j]])
  trHb.2[j]=sum(e.s[[j]]*B.tilde[[j]]*(1-B.tilde[[j]])*(1-2*B.tilde[[j]])*
  y.s[[j]]*x.b[[j]]^2)
  trH.2[j]=trHbeta.2[j]+trHb.2[j]
}
trHbeta.0=sum(sum2)
trH.0=trHbeta.0+sum(trHb.0)
d1.tau=c(rep(0,p))
for (j in 1:p){
  part1.3[j]=-(r+p*q)*trH.1[j]/(2*trH.0)
  d1.tau[j]=part1.1[j]+part1.2[j]+part1.3[j]
}
S2=c(rep(0,p))
part2.1=c(rep(0,p))
part2.2=c(rep(0,p))
part2.3=c(rep(0,p))
B.2=c(rep(0,p))
d2.tau=c(rep(0,p))
for (j in 1:p){
  S2[j]=0
  for (s in 1:M){ S2[j]=N[s]*(xi[j]/(xi[j]+s-1)^2)+S2[j]}
  part2.1[j]=-S2[j]+sum(z.s[[j]]*B[[j]])-sum((z.s[[j]]-e.s[[j]]*mu[[j]])*
  B[[j]]*(1-B[[j]]))+part1.1[j]
  B.2[j]=-2*b*((p-1)*A-4*(tau[j]-a)^2)/(p^2*A^3)+z0*exp(a)*(exp(2*a)-
  4*z0*exp(a)+z0^2)/(p^2*(exp(a)+z0)^4)
  part2.2[j]=-2*b*((p-1)*A-2*(tau[j]-a)^2)/(p*A)^2 -2*z0*exp(a)/
  (p*(exp(a)+z0))^2-(B.2[j]*B.0[j]-B.1[j]^2)/(2*B.0[j]^2)
  part2.3[j]=-(r+p*q)*(trH.2[j]*trH.0-trH.1[j]^2)/(2*trH.0^2)
  d2.tau[j]=part2.1[j]+part2.2[j]+part2.3[j]
}
}

```

