

VALÉRIE BÉRUBÉ

**MODÈLES AVANCÉS EN RÉGRESSION
APPLIQUÉS À LA TARIFICATION IARD**

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

FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

2007

©Valérie Bérubé, 2007

Résumé

En statistique, les modèles linéaires généralisés (GLM) sont utilisés pour développer des modèles pouvant s'adapter à plusieurs types de données. Cependant, leur présence en actuariat et dans les compagnies d'assurance est plus récente.

Dans ce mémoire, on présente d'abord brièvement des méthodes traditionnelles de tarification et de crédibilité. Par la suite, on décrit de façon exhaustive les modèles linéaires généralisés, auxquels on ajoute des effets aléatoires pour qu'ils soient applicables aux contextes de données longitudinales qui sont répandus en actuariat. On termine en démontrant certains liens existant entre la théorie statistique et la pratique actuarielle et un exemple avec les données de la Société d'Assurance Automobile du Québec (SAAQ) est présenté.

Avant-propos

Je tiens d'abord à remercier mes directeurs de recherche, Hélène Cossette et Étienne Marceau, sans qui ce projet de maîtrise aurait été impossible. Leur support tant académique que personnel m'ont permis de parfaire mes connaissances et de développer des techniques de recherche que je n'aurais jamais aussi bien apprises sur le marché du travail. Ils ont fait de moi la personne que je suis aujourd'hui et je leur en serai éternellement reconnaissante. Je dois aussi souligner leur patience et leur support lors de mon départ pour Montréal et la fin de ma maîtrise à distance. Ils sont demeurés présents et m'ont aidé à bien terminer malgré les nouvelles difficultés de l'étude et du travail.

Je voudrais aussi souligner l'apport considérable à ce travail de Thierry Duchesne, professeur au département de mathématique et statistique. D'abord avec son cours gradué portant sur la régression, il m'a passionné pour ce sujet et donné envie d'en apprendre d'avantage. Je tiens aussi à le remercier d'avoir accepté de me donner une lecture dirigée sur le sujet malgré son horaire très chargé. Finalement, merci d'avoir accepté de faire partie du comité d'évaluation de mon mémoire.

À toute l'équipe de tarification de la SAAQ, merci de m'avoir intégré aussi rapidement et d'avoir fait de la fin de ma maîtrise une expérience de travail des plus enrichissantes. En particulier, merci à Sébastien pour ses idées et sa programmation hors pair avec qui le travail en équipe s'est avéré être d'une efficacité incroyable et aussi à Josée d'avoir réussi à nous empêcher de parler continuellement de travail durant les pauses.

Je tiens à remercier particulièrement toute ma famille pour le soutien financier et les encouragements constants durant ces longues années d'études qui arriveront bientôt à leur fin. Merci aussi à mes amis et à mon copain qui ont su m'apporter le support nécessaire à la rédaction et le divertissement essentiel à la survie. Ce projet peut sembler être individuel, mais il n'aurait

pu être complété sans vous.

Je voudrais finalement remercier les différents organismes qui m'ont donné du soutien financier sous forme de bourse, soit la Chaire en actuariat Industrielle Alliance, la Banque Nationale, Hydro-Québec et le CRSNG.

Table des matières

Résumé	iii
Avant-propos	v
Liste des tableaux	xi
Table des figures	xiii
Introduction	1
1 Méthodes de tarification traditionnelles	5
1.1 Analyse univariée (one-way analysis)	6
1.2 Procédure du biais minimum	7
1.2.1 Description de la méthode originale	7
1.2.2 Autres modèles de base	10
1.2.3 Fonction du maximum de vraisemblance	11
1.2.4 Limitations	12
2 Résumé des notions de base en crédibilité	13
2.1 Primes de crédibilité	13
2.2 Modèle classique de Bühlmann	14
2.3 Modèle de Bühlmann-Straub	19
3 Notions de régression	23
3.1 Introduction	23
3.2 Introduction aux modèles linéaires généralisés	23
3.3 Modélisation et estimation	26
3.3.1 Estimation basée sur la fonction de vraisemblance	27
3.3.2 Autres méthodes d'estimation	31

3.4	Tests d'hypothèses	33
3.4.1	Fonction de vraisemblance spécifiée	33
3.4.2	Statistique t de Student	35
3.5	Modèles pour données de fréquence	35
3.5.1	Modèles binomiale négative NB1 et NB2	37
3.5.2	Tests de surdispersion	38
3.5.3	Introduction aux modèles mixtes pour données de fréquence	40
3.5.4	Théorème du double croisement	42
3.5.5	Terme d'offset	44
3.6	Exemple des visites chez le médecin	45
3.7	Modèles mixtes	47
3.7.1	Prévision	49
3.7.2	Composantes de variance	50
3.8	Modèles linéaires généralisés mixtes	51
3.8.1	Le pseudo-modèle	52
3.8.2	Estimateurs de la variabilité	54
3.9	Exemple des crises d'épilepsie	55
3.9.1	Modèle de base	56
3.9.2	L'approche GEE	56
3.9.3	Le modèle GLMM	57
4	Tarifification basée sur les GLM	59
4.1	Connections entre la méthode du biais minimum et les GLM	59
4.1.1	Modèle de Poisson multiplicatif	59
4.1.2	Modèle de régression linéaire	60
4.1.3	Autres modèles linéaires généralisés	63
4.2	Connections entre la crédibilité sur les montants moyens de sinistres et les modèles mixtes	65
4.2.1	Modèle de Bühlmann	66
4.2.2	Modèle avec poids de Bühlmann-Straub	68
4.2.3	Estimation des composantes de variances	69
5	Exemple pratique	71
5.1	Contexte et données	72
5.1.1	Variables catégoriques	73
5.1.2	Variables continues	75
5.2	Modèles linéaires généralisés appliqués aux données	77

<i>TABLE DES MATIÈRES</i>	ix
5.2.1 Distribution de fréquence	78
5.2.2 Distribution pour la sévérité	80
5.2.3 Modèle final	82
Conclusion	87
Bibliographie	92

Liste des tableaux

1.1	Exposition du portefeuille par classe	6
1.2	Montants moyens de sinistres par classe	6
3.1	Distribution de la fréquence des visites	46
3.2	Définitions des variables exogènes	47
3.3	Coefficients et écarts-type pour les modèles de Poisson, NB1 et NB2	48
5.1	Niveaux possibles des variables utilisé	80
5.2	Relativités selon les différents paramètres	84

Table des figures

3.1	Exemple avec un seul croisement, $f(y \mu)$ est en traitillé	44
3.2	Double croisements, $f(y \mu)$ est en traitillé	45
5.1	Distribution de l'exposition selon le niveau des primes	85

Introduction

Lorsqu'un individu fait face à un risque, il cherche généralement une façon de se protéger contre l'éventualité d'une perte, ce qui lui offre sécurité et tranquillité d'esprit. Les compagnies d'assurance comblent ainsi ce besoin de sécurité en assumant une partie des pertes en échange d'un montant, payé par la personne qui désire la couverture. Il revient donc à la compagnie d'assurance la tâche d'évaluer le risque auquel fait face l'individu et de le quantifier. Par la suite, l'individu peut assurer les pertes découlant d'un événement défini dans la protection en échange de la prime d'assurance proposée. Dans ce mémoire, on décrit différentes façons pour les assureurs de quantifier les risques.

Le premier chapitre est consacré à la description des méthodes actuelles de tarification. Tout d'abord, on présente l'analyse unilatérale, qui se trouve à être la toute première méthode de tarification et où on trouve des relativités de classe qui n'incluent pas les interactions présentes entre les facteurs utilisés. Pour palier à ces lacunes, les techniques du biais minimum ont été introduites. Cette méthode récursive a pour but d'éliminer les effets des autres variables et de bien isoler chacun des facteurs séparément. Les premières procédures du biais minimum étaient non paramétriques, mais des méthodes basées sur la fonction de vraisemblance, donc nécessitant des hypothèses quant à la distribution sous-jacente, ont fait leur apparition plus tard dans la littérature. Par la suite, on décrit les modèles de base utilisés en crédibilité, soit les modèles de Bühlmann et Bühlmann-Straub. La théorie de la crédibilité permet d'obtenir une prime qui est une combinaison linéaire de la prime collective et de l'expérience de l'assuré ou du groupe. La proportion qui leur est allouée dépend du niveau de crédibilité accordé à chacun des assurés.

Dans le chapitre 3, on décrit les modèles linéaires généralisés (GLM). On débute par la régression linéaire, qui suppose une distribution normale des termes d'erreurs. Par la suite, on généralise le principe de la régression

pour inclure d'autres distributions pour les erreurs, soit celles de la famille exponentielle. Ceci permet de trouver des modèles s'appliquant aux données discrètes et aux ratios, ce qui n'était pas possible avec la régression linéaire. Une fois que les modèles ont été présentés, on précise les méthodes pour estimer les paramètres et faire des tests d'hypothèses. Étant donné qu'une bonne partie du travail de l'actuaire consiste à modéliser le nombre de sinistres, on consacre une section particulière aux modèles pour données de fréquence. On compare alors la distribution de Poisson avec deux façon différentes de présenter la loi binomiale négative, qui peuvent remplacer la distribution de Poisson lorsqu'elle convient mal aux données. On conclut la section consacrée aux modèles pour données de fréquence avec un exemple statistique illustrant les différents concepts de la section.

On a débuté le chapitre en présentant la régression linéaire, qu'on a ensuite généralisée pour permettre l'utilisation d'un plus grand nombre de distributions. Maintenant, on travaille avec la distribution normale, à laquelle on ajoute des coefficients aléatoires différents pour chacun des individus pour donner les modèles mixtes. Pour ce faire, on doit être en présence de données longitudinales, c'est-à-dire des données pour des individus qui sont suivis dans le temps. Les effets aléatoires ajoutés parviennent à distinguer les comportements différents d'un individu à l'autre, mais constants dans le temps. Tout comme on a pu étendre la théorie des modèles linéaires aux GLM, on généralise les GLM et les modèles mixtes en une grande catégorie, soit les modèles linéaires généralisés mixtes (GLMM). On peut alors modéliser des données de fréquences longitudinales ou d'autres données provenant d'une distribution de la famille exponentielle qui sont collectées sur plusieurs périodes. Pour bien savoir comment mettre en application les GLMM on fait un exemple statistique et on explore les différentes méthodes pour modéliser les GLMM.

Maintenant que toute la théorie nécessaire a été vue, on peut lier les modèles statistiques aux contextes de tarification présentés. En effet, il s'avère que certains modèles de biais minimum sont des cas particuliers de modèles linéaires généralisés. On présente donc ces cas particuliers et quels GLM donnent les mêmes relativités. Par la suite, on exprime les modèles de Bühlmann et Bühlmann-Straub avec des modèles mixtes pour les montants de sinistres.

Finalement, on termine avec une tarification complète des données de dommages corporels de la Société d'Assurance Automobile du Québec (SAAQ) à l'aide des GLM. On explique de façon détaillée le cheminement utilisé et

INTRODUCTION

3

les résultats obtenus.

Chapitre 1

Méthodes de tarification traditionnelles

Pour ce premier chapitre, on fait la description de certaines méthodes qui sont ou étaient utilisées en tarification d'assurance générale. Il y a deux étapes principales lorsque les actuaires des compagnies d'assurance déterminent les primes qui seront payées par les assurés. Tout d'abord, ils doivent trouver le montant global de primes à amasser afin de payer les sinistres qui surviendront au courant de l'année. Par la suite, ils doivent choisir les caractéristiques des assurés qui déterminent le risque de ces derniers afin de faire payer une prime adéquate à chacun. On appelle aussi ces caractéristiques des facteurs de tarification, puisque les primes sont directement déterminées à partir de ces éléments. De cette façon, une prime de base est calculée, à laquelle on ajoute différents facteurs qui correspondent aux caractéristiques particulières de l'assuré. Ces facteurs, les relativités, sont déterminés selon plusieurs méthodes qu'on présente et analyse tout au long de ce mémoire. On présente dans la section suivante une méthode statistique permettant de trouver des relativités dans un contexte où les facteurs de tarification sont indépendants et où la distribution est uniforme. Par la suite, on présente la méthode du biais minimum introduite par Bailey et Simon (1960) qui permet de réduire au minimum les corrélations entre les facteurs.

1.1 Analyse univariée (one-way analysis)

L'analyse univariée, dite *one-way analysis* est une méthode de base permettant d'avoir un premier aperçu des relations entre les classes d'assurés. La littérature actuarielle y fait souvent référence afin de comparer les nouvelles méthodes de tarification. Comme son nom l'indique, l'analyse univariée calcule les relativités pour chacun des facteurs séparément, sans tenir compte des autres critères de tarification.

Pour déterminer les relativités des classes associées à un certain facteur, on doit comparer les montants de sinistres moyens de ses classes, pondérés par leur exposition. Par exemple, dans un contexte d'assurance automobile, on suppose qu'un assureur a divisé son territoire en deux catégories, soit rural et urbain. De plus, les hommes ont des taux différents des femmes. On a donc quatre classes différentes de taux, qui ont chacune une exposition donnée dans le tableau 1.1. Selon le contexte de tarification, l'exposition peut être donnée par différentes mesures. Par exemple, en assurance automobile, elle est souvent égale au nombre de véhicules par classe. En assurance habitation, il peut être préférable d'utiliser une unité d'exposition pour chaque tranche de 10 000 \$ d'assurance et ainsi de suite.

	Urbain	Rural	Total
Hommes	200	100	300
Femmes	100	200	300
Total	300	300	600

TAB. 1.1 – Exposition du portefeuille par classe

Il est aussi important d'avoir les montants moyens des sinistres par classe, ou les montants totaux dont on peut ensuite déduire les montants moyens en utilisant l'exposition. Ils sont donnés dans le tableau 1.2. Pour calculer

	Urbain	Rural	Total
Hommes	400	200	333,33
Femmes	200	100	133,33
Total	333,33	133,33	233,33

TAB. 1.2 – Montants moyens de sinistres par classe

la relativité du territoire urbain par rapport à celui du rural, on ne doit

pas tenir compte du sexe des assurés. On calcule donc les montants moyens de sinistres du territoire en regroupant les hommes et les femmes. Pour le territoire urbain, on a des sinistres moyens de

$$\frac{200 * 400 + 100 * 200}{200 + 100} = 333,33,$$

et on trouve de la même façon un montant de 133,33 \$ pour le territoire rural. La relativité de territoire serait dans cet exemple de $333,33/133,33 = 2,5$. On trouve aussi une relativité pour le sexe de 2,5 des hommes par rapport aux femmes. En utilisant comme cellule de référence les femmes de milieu rural, la méthode univariée conduit aux relativités suivantes pour les quatre classes :

	Urbain	Rural
Hommes	6,25	2,5
Femmes	2,5	1

Cependant, en regardant les sinistres par classe plutôt que par facteur indépendant comme pour la méthode univariée, on constate que pour chacun des territoires, les hommes ont une sévérité deux fois plus élevée que les femmes (400/200 pour le territoire urbain et 200/100 pour le rural). De plus, pour tous les sexes, le territoire urbain a une sévérité deux fois plus élevée que le rural. On remarque donc que les relativités sont biaisées pour les portefeuilles où la distribution de l'exposition n'est pas uniforme. Comme la technique ne permet pas la correction des relativités dans les cas où les assurés ne sont pas répartis uniformément, elle n'est plus utilisée par les assureurs. Elle est cependant très simple et permet d'expliquer les bases de la tarification.

1.2 Procédure du biais minimum

1.2.1 Description de la méthode originale

La technique du biais minimum a été introduite dans la littérature par Bailey et Simon (1960) dans les Proceedings de la Casualty Actuarial Society (PCAS). Son but fondamental est de trouver les relativités d'un système de classification à plusieurs niveaux qui représentent uniquement le risque en question, c'est-à-dire qui élimine le plus possible la dépendance entre les

8 CHAPITRE 1. MÉTHODES DE TARIFICATION TRADITIONNELLES

différents facteurs. On en déduit donc que la procédure du biais minimum n'a aucun avantage pour un système de tarification qui ne repose que sur un facteur, puisqu'il ne peut pas y avoir de corrélation entre les différents facteurs. Par contre, pour les tarifications faites en assurance automobile ou habitation, le biais minimum peut apporter des avantages considérables à l'approche traditionnelle unilatérale. Récemment, Feldblum et Brosius (2002) ont publié un guide très élaboré afin de bien décrire la méthode et en faciliter son application.

Dans leur article, Bailey et Simon (1960), définissent quatre critères qui doivent être respectés par les relativités finales choisies :

- Critère 1. Elles doivent reproduire l'expérience pour chacune des combinaisons de classes (les combinaisons âge/territoire par exemple) et pour l'expérience globale.
- Critère 2. Elles doivent refléter la crédibilité relative des différentes cellules.
- Critère 3. Elles doivent produire des écarts minimaux avec les données brutes pour le plus grand nombre de personnes possibles.
- Critère 4. Elles doivent produire un taux pour chacun des sous groupes qui est assez près de l'expérience pour pouvoir dire que les différences peuvent être le résultat de la chance.

La meilleure combinaison de relativités est celle qui réussira à respecter le plus possible les critères, en particulier le critère 3, qui demande d'avoir le plus petit biais possible pour chacune des classes. Par la suite, si plusieurs ensembles de relativités respectent tous les critères, le choix peut être basé sur des techniques plus subjectives, telles que la simplicité d'application, la ressemblance avec les relativités actuelles, la facilité de compréhension pour le personnel non-technique ou simplement par préférence de l'actuaire.