```

#####
##calcul des derivees de beta
sum1.1=matrix(0,r,p)
for (j in 1:p){
  mu[[j]]=as.vector(mu[[j]])
  B[[j]]=as.vector(B[[j]])
  for(i in 1:n[j]){
    sum1.1[,j]=(z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*x.s[[j]][i]+sum1.1[,j]
  }
  d1.beta=c(rep(0,r))
  for (k in 1:r){ d1.beta[k]=sum(sum1.1[k,j])}
  liste1=list(matrix(0,r,r))
  sum1.2=list(matrix(0,r,r))
  while(length(sum1.2)<p){ sum1.2=c(sum1.2,liste1)}
  d2.beta=matrix(0,r,r)
  for (j in 1:p){
    for(i in 1:n[j]){
      sum1.2[[j]]=-e.s[[j]][i]*B[[j]][i]*lam[[j]][i]*x.s[[j]][i]*%*%t(x.s[[j]][i,]+
      sum1.2[[j]])
    }
    d2.beta=sum1.2[[j]]+d2.beta
  }
#####
##calcul des derivees des b_j
d1.b=c(rep(0,p))
d2.b=c(rep(0,p))
trNb.0=c(rep(0,p))
trNb.1=c(rep(0,p))
trNb.2=c(rep(0,p))
Nb=c(rep(0,p))
q=1
m0=2
Lam0=1
for (j in 1:p){
  Nb[j]=q*(m0+p)*(Lam0^(-1)*(q+sum(b.prev*Lam0^(-1)*b.prev))-2*Lam0^(-1)*b.prev[j]*
  b.prev[j]*Lam0^(-1))/(q+sum(b.prev*Lam0^(-1)*b.prev))^2
  trNb.0[j]=Nb[j]
  trNb.1[j]=-q*(m0+p)*2*b.prev[j]*(2+q+sum(b.prev^2))/(q+sum(b.prev^2))^3
  trNb.2[j]=-2*q*(m0+p)*((2+q+sum(b.prev^2)+2*b.prev[j]^2)*(q+sum(b.prev^2))-
  6*b.prev[j]^2*(2+q+sum(b.prev^2)))/(q+sum(b.prev^2))^4
  d1.b[j]=sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]]*x.b[[j]])-q*(m0+p)*
  (Lam0^(-1)*b.prev[j])/(q+sum(b.prev*Lam0^(-1)*b.prev))
  d2.b[j]=-sum(e.s[[j]]*B[[j]]*lam[[j]]*x.b[[j]]*x.b[[j]])-Nb[j]
}
#####
##matrice hessienne
d2.beta.b=matrix(0,r,p)
d2.beta.tau=matrix(0,r,p)
for (j in 1:p){
  for(i in 1:n[j]){
    d2.beta.b[,j]=(-z.s[[j]][i]-B[[j]][i]*(e.s[[j]][i]*mu[[j]][i]-
    z.s[[j]][i])*B[[j]][i]*x.s[[j]][i]*x.b[[j]][i]+d2.beta.b[,j])
    d2.beta.tau[,j]=(z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*
    (1-B[[j]][i])*x.s[[j]][i]+d2.beta.tau[,j])
  }
  d2.b.b=matrix(0,p,p)
  trNb.b=matrix(0,p,p)
  d2.b.tau=matrix(0,p,p)
  for (j in 1:p){
    for (k in 1:p){
      trNb.b[j,k]=-2*q*(m0+p)*((2*b.prev[j]*b.prev[k])*(q+sum(b.prev^2))-
      6*b.prev[j]*b.prev[k]*(2+q+sum(b.prev^2)))/(q+sum(b.prev^2))^4
      d2.b.b[j,k]=2*q*(m0+p)*(Lam0^(-1)*b.prev[j]*b.prev[k]*Lam0^(-1))/
      (q+sum(b.prev*Lam0^(-1)*b.prev))^2-(r+p*q)*(trNb.b[j,k]*trH.0-
      trNb.1[j]*trNb.1[k])/(trH.0^2)
      d2.b.tau[j,j]=sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]]*(1-B[[j]])*x.b[[j]])
    }
  }
}

```