Tout comme les primes qui peuvent être déterminées sur une base additive ou multiplicative, le principe du biais minimum peut aussi être appliqué selon ces deux principes. En effet, pour déterminer une prime, on peut considérer qu'elle est donnée par une prime de base à laquelle on ajoute des montants selon les caractéristiques de l'assuré (additive), ou à laquelle on choisit de multiplier des facteurs selon ces mêmes caractéristiques (multiplicative). À ces deux types de primes, on peut associer diverses fonctions de biais qui conduisent à différentes formules pour déterminer les relativités. Bailey et Simon (1960) suggèrent la déviation moyenne et la statistique du khi-carré comme fonction de biais pour déterminer le meilleur modèle. On peut alors simplement poser la dérivée de la fonction de biais égale à zéro pour trou-

ver une formule itérative qui mène aux relativités. D'autres méthodes ont aussi été publiées, comme les moindres carrés et les relativités basées sur le maximum de vraisemblance introduites par Brown (1988). Cependant, la distribution des sinistres doit être spécifiée pour utiliser cette méthode.

Pour donner un exemple de développement de relativités, on utilise la fonction de biais du khi-carré. On considère un système de classification à deux variables, soit le territoire et une variable représentant le mérite du conducteur, comme le nombre d'années sans accident. La relativité pour le i^{e} territoire ($i = 1, 2, \dots, 5$) est donnée par x_i et celle pour la j^{e} classe de mérite ($j = 1, 2, \dots, 4$) par y_j . On note r_{ij} la relativité observée pour une personne de territoire i et de classe de mérite j . Finalement, n_{ij} est l'exposition de la cellule (i, j) et K est une constante fixée selon les données utilisées. Pour le modèle multiplicatif, on a la fonction suivante à minimiser :

$$\chi^2 = K \sum_{i,j} n_{ij} x_i y_j \left(\frac{r_{ij} - x_i y_j}{x_i y_j} \right)^2. \quad (1.1)$$

On constate donc que les relativités doivent absolument respecter le critère 2 qui stipule que la crédibilité relative de chacun des groupes doit être définie. Pour minimiser les écarts (critère 3), il suffit de poser les dérivées partielles égales à zéro. On obtient donc

$$x_i = \left(\frac{\sum_j \frac{n_{ij} r_{ij}^2}{y_j}}{\sum_j n_{ij} y_j} \right)^{\frac{1}{2}}$$

et

$$y_j = \left(\frac{\sum_i \frac{n_{ij} r_{ij}^2}{x_i}}{\sum_i n_{ij} x_i} \right)^{\frac{1}{2}}.$$

Avec cinq territoires et quatre classes de mérite, on a neuf équations et neuf inconnues. Comme les équations sont présentées sous une forme assez complexe, il est plus simple de procéder par itération. On peut utiliser comme point de départ les relativités trouvées avec la méthode traditionnelle de l'analyse unilatérale, pour en déduire une première approximation des x_i , qu'on utilise ensuite pour trouver les y_j . On itère jusqu'à ce que la différence entre deux ensembles de relativités soit nulle.

Pour ce qui est des relativités additives, on transforme (1.1) de façon à exprimer les coefficients sous une forme additive :

$$\chi^2 = K \sum_{i,j} \frac{n_{ij} (r_{ij} - x_i - y_j)^2}{x_i + y_j}.$$

Poser les dérivées partielles égales à zéro donne

$$\frac{\partial \chi^2}{\partial x_i} = K \sum_j n_{ij} - K \sum_j \frac{n_{ij} r_{ij}^2}{(x_i + y_j)^2} = 0,$$

et on peut utiliser Newton-Raphson pour trouver la correction Δx_i à être ajoutée à x_i :

$$\Delta x_i = - \frac{\partial \chi^2 / \partial x_i}{\partial^2 \chi^2 / \partial x_i^2} = \frac{\sum_j n_{ij} \left(\frac{r_{ij}}{x_i + y_j} \right)^2 - \sum_j n_{ij}}{2 \sum_j n_{ij} \left(\frac{r_{ij}}{x_i + y_j} \right)^2 \left(\frac{1}{x_i + y_j} \right)}. \quad (1.2)$$

Les expressions des Δy_j sont les mêmes, à l'exception des sommations qui sont effectuées sur les i à la place des j . Pour trouver les x_i et les y_j de façon itérative, on pose des relativités de départ, on utilise ces valeurs pour trouver les Δx_i et les Δy_j , valeurs qu'on ajoute respectivement aux x_i et aux y_j pour finalement utiliser ces nouvelles valeurs dans (1.2). Pour la k^e itération, on a $x_i^k = x_i^{k-1} + \Delta x_i^{k-1}$ et on refait l'opération jusqu'à ce que Δx_i et Δy_j soient nuls.

1.2.2 Autres modèles de base

Les relativités pour les autres fonctions de biais $f(x_i, y_j)$ sont développées de la même façon. La déviance moyenne pour la classe i à minimiser est

$$\frac{\sum_j n_{ij} (r_{ij} - f(x_i, y_j))}{\sum_j n_{ij} r_{ij}},$$

où $f(x_i, y_j)$ est égal à $x_i y_j$ pour le modèle multiplicatif et $x_i + y_j$ pour le modèle additif. On obtient donc respectivement les relativités x_i suivantes pour les modèles multiplicatifs et additifs :

$$x_i = \frac{\sum_j n_{ij} r_{ij}}{\sum_j n_{ij} y_j} \quad (1.3)$$

et

$$x_i = \frac{\sum_j n_{ij}(r_{ij} - y_j)}{\sum_j n_{ij}}. \quad (1.4)$$

Pour ce qui de la fonction de biais des moindres carrés avec une prime multiplicative, l'équation à minimiser est

$$\sum_{i,j} n_{ij}(r_{ij} - x_i y_j)^2,$$

ce qui mène aux relativités x_i suivantes :

$$x_i = \frac{\sum_j n_{ij} r_{ij} y_j}{\sum_j n_{ij} y_j^2}.$$

1.2.3 Fonction du maximum de vraisemblance

Une alternative à minimiser le biais existant entre les relativités observées et celles du modèle consiste à maximiser la fonction de vraisemblance. Pour ce faire, on doit spécifier une distribution pour les sinistres et ensuite trouver les relativités qui maximisent la vraisemblance. On doit aussi supposer que les r_{ij} sont indépendants. Par exemple, si les sinistres peuvent être modélisés par une distribution exponentielle de moyenne $n_{ij} E[r_{ij}] = n_{ij} x_i y_j$ (donc avec une forme multiplicative), alors la fonction de vraisemblance à maximiser est donnée par

$$\begin{aligned} L &= \prod_j \prod_i [f(x_i, y_j)]^{n_{ij}} \\ &= \prod_j \prod_i \left(\frac{1}{x_i y_j} \exp \left\{ -\frac{r_{ij}}{x_i y_j} \right\} \right)^{n_{ij}}. \end{aligned} \quad (1.5)$$

On peut, de façon équivalente, maximiser la fonction de log-vraisemblance :

$$l = \ln L = - \sum_j \sum_i \left(n_{ij} (\ln x_i + \ln y_j) + \frac{n_{ij} r_{ij}}{x_i y_j} \right). \quad (1.6)$$

Il est assez difficile de résoudre ce système analytiquement, puisque dans la plupart des cas, on a un nombre élevé d'équations. En effet, si la tarification

est basée, comme dans l'exemple précédent, sur cinq territoires et quatre classes de mérite, on se retrouve encore avec neuf équations à résoudre. Il est alors préférable de conserver une méthode par itérations. Si on pose les dérivés partielles de (1.6) égales à zéro, on obtient les relativités suivantes :

$$x_i = \frac{\sum_j \frac{n_{ij} r_{ij}}{y_j}}{\sum_j n_{ij}}.$$

Il est donc possible d'obtenir plusieurs ensembles de relativités avec cette méthode, puisqu'on peut utiliser plusieurs distributions, en plus d'utiliser une prime additive ou multiplicative.

1.2.4 Limitations

Le principal inconvénient de la technique du biais minimum est qu'il est impossible de déterminer si les facteurs qu'on utilise sont vraiment ceux qui expliquent les coûts. La méthode trouve des relativités, mais le choix des variables doit être fait sur une base complètement intuitive, sans fondement statistique. De plus, comme il n'y a aucune structure statistique en arrière du modèle, on ne peut pas avoir d'intervalle de confiance pour les paramètres ni aucun critère sur la qualité du modèle.

Par contre, en utilisant la fonction de vraisemblance, qui est en soi une extension du biais minimum puisqu'il n'y a plus de fonction de biais d'impliquée, on arrive à combler ces lacunes. On démontre au chapitre 4 comment on peut exprimer certains modèles de biais minimum en utilisant la fonction de vraisemblance.

Chapitre 2

Résumé des notions de base en crédibilité

2.1 Primes de crédibilité

La théorie de la crédibilité permet à l'assureur de déterminer la prime pour le contrat d'un assuré en se basant en partie sur son expérience et en partie sur l'expérience du portefeuille. En effet, un assureur qui chargerait à tous ses clients la moyenne des coûts du groupe perdrait les assurés ayant le moins de réclamations, puisque ceux-ci trouveraient facilement une meilleure prime chez un compétiteur, laissant ainsi à l'assureur les risques les plus mauvais. Un objectif des assureurs est donc de trouver la prime la plus juste possible afin d'être compétitif pour les meilleurs risques et couvrir les pertes de façon adéquate pour les mauvais. Bühlmann (1967) propose un modèle non-paramétrique permettant d'obtenir la meilleure prime linéaire au sens des moindres carrés estimant la prime de risque. Pour généraliser ce modèle, Bühlmann et Straub (1970) ont introduit des poids qui permettent de considérer l'importance d'un assuré dans la tarification. Par exemple, pour un régime d'assurance collective, on pourrait prendre la masse salariale comme poids, ce qui ferait en sorte que l'expérience d'un employeur ayant une masse salariale très élevée aurait un poids plus élevé dans le modèle. On approfondit ces deux modèles dans les sections suivantes, en se basant sur Bühlmann (2005), Cosette (1993), Dannenburg et coll. (1996), Goovaerts et Hoogstad (1987) et Goulet (1994).

2.2 Modèle classique de Bühlmann

Comme mentionné à la section 2.1, le modèle classique de Bühlmann nous permet de trouver une prime linéaire sans faire d'hypothèse quant à la distribution des variables aléatoires utilisées pour modéliser les données. Pour ce faire, on doit cependant émettre des hypothèses sur les relations existant entre ces variables. Soit X_{it} , le montant total des sinistres pour l'assuré i ($i = 1, \dots, n$) à la période t ($t = 1, \dots, T$). Si Θ_i est le paramètre de risque pour l'assuré i et X_i est le vecteur des T observations pour l'individu i , alors on suppose que les contrats (les paires X_i, Θ_i) sont indépendants et identiquement distribués. Aussi, pour un individu i , les variables aléatoires conditionnelles $X_{i1}|\Theta_i = \theta_i, X_{i2}|\Theta_i = \theta_i, \dots, X_{iT}|\Theta_i = \theta_i$ sont indépendantes et identiquement distribuées. On a donc l'indépendance entre les contrats et pour un même contrat dans le temps. La seconde hypothèse indique aussi la présence d'homogénéité dans le temps, c'est-à-dire que le comportement de l'assuré demeure le même avec les années. Pour ce qui est du paramètre de risque, il peut être défini comme étant le risque global que représente l'assuré. Le paramètre n'est pas observable, mais il détermine directement le niveau des sinistres qu'il subira. Par exemple, dans un cas simple, on pourrait considérer que les montants de sinistres sont de distribution exponentielle, tel que

$$X_{it}|\theta_i \sim \text{Exponentielle}(\theta_i),$$

et que le paramètre de risque suit une loi discrète telle que

$$P(\Theta = \theta_i) = \begin{cases} 0,3, & \theta_i = 20 \\ 0,6, & \theta_i = 50 \\ 0,1, & \theta_i = 100. \end{cases}$$

De cette façon, la loi exponentielle a un paramètre différent selon le niveau de risque de l'assuré i . Bien que l'illustration ait été faite avec une situation discrète pour le paramètre de risque, ce dernier peut aussi avoir une distribution continue.

Pour le modèle de Bühlmann, la prime pour la période $T + 1$ est donnée par

$$\begin{aligned} \widehat{\mu(\Theta_i)} &= \frac{T}{T + s^2/a} \bar{X}_i + \frac{s^2/a}{T + s^2/a} \bar{X} \\ &= z\bar{X}_i + (1 - z)\bar{X}, \end{aligned} \tag{2.1}$$

où

$$\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it},$$

$$\bar{X} = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n X_{it}$$

et

$$s^2 = E[\sigma^2(\Theta_i)] = E[\text{Var}(X_{it}|\Theta_i)]$$

$$a = \text{Var}[\mu(\Theta_i)] = \text{Var}[E(X_{it}|\Theta_i)].$$

On constate donc que s^2 représente la variabilité moyenne dans le temps de l'exposition d'un assuré, alors que le a donne la variabilité entre les primes de risque des assurés. Ainsi, un a qui tend vers l'infini fait tendre le facteur de crédibilité vers 1, puisque les primes de risque ($\mu(\theta_i)$) sont trop différentes les unes des autres pour utiliser la prime collective. À l'opposé, un a très petit indique que les primes de risque $\mu(\theta_i)$ sont très semblables et ne varient pas beaucoup, ce qui fait en sorte que le z tend vers zéro. De plus, un s^2 qui tend vers l'infini fait tendre le z vers 0, puisque l'expérience des assurés est trop instable pour s'y fier. Un petit s^2 indique que l'expérience dans le temps des assurés est très stable et peu volatile et est donc très crédible.

Pour arriver à la prime linéaire à variance minimale donnée en (2.1), on doit considérer un assuré et définir $\widehat{\mu(\Theta_i)}$ comme :

$$\widehat{\mu(\Theta)} = c_0 + \sum_{t=1}^T c_t X_t.$$

On cherche les c_t , $t = 0, \dots, T$ qui minimisent l'erreur quadratique moyenne, soit

$$\min_{c_0, \dots, c_T} E[(\mu(\Theta) - \widehat{\mu(\Theta)})^2].$$

Pour ce faire, on pose les $T + 1$ dérivés partielles par rapport aux c_t à zéro et on résoud le système d'équations. Tout d'abord, pour c_0 , on a

$$\frac{\partial}{\partial c_0} E \left[\left(\mu(\Theta) - c_0 - \sum_{t=1}^T c_t X_t \right)^2 \right] = E \left[-2 \left(\mu(\Theta) - c_0 - \sum_{t=1}^T c_t X_t \right) \right] = 0.$$

16 CHAPITRE 2. RÉSUMÉ DES NOTIONS DE BASE EN CRÉDIBILITÉ

Étant donné que $E[\mu(\Theta)] = E[E(X_t|\Theta)] = E[X_t]$, on arrive à

$$c_0 = \left(1 - \sum_{t=1}^T c_t\right) E[\mu(\Theta)].$$

Maintenant qu'on a une expression pour c_0 , on peut l'insérer dans l'équation de départ pour effectuer les autres dérivés. Pour $j = 1, \dots, T$, on a

$$\begin{aligned} \frac{\partial}{\partial c_j} E \left[\left(\mu(\Theta) - E[\mu(\Theta)] - \sum_{t=1}^T c_t (X_t - E[\mu(\Theta)]) \right)^2 \right] &= 0 \\ \Leftrightarrow \\ E \left[-2 \left(\mu(\Theta) - E[\mu(\Theta)] - \sum_{t=1}^T c_t (X_t - E[\mu(\Theta)]) \right) (X_j - E[\mu(\Theta)]) \right] &= 0, \end{aligned}$$

ce qui peut être exprimé comme

$$\begin{aligned} E \left[(\mu(\Theta) - E[\mu(\Theta)])(X_j - E[\mu(\Theta)]) \right] &= \sum_{t=1}^T c_t E \left[(X_t - E[\mu(\Theta)])(X_j - E[\mu(\Theta)]) \right] \\ \text{Cov}[\mu(\Theta), X_j] &= \sum_{t=1}^T c_t \text{Cov}(X_t, X_j), \end{aligned} \quad (2.2)$$

pour $j = 1, \dots, T$. Dans le but de déterminer les coefficients c_j , on doit trouver les formes des différentes covariances. Pour ce faire, on conditionne sur le paramètre de risque Θ . On trouve alors

$$\begin{aligned} \text{Cov}[\mu(\Theta), X_j] &= \text{Cov}[E(\mu(\Theta)|\Theta), E(X_j|\Theta)] + E[\text{Cov}(\mu(\Theta)|\Theta, X_j|\Theta)] \\ &= \text{Var}[\mu(\Theta)] + 0 \\ &= a, \end{aligned}$$

car $\mu(\Theta)|\Theta$ est une constante. On a aussi besoin de

$$\begin{aligned} \text{Cov}[X_t, X_j] &= \text{Cov}[E(X_t|\Theta), E(X_j|\Theta)] + E[\text{Cov}(X_t|\Theta, X_j|\Theta)] \\ &= \text{Var}[\mu(\Theta)] + \delta_{jt} E[\sigma^2(\Theta)] \\ &= a + \delta_{jt} s^2, \end{aligned}$$

où δ_{jt} est le delta de Kronecker, qui est égal à 1 si $t = j$ et à 0 dans les autres cas. En effet, on a déjà mentionné que les sinistres de deux périodes

différentes sont indépendants lorsqu'on connaît le paramètre de risque. En remplaçant ces expressions dans (2.2), on obtient

$$a = a \sum_{t=1}^T c_t + c_j s^2,$$

pour $j = 1, \dots, T$. On constate que les T équations sont identiques et donc que $c_1 = c_2 = \dots = c_T$. Ainsi, on a

$$a = aTc_1 + c_1 s^2,$$

ce qui est équivalent à

$$c_1 = \frac{a}{aT + s^2} = c_2 = \dots = c_T.$$

On peut donc conclure que, pour l'assuré i , la prime linéaire qui minimise l'erreur quadratique moyenne est exprimée comme

$$\begin{aligned} \widehat{\mu(\Theta_i)} &= \left(1 - \sum_{t=1}^T \frac{a}{aT + s^2}\right) E[\mu(\Theta_i)] + \frac{a}{aT + s^2} \sum_{t=1}^T X_{it} \\ &= \frac{Ta}{aT + s^2} \sum_{t=1}^T \frac{X_{it}}{T} + \left(1 - \frac{aT}{aT + s^2}\right) E[\mu(\Theta_i)] \\ &= z\bar{X}_i + (1 - z)E[\mu(\Theta_i)]. \end{aligned}$$

Plus le facteur de crédibilité z est élevé, plus le poids accordé à l'expérience individuelle sera grand. À l'opposé, un petit z entraînera un poids plus important à la prime collective.

Les estimateurs sans biais proposés pour les paramètres s^2 et a sont les suivants :

$$\hat{s}^2 = \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T (X_{it} - \bar{X}_i)^2$$

et

$$\hat{a} = \frac{1}{n-1} \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \frac{\hat{s}^2}{T}.$$

De plus, on estime la prime collective par l'estimateur sans biais \bar{X} .

Un estimateur sans biais doit satisfaire la condition suivante :

$$E[\hat{\Theta}] = \Theta,$$

où $\hat{\Theta}$ est un estimateur du paramètre quelconque Θ . Par exemple, on peut démontrer que \hat{s}^2 est un estimateur sans biais comme suit :

$$\begin{aligned} E[\hat{s}^2] &= \frac{1}{n(T-1)} E \left[\sum_{i=1}^n \sum_{t=1}^T (X_{it} - \bar{X}_i)^2 \right] \\ &= \frac{1}{n(T-1)} E \left[\sum_{i=1}^n \sum_{t=1}^T \left((X_{it} - E[\mu(\Theta)]) - (\bar{X}_i - E[\mu(\Theta)]) \right)^2 \right] \\ &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T \left[E \left[(X_{it} - E[\mu(\Theta)])^2 \right] \right. \\ &\quad \left. - 2E \left[(X_{it} - E[\mu(\Theta)])(\bar{X}_i - E[\mu(\Theta)]) \right] \right. \\ &\quad \left. + E \left[(\bar{X}_i - E[\mu(\Theta)])^2 \right] \right] \\ &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T \text{Var}[X_{it}] - 2\text{Cov}[X_{it}, \bar{X}_i] + \text{Var}[\bar{X}_i]. \end{aligned}$$

Par la suite, on peut montrer que

$$\text{Var}[X_{it}] = a + s^2,$$

$$\text{Cov}[X_{it}, \bar{X}_i] = a + \frac{s^2}{T}$$

et

$$\text{Var}[\bar{X}_i] = a + \frac{s^2}{T}.$$

Avec ces éléments, on trouve que

$$\begin{aligned}
 E[\hat{s}^2] &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T \left(a + s^2 - a - \frac{s^2}{T} \right) \\
 &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T \frac{T-1}{T} s^2 \\
 &= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T s^2 \\
 &= s^2.
 \end{aligned}$$

De la même façon, on peut démontrer que l'estimateur proposé pour estimer a est sans biais.

Le principal désavantage de ce modèle est qu'il accorde le même poids à tous les contrats. Cela peut être approprié lorsqu'un contrat couvre une voiture en assurance automobile, mais lorsqu'il s'agit d'assurance collective et que le contrat couvre un nombre variable d'employés, il devient préférable d'accorder une crédibilité différente à chacun des contrats, selon leur taille. Le modèle proposé par Bühlmann et Straub (1970), qui introduit des poids permettant de relativiser l'importance des contrats au sein du portefeuille, sera considéré dans la prochaine sous-section.

2.3 Modèle de Bühlmann-Straub

Pour palier au manque de flexibilité du modèle de Bühlmann qui requiert des poids égaux pour tous les contrats, on présente le modèle de Bühlmann-Straub qui introduit des poids w_{it} qui varient pour tous les contrats selon l'importance des assurés au sein du portefeuille et dans le temps. Les hypothèses de base sont donc modifiées, puisque les sinistres de chacun des contrats ne sont plus identiquement distribués. En effet, il est normal pour un assuré avec une exposition deux fois plus élevée qu'un autre risque d'avoir deux fois plus de sinistres. Il n'est donc pas approprié de considérer que les contrats ont des sinistres provenant de la même distribution. Pour ce qui est de la première hypothèse du modèle de Bühlmann, on conserve l'indépendance pour les paires (X_{it}, Θ_i) , mais il n'y a que les variables Θ_i qui sont identiquement distribuées. L'indépendance est conservée pour la seconde hy-

pothèse, mais l'homogénéité est restreinte à la valeur espérée des observations, soit $E[X_{it}|\Theta_i]$. De plus, la variance conditionnelle des observations est donnée par

$$\text{Var}[X_{it}|\Theta_i] = \frac{\sigma^2(\Theta_i)}{w_{it}}, \quad (2.3)$$

donc elle est inversement proportionnelle à la taille d'un assuré. Dans ce modèle, $\sigma^2(\Theta_i)$ représente la variance pour une unité.

On constate que, en général, un «gros» assuré aura une expérience plus stable, donc une variance globale plus faible, ce qui est cohérent avec l'expression de la variance en (2.3). La prime linéaire qui minimise l'erreur quadratique est de la même forme que dans le modèle de Bühlmann, à l'exception du facteur de crédibilité z_i qui tient compte des poids de l'assuré i et est donc différent pour chacun des assurés. Pour bien définir le modèle, on a besoin de

$$\begin{aligned} w_{..} &= \sum_{i=1}^n w_i = \sum_{i=1}^n \sum_{t=1}^T w_{it} \\ z &= \sum_{i=1}^n z_i \\ X_{iw} &= \sum_{t=1}^T \frac{w_{it}}{w_i} X_{it} \\ X_{w.} &= \sum_{i=1}^n \frac{w_i}{w_{..}} X_{iw} \\ X_{zw} &= \sum_{i=1}^n \frac{z_i}{z} X_{iw}. \end{aligned}$$

Étant donné que les poids des assurés pour chacune des années sont différents, on obtient un facteur de crédibilité différent pour chacun des assurés dans l'expression de la prime, soit

$$z_i = \frac{w_i}{w_i + s^2/a},$$

où $a = \text{Var}[\mu(\Theta_i)]$ et $s^2 = E[\sigma^2(\Theta_i)]$. On constate alors qu'on peut retrouver le modèle de Bühlmann en posant $w_{it} = 1, \forall i$ et $\forall t$. On a alors $w_i = T$ pour tous les i et on arrive au z de (2.1).

En utilisant la même méthode que pour le modèle de Bühlmann, on a que la prime de l'assuré i à la période $T + 1$ est donnée par

$$P_{i,T+1} = z_i X_{iw} + (1 - z_i) E[\mu(\Theta_i)].$$

Pour estimer $E[\mu(\Theta_i)]$, on se sert de X_{zw} . On pourrait aussi utiliser X_{ww} , mais parmi les estimateurs sans biais de $E[\mu(\Theta_i)]$ de la forme $\sum c_{ij} X_{ij}$, X_{zw} est celui qui a la variance minimale.

Les estimateurs des paramètres a et s^2 sont aussi quelque peu différents de ceux du modèle de Bühlmann, dû à l'introduction des poids. Pour s^2 , il a été proposé

$$\hat{s}^2 = \frac{1}{T(n-1)} \sum_{i=1}^n \sum_{t=1}^T w_{it} (X_{it} - X_{iw})^2,$$

et pour estimer a ,

$$\hat{a} = \frac{w_{..}}{w_{..}^2 - \sum_{i=1}^n w_i^2} \left(\sum_{i=1}^n w_i (X_{iw} - X_{ww})^2 - (n-1) \hat{s}^2 \right)$$

a été proposé. Ces estimateurs sont sans biais, mais il peut arriver que \hat{a} soit négatif. On peut alors prendre le maximum entre \hat{a} et 0, mais cela introduit un biais.

Un pseudo-estimateur, celui de Bischel-Straub (1976), est proposé pour palier à cette lacune. Il n'est jamais négatif et est sans biais :

$$\tilde{a} = \frac{1}{n-1} \sum_{i=1}^n z_i (X_{iw} - X_{zw})^2.$$

\tilde{a} est qualifié de pseudo-estimateur, car on a besoin d'un estimé de a pour déterminer les z_i . Parmi les estimateurs de la forme

$$\sum_{i=1}^n c_i (X_{iw} - X_{zw})^2,$$

\tilde{a} est celui qui a la variance minimale. Cependant, il a le désavantage d'être trouvé par des méthodes numériques jusqu'à la convergence si on l'utilise pour estimer a .

Un avantage de ces modèles est qu'on peut les appliquer tant au nombre de sinistres qu'aux montants moyens ou totaux. En effet, les estimateurs choisis pour ces paramètres sont entièrement non-paramétriques, ce qui fait en sorte qu'on peut très bien décider que les X_{it} sont des nombres ou des montants de sinistre. Par ailleurs, à part le paramètre de risque qui n'est pas quantifiable, il est impossible d'intégrer dans le modèle les caractéristiques de l'assuré comme l'âge ou le sexe du conducteur dans le cas de l'assurance automobile. Il serait donc intéressant d'introduire une catégorie de modèles qui permettent d'intégrer les informations disponibles sur les individus.

On mentionne que les modèles de Bühlmann et de Bühlmann-Straub sont présentés notamment dans Klugman et coll. (2004).

Chapitre 3

Notions de régression

3.1 Introduction

Le but de ce mémoire étant d'appliquer des modèles statistiques de régression avancés à des situations de tarification actuarielle, il est important de bien se familiariser avec les techniques de régression afin de les comprendre et de percevoir leurs utilisations possibles. On débute la présentation des modèles statistiques avec la régression linéaire, à laquelle on ajoute des composantes aléatoires. Par la suite, on s'intéresse aux modèles où la variable réponse n'est pas de distribution normale, puisqu'il existe plusieurs situations actuarielles où les données sont de nature discrète. Pour terminer, comme dans le cas de la régression normale, on ajoute des effets aléatoires.

Les notions générales de ce chapitre sont basées sur Cameron et Trivedi (1998), Diggle et coll. (2002), McCullagh et Nelder (1997) et Winkelmann (2000). On peut aussi trouver des applications des notions de régression avancées dans Denuit et Charpentier (2004, 2005) et Gouieroux (1999).

3.2 Introduction aux modèles linéaires généralisés

La régression linéaire peut être appliquée dans un contexte où il y a présence d'observations qui semblent être expliquées en partie par des variables disponibles pour chacune des observations. Par exemple, il est possible de construire un modèle qui permet d'expliquer le comportement aléatoire des

montants de sinistres en fonction de l'âge du conducteur, de son sexe, du type de véhicule, etc. Dans ce contexte, les montants de sinistres sont représentés par Y_i et les caractéristiques de l'assuré i par le vecteur de variables \mathbf{x}_i .

On considère une variable endogène, ou dépendante, Y qui est définie en partie par un certain nombre p' de variables exogènes, $x_1, \dots, x_{p'}$, pour les n unités indépendantes. En regroupant toutes les informations pour le i^{e} individu en un vecteur $\mathbf{x}'_i = (1, x_{i,1}, \dots, x_{i,p'})$, on représente les informations pour les n unités en une matrice $n \times p$, où $p = p' + 1$:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p'} \\ 1 & x_{2,1} & \cdots & x_{2,p'} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p'} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_n \end{pmatrix} = (1, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{p'}).$$

Le vecteur de 1 est essentiel pour permettre au modèle d'avoir une ordonnée à l'origine. L'ordonnée à l'origine est la moyenne de la variable endogène lorsque toutes les variables exogènes sont égales à zéro. Il est aussi possible de regrouper les variables exogènes en vecteurs colonnes, $\mathbf{x}_i = (x_{1,i}, \dots, x_{p',i})'$. Considérant que les variables exogènes n'ont pas toutes la même importance et la même relation avec la variable endogène Y , on introduit des coefficients β qui représentent la relation existante entre Y et les \mathbf{x}_i . En effet, un modèle qui explique des coûts de sinistres peut inclure comme variable l'âge de l'assuré et son nombre d'années d'expérience, alors que les coûts ne sont pas simplement exprimés comme la somme de ces deux variables. En introduisant des coefficients qui sont associés à l'âge et au nombre d'années d'expérience, il est possible de bien définir le lien existant entre les coûts et les variables. La variable endogène n'étant jamais expliquée entièrement par les variables exogènes dans la réalité, on doit ajouter un terme d'erreur ε_i associé à chacune des observations. La variable dépendante est donc expliquée par

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \quad (3.1)$$

où $\mathbf{Y} = (Y_1, \dots, Y_n)'$, $\beta = (\beta_0, \beta_1, \dots, \beta_{p'})'$ et $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$.

Pour le i^{e} individu, on définit le prédicteur linéaire η_i par

$$\eta_i = \mathbf{x}'_i \beta,$$

qui correspond à une combinaison linéaire des variables exogènes et de leur coefficient respectif. En régression linéaire comme en (3.1), $E[Y] = \eta$, mais

il n'est pas toujours de même pour d'autres types de modèles qu'on présente plus tard dans cette section. Le prédicteur linéaire est donc utile pour définir la relation entre la variable endogène et les variables indépendantes. En posant $E(Y) = \mu$, on introduit la relation entre l'espérance de Y et le prédicteur linéaire comme

$$\mu = g^{-1}(\eta),$$

où $g(\cdot)$ est appelé la fonction de lien. Parmi les fonctions de lien les plus courantes, on retrouve le lien identité, $g(u) = u$, le lien log, $g(u) = \ln(u)$, le lien logit, $g(u) = \ln(u/(1-u))$ et le lien probit, $g(u) = \Phi(u)$, où $\Phi(z) = P(Z \leq z)$ et Z obéit à une loi normale standard.

Dans tous les cas vus précédemment, on a supposé que la variable endogène Y était de distribution normale. Cependant, étant donné que la distribution normale est continue, plusieurs situations ne peuvent pas être modélisées de cette façon. On a par exemple les nombres de sinistres, le nombre d'éléments réussis à un test et la guérison ou la rechute d'une maladie qui nécessiteraient d'autres lois que la normale et d'autres fonctions de lien que le lien identité. On introduit donc les modèles linéaires généralisés (GLM) qui permettent d'utiliser des techniques de régression en autant que la variable endogène provienne d'une distribution de la famille exponentielle.

En effet, la distribution d'une variable Y est dite de famille exponentielle si elle peut être écrite sous la forme

$$f_Y(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\},$$

où $a(\cdot)$, $b(\cdot)$ et $c(\cdot)$ sont des fonctions quelconques. On a alors que, pour toutes les distributions de la famille exponentielle,

$$E[Y] = b'(\theta)$$

et

$$\text{Var}[Y] = b''(\theta).$$

Si ϕ est connu, le paramètre θ peut être identifié comme étant le paramètre canonique et le paramètre ϕ est celui de dispersion. On montre facilement

que la loi normale fait partie de cette famille. En effet,

$$\begin{aligned} f_Y(y; \theta, \phi) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(y - \mu)^2}{2\sigma^2} \right\} \\ &= \exp \left\{ \frac{y\mu - \mu^2/2}{\sigma^2} - \frac{1}{2} \left(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right) \right\}. \end{aligned}$$

On peut donc en déduire que :

$$\begin{aligned} a(\phi) &= \phi \\ b(\theta) &= \theta^2/2 \\ c(y, \phi) &= -\frac{1}{2} \left(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right), \end{aligned}$$

où $\theta = \mu$ et $\phi = \sigma^2$. On confirme donc que $E[Y] = 2\mu/2 = \mu$ et $\text{Var}[Y] = (1)\sigma^2 = \sigma^2$.

Les lois de Poisson, binomiale et binomiale négative sont des exemples de lois discrètes qui font partie de cette famille. On peut exprimer la loi de Poisson sous la forme

$$f_Y(y; \mu) = \exp(y \ln(\mu) - \mu - \ln y!). \quad (3.2)$$

Son paramètre canonique est de $\theta = \ln \mu$ et le paramètre de dispersion ϕ est égal à 1. On voit aussi que $b(\theta) = b(\ln(\mu)) = \mu$, donc $b(\theta) = \exp(\theta)$.

On présente dans la section suivante deux approches générales souvent utilisées pour faire l'estimation des modèles linéaires généralisés (GLM). On s'intéresse d'abord à l'estimation basée sur le maximum de la fonction de vraisemblance, puis à la méthode de quasi-vraisemblance, une analyse qui permet à certains aspects de la distribution de ne pas être spécifiés, parfois même erronés.

3.3 Modélisation et estimation

Dans un premier temps, on présente une méthode basée sur la fonction de vraisemblance qui suppose que la distribution d'où proviennent les observations est connue avec certitude. Cette condition étant assez stricte, on étudie par la suite comment on peut l'alléger et on regarde l'effet sur les estimations de façon générale et pour le cas spécial de la famille exponentielle. En effet,

pour une distribution de la famille exponentielle, on a seulement besoin de bien spécifier l'espérance pour obtenir des estimateurs convergents.

Un estimateur est dit convergent s'il est très près de la vraie valeur du paramètre pour un échantillon de grande taille, ce qui peut s'écrire comme

$$\lim_{n \rightarrow \infty} P[|\hat{\theta}_n - \theta^*| > \varepsilon] = 0, \quad \forall \varepsilon > 0,$$

où $\hat{\theta}_n$ est l'estimateur trouvé pour un échantillon de taille n et θ^* est le vrai paramètre.