```

#####
##calcul des derivees de beta
sum1.1=matrix(0,r,p)
for (j in 1:p){
  mu[[j]]=as.vector(mu[[j]])
  B[[j]]=as.vector(B[[j]])
  for(i in 1:n[j]){
    sum1.1[,j]=(z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*x.s[[j]][i]+sum1.1[,j]
  }
  d1.beta=c(rep(0,r))
  for (k in 1:r){ d1.beta[k]=sum(sum1.1[k,j])}
  liste1=list(matrix(0,r,r))
  sum1.2=list(matrix(0,r,r))
  while(length(sum1.2)<p){ sum1.2=c(sum1.2,liste1)}
  d2.beta=matrix(0,r,r)
  for (j in 1:p){
    for(i in 1:n[j]){
      sum1.2[[j]]=-e.s[[j]][i]*B[[j]][i]*lam[[j]][i]*x.s[[j]][i]*%*%t(x.s[[j]][i,]+
      sum1.2[[j]])
    }
    d2.beta=sum1.2[[j]]+d2.beta
  }
#####
##calcul des derivees des b_j
d1.b=c(rep(0,p))
d2.b=c(rep(0,p))
trNb.0=c(rep(0,p))
trNb.1=c(rep(0,p))
trNb.2=c(rep(0,p))
Nb=c(rep(0,p))
q=1
m0=2
Lam0=1
for (j in 1:p){
  Nb[j]=q*(m0+p)*(Lam0^(-1)*(q+sum(b.prev*Lam0^(-1)*b.prev))-2*Lam0^(-1)*b.prev[j]*
  b.prev[j]*Lam0^(-1))/(q+sum(b.prev*Lam0^(-1)*b.prev))^2
  trNb.0[j]=Nb[j]
  trNb.1[j]=-q*(m0+p)*2*b.prev[j]*(2+q+sum(b.prev^2))/(q+sum(b.prev^2))^3
  trNb.2[j]=-2*q*(m0+p)*((2+q+sum(b.prev^2)+2*b.prev[j]^2)*(q+sum(b.prev^2))-
  6*b.prev[j]^2*(2+q+sum(b.prev^2)))/(q+sum(b.prev^2))^4
  d1.b[j]=sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]]*x.b[[j]])-q*(m0+p)*
  (Lam0^(-1)*b.prev[j])/(q+sum(b.prev*Lam0^(-1)*b.prev))
  d2.b[j]=-sum(e.s[[j]]*B[[j]]*lam[[j]]*x.b[[j]]*x.b[[j]])-Nb[j]
}
#####
##matrice hessienne
d2.beta.b=matrix(0,r,p)
d2.beta.tau=matrix(0,r,p)
for (j in 1:p){
  for(i in 1:n[j]){
    d2.beta.b[,j]=(-z.s[[j]][i]-B[[j]][i]*(e.s[[j]][i]*mu[[j]][i]-
    z.s[[j]][i])*B[[j]][i]*x.s[[j]][i]*x.b[[j]][i]+d2.beta.b[,j])
    d2.beta.tau[,j]=(z.s[[j]][i]-e.s[[j]][i]*mu[[j]][i])*B[[j]][i]*
    (1-B[[j]][i])*x.s[[j]][i]+d2.beta.tau[,j])
  }
  d2.b.b=matrix(0,p,p)
  trNb.b=matrix(0,p,p)
  d2.b.tau=matrix(0,p,p)
  for (j in 1:p){
    for (k in 1:p){
      trNb.b[j,k]=-2*q*(m0+p)*((2*b.prev[j]*b.prev[k])*(q+sum(b.prev^2))-
      6*b.prev[j]*b.prev[k]*(2+q+sum(b.prev^2)))/(q+sum(b.prev^2))^4
      d2.b.b[j,k]=2*q*(m0+p)*(Lam0^(-1)*b.prev[j]*b.prev[k]*Lam0^(-1))/
      (q+sum(b.prev*Lam0^(-1)*b.prev))^2-(r+p*q)*(trNb.b[j,k]*trH.0-
      trNb.1[j]*trNb.1[k])/(trH.0^2)
      d2.b.tau[j,j]=sum((z.s[[j]]-e.s[[j]]*mu[[j]])*B[[j]]*(1-B[[j]]*x.b[[j]]))
    }
  }
}

```

```

#####
###maximisation de beta
tol=0.001
diff=1
it=0
maxit=10
beta.prev0=beta.prev
theta0=c(beta.prev,tau0,b.prev)
while (diff > tol & it < maxit ){
  Hess=mat.hess(theta0)
  Grad=grad.theta(theta0)
  Hess.beta=Hess[c(1:r),c(1:r)]
  Grad.beta=Grad[1:r]
  invH.beta=solve(Hess.beta)
  beta.prev1=beta.prev0-invH.beta*%Grad.beta
  diff=max(abs(beta.prev0-beta.prev1))
  it<- it+1
  beta.prev0<- beta.prev1
  theta0<- c(beta.prev0,tau0,b.prev)
}
beta.prev00=beta.prev0
#####
###maximisation des b_j
tol=0.001
diff=1
it=0
maxit=90
maxit2= 100
b.prev1= rep(0, p)
b.prev00=b.prev
b.prev0=b.prev
theta0= c(beta.prev0,tau0,b.prev)
while (diff > tol & it < maxit ){
  for ( i in seq(1, p) ) {
    diff2= 1
    it2= 0
    while( diff2 > tol & it2 < maxit2 ){
      Hess=mat.hess(theta0)
      Hess.b=Hess[(r+p+i),(r+p+i)]
      Grad.b=Grad[(r+p+i)]
      invH.b= 1 / Hess.b
      b.prev1[i]=b.prev0[i]-invH.b*%Grad.b
      diff2= abs( b.prev0[i] - b.prev1[i] )
      it2= it2 + 1
      b.prev0[i]= b.prev1[i]
      theta0<- c(beta.prev0,tau0,b.prev0)}}
    diff=max(abs(b.prev00-b.prev1))
    it<- it+1
    b.prev0<- b.prev1
    b.prev00= b.prev1
    theta0<- c(beta.prev0,tau0,b.prev0)
  }
  beta.prev0[1]=beta.prev[1]
  theta0<- c(beta.prev0,tau0,b.prev0)
  matif=mat.hess(theta0)
  var.asp=solve(-matif)
  nm=r+2*p
  var.par=rep(0,nm)
  for (j in 1:nm){var.par[j]=var.asp[j,j]}
  est.par= c(beta.prev00,exp(tau0))
  var.xi=(exp(var.par[(r+1):(r+p)])-1)*(exp(tau0+var.par[(r+1):(r+p)]/2))^2
#####
###moyenne et variance des mu_ij
eta.u=split(rep(0,l),group)
mu.u=split(rep(0,l),group)
B.u=split(rep(0,l),group)

```

```

lam.u=split(rep(0,l),group)
sigm.u=split(rep(0,l),group)
var.lmu=split(rep(0,l),group)
mu.h=split(rep(0,l),group)
var.mu.h=split(rep(0,l),group)
for (j in 1:p){
eta.u[[j]]=x.s[[j]]%% beta.prev00 + x.b[[j]]*b.prev0[j]
mu.u[[j]]=exp(eta.u[[j]])
B.u[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu[[j]])
#B.u[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu.h[[j]])#p=1
lam.u[[j]]=(1-B.u[[j]])*y.s[[j]]+B.u[[j]]*mu[[j]]
sigm.u[[j]]=sqrt(lam.u[[j]]*(1-B.u[[j]])/e.s[[j]])
for(i in 1:n[j]){
var.lmu[[j]][i]=t(x.s[[j]][i,])%% var.asp[c(1,r),c(1,r)]%%x.s[[j]][i,]+
x.b[[j]][i]*var.asp[(r+p+), (r+p+)]*x.b[[j]][i+2]*t(x.s[[j]][i,])%%
var.asp[c(1,r),r+p+j]*x.b[[j]][i]
#var.lmu[[j]][i]=t(x.s[[j]][i,])%% var.asp[c(1,r),c(1,r)]%%x.s[[j]][i,]#p==1
mu.h[[j]]=exp(eta.u[[j]]+var.lmu[[j]]/2)
var.mu.h[[j]]=mu.h[[j]]^2*(exp(var.lmu[[j]])-1)
}
#####
###moyenne et variance des lambda_ij
b.h=split(rep(0,l),group)
omg=split(rep(0,l),group)
B.h=split(rep(0,l),group)
a1.h=split(rep(0,l),group)
a2.h=split(rep(0,l),group)
var.B.h=split(rep(0,l),group)
E1B1=split(rep(0,l),group)
E0B2=split(rep(0,l),group)
E1B2=split(rep(0,l),group)
E2B2=split(rep(0,l),group)
for (j in 1:p){
for(i in 1:n[j]){
omg[[j]][i]=var.asp[r+j,r+j]+var.lmu[[j]][i]-2*(var.asp[r+j,c(1,r)]%%
x.s[[j]][i,]+var.asp[r+j,r+p+j]*x.b[[j]][i])
b.h[[j]][i]= (var.asp[r+j,c(1,r)]%%x.s[[j]][i,]+var.asp[r+j,r+p+j]*
x.b[[j]][i]-var.lmu[[j]][i])/omg[[j]][i]
B.h[[j]]=exp(tau0[j])/(exp(eta.u[[j]]+log(e.s[[j]]))+exp(tau0[j]))
a1.h[[j]]=1/(omg[[j]]*(1-b.h[[j]]))
a2.h[[j]]=1/(omg[[j]]*b.h[[j]])
var.B.h[[j]]=b.h[[j]]*(1-b.h[[j]])/(a1.h[[j]]+a2.h[[j]]+1)
E1B1[[j]]=(a1.h[[j]]+b.h[[j]])/(a1.h[[j]]+a2.h[[j]])
E0B2[[j]]=(a1.h[[j]])*(a1.h[[j]]+1)/((a1.h[[j]]+a2.h[[j]])*
(a1.h[[j]]+a2.h[[j]]+1))
E1B2[[j]]=(a1.h[[j]]+b.h[[j]])*(a1.h[[j]]+b.h[[j]]+1)/((a1.h[[j]]+
a2.h[[j]])*(a1.h[[j]]+a2.h[[j]]+1))
E2B2[[j]]=(a1.h[[j]]+2*b.h[[j]])*(a1.h[[j]]+2*b.h[[j]]+1)/((a1.h[[j]]+
a2.h[[j]])*(a1.h[[j]]+a2.h[[j]]+1))
}
lam.h=split(rep(0,l),group)
var.lam.h=split(rep(0,l),group)
se.lam.h=split(rep(0,l),group)
var.lam.h3=split(rep(0,l),group)
for (j in 1:p){
#B.u[[j]]=xi[j]/(xi[j]+e.s[[j]]*mu[[j]])
#lam.h[[j]]=(1-B.u[[j]])*y.s[[j]]+B.u[[j]]*mu.h[[j]]
#se.lam.h[[j]]=sqrt(lam.h[[j]]*(1-B.h[[j]])/e.s[[j]])
#lam.h[[j]]=(1-B.u[[j]])*y.s[[j]]+B.u[[j]]*mu[[j]]
var.lam.h3[[j]]=(var.mu.h[[j]]+mu.h[[j]]*E2B2[[j]]+y.s[[j]]^2*E0B2[[j]]-
2*y.s[[j]]*mu.h[[j]]*E1B2[[j]]-(y.s[[j]]*B.h[[j]]-mu.h[[j]]*E1B1[[j]])^2
var.lam.h[[j]]=(1-2*B.h[[j]]+E0B2[[j]])*y.s[[j]]/e.s[[j]]+mu.h[[j]]*
(E1B1[[j]]-E1B2[[j]])/e.s[[j]]+var.lam.h3[[j]]
if (length(y.s[[j]])<3){var.lam.h[[j]]=lam.u[[j]]*(1-B.u[[j]])/e.s[[j]]}
se.lam.h[[j]]=sqrt(var.lam.h[[j]])
}

```