3.3.1 Estimation basée sur la fonction de vraisemblance

Les méthodes d'estimation basées sur la fonction de vraisemblance sont des méthodes pour lesquelles la fonction de densité ou de masse de probabilité de la variable dépendante est spécifiée. On sait que la distribution des y_i sachant le vecteur des variables indépendantes et le vecteur de paramètres inconnus θ est $f(y_i | \mathbf{x}_i, \theta)$, et on considère que les $y_i | \mathbf{x}_i, \theta$ sont indépendants.

Le principe du maximum de vraisemblance consiste à choisir pour estimateurs de θ , les valeurs qui maximisent la probabilité jointe d'observer les valeurs échantillonales y_1, \dots, y_n . Cette probabilité est appelée fonction de vraisemblance et est définie par

$$L(\theta) = \prod_{i=1}^n f(y_i | \mathbf{x}_i, \theta).$$

Il peut être difficile de travailler avec un produit de fonctions de densité et comme le θ qui maximise $L(\theta)$ est le même qui maximise son logarithme, il est préférable d'utiliser la fonction de log-vraisemblance, définie par

$$\mathcal{L}(\theta) = \ln L(\theta) = \sum_{i=1}^n \ln f(y_i | \mathbf{x}_i, \theta).$$

Les estimateurs du maximum de vraisemblance, $\hat{\theta}_{MV}$, sont les racines de la dérivée de la fonction de log-vraisemblance,

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln f_i}{\partial \theta},$$

où $f_i = f(y_i | \mathbf{x}_i, \boldsymbol{\theta})$. Comme la distribution asymptotique des estimateurs est basée sur l'hypothèse que la distribution est bien spécifiée, les y_i ont comme densité $f(y_i | \mathbf{x}_i, \boldsymbol{\theta}^*)$, où $\boldsymbol{\theta}^*$ regroupe les vrais paramètres de la distribution.

Cette technique permet de trouver des estimateurs qui sont dits convergents. Maintenant que ces estimateurs convergents sont trouvés pour les paramètres $\boldsymbol{\theta}^*$, on s'intéresse à leur variance. On trouve que, si la distribution est bien spécifiée, cette matrice de variance-covariance est asymptotiquement donnée par l'inverse de

$$\mathbf{A} = - \lim_{n \rightarrow \infty} \frac{1}{n} E \left[\sum_{i=1}^n \frac{\partial^2 \ln f_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}^*} \right].$$

Il en résulte que la distribution asymptotique de $\hat{\boldsymbol{\theta}}_{MV}$, le paramètre estimé selon la méthode du maximum de vraisemblance, est donnée par

$$\hat{\boldsymbol{\theta}}_{MV} \sim N(\boldsymbol{\theta}^*, \mathbf{A}^{-1}).$$

Lorsqu'on évalue \mathbf{A} à $\hat{\boldsymbol{\theta}}$ et qu'on en prend l'espérance, on obtient la matrice d'information de Fisher $\mathcal{I}(\boldsymbol{\theta})$ qui est souvent utilisée comme référence dans la littérature.

Étant donné la forme particulière de la fonction de densité des membres de la famille exponentielle, la fonction de log-vraisemblance, $\mathcal{L}(\boldsymbol{\theta}, \phi; y)$ est une fonction très utile pour ces modèles. Elle peut être exprimée sous la forme

$$\mathcal{L}(\boldsymbol{\theta}, \phi) = \frac{y\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\phi)} + c(y, \phi). \quad (3.3)$$

La fonction de log-vraisemblance permet entre autre de trouver les expressions de l'espérance et de la variance de la variable endogène Y . On suppose en premier lieu que le paramètre ϕ est connu.

On trouve $E(Y)$ en dérivant la fonction de log-vraisemblance donnée en (3.3) par rapport à $\boldsymbol{\theta}$. D'abord, on a que

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_Y(y; \boldsymbol{\theta}, \phi) &= \frac{\frac{\partial}{\partial \boldsymbol{\theta}} f_Y(y; \boldsymbol{\theta}, \phi)}{f_Y(y; \boldsymbol{\theta}, \phi)} \\ &= \frac{y - b'(\boldsymbol{\theta})}{a(\phi)}, \end{aligned} \quad (3.4)$$

où $b'(\theta)$ est la dérivée de $b(\theta)$ par rapport à θ . On peut réécrire (3.4) comme suit :

$$\frac{\partial}{\partial \theta} f_Y(y; \theta; \phi) = \frac{y - b'(\theta)}{a(\phi)} f_Y(y; \theta; \phi). \quad (3.5)$$

En intégrant de chaque côté (3.5), on obtient

$$\int_{D_Y} \frac{\partial}{\partial \theta} f_Y(y; \theta; \phi) = \int_{D_Y} \frac{y - b'(\theta)}{a(\phi)} f_Y(y; \theta; \phi).$$

Comme l'intégrale sur le domaine de Y de la fonction de densité de Y est égale à 1, on trouve que

$$0 = \frac{E(Y) - b'(\theta)}{a(\phi)},$$

donc que

$$E(Y) = b'(\theta).$$

De la même façon, on peut trouver que

$$0 = -\frac{b''(\theta)}{a(\phi)} + \frac{\text{Var}(Y)}{a^2(\phi)},$$

ce qui donne l'expression de la variance de Y

$$\text{Var}(Y) = b''(\theta)a(\phi).$$

Par exemple, dans le cas où $Y \sim \text{Poisson}(\mu)$, on a trouvé en (3.2) que $b(\theta) = \exp(\theta)$, (car $\exp \ln(\mu) = \mu$) ce qui mène à

$$\begin{aligned} E(Y) &= \frac{d}{d\theta} \exp(\theta) \\ &= \exp(\theta) \\ &= \mu. \end{aligned}$$

Comme $a(\phi) = 1$, on trouve que la variance est égale à l'espérance.

Exemple pour le modèle de Poisson

On applique les résultats de la section précédente en supposant que les y_i obéissent à une loi de Poisson. D'abord, la fonction de densité des y_i sachant la valeur des variables exogènes est donnée par

$$f(y_i | \mathbf{x}_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!},$$

où $\mu_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ et $\boldsymbol{\beta}$ est un vecteur de p paramètres. On peut donc exprimer la moyenne des y_i comme une fonction des variables exogènes. De plus, on obtient une fonction de vraisemblance égale à

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \exp\{y_i(\mathbf{x}'_i \boldsymbol{\beta}) - \exp(\mathbf{x}'_i \boldsymbol{\beta})\} - y_i!.$$

La fonction de log-vraisemblance est donc

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^n \{y_i(\mathbf{x}'_i \boldsymbol{\beta}) - \exp(\mathbf{x}'_i \boldsymbol{\beta}) - \ln y_i!\}.$$

En prenant les dérivées partielles par rapport aux β_j , on obtient que les $\hat{\boldsymbol{\beta}}_{MV}$ sont les solutions du système d'équation

$$\sum_{i=1}^n \{y_i - \exp(\mathbf{x}'_i \boldsymbol{\beta})\} \mathbf{x}_i = \mathbf{0}.$$

En dérivant une seconde fois par rapport au vecteur des β_j , on trouve que \mathbf{A} est égal à

$$\mathbf{A} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i \exp(\mathbf{x}'_i \boldsymbol{\beta}^*) \mathbf{x}_i \mathbf{x}'_i, \quad (3.6)$$

où $\boldsymbol{\beta}^*$ est le vrai paramètre de la distribution. On obtient finalement la distribution asymptotique de $\hat{\boldsymbol{\beta}}_{MV}$

$$\hat{\boldsymbol{\beta}}_{MV} \sim N \left(\boldsymbol{\beta}^*, \left(\sum_{i=1}^n \exp(\mathbf{x}'_i \boldsymbol{\beta}^*) \mathbf{x}_i \mathbf{x}'_i \right)^{-1} \right).$$

3.3.2 Autres méthodes d'estimation

Lorsque la densité de la variable endogène est erronée, les estimateurs du maximum de vraisemblance ne sont plus en général convergents et ils doivent être ajustés, ce qui rend leur forme plus complexe. Pour les modèles linéaires généralisés qui requièrent que les variables dépendantes proviennent d'une distribution de la famille exponentielle, la convergence est conservée même dans le cas où la distribution est erronée. La seule condition requise pour la convergence est que l'espérance conditionnelle soit bien spécifiée.

Un exemple simple de ce phénomène réside dans la régression linéaire, où les $\hat{\beta}$ sont convergents même si les données n'obéissent pas à une loi normale ou s'il y a présence d'hétéroscédasticité. La seule condition est que $E[y_i | \mathbf{x}_i] = \mathbf{x}_i' \beta^*$. Par contre, même si les estimateurs sont convergents, si la variance de la variable endogène est mal spécifiée, l'inférence statistique sera invalide, dû aux statistiques t erronées et aux écarts-type. C'est pour cela qu'en régression linéaire, on utilise les moindres carrés pondérés à la place des moindres carrés ordinaires. On présente maintenant une autre technique des moindres carrés qui produit des inférences valides.

Moindres carrés pondérés

On utilise un exemple de régression linéaire pour montrer les conséquences d'une mauvaise spécification de la variance des observations. On suppose que $\text{Var}(y_i) = v_i$, où v_i est connu. Les variances des observations peuvent donc être différentes d'une observation à l'autre. Les estimateurs des moindres carrés pondérés correspondent à la solution du système d'équations suivant :

$$\sum_{i=1}^n \frac{1}{v_i} (y_i - \mathbf{x}_i' \beta) \mathbf{x}_i = \mathbf{0}.$$

Si toutes les variances sont identiques, (présence d'homoscédasticité) on revient à notre condition de premier ordre énoncée à la section 3.3.1. La matrice diagonale \mathbf{V} des v_i , est appelée la matrice de travail pour la variance, car on la considère comme connue mais elle est peut-être erronée. À l'aide de quelques manipulations algébriques simples, on trouve que

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

Si la vraie matrice de variance des y_i est la matrice diagonale Ω , on trouve que la variance des $\hat{\beta}$ est donnée par

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\Omega\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \quad (3.7)$$

Si la vraie variance des y_i (la variance trouvée avec la vraie fonction de densité) est la même que la matrice de travail pour la variance, (3.7) devient

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}.$$

On trouve donc des estimateurs convergents et des inférences valides avec comme condition que les deux premiers moments soient bien spécifiés.

Famille exponentielle

De façon générale, pour les distributions de la famille exponentielle, on a que

$$\sqrt{n}(\hat{\beta} - \beta^*) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}),$$

ce qui signifie que $\sqrt{n}(\hat{\beta} - \beta^*)$ converge en distribution vers une loi normale. On rappelle que X_n converge en distribution vers X si

$$\lim_{n \rightarrow \infty} P(X_n \leq x) = P(X \leq x),$$

pour tout $x \in \mathbb{R}$. En considérant que $E[y_i|\mathbf{x}_i] = \mu_i$ et que $\text{Var}[y_i|\mathbf{x}_i] = \omega_i$, on a que \mathbf{A} est définie comme

$$\mathbf{A} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{1}{v_i} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta'} \Bigg|_{\beta^*},$$

et \mathbf{B} comme

$$\mathbf{B} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\omega_i}{v_i^2} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta'} \Bigg|_{\beta^*}.$$

Pour bien illustrer ce résultat, on prend l'exemple de la régression de Poisson, où $\mu_i = \exp(\mathbf{x}_i'\beta)$. L'équidispersion du modèle indique que $v_i = \mu_i$, car la variance d'une loi de Poisson est égale à son espérance. On obtient

alors que

$$\begin{aligned} \mathbf{A} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{1}{v_i} \mu_i^2 \mathbf{x}_i \mathbf{x}_i' \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mu_i \mathbf{x}_i \mathbf{x}_i', \end{aligned}$$

qui est identique à (3.6). De plus, \mathbf{B} est donné par

$$\begin{aligned} \mathbf{B} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\omega_i}{v_i^2} \mu_i^2 \mathbf{x}_i \mathbf{x}_i' \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i'. \end{aligned}$$

On constate que si la distribution est exacte, alors $\omega_i = v_i$ et on revient à $\widehat{\beta} \sim N(\beta^*, \mathbf{A}^{-1})$. Si, par contre, $\omega_i \neq v_i$, donc si la distribution supposée n'est pas celle qui représente réellement les données, les simplifications ne sont plus possibles et on doit utiliser $\text{Var}(\widehat{\beta}) = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$.

3.4 Tests d'hypothèses

3.4.1 Fonction de vraisemblance spécifiée

On étudie trois méthodes statistiques classiques pour tester des hypothèses lorsque la fonction de vraisemblance est spécifiée ; le test du rapport des vraisemblances, le test de Wald et le test des multiplicateurs de Lagrange, aussi appelé test score. Ces tests sont utiles pour vérifier si un certain sous-modèle est préférable au modèle complet. On doit spécifier un vecteur de restrictions qui identifie le modèle réduit. Par exemple, si on veut avoir $\theta_1 = 5$, on considère la l^{e} restriction $r_l(\theta) = \theta_1 - 5 = 0$ dans l'hypothèse nulle, contre $\theta_1 - 5 \neq 0$. En considérant toutes les restrictions voulues, on obtient

$$H_0 : \mathbf{r}(\theta^*) = \mathbf{0},$$

contre

$$H_a : \mathbf{r}(\theta^*) \neq \mathbf{0},$$

où \mathbf{r} est la matrice des h restrictions. En imposant de telles restrictions, on obtient un modèle à tester qui constitue en fait un sous-modèle du modèle original. Il est donc approprié de faire des tests en utilisant les fonctions de log-vraisemblance.

Si $L(\boldsymbol{\theta})$ représente la fonction de vraisemblance, on a que $\hat{\boldsymbol{\theta}}_u$ est l'estimateur qui maximise la log-vraisemblance non restreinte $\mathcal{L}(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta})$. Pour sa part, $\tilde{\boldsymbol{\theta}}_r$ est l'estimateur qui maximise la fonction de log-vraisemblance en tenant compte des restrictions imposées sous H_0 . De façon équivalente, $\tilde{\boldsymbol{\theta}}_r$ maximise $\mathcal{L}(\boldsymbol{\theta}) - \lambda' \mathbf{r}(\boldsymbol{\theta})$, où λ est un vecteur de h multiplicateurs lagrangiens. Asymptotiquement, si le modèle est bien spécifié, les trois tests présentés dans cette section suivent une loi $\chi^2(h)$. On rejette H_0 au niveau de signification α si le test statistique est supérieur à $\chi^2(h; \alpha)$.

Le test du rapport des vraisemblances s'énonce comme suit

$$T_{LR} = -2[\mathcal{L}(\tilde{\boldsymbol{\theta}}_r) - \mathcal{L}(\hat{\boldsymbol{\theta}}_u)].$$

De cette façon, si H_0 est vraie, $T_{LR} \simeq 0$ puisque les restrictions n'entraînent pas de perte importante d'information. Puisqu'un modèle réduit compte moins de paramètres, donc moins d'incertitude causée par l'estimation de ces paramètres, on veut tester quel est le meilleur compromis entre la perte d'information et l'incertitude. Finalement, on nomme ce test «rapport» des vraisemblances, car le logarithme d'un quotient revient à soustraire les logarithmes. On énonce donc ce test de façon équivalente comme $T_{LR} = -2 \ln \left(\frac{L(\tilde{\boldsymbol{\theta}}_r)}{L(\hat{\boldsymbol{\theta}}_u)} \right)$.

Le test de Wald est défini comme

$$T_W = \mathbf{r}(\hat{\boldsymbol{\theta}}_u)' \left\{ \frac{\partial \mathbf{r}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}_u} \left[\frac{1}{n} \hat{\mathbf{A}}(\hat{\boldsymbol{\theta}}_u)^{-1} \right] \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Big|_{\hat{\boldsymbol{\theta}}_u} \right\}^{-1} \mathbf{r}(\hat{\boldsymbol{\theta}}_u),$$

où $\hat{\mathbf{A}}(\hat{\boldsymbol{\theta}}_u)$ est un estimateur convergent de la matrice de variance définie en (3.6), évalué à $\hat{\boldsymbol{\theta}}_u$. Ce test évalue à quel point $\mathbf{r}(\hat{\boldsymbol{\theta}}_u)$ est près de la valeur hypothétique de $\mathbf{0}$ sous l'hypothèse nulle. En prenant l'expansion de Taylor de $\mathbf{r}(\hat{\boldsymbol{\theta}}_u)$ autour de $\boldsymbol{\theta}^*$, il peut être démontré que $\mathbf{r}(\hat{\boldsymbol{\theta}}_u)$ tend asymptotiquement vers une normale de moyenne nulle et de variance $\frac{\partial \mathbf{r}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}_u} \left[\frac{1}{n} \hat{\mathbf{A}}(\hat{\boldsymbol{\theta}}_u)^{-1} \right] \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Big|_{\hat{\boldsymbol{\theta}}_u}$.

Le test des multiplicateurs de Lagrange (ou test score) est défini comme

$$T_{LM} = \mathbf{s}'(\boldsymbol{\theta}) \Big|_{\tilde{\boldsymbol{\theta}}_r} \left[\frac{1}{n} \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}_r)^{-1} \right] \mathbf{s}(\boldsymbol{\theta}) \Big|_{\tilde{\boldsymbol{\theta}}_r},$$

où $\tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}_r)$ est l'estimateur convergent de la matrice de variance (3.6) évalué à $\tilde{\boldsymbol{\theta}}_r$ et où $\mathbf{s}(\boldsymbol{\theta})$ est le vecteur score défini par

$$\mathbf{s}(\boldsymbol{\theta}) = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \frac{\partial \ln f_i}{\partial \boldsymbol{\theta}}.$$

Pour l'estimateur du maximum de vraisemblance non restreint $\hat{\boldsymbol{\theta}}_u$, le vecteur score donne $\mathbf{0}$. Si les restrictions n'entraînent pas une perte importante d'information, $\mathbf{s}(\tilde{\boldsymbol{\theta}}_r)$ devrait se rapprocher de $\mathbf{0}$.

3.4.2 Statistique t de Student

Les statistiques t sont utilisées pour déterminer si les paramètres qui sont inclus dans le modèle sont vraiment significatifs. Un avantage considérable de ce test par rapport à ceux étudiés à la section précédente est qu'il ne nécessite pas l'exactitude de la distribution utilisée aux fins de modélisation. En effet, pour que le test soit valide, il faut seulement que les deux premiers moments soient bien spécifiés. Par exemple, pour tester si un certain paramètre estimé $\hat{\beta}$ est égal à un β^* fixé, on effectue le test suivant

$$t_{N-p-1} = \frac{\hat{\beta} - \beta^*}{s_{\hat{\beta}}},$$

où N est le nombre d'observations, p est le nombre de paramètres et $s_{\hat{\beta}}$ est l'écart-type du paramètre $\hat{\beta}$. Si t_{N-p-1} est plus grand que la valeur critique t_c provenant d'une table de Student avec le bon nombre de degrés de liberté, alors on rejette l'hypothèse nulle que le paramètre β est égal à β^* . En utilisant β^* égal à zéro, on peut tester si les paramètres sont significatifs dans le modèle.

3.5 Modèles pour données de fréquence

Les seules distributions discrètes qui font partie de la famille exponentielle sont celles de Poisson, binomiale et binomiale négative. Lorsqu'il s'agit de modéliser un nombre de sinistres, la loi de Poisson est utilisée dans bien des cas et ce, même lorsque l'espérance n'est pas égale à la variance. Cependant, il est alors impossible d'utiliser les inférences statistiques. Avec la méthode

du maximum de vraisemblance, les paramètres sont toujours convergents, mais les écart-types ne sont pas valides si la variance conditionnelle n'est pas bien spécifiée. Dans un cas où on utilise la loi de Poisson sans avoir l'équidispersion, la variance conditionnelle ne peut pas être bien spécifiée, puisqu'elle n'est pas égale à l'espérance. Parce que la loi de Poisson est un cas particulier de la binomiale négative, qui ne requiert pas l'équidispersion, on peut utiliser nos inférences en corrigeant nos écart-types. On peut alors bien spécifier la variance avec ce modèle et identifier, à l'aide des tests étudiés en 3.4.1, le modèle le plus approprié.

Lors de l'application du modèle de Poisson, la surdispersion qu'on observe provient souvent du fait que les variables exogènes du modèle ne reflètent pas entièrement la variabilité de la variable endogène. Une solution possible est d'introduire un terme d'erreur multiplicatif dans l'espérance conditionnelle, ce qui conduit aux modèles mixtes.

En plus de la surdispersion, il y a diverses situations où la loi de Poisson ne peut tout simplement pas représenter de pareilles données dans sa forme décrite précédemment. Quelques exemples sont donnés ci-dessous :

1. La troncation et la censure font en sorte que les plus petites observations peuvent être exclues, ou alors qu'il y ait une classe comprenant un nombre d'observations plus grand qu'un certain seuil sans toutefois connaître leur véritable valeur.
2. Le problème d'un surplus de zéro survient lorsque pour une certaine valeur comprise sur le support de la variable dépendante, souvent le zéro, on observe beaucoup plus de cette valeur que dans une distribution de Poisson habituelle. Ce problème est fréquent lors de convolutions de Poisson.
3. Lorsque la distribution empirique est bimodale ou multimodale, une extension de la loi de Poisson peut être désirable. En effet, une distribution bimodale peut être causée par deux populations différentes mises ensemble et pourra être représentée par un mélange.

Comme le modèle de régression de Poisson a déjà été présenté en exemple en 3.3.1, cette section se limite aux modèles binomiale négative et les tests de surdispersion. Ensuite, on considère les cas où les variables exogènes n'expliquent pas l'ensemble de la variable endogène en créant ainsi de l'hétérogénéité en section 3.5.3. On poursuit avec la section 3.5.4 qui présente le théorème du double croisement à l'origine présenté par Shaked (1980) afin de

comparer le comportement d'une distribution conditionnelle avec une marginale.

3.5.1 Modèles binomiale négative NB1 et NB2

Comme les données rejettent pratiquement toujours l'hypothèse d'équidispersion, il est intéressant d'utiliser un modèle qui est plus souple face à cette hypothèse. On conserve donc la moyenne μ_i égale à $\exp(\mathbf{x}_i'\boldsymbol{\beta})$, mais on définit la variance comme une fonction quelconque de μ_i et du paramètre de dispersion α ,

$$\text{Var}(y_i|\mathbf{x}_i) = \omega_i = \omega(\mu_i, \alpha).$$

On spécifie souvent la variance comme

$$\omega_i = \mu_i + \alpha\mu_i^p,$$

où p est une constante fixée. On se limite en général à $p \in \{1, 2\}$.

Le modèle binomial négatif NB1 pose $p = 1$. On obtient donc une variance qui est une fonction linéaire de l'espérance :

$$\begin{aligned}\omega_i &= (1 + \alpha)\mu_i \\ &= \phi\mu_i,\end{aligned}$$

où $\phi = 1 + \alpha$. La seconde notation est utilisée dans le contexte des GLM et ϕ est le paramètre de dispersion.

Le modèle NB2 pose quant à lui $p = 2$. On se retrouve donc avec la fonction de variance

$$\omega_i = \mu_i + \alpha\mu_i^2.$$

Le paramètre de dispersion α reste à être estimé.

Comme mentionné au début de la section 3.3.2, les estimateurs des β obtenus à l'aide de la méthode du maximum de vraisemblance sont convergents si l'espérance est bien spécifiée et ce, même dans les cas où on n'a pas l'équidispersion. On exprime la variance des $\hat{\beta}$ avec notre modèle NB1 en fonction de la variance des $\hat{\beta}$ trouvée avec le modèle de Poisson. En effet, comme le modèle de Poisson exige l'équidispersion pour avoir des inférences valides et non le modèle de la binomiale négative NB1, on trouve le paramètre de dispersion ϕ qui les lie. En se basant sur la section 3.3.2 et en considérant que les variances de travail v_i sont égales aux μ_i , on a que l'estimateur de la

variance des β selon la méthode du maximum de vraisemblance (MLE) est donné par

$$\text{Var}_{\text{PMLE}}(\hat{\beta}) = \left(\sum_{i=1}^n \mu_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left(\sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i' \right) \left(\sum_{i=1}^n \mu_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1}. \quad (3.8)$$

Comme $\omega_i = \phi \mu_i$ dans le cas de la NB1 et seulement μ_i dans le cas de la Poisson, on trouve facilement que

$$\text{Var}_{\text{NB1}}(\hat{\beta}) = \phi \text{Var}_{\text{Pois}}(\hat{\beta}).$$

Il est simple d'estimer les variances des coefficients pour le modèle de Poisson et de les multiplier par $\hat{\phi}$ s'il y a surdispersion. Puisque $E[(y_i - \mu_i)^2] = \phi \mu_i$, l'estimateur standard pour ϕ est donné par

$$\hat{\phi}_{\text{NB1}} = \frac{1}{n - k} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i},$$

où k est le nombre de variable exogènes utilisées dans le modèle. Pour le modèle NB2, il est impossible de simplifier l'expression de la variance, ni de l'exprimer en fonction de la variance du modèle de Poisson. On doit donc travailler avec l'expression donnée en (3.8) :

$$\text{Var}_{\text{NB2}}(\hat{\beta}) = \left(\sum_{i=1}^n \mu_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left(\sum_{i=1}^n (\mu_i + \alpha \mu_i^2) \mathbf{x}_i \mathbf{x}_i' \right) \left(\sum_{i=1}^n \mu_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1}.$$

Un des estimateurs possibles de $\hat{\alpha}$ est trouvé de la même façon que le paramètre de dispersion du modèle NB1 et est égal à

$$\hat{\alpha}_{\text{NB2}} = \frac{1}{n - k} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i}{\hat{\mu}_i^2}.$$

3.5.2 Tests de surdispersion

Les conséquences résultant du non-respect de l'hypothèse d'équidispersion sont semblables à celles découlant de l'hétéroscédasticité en régression. Cependant, l'ampleur des effets sur les écarts-type et les statistiques t peut être considérable. Par exemple, si on a dans les faits que $\omega_i = 9\mu_i$, alors on

sait d'après la section précédente que la vraie variance des $\hat{\beta}$ est neuf fois supérieure à celle calculée, et les statistiques t doivent être réduites par un facteur trois.

Des données sont considérées surdispersées si leur variance conditionnelle est supérieure à leur espérance conditionnelle. Dans le même ordre d'idées, il se peut que des données soient sousdispersées si leur variance conditionnelle est inférieure à leur espérance conditionnelle. On a vu en 3.5.1 que la distribution binomiale négative a une variance supérieure à sa moyenne. Pour ce qui est de la sousdispersion, elle peut être traitée en utilisant un modèle binomial.

Comme on a que la variance du modèle NB1 est donnée par $\phi \text{Var}_{\text{Pois}}$, on peut tester la surdispersion de ce modèle en utilisant l'hypothèse nulle $H_0 : \phi = 1$ et en faisant un test du rapport des vraisemblances. Comme expliqué en section 3.4.1, on doit d'abord trouver la valeur maximisée de la fonction de log-vraisemblance en fixant le paramètre ϕ à 1 (l_0), qu'on compare avec la valeur maximisée de la fonction de log-vraisemblance sans restriction (l_1) de la façon suivante :

$$T = 2[l_1 - l_0].$$

Si $P(\chi_{(1)}^2 > T) < \delta$, où δ est notre seuil de tolérance, alors on rejette l'hypothèse nulle et on doit utiliser la binomiale négative plutôt que la Poisson.

Pour ce qui est du modèle NB2, on a que la distribution de Poisson est en fait un cas limite lorsque $\alpha = 0$. En effet, la fonction de masse de la binomiale négative peut être écrite comme

$$h(y|\mu) = \binom{y + \alpha^{-1} - 1}{\alpha^{-1} - 1} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^y.$$

Lorsque $\alpha \rightarrow 0$, la limite de cette fonction est donnée par

$$h(y|\mu) = \frac{\mu^y e^{-\mu}}{y!}.$$

On peut donc tester l'hypothèse nulle $H_0 : \alpha = 0$ et la contre-hypothèse que $\alpha > 0$ avec un test du rapport des vraisemblances ou un test de Wald comme vu à la section 3.4.1. Comme α ne peut pas être négatif et qu'on doit faire un test unilatéral, on ne peut pas considérer que la distribution de ces statistiques est Khi-carrée ; on doit faire un mélange de Khi-carré. À un seuil

δ , on peut rejeter l'hypothèse nulle si

$$\begin{aligned} 0,5P(\chi_{(0)}^2 > x) + 0,5P(\chi_{(1)}^2 > x) < \delta \\ \Leftrightarrow \\ 0,5P(\chi_{(1)}^2 > x) < \delta, \end{aligned}$$

puisque la loi χ^2 avec 0 degré de liberté est dégénérée à 0.

3.5.3 Introduction aux modèles mixtes pour données de fréquence

On introduit dans cette section les modèles pour tenir compte de l'effet de l'hétérogénéité. Les modèles mixtes seront couverts plus longuement à la section 3.7. Dans un modèle de Poisson sans hétérogénéité, la distribution des y_i est conditionnelle sur les réalisations \mathbf{x}_i . Dans les modèles mixtes, on introduit une variable aléatoire ν_i qui représente l'hétérogénéité qui n'est pas observée dans les données de l'individu i . On suppose ainsi qu'il reste, dans chacune des observations, une composante aléatoire qu'on ne peut pas modéliser à l'aide des variables du modèle. Pour pouvoir utiliser un tel modèle, on doit être en mesure d'identifier la relation existante entre les y_i , les \mathbf{x}_i et les ν_i . Le terme aléatoire est souvent introduit de façon multiplicative à l'élément μ_i de la régression de Poisson. On se retrouve ainsi avec

$$\begin{aligned} E(y_i|\mathbf{x}_i, \nu_i) &= \exp(\mathbf{x}_i\boldsymbol{\beta})\nu_i \\ &= \exp(\mathbf{x}_i\boldsymbol{\beta} + \ln(\nu_i)). \end{aligned} \tag{3.9}$$

Considérer l'hétérogénéité de façon multiplicative permet de ne pas violer l'hypothèse de non-négativité des y_i .

Surdispersion due à l'hétérogénéité non observée

Lorsqu'on arrive à la conclusion qu'il est nécessaire d'introduire une composante aléatoire associée à chaque observation, il va sans dire que la variance des y_i ne peut qu'augmenter. Par conséquent, on ne peut plus utiliser un modèle où l'espérance est égale à la variance. En remplaçant $\exp(\mathbf{x}_i'\boldsymbol{\beta})$ par μ_i on peut reformuler (3.9) comme

$$\mu_i^* = E(y_i|\mathbf{x}_i, \nu_i) = \mu_i\nu_i,$$

et voir le terme d'hétérogénéité non observée $\nu_i = \exp(\varepsilon_i)$ comme un terme manquant parmi nos variables exogènes. Il est cependant important de bien faire la distinction entre la variabilité de notre variable endogène et l'effet aléatoire de ses nouvelles composantes ν_i (ou ε_i). Il est courant de supposer les ν_i indépendants et identiquement distribués (iid). Par exemple, il est plausible de supposer que les ε_i suivent une loi normale de moyenne 0 et de variance σ^2 , résultant en de nouvelles équations pour les moments. En posant $E(y_i|\mathbf{x}_i, \nu_i) = \text{Var}(y_i|\mathbf{x}_i, \nu_i) = \mu_i$ comme dans le cas de la Poisson et en considérant $E(\nu_i) = 1$ et $\text{Var}(\nu_i) = \sigma_\nu^2$ on obtient

$$\begin{aligned} E(y_i|\mathbf{x}_i) &= \mu_i \\ \text{Var}(y_i|\mathbf{x}_i) &= E[\text{Var}(y_i|\mathbf{x}_i, \nu_i)] + \text{Var}[E(y_i|\mathbf{x}_i, \nu_i)] \\ &= \mu_i + \mu_i^2 \sigma_\nu^2 > E(y_i|\mathbf{x}_i). \end{aligned}$$

Un fait intéressant à noter est que la forme de la variance des y_i est la même que dans la cas d'une binomiale négative NB2 présentée à la section 3.5.1.

On peut aussi trouver une forme entièrement paramétrique pour $y|\mu_i$ en spécifiant la distribution de ν_i . En plus de spécifier la distribution de $y_i|\mu_i, \nu_i$, si on spécifie celle de ν_i , on trouve la densité marginale de $y|\mu_i$

$$h(y|\mu) = \int f(y|\mu, \nu)g(\nu)d\nu. \quad (3.10)$$

Lorsque $h(\cdot)$ est de la même famille que $f(\cdot)$, on dit que $f(\cdot)$ et $g(\cdot)$ sont des conjuguées naturelles. Le résultat donnant une distribution facile à appliquer, il est approprié de rechercher le plus possible des distributions qui sont des conjuguées naturelles. Par exemple, si $f(\cdot)$ est de distribution Poisson avec paramètre $\mu_i^* = \mu_i \nu_i$ et que $g(\cdot)$ est de distribution gamma, on a

$$\begin{aligned} f(y; \mu, \nu) &= \frac{(\mu\nu)^y e^{-\mu\nu}}{y!} \\ g(\nu; \delta, \lambda) &= \frac{\lambda^\delta}{\Gamma(\delta)} \nu^{\delta-1} e^{-\lambda\nu}. \end{aligned}$$

En appliquant (3.10), on obtient

$$\begin{aligned} h(y|\mu) &= \frac{\mu^y \lambda^\delta}{y! \Gamma(\delta)} \int \nu^{y+\delta-1} e^{-\nu(\lambda+\mu)} d\nu \\ h(y|\mu) &= \binom{y+\delta-1}{\delta-1} \left(\frac{\lambda}{\lambda+\mu} \right)^\delta \left(\frac{\mu}{\lambda+\mu} \right)^y, \end{aligned}$$

ce qui donne une distribution binomiale négative. En posant $E(\nu) = 1$ comme précédemment, on obtient $\delta = \lambda$, que l'on pose égal à α^{-1} . On trouve ainsi

$$h(y|\mu) = \binom{y + \alpha^{-1} - 1}{\alpha^{-1} - 1} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^y,$$

qui correspond à la binomiale négative NB2 trouvée en début de section.

Proportion de zéro due à l'hétérogénéité non observée

Une autre conséquence de l'hétérogénéité non observée est la proportion d'observations égales à zéro qui augmente. Indépendamment de la forme de la distribution des ν_i , et sachant que cette distribution n'est pas dégénérée, il est vrai que

$$h(y_i = 0|\mu_i) \geq f(y_i = 0|\mu_i, \nu_i).$$

Mullahy(1997) a démontré que la propriété de $f(y_i|\mu_i, \nu_i)$ qui mène à ces résultats est sa stricte convexité en μ . Aussi, dans la majorité des cas, il en résulte que la fréquence d'observations égales à 1 est moins élevée dans le mélange que dans la distribution originale. Finalement, on observe que le mélange possède une queue à droite plus épaisse que $f(y_i|\mu_i, \nu_i)$. Ces résultats représentent un cas spécial du résultat général pour les mélanges de la famille exponentielle présenté dans le théorème du double croisement à la section suivante.