```

#####ouput
  se.par=sqrt(c(var.par[1:r],var.xi))
param=cbind(round(est.par,3),round(se.par,3))
dimnames(param)=list(c('beta0','beta1','xi1','xi2','xi3','xi4','xi5','xi6','xi7',
'xi8'),c('estimate','se'))
B.u=unsplit(B.u,group)
lam.h=unsplit(lam.h,group)
se.lam.h=unsplit(se.lam.h,group)
sigm.u=unsplit( sigm.u,group)

mB.u=matrix(0,l/2,3)
mB.u[,1]=c(rep(1,5),rep(3,7),5,6,rep(8,3),rep(9,11),rep(10,9),rep(11,3))
mB.u[,2]=B.u[1:(l/2)]
  mB.u[,3]=B.u[(l/2+1):l]
colnames(mlam.h)=c('AE','treatment','control')

mlam.h=matrix(0,l/2,3)
mlam.h[,1]=c(rep(1,5),rep(3,7),5,6,rep(8,3),rep(9,11),rep(10,9),rep(11,3))
mlam.h[,2]=lam.h[1:(l/2)]
  mlam.h[,3]=lam.h[(l/2+1):l]
colnames(mlam.h)=c('AE','treatment','control')

mse.lam.h=matrix(0,l/2,3)
mse.lam.h[,1]=c(rep(1,5),rep(3,7),5,6,rep(8,3),rep(9,11),rep(10,9),rep(11,3))
mse.lam.h[,2]=se.lam.h[1:(l/2)]
mse.lam.h[,3]=se.lam.h[(l/2+1):l]
colnames(mse.lam.h)=c('AE','treatment','control')

out=list(param,round(mB.u,3),round(mlam.h,3),round(mse.lam.h,3))
names(out)=c('parametres','B_ij','mean lam_ij','se lam_ij')
out

```