3.5.4 Théorème du double croisement

On démontre dans cette section le théorème du double croisement énoncé dans Shaked (1980). Tout d'abord, à des fins de simplicité, on travaille seulement avec la famille exponentielle à un paramètre. On considère la distribution conditionnelle

$$f(y|\theta) = \exp \{a(\theta) + c(\theta)y\}, \quad (3.11)$$

et sa distribution marginale

$$h(y) = \int_{D_\theta} \exp \{a(\theta) + c(\theta)y\} \pi(\theta) d\theta. \quad (3.12)$$

Posons y , une variable aléatoire continue ou discrète avec (3.11) comme fonction de densité, $E[\theta] = 1$ et $\text{Var}[\theta] = \sigma^2 > 0$. Alors, la distribution marginale

(3.12) aura des queues plus épaisses que $f(y|\theta)$ et $h(y) - f(y|\theta)$ aura une séquence de signes de $\{+, -, +\}$ plus y augmente. En effet, si μ est l'espérance du paramètre θ définie comme

$$\mu = \int_{D_\theta} \theta \pi(\theta) d\theta,$$

on peut utiliser

$$f(y|\mu) = \exp \{a(\mu) + c(\mu)y\}$$

comme approximation à $h(y)$. On étudie donc le comportement des signes de la fonction $h(y) - f(y|\mu)$, en supposant que la distribution marginale et le mélange ont la même espérance

$$\int y f(y|\mu) dy = \int y h(y) dy.$$

On peut se définir la fonction $r(y)$ qui aura le même nombre de changements de signes que $h - f$:

$$\begin{aligned} r(y) &= \frac{h(y)}{f(y|\mu)} - 1 \\ &= \int_{D_\theta} \exp\{(a(\theta) - a(\mu)) + (c(\theta) - c(\mu))y\} \pi(\theta) d\theta - 1. \end{aligned}$$

Comme la fonction $r(y)$ est une fonction convexe, il peut y avoir au maximum deux changements de signes. Comme toutes les fonctions de densité somment à 1 et que les deux fonctions sont différentes, il est impossible qu'il n'y ait aucun changement de signes. Pour prouver qu'il n'y en pas qu'un seul, on suppose au contraire qu'il y en a exactement un et on démontre que cette situation est impossible. Si on regarde le cas où la séquence des signes est donnée par $\{+, -\}$, on a que la fonction $h(y)$ doit prendre des valeurs plus élevées que $f(y|\mu)$, comme sur le graphique 3.1, et que les fonctions de répartition ne se croisent pas. Si $H(y) = \int h(y) dy$ et $F(y|\mu) = \int f(y|\mu) dy$, on a que

$$F(y|\mu) < H(y).$$

Ce résultat, combiné avec le fait que $f(y|\mu) \neq h(y)$ implique que

$$\int y h(y) dy < \int y f(y|\mu) dy,$$

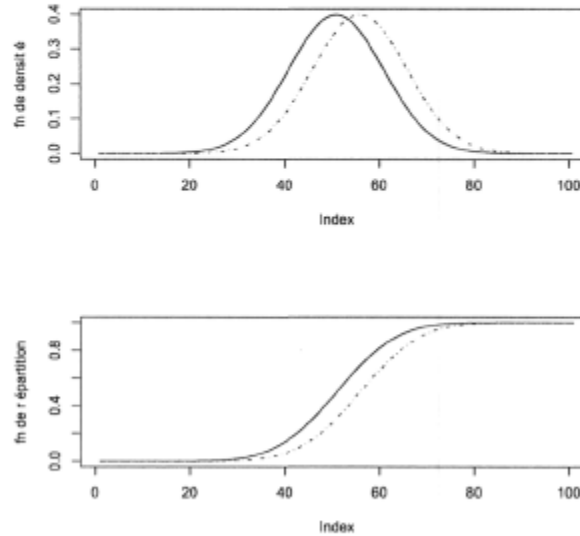


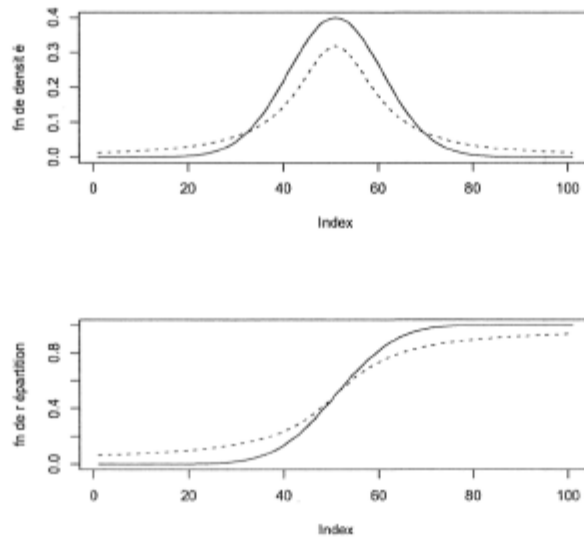
FIG. 3.1 – Exemple avec un seul croisement, $f(y|\mu)$ est en traitillé

ce qui contredit l'hypothèse de départ d'égalité des espérances. On peut arriver au même résultat pour la séquence de signes $\{-, +\}$. Il ne reste plus que l'hypothèse des deux croisements, qu'on confirme avec le graphique 3.2. Comme $r(y)$ est convexe, la séquence des signes est $\{+, -, +\}$.

Il est intéressant de considérer la binomiale négative comme un mélange de Poisson/Gamma. Ainsi, on conclut du théorème des deux croisements que la proportion de zéros est plus élevée dans la binomiale négative, en plus d'avoir une queue plus épaisse à droite.

3.5.5 Terme d'offset

Un terme d'offset est une variable qu'on ajoute au modèle, mais dont on pose le coefficient égal à 1 au lieu de l'estimer. Il peut être très utile dans des contextes de données de dénombrement où la variable réponse Y n'a pas la même exposition pour toutes les observations. Par exemple, en assurance automobile, un assuré qui a annulé sa police d'assurance après six mois a une probabilité d'avoir un sinistre deux fois inférieure à celle d'un assuré

FIG. 3.2 – Double croisements, $f(y|\mu)$ est en traitillé

avec les mêmes caractéristiques qui a eu une exposition de douze mois. En incorporant le logarithme du nombre de mois en offset, il est possible de tenir compte de cette différence d'exposition.

Par exemple, pour un modèle qui n'a que les variables âge (x_1) et sexe (x_2) en plus du terme d'offset ($\ln(\text{expo})$), on a que les nombres de sinistres prédits sont donnés par

$$\begin{aligned} y_i &= \exp\{\ln(\text{expo}) + \beta_1 x_1 + \beta_2 x_2\} \\ &= \text{expo} * \exp(\beta_1 x_1 + \beta_2 x_2), \end{aligned}$$

où «expo» est l'exposition de la cellule.

3.6 Exemple des visites chez le médecin

On considère dans cette section un exemple sur le nombre de visites chez le médecin durant deux semaines pour un échantillon de 5190 adultes étudié

par la *Australian Health Survey* en 1977-1978. Cet exemple est présenté dans les chapitre 3 et 4 de Cameron & Trivedi (1998). Le tableau 3.1 présente les valeurs observées de notre variable dépendante DVISITS, qui est le nombre de visites chez le médecin durant les deux semaines de l'étude. On trouve d'abord

Nombre de visites :	0	1	2	3	4
Fréquence :	4141	782	174	30	24
Fréquence relative :	0,798	0,151	0,033	0,006	0,005
Nombre de visites :	5	6	7	8	9
Fréquence :	9	12	12	5	1
Fréquence relative :	0,002	0,002	0,002	0,001	0,000

TAB. 3.1 – Distribution de la fréquence des visites

la moyenne des observations et leur variance, qui sont respectivement égales à 0,302 et 0,637. On constate donc que les données sont surdispersées, mais il est possible que cette surdispersion disparaisse avec l'ajout de l'effet des variables exogènes. Les variables utilisées dans le modèle sont définies dans le tableau 3.2. On trouve maintenant les $\hat{\beta}$ associés à chacune des variables exogènes qui maximisent la fonction de vraisemblance. Une fois ces paramètres estimés, on en déduit le paramètre de dispersion $\hat{\phi} = 1,328$ qui détermine directement le $\hat{\alpha} = 0,328$ du modèle NB1. Pour estimer le paramètre de dispersion α de la NB2, on utilise

$$\hat{\alpha}_{NB2} = \frac{1}{n-k} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i}{\hat{\mu}_i^2}.$$

En considérant nos $k = 13$ variables, on obtient $\hat{\alpha}_{NB2} = 0,2864$.

Une fois que les coefficients sont estimés, on détermine quelle distribution est la plus appropriée en faisant un test de surdispersion. On sait que les données étaient *a priori* surdispersées, mais l'ajout de l'effet des variables exogènes a peut-être changé la situation; il est donc préférable de faire un test statistique pour s'en assurer. Comme le cas de la Poisson est un cas particulier des binomiales négatives utilisées ici, on peut faire un test du rapport des vraisemblances. Pour le modèle NB1, on obtient une statistique $T_{LR} = 2(2814,7 - 2119,9) = 1389,6$, ce qui est largement supérieur à $\chi_{0,99}^2(1) = 6,64$ pour rejeter l'hypothèse nulle d'équidispersion. Pour ce qui

Variable	Définition
SEX	Variable indicatrice (V.i.) égale à 1 si le sujet est féminin
AGE	Âge en années divisé par 100
AGESQ	AGE au carré
INCOME	Revenu annuel en dizaine de milliers de dollars
LEVYPLUS	V.i. égale à 1 si le sujet a une assurance privée
FREEPOOR	V.i. égale à 1 si le sujet reçoit de l'assurance santé gouvernementale gratuite due à un faible revenu
FREEREPA	V.i. égale à 1 si le sujet reçoit de l'assurance santé gouvernementale gratuite due à son âge avancé, son incapacité ou son statut de vétéran
ILLNESS	Nombre de maladies dans les deux dernières semaines
ACTDAYS	Nombre de jours à activités réduites dans les deux dernières semaines dû à une maladie ou une blessure
HSCORE	Score d'un questionnaire général sur la santé utilisant la méthode de Goldberg
CHCOND1	V.i. égale à 1 si le sujet a une condition chronique qui ne limite pas ses activités
CHCOND2	V.i. égale à 1 si le sujet a une condition chronique qui limite ses activités

TAB. 3.2 – Définitions des variables exogènes

est du modèle NB2, on a la statistique $T_{LR} = 2(2814,7 - 2657,9) = 313,6$, ce qui est aussi supérieur à $0,5\chi_{0,99}^2(1) = 3,32$. On peut confirmer ce résultat avec la statistique de Wald, donnée par $T_W = 1,077/1,03 = 10,46$, qui est supérieure à $z_{0,99} = 2,33$. On peut donc conclure que ces données sont surdispersées et qu'un modèle binomiale négative est largement plus approprié.

3.7 Modèles mixtes

Les modèles vus précédemment sont utiles dans de nombreuses situations, mais ils supposent que les variables endogènes ne sont pas corrélées. Cependant, dans certaines situations, les variables y_i sont corrélées entre elles. Par exemple, les données longitudinales sont des observations prises sur des unités pour plusieurs périodes de temps. Comme les mesures proviennent de la même unité, il y a de fortes probabilités que ces mesures soient corrélées.

Variables	Coefficients			Écartst-type		
	Poisson	NB1	NB2	Poisson	NB1	NB2
Intercept	-2,224	-2,224	-2,190	0,190	0,219	0,234
SEX	0,157	0,157	0,217	0,056	0,065	0,069
AGE	1,056	1,056	-0,216	1,001	1,153	1,281
AGESQ	-0,849	-0,849	0,609	1,078	1,242	1,406
INCOME	-0,205	-0,205	-0,142	0,088	0,102	0,108
LEVYPLUS	0,123	0,123	0,118	0,072	0,083	0,086
FREEPOOR	-0,440	-0,440	-0,497	0,180	0,207	0,207
FREEREPA	0,080	0,080	0,145	0,092	0,106	0,117
ILLNESS	0,187	0,187	0,214	0,018	0,021	0,024
ACTDAYS	0,127	0,127	0,144	0,005	0,006	0,008
HSCORE	0,030	0,030	0,038	0,010	0,012	0,014
CHCOND1	0,114	0,114	0,099	0,066	0,077	0,079
CHCOND2	0,141	0,141	0,190	0,083	0,096	0,104
α	-	0,328	1,077	-	-	0,103
$-\ln L$	2814,7	2119,9	2657,9			

TAB. 3.3 – Coefficients et écarts-type pour les modèles de Poisson, NB1 et NB2

Par contre, toutes les données ne sont pas corrélées, puisque, comme en régression linéaire, les observations provenant d'individus différents ne sont pas nécessairement corrélées. Le modèle de régression linéaire mixte général permet l'ajout d'un vecteur d'effets aléatoires non-observables qui introduisent une corrélation entre les données provenant d'une même unité. On peut donc exprimer notre modèle sous forme matricielle par

$$Y = X\beta + Z\gamma + \varepsilon. \quad (3.13)$$

On constate que le modèle est identique à celui utilisé en régression linéaire, avec le coefficient $Z\gamma$ en plus. Comme le vecteur des effets aléatoires γ est présent pour introduire de la corrélation entre certains termes, la matrice Z est une matrice de schéma utilisée pour bien spécifier les corrélations. Les hypothèses sur ε sont les mêmes, soient $E[\varepsilon] = 0$ et $\text{Var}[\varepsilon] = \mathbf{V}$, et on a pour le vecteur d'effets aléatoires que $E[\gamma] = 0$ et que $\text{Var}[\gamma] = \mathbf{D}$. Finalement, ε

et γ sont indépendants et on a que

$$\text{Var} \begin{pmatrix} \varepsilon \\ \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}.$$

Avec ces informations, on peut facilement trouver l'espérance et la variance du vecteur \mathbf{Y} données par

$$\begin{aligned} E[\mathbf{Y}] &= \mathbf{X}\beta \\ \text{Var}[\mathbf{Y}] &= \mathbf{ZDZ}' + \mathbf{V} = \Sigma. \end{aligned}$$

Contrairement aux cas de régression linéaire ou aux modèles linéaires généralisés, on ne peut pas estimer directement les β sans avoir d'abord estimé les composantes de variance. D'emblée, en utilisant la méthode des moindres carrés généralisés et en considérant les composantes de variance comme connues, on obtient que

$$\hat{\beta} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}.$$

On peut facilement comprendre la provenance de cette équation en faisant un parallèle avec la méthode des moindres carrés pondérés. En effet, on peut représenter notre modèle mixte par

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon^*,$$

où $E[\varepsilon^*] = 0$ et $\text{Var}[\varepsilon^*] = \Sigma$.

En utilisant les estimés des composantes de variance dont la forme sera déterminée dans la sous-section 3.7.2, on a

$$\hat{\beta} = (\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{Y}. \quad (3.14)$$

3.7.1 Prévision

On traite dans cette section du problème général des prévisions dans un contexte de données longitudinales, tel que décrit dans Frees et coll. (1999, 2001). On peut ainsi utiliser ces notions et les appliquer aux modèles de crédibilité et de tarification, puisqu'un objectif important de l'assurance consiste à prévoir de façon la plus précise possible le futur. Donc, avec un modèle linéaire mixte comme présenté en (3.13), on peut vouloir prédire une variable

quelconque W , de moyenne $E[W] = \lambda' \beta$ et de variance $\text{Var}[W] = \sigma_W^2$. On désigne le vecteur des covariances entre \mathbf{Y} et W comme $\text{Cov}(W, \mathbf{Y}) = \text{Cov}_{W, \mathbf{Y}}$. Il est montré dans Frees et coll. (1999) que le meilleur prédicteur de W qui minimise l'erreur quadratique moyenne est donné par

$$W^* = E[W] + \text{Cov}_{W, \mathbf{Y}} \Sigma^{-1} [\mathbf{Y} - E(\mathbf{Y})] = \lambda' \beta + \text{Cov}_{W, \mathbf{Y}} \Sigma^{-1} (\mathbf{Y} - \mathbf{X} \beta).$$

Pour en arriver au meilleur prédicteur linéaire non-biaisé (BLUP), on remplace β par ses meilleurs estimateurs linéaires non-biaisés,

$$W_{BLUP} = \lambda' \hat{\beta} + \text{Cov}_{W, \mathbf{Y}} \Sigma^{-1} (\mathbf{Y} - \mathbf{X} \hat{\beta}). \quad (3.15)$$

On peut utiliser l'expression trouvée en (3.15) pour prédire une observation d'un effet aléatoire pour un individu donné. En posant $W = \mathbf{c}' \gamma_i$, on a que $E[W] = 0$, et on en déduit que $\lambda = \mathbf{0}$. Aussi, on a que $\text{Cov}(W, \mathbf{Y}_j) = \mathbf{c}' \mathbf{DZ}'_i$ pour $j = i$ et $\text{Cov}(W, \mathbf{Y}_j) = \mathbf{0}$ pour $i \neq j$. On a donc

$$W_{BLUP} = \mathbf{c}' \mathbf{DZ}'_i \Sigma_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}),$$

et

$$\hat{\gamma}_{i, BLUP} = \mathbf{DZ}'_i \Sigma_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}). \quad (3.16)$$

Avec cette prévision et un estimé de β , il est possible de trouver une prévision pour \mathbf{Y} .

3.7.2 Composantes de variance

On utilise deux méthodes afin d'estimer les composantes de variance, soit la méthode du maximum de vraisemblance et la méthode du maximum de vraisemblance restreint (REML). C'est cette dernière méthode qui est utilisée par défaut dans SAS et dans R. Pour ce faire, on rappelle la fonction de densité de \mathbf{X} , où \mathbf{X} suit une loi normale multivariée ($\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \|\boldsymbol{\Sigma}\|}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

où $\|x\|$ est le déterminant de x . Pour la méthode du maximum de vraisemblance, on obtient la fonction de log-vraisemblance suivante

$$\begin{aligned} \mathcal{L}(y; \mathbf{D}, \mathbf{V}) &= -\frac{1}{2} \ln \|\boldsymbol{\Sigma}\| - \frac{1}{2} (\mathbf{Y} - \mathbf{X} \hat{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\beta}) - \frac{n}{2} \ln(2\pi) \\ &= -\frac{1}{2} \ln \|\boldsymbol{\Sigma}\| - \frac{1}{2} \mathbf{r}' \boldsymbol{\Sigma}^{-1} \mathbf{r} - \frac{n}{2} \ln(2\pi), \end{aligned}$$

où $\mathbf{r} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{Y}$. Pour maximiser la vraisemblance, on peut utiliser des méthodes numériques telles que Newton-Raphson ou Fisher scoring. Il existe quelques cas particuliers où on peut obtenir une expression qu'on peut résoudre sans utiliser de méthodes numériques.

Comme l'estimateur du maximum de vraisemblance est un estimateur biaisé négativement, il est intéressant de trouver l'estimateur des composantes de variance selon la méthode REML qui donne souvent un estimateur non-biaisé. Pour ce faire, on doit maximiser une vraisemblance où certains éléments ont été marginalisés et ignorés dans les calculs. On a ainsi

$$\mathcal{L}_{REML}(y; \mathbf{D}, \mathbf{V}) = -\frac{1}{2} \ln \|\boldsymbol{\Sigma}\| - \frac{1}{2} \ln \|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\| - \frac{1}{2} \mathbf{r}'\boldsymbol{\Sigma}^{-1}\mathbf{r} - \frac{n-p}{2} \ln(2\pi),$$

où p est le rang de la matrice \mathbf{X} . On peut aussi utiliser les méthodes numériques de Newton-Raphson et de Fisher pour trouver nos estimateurs. Pour trouver les points de départ des algorithmes, des méthodes de moments sont généralement utilisées.

3.8 Modèles linéaires généralisés mixtes

Comme leur nom l'indique, les modèles linéaires généralisés mixtes sont des modèles linéaires généralisés auxquels on a ajouté des effets aléatoires. Ces modèles sont souvent utilisés dans des contextes de données longitudinales, où des individus sont suivis dans le temps. En utilisant la même notation que pour les GLM, on a l'expression de l'espérance conditionnelle suivante

$$E[Y|\boldsymbol{\gamma}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}) = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu},$$

où $\boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{G})$. Il est important de noter que les $\boldsymbol{\beta}$ des GLMM n'ont pas la même signification que ceux des GLM. En effet, pour les modèles linéaires généralisés mixtes, on a que les paramètres représentent des effets conditionnels, alors que ce sont plutôt des effets marginaux pour les GLM. Ainsi, leur effet est sur un groupe en particulier plutôt que sur l'ensemble des données. On a que la variance des observations conditionnelle aux $\boldsymbol{\gamma}$ est donnée par $\text{Var}[Y|\boldsymbol{\gamma}] = \mathbf{A}^{1/2}\mathbf{R}\mathbf{A}^{1/2}$. \mathbf{A} est une matrice diagonale contenant les fonctions de variance du modèle et \mathbf{R} est une matrice de corrélations du modèle, généralement donnée par $\phi\mathbf{I}$. Par exemple, pour le modèle de Poisson, les éléments de la matrice \mathbf{A} sont égaux à μ et le paramètre de dispersion ϕ est fixé à 1. On peut tester la surdispersion en demandant que le paramètre ϕ soit variable.

3.8.1 Le pseudo-modèle

Une façon possible d'obtenir les estimateurs de ce modèle est de le «linéariser» à l'aide du développement en série de Taylor par rapport à $\tilde{\beta}$ et à $\tilde{\gamma}$:

$$g^{-1}(\boldsymbol{\eta}) \doteq g^{-1}(\tilde{\boldsymbol{\eta}}) + \tilde{\Delta}\mathbf{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \tilde{\Delta}\mathbf{Z}(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}),$$

où

$$\tilde{\Delta} = \left(\frac{\partial g^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right)_{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}}.$$

En réorganisant les termes, on peut isoler un modèle linéaire mixte

$$\tilde{\Delta}^{-1}(\boldsymbol{\mu} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\boldsymbol{\gamma}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}.$$

Comme la partie de gauche est l'espérance conditionnelle de

$$\tilde{\Delta}^{-1}(\mathbf{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\boldsymbol{\gamma}},$$

on utilise cette expression pour transformer notre variable endogène et travailler avec un modèle linéaire. On définit notre nouvelle variable endogène \mathbf{P} comme

$$\mathbf{P} \equiv \tilde{\Delta}^{-1}(\mathbf{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\boldsymbol{\gamma}},$$

La variance de notre variable endogène transformée \mathbf{P} est donnée par

$$\text{Var}[\mathbf{P}|\boldsymbol{\gamma}] = \tilde{\Delta}^{-1} \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2} \tilde{\Delta}^{-1}.$$

On peut donc exprimer la variable \mathbf{P} sous la forme d'un modèle linéaire mixte :

$$\mathbf{P} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

où on suppose que $\boldsymbol{\epsilon} \sim N(0, \text{Var}[\mathbf{P}|\boldsymbol{\gamma}])$.

Pour trouver les estimateurs des paramètres, on utilise la méthode du maximum de vraisemblance ou la méthode du maximum de vraisemblance restreint comme vu à la section précédente. La variance marginale de notre pseudo-modèle linéaire est donnée par

$$\mathbf{V}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \tilde{\Delta}^{-1} \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2} \tilde{\Delta}^{-1},$$

où $\boldsymbol{\theta}$ est le vecteur des paramètres inconnus dans \mathbf{G} et \mathbf{R} . On obtient les fonctions de log-vraisemblance à maximiser suivantes

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \mathbf{p}) &= -\frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2} \mathbf{r}' \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{r} - \frac{n}{2} \ln(2\pi) \\ \mathcal{L}_{REML}(\boldsymbol{\theta}, \mathbf{p}) &= -\frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2} \mathbf{r}' \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{r} - \frac{1}{2} \ln |\mathbf{X}' \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{X}| - \frac{n-k}{2} \ln(2\pi),\end{aligned}$$

où $\mathbf{r} = \mathbf{p} - \mathbf{X}(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{p}$ et k est le rang de la matrice \mathbf{X} . À la convergence, les estimateurs des effets fixes et des effets aléatoires sont donnés par

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{p}$$

et par

$$\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}} \mathbf{Z}' \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} \hat{\mathbf{r}}.$$

Si la distribution de la famille exponentielle utilisée comprend un paramètre d'échelle $\phi \neq 1$, alors on estime ce paramètre avec le logarithme de la pseudo-vraisemblance. À cette fin, on définit

$$\mathbf{V}(\boldsymbol{\theta}^*) = \mathbf{Z} \mathbf{G}^* \mathbf{Z}' + \tilde{\Delta}^{-1} \mathbf{A}^{1/2} \mathbf{R}^* \mathbf{A}^{1/2} \tilde{\Delta}^{-1},$$

où $\boldsymbol{\theta}^*$ est un vecteur des composantes \mathbf{G}^* et \mathbf{R}^* reparamétrisés pour faire ressortir le paramètre ϕ . En effet, si \mathbf{G} a une structure de composantes de variance et que $\mathbf{R} = \phi \mathbf{I}$, alors $\boldsymbol{\theta}^*$ contient le ratio des composantes de variance avec ϕ et $\mathbf{R}^* = \mathbf{I}$. La solution de $\hat{\phi}$ est alors donnée par

$$\hat{\phi} = \frac{\hat{\mathbf{r}}' \mathbf{V}(\hat{\boldsymbol{\theta}}^*) \hat{\mathbf{r}}}{m},$$

où $m = n$ pour la méthode du maximum de vraisemblance et $m = n - k$ pour la méthode du maximum de vraisemblance restreint. En substituant ce $\hat{\phi}$ dans notre fonction de log pseudo-vraisemblance, on peut trouver les estimés des paramètres qui restent en maximisant

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}^*, \mathbf{p}) &= -\frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\theta}^*)| - \frac{n}{2} \ln \{\mathbf{r}' \mathbf{V}(\boldsymbol{\theta}^*)^{-1} \mathbf{r}\} - \frac{n}{2} \{1 + \ln(2\pi/n)\} \\ \mathcal{L}_{REML}(\boldsymbol{\theta}^*, \mathbf{p}) &= -\frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\theta}^*)| - \frac{n-k}{2} \ln \{\mathbf{r}' \mathbf{V}(\boldsymbol{\theta}^*)^{-1} \mathbf{r}\} \\ &\quad - \frac{1}{2} \ln |\mathbf{X}' \mathbf{V}(\boldsymbol{\theta}^*)^{-1} \mathbf{X}| - \frac{n-k}{2} [1 + \ln\{\pi/(n-k)\}].\end{aligned}$$

3.8.2 Estimateurs de la variabilité

On peut estimer la matrice de variances-covariances de $\hat{\beta}$ et de $\hat{\gamma} - \gamma$ à l'aide des équations du pseudo-model mixte. Tout d'abord, on définit la matrice \mathbf{S} comme

$$\mathbf{S} \equiv \widehat{\text{Var}}[\mathbf{P}|\gamma] = \tilde{\Delta}^{-1} \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2} \tilde{\Delta}^{-1},$$

où les composantes de la partie de droite sont estimés à l'aide du maximum de vraisemblance ou du maximum de vraisemblance restreint. Les équations du pseudo-model mixte sont données par

$$\begin{bmatrix} \mathbf{X}'\mathbf{S}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{S}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{S}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z} + \mathbf{G}(\hat{\theta})^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{S}^{-1}\mathbf{p} \\ \mathbf{Z}'\mathbf{S}^{-1}\mathbf{p} \end{bmatrix}.$$

On peut donc estimer $\hat{\beta}$ et $\hat{\gamma}$ à partir des équations du pseudo-model. Par exemple, pour trouver la variance de $\hat{\beta}$, on isole $\hat{\gamma}$ dans la seconde équation, que l'on substitue dans la première. L'expression de $\hat{\gamma}$ est donnée par

$$\hat{\gamma} = \left[\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z} + \mathbf{G}(\hat{\theta})^{-1} \right]^{-1} \mathbf{Z}'\mathbf{S}^{-1}(\mathbf{p} - \mathbf{X}\hat{\beta}),$$

que l'on remplace dans

$$\mathbf{X}'\mathbf{S}^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{S}^{-1}\mathbf{Z}\hat{\gamma} = \mathbf{X}'\mathbf{S}^{-1}\mathbf{p},$$

pour finalement obtenir

$$\mathbf{X}'\mathbf{S}^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{S}^{-1}\mathbf{Z} \left[\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z} + \mathbf{G}(\hat{\theta})^{-1} \right]^{-1} \mathbf{Z}'\mathbf{S}^{-1}(\mathbf{p} - \mathbf{X}\hat{\beta}) = \mathbf{X}'\mathbf{S}^{-1}\mathbf{p}$$

$$\mathbf{X}'\mathbf{S}^{-1}(\mathbf{I} - \mathbf{W})\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{S}^{-1}(\mathbf{I} - \mathbf{W})\mathbf{p},$$

où $\mathbf{W} = \mathbf{Z} \left[\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z} + \mathbf{G}(\hat{\theta})^{-1} \right]^{-1}$. En simplifiant, on obtient que $\mathbf{X}\hat{\beta} = \mathbf{p}$, donc que $\widehat{\text{Var}}[\mathbf{X}\hat{\beta}] = \widehat{\text{Var}}[\mathbf{P}]$. La variance estimée de $\hat{\beta}$ est donc l'inverse généralisé¹ de $\mathbf{X}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{X}$:

$$\widehat{\text{Var}}[\hat{\beta}] = (\mathbf{X}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{X})^{-}.$$

¹L'inverse généralisé de la matrice A de dimension $m \times n$ est la matrice A^{-} telle que $AA^{-}A = A$. Lorsque la matrice A est de dimension $n \times n$ et est non-singulière, l'inverse généralisé unique est donné par A^{-1} .

La matrice de variance-covariance est donc égale à

$$\begin{bmatrix} \hat{\Omega} & -\hat{\Omega}\mathbf{X}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{Z}\mathbf{G}(\hat{\theta}) \\ -\mathbf{G}(\hat{\theta})\mathbf{Z}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{X}\hat{\Omega} & \mathbf{M} + \mathbf{G}(\hat{\theta})\mathbf{Z}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{X}\hat{\Omega}\mathbf{X}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{Z}\mathbf{G}(\hat{\theta}) \end{bmatrix},$$

où $\hat{\Omega} = (\mathbf{X}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{X})^{-1}$ et $\mathbf{M} = [\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z} + \mathbf{G}(\hat{\theta})^{-1}]^{-1}$.

Cette matrice tend par contre à sous-estimer la variabilité de $\hat{\beta}$ et de $\hat{\gamma}$, puisqu'elle ne considère pas l'incertitude de l'estimation de \mathbf{G} et \mathbf{Z} . Des facteurs d'inflation pour ajuster les valeurs ont été proposés dans la littérature, mais la correction est petite pour des jeux de données qui sont bien équilibrés. La procédure Glimmix de SAS n'en utilise aucun par défaut.

3.9 Exemple des crises d'épilepsie

On poursuit notre analyse en regardant les conséquences possibles pouvant survenir lorsque la corrélation existant entre les données prises sur le même individu n'est pas considérée. Pour ce faire, on utilise un exemple tiré du livre de Diggle, Liang & Zeger (1994), considérant des données sur les crises d'épilepsie de 58 individus lors de cinq périodes consécutives (une première période initiale de huit semaines suivie de quatre périodes de deux semaines). La variable traitement `trt` indique si le patient a reçu le traitement (1) ou un placebo (0). On se définit aussi une variable indicatrice `x1` égale à 0 si on est dans la période initiale de huit semaines et égale à 1 si on est dans une des périodes subséquentes. Finalement, on introduit un terme d'offset correspondant au logarithme du nombre de semaines écoulées (`logtime`), car on considère que le nombre de crises d'épilepsie est directement proportionnel au nombre de semaines considérées. Pour bien voir les conséquences de la négligence de la corrélation, on présente d'abord le meilleur GLM trouvé avec les données. Par la suite, on utilise l'approche GEE, qui est une méthode qui consiste à introduire de la corrélation entre les observations d'un même sujet, sans toutefois spécifier la structure de cette corrélation. On termine avec un GLMM, où les effets aléatoires seront bien spécifiés.

3.9.1 Modèle de base

Le premier modèle qui est ajusté est une régression de Poisson, comme il a été fait dans l'exemple de la section 3.6 :

$$Y_{ij} = \exp(\ln t_{ij} + \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{i2} + \beta_3 x_{ij1} x_{i2}),$$

pour $i = 1, \dots, 58$ et $j = 0, 1, 2, 3, 4$. Dans ce modèle, Y_{ij} est le nombre de crises pour l'individu i à la période j , t_{ij} représente le temps en semaines, x_{ij1} est la variable `x1` et x_{i2} est la variable `trt`. On arrive aux coefficients suivants, tous significatifs à 5% :

$$Y_{ij} = \exp(\ln t_{ij} + 1,3476 + 0,1108x_{ij1} - 0,1080x_{i2} - 0,3016x_{ij1}x_{i2}).$$

Selon ce modèle, x_{i2} est une variable significative et on conclut que les patients ayant reçu le traitement ont moins de crises que ceux qui ne l'ont pas reçu, puisque le coefficient est négatif. Par contre, la dispersion du modèle est particulièrement élevée et rend le modèle inacceptable. En effet, on trouve un $\hat{\phi} = 8,4371$. On change donc la distribution de Poisson pour une binomiale négative, en conservant le lien log. Les coefficients restent les mêmes, mais étant donné que le paramètre de dispersion n'est plus fixé à 1 comme il est nécessaire dans le modèle de Poisson, les écart-types augmentent considérablement et la variable `trt` est maintenant la moins significative, avec une valeur p de 0,6394. Déjà, en corrigeant les écart-types pour la surdispersion, on trouve que le traitement n'est pas significatif. On considère maintenant l'approche GEE.

3.9.2 L'approche GEE

L'approche avec les équations d'estimation généralisées (GEE) est une méthode qui permet d'avoir des inférences valides, sans toutefois avoir la tâche de déterminer précisément la forme de la corrélation entre les observations comme pour les modèles mixtes. En effet, avec les GEE, nous corrigeons les écart-types en ajoutant au modèle une corrélation sous-jacente aux données, mais invisible dans le modèle. Encore une fois, on conserve les mêmes estimés des β , mais la variable `trt` n'est plus significative avec une valeur p de 0,5770 et on conclut que le traitement n'est pas significativement différent du placebo.

3.9.3 Le modèle GLMM

Pour déterminer notre modèle linéaire généralisé mixte, on utilise la même procédure que pour les modèles mixtes. D'abord, on inclut dans le modèle tous les effets fixes, auxquels on ajoute ensuite les effets aléatoires. Par la suite, on teste les différentes formes possibles pour la matrice de covariances des effets aléatoires. La meilleure forme de la matrice est celle qui donnera les plus petites variances possibles pour les effets aléatoires. Pour la déterminer, on peut se baser sur le critère d'information d'Akaike (AIC), ou le critère de Bayes (BIC) qu'on doit minimiser. Une fois que la structure de variance qui s'applique le mieux aux données est déterminée, il est temps de réduire les effets aléatoires pour n'avoir que ceux qui sont significatifs. Pour ce faire, on fait un test du ratio de vraisemblance en testant l'hypothèse nulle H_0 que tous les effets sont significatifs, avec la contre-hypothèse qu'il y en a un qu'on peut retirer. Avec l'exemple des crises d'épilepsie, on trouve que la fonction de vraisemblance du modèle avec des effets aléatoires pour l'intercept et les variables `trt` et `x1` est exactement la même que le modèle sans `trt`. On peut donc retirer l'effet aléatoire sur `trt`. Cependant, en essayant de retirer l'effet aléatoire sur `x1`, on remarque une baisse significative de la fonction de vraisemblance. Il est donc préférable de conserver les effets aléatoires pour l'intercept et la variable `x1`. Notre modèle GLMM est donc donné par

$$Y_{ij} = \exp(\ln t_{ij} + 1, 1097 + \gamma_{i0} + 0, 0148x_{ij1} + \gamma_{i1} - 0, 0184x_{i2} - 0, 3091x_{ij1}x_{i2}).$$

Finalement, comme tous les autres modèles testés, la variable `trt` est la moins significative, avec une valeur p de 0,9230. En négligeant la présence de corrélation entre les observations d'un même individu, on serait en mesure de conclure que le traitement produit un effet significatif, alors qu'en réalité, aucune différence entre le traitement et le placebo est remarquée.

Chapitre 4

Tarification basée sur les GLM

4.1 Connexions entre la méthode du biais minimum et les GLM

On a étudié à la section 1.2.3 une méthode de biais minimum qui reposait sur la maximisation de la fonction de vraisemblance plutôt que sur la minimisation d'une fonction de biais, tout en conservant le même principe de base. En associant les méthodes de biais minimum à un GLM, on peut voir les hypothèses statistiques sous-jacentes et mieux juger de la qualité du modèle, chose très difficile avec le biais minimum. Les GLM offrent aussi une méthode alternative pour calculer les relativités, souvent plus efficace que les itérations du biais minimum. Dans ce chapitre, on démontre en se basant sur Midenhall (1999) comment un certain modèle traditionnel de biais minimum arrive au même résultat qu'une technique du maximum de vraisemblance. Par la suite, on identifie quels modèles traditionnels sont des cas particuliers des GLM.

4.1.1 Modèle de Poisson multiplicatif

On reproduit dans cette section le modèle de déviance moyenne combiné avec une prime multiplicative à l'aide de la fonction de vraisemblance de la loi de Poisson. Comme introduit à la section 1.2.3, on a $n_{ij}E[r_{ij}] = n_{ij}x_i y_j = \mu_{ij}$, où μ_{ij} est le paramètre de la loi de Poisson. La fonction de vraisemblance est

donc donnée par

$$L = \prod_j \prod_i \frac{e^{-\mu_{ij}} (\mu_{ij})^{n_{ij} r_{ij}}}{(n_{ij} r_{ij})!},$$

et la fonction de log-vraisemblance à maximiser est

$$l = \sum_j \sum_i \{(n_{ij} r_{ij}) [\ln x_i + \ln y_j + \ln n_{ij}] - x_i y_j n_{ij} - \ln(n_{ij} r_{ij})!\}.$$

Finalement, la dérivée partielle par rapport à x_k que l'on pose égale à zéro donne les relativités suivantes pour x_i :

$$\begin{aligned} \frac{\partial l}{\partial x_k} = 0 &\Rightarrow \sum_j \left\{ \frac{n_{kj} r_{kj}}{x_k} - n_{kj} y_j \right\} = 0 \\ &\Rightarrow x_i = \frac{\sum_j n_{ij} r_{ij}}{\sum_j n_{ij} y_j}. \end{aligned} \quad (4.1)$$

On constate que (4.1) est équivalent à (1.3) et donc le modèle basé sur la fonction de vraisemblance d'une loi de Poisson reproduit les résultats de la déviance moyenne combinée à une prime multiplicative de Bailey et Simon (1960).

4.1.2 Modèle de régression linéaire

Pour faire le lien entre les modèles linéaires généralisés, on commence par exprimer le modèle de biais minimum avec une prime additive sous une forme matricielle comme dans un modèle de régression linéaire. On cherche donc à retrouver la forme

$$\mathbf{r} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où \mathbf{r} est le vecteur des relativités observées. Étant donné que les paramètres de notre modèle sont les relativités x_i et y_j , on a que le vecteur $\boldsymbol{\beta}$ est donné par

$$\boldsymbol{\beta} = \begin{pmatrix} x_1 \\ \vdots \\ x_{n_1} \\ y_1 \\ \vdots \\ y_{n_2} \end{pmatrix},$$

4.1. CONNECTIONS ENTRE LA MÉTHODE DU BIAIS MINIMUM ET LES GLM61

où n_1 est le nombre de x_i et n_2 le nombre de y_j . La matrice \mathbf{X} est quand à elle définie comme

$$\mathbf{X} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{I} \\ \vdots & \vdots \\ \mathbf{A}_{n_1} & \mathbf{I} \end{pmatrix},$$

où

$$\mathbf{A}_i = \begin{pmatrix} 0 & \dots & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & \dots & 0 \end{pmatrix}$$

de dimension $n_2 \times n_1$ et avec \mathbf{I} de dimension $n_2 \times n_2$, pour nous donner le résultat suivant :

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X} \begin{pmatrix} x_1 \\ \vdots \\ x_{n_1} \\ y_1 \\ \vdots \\ y_{n_2} \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_1 + y_{n_2} \\ x_2 + y_1 \\ \vdots \\ x_2 + y_{n_2} \\ \vdots \\ x_{n_1} + y_1 \\ \vdots \\ x_{n_1} + y_{n_2} \end{pmatrix},$$

de dimension $n \times 1$, où $n = n_1 n_2$. Maintenant, on utilise les formules de régression linéaire pondérée pour trouver la solution des $\hat{\boldsymbol{\beta}}$

$$\mathbf{X}'\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{W}\mathbf{r},$$

où $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$. Les matrices de poids sont définies comme

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{W}_{n_1} \end{pmatrix},$$

avec

$$\mathbf{W}_1 = \begin{pmatrix} w_{i1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_{in_2} \end{pmatrix}.$$

On remarque que \mathbf{W} est de dimension $n \times n$, alors que \mathbf{W}_1 est de dimension $n_2 \times n_2$. Puisque les matrices \mathbf{X} et \mathbf{W} sont des matrices en bloc, il est facile de trouver la forme de $\mathbf{X}'\mathbf{W}\mathbf{X}$, donnée par

$$\mathbf{X}'\mathbf{W}\mathbf{X} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D} \end{pmatrix},$$

de dimension $(n_1 + n_2) \times (n_1 + n_2)$. Les composantes de la matrice et leur dimension sont

$$\mathbf{B} = \begin{pmatrix} \sum_j w_{1j} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum_j w_{n_1j} \end{pmatrix}_{n_1 \times n_1},$$

$$\mathbf{C} = \begin{pmatrix} w_{11} & \dots & w_{1n_2} \\ \vdots & & \vdots \\ w_{n_11} & \dots & w_{n_1n_2} \end{pmatrix}_{n_1 \times n_2},$$

et

$$\mathbf{D} = \begin{pmatrix} \sum_i w_{i1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum_i w_{in_2} \end{pmatrix}_{n_2 \times n_2}.$$

4.1. CONNECTIONS ENTRE LA MÉTHODE DU BIAIS MINIMUM ET LES GLM63

Il en résulte que

$$\begin{aligned}
 \mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}'\mathbf{W}\mathbf{X} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{y} \\ \mathbf{C}'\mathbf{x} + \mathbf{D}\mathbf{y} \end{pmatrix} \\
 &= \begin{pmatrix} x_1 \sum_j w_{1j} + \sum_j w_{1j}y_j \\ \vdots \\ x_{n_1} \sum_j w_{n_1j} + \sum_j w_{n_1j}y_j \\ \sum_i w_{i1}x_i + y_1 \sum_i w_{i1} \\ \vdots \\ \sum_i w_{in_2}x_i + y_{n_2} \sum_i w_{in_2} \end{pmatrix}. \tag{4.2}
 \end{aligned}$$

La dernière étape consiste à combiner chacune des lignes de (4.2) avec la partie de droite de notre équation de départ, donnée par

$$\mathbf{X}'\mathbf{W}\mathbf{r} = \begin{pmatrix} \sum_j w_{1j}r_{1j} \\ \vdots \\ \sum_j w_{n_1j}r_{n_1j} \\ \sum_i w_{i1}r_{i1} \\ \vdots \\ \sum_i w_{in_2}r_{in_2} \end{pmatrix}.$$

On retrouve les équations de (1.4), soit

$$x_i = \frac{\sum_j w_{ij}(r_{ij} - y_j)}{\sum_j w_{ij}},$$

où les poids w_{ij} sont les n_{ij} . Par conséquent, si l'hypothèse de normalité des résidus est raisonnable, on peut utiliser la régression linéaire au lieu de la méthode de déviance moyenne avec la prime additive et profiter des avantages statistiques de la régression. En effet, avec un modèle linéaire, on peut avoir des écarts-type et des intervalles de confiance, ce qui est impossible avec la méthode de biais minimum traditionnelle.

4.1.3 Autres modèles linéaires généralisés

On présente maintenant d'autres distributions membres de la famille exponentielle associées à une fonction de lien qui reproduisent des relativités

du biais minimum.

Poisson – Lien logarithmique

On suppose que la distribution des r_{ij} est Poisson tel que

$$E[r_{ij}] = \mu_{ij} = \exp(x_i + y_j) = a_i b_j,$$

où $a_i = e^{x_i}$ et $b_j = e^{y_j}$. On constate alors que le lien logarithmique transforme nos paramètres de la forme additive en prime multiplicative. En prenant le logarithme de la fonction du maximum de vraisemblance donnée en (1.5), on obtient la fonction suivante à maximiser :

$$l = \sum_i \sum_j n_{ij} r_{ij} \ln \mu_{ij} - n_{ij} \mu_{ij} - n_{ij} \ln r_{ij}!,$$

qu'on peut dériver par rapport à x_k de la façon suivante :

$$\begin{aligned} \frac{\partial l}{\partial x_k} &= \frac{\partial l}{\partial \mu_k} \frac{\partial \mu_k}{\partial x_k} = \sum_j \left(\frac{n_{kj} r_{kj}}{\mu_{kj}} - n_{kj} \right) \mu_{kj} \\ &= \sum_j (n_{kj} r_{kj} - n_{kj} a_k b_j). \end{aligned}$$

Une fois cette dérivée posée égale à zéro, on retrouve les relativités de la fonction de biais de déviance moyenne pour la prime multiplicative qui est donnée par

$$a_i = \frac{\sum_j n_{ij} r_{ij}}{\sum_j n_{ij} b_j}.$$

Normal – Lien logarithmique

On peut appliquer les mêmes étapes qu'avec la loi de Poisson à la distribution normale. On obtient la fonction de logvraisemblance à maximiser suivante :

$$l = - \sum_i \sum_j \frac{n_{ij} (r_{ij} - \mu_{ij})^2}{2\sigma^2} - n_{ij} \ln \sigma \sqrt{2\pi},$$

4.2. CONNECTIONS ENTRE LA CRÉDIBILITÉ SUR LES MONTANTS MOYENS DE SINISTRI

qui, une fois dérivée par rapport à x_k donne

$$\begin{aligned}\frac{\partial l}{\partial x_k} &= \frac{\partial l}{\partial \mu_k} \frac{\partial \mu_k}{\partial x_k} = \sum_j \frac{n_{kj}(r_{kj} - \mu_{kj})}{\sigma^2} \mu_{kj} \\ &= \frac{1}{\sigma^2} \sum_j n_{kj} r_{kj} a_k b_j - n_{kj} a_k^2 b_j^2.\end{aligned}$$

Les relativités obtenues sont les mêmes que le modèle multiplicatif des moindres carrés :

$$a_i = \frac{\sum_j n_{ij} r_{ij} b_j}{\sum_j n_{ij} b_j^2}.$$

On constate à ce point que ce ne sont pas toutes les fonctions du biais minimum qui ont un GLM équivalent, comme c'est le cas pour les fonctions de biais du Khi-carré. D'un autre côté, en choisissant d'autres combinaisons de distributions et de fonctions de lien, on peut trouver une multitude de relativités qui n'ont pas de formule de biais classique. Ces relativités ont toutefois des hypothèses statistiques sous-jacentes et permettent ainsi encore une fois de déterminer si les classes sont vraiment significatives et de déduire des intervalles de confiance pour les paramètres.

4.2 Connexions entre la crédibilité sur les montants moyens de sinistres et les modèles mixtes

Maintenant que les modèles mixtes ont été introduits, on est en mesure de démontrer que certains modèles de crédibilité en sont des cas particuliers qui peuvent être exprimés sous la même forme que les autres modèles de régression. On traite dans ce chapitre uniquement de la crédibilité sur les montants moyens de sinistres. Les autres cas débordent du sujet de ce mémoire.

On applique les modèles de Bühlmann et Bühlmann-Straub aux montants de sinistres et on les compare à des modèles linéaires. On utilise la distribution normale pour trouver la forme de la prime et pour les méthodes d'estimation des paramètres. L'hypothèse normale est justifiée notamment dans le contexte de l'assurance collective. L'assuré correspond au régime d'assurance collective et on suppose un nombre assez élevé de participants. Alors,

les coûts totaux pour l'ensemble du groupe sont donnés par la somme de ceux des participants et, dans ces circonstances, la distribution normale est justifiée par le théorème central limite.

4.2.1 Modèle de Bühlmann

On cherche à retrouver la prime pour la période $T + 1$ du modèle de Bühlmann avec un modèle de la forme suivante :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$$

Pour l'assuré i à la période t , on souhaite avoir la relation

$$y_{it} = \beta + \gamma_i + \varepsilon_{it},$$

où β est la moyenne de tous les sinistres et γ_i est la déviation moyenne pour l'assuré i . Cela dit, on peut représenter tous les assurés pour toutes les périodes avec les matrices \mathbf{X} et \mathbf{Z} suivantes :

$$\mathbf{X} = (1, 1, \dots, 1)'$$

de dimension $nT \times 1$ et

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_n \end{pmatrix}$$

de dimension $nT \times n$. Les \mathbf{z}_i sont des vecteurs colonnes de T éléments tous égaux à 1. On constate alors que l'espérance des y_{it} est égale à β et l'espérance d'un assuré en particulier étant donné son paramètre de risque $E[y_{it}|\gamma_i]$ est égale à $\beta + \gamma_i$. Pour ce qui est de la variance des sinistres d'un assuré i , elle est donnée par

$$\text{Var}[\mathbf{y}_i] = \boldsymbol{\Sigma}_i = \mathbf{z}_i D \mathbf{z}_i' + \mathbf{v}_i,$$

où $D = \sigma_\gamma^2$, $\mathbf{v}_i = \text{Var}[\boldsymbol{\varepsilon}_i] = \sigma_\varepsilon^2 \mathbf{I}_T$ et \mathbf{I}_T est une matrice identité $T \times T$.

Pour trouver des estimés pour nos paramètres, on considère d'abord que σ_γ^2 et σ_ε^2 sont connus. On utilise maintenant la formule développée pour la régression linéaire pondérée à la section 3.3.2 :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}). \quad (4.3)$$

4.2. CONNECTIONS ENTRE LA CRÉDIBILITÉ SUR LES MONTANTS MOYENS DE SINISTRI

Des manipulations algébriques simples nous mènent à l'estimateur de β suivant :

$$\hat{\beta} = \frac{n \sum_i \sum_t y_{it}}{n^2 T} = \bar{y}.$$

Avec l'estimateur pour le paramètre β , on est en mesure de déduire $\hat{\gamma}$ à l'aide de (3.16) :

$$\hat{\gamma}_{BLUP} = DZ' \Sigma^{-1} (\mathbf{Y} - \mathbf{X} \hat{\beta}). \quad (4.4)$$

Encore une fois, il suffit de résoudre le système d'équations pour obtenir la solution suivante :

$$\begin{aligned} \hat{\gamma}_{i,BLUP} &= \frac{T \sigma_\gamma^2}{T \sigma_\gamma^2 + \sigma_\varepsilon^2} (\bar{y}_i - \bar{y}) \\ &= \frac{T}{T + \sigma_\varepsilon^2 / \sigma_\gamma^2} (\bar{y}_i - \bar{y}), \end{aligned}$$

où $\bar{y}_i = (1/T) \sum_{t=1}^T y_{it}$.

Finalement, on peut aussi se servir des modèles mixtes pour trouver la prime d'un assuré i pour la période $T + 1$ en estimant

$$W = E[y_{i,T+1} | \gamma_i] = x_{i,T+1} \beta + z_{i,T+1} \gamma_i.$$

On a vu que la meilleure prévision linéaire non biaisée (BLUP) était le résultat de $W_{BLUP} = x_{i,T+1} \hat{\beta} + z_{i,T+1} \hat{\gamma}_{i,BLUP}$:

$$\begin{aligned} W_{BLUP} &= \bar{y} + \frac{T}{T + \sigma_\varepsilon^2 / \sigma_\gamma^2} (\bar{y}_i - \bar{y}) \\ &= \frac{T}{T + \sigma_\varepsilon^2 / \sigma_\gamma^2} \bar{y}_i + \left(1 - \frac{T}{T + \sigma_\varepsilon^2 / \sigma_\gamma^2} \right) \bar{y} \\ &= \zeta \bar{y}_i + (1 - \zeta) \bar{y}, \end{aligned}$$

où $\zeta = \frac{T}{T + \sigma_\varepsilon^2 / \sigma_\gamma^2} (\bar{y}_i - \bar{y})$.

On arrive donc à la même forme de la prime qu'avec le modèle de Bühlmann. Cependant, il ne faut pas perdre de vue le fait qu'on a trouvé nos estimateurs en considérant les composantes de variances comme connues, alors que pour obtenir la prime définitive, on doit les estimer. Avec la méthode

de Bühlmann, on utilise \hat{s}^2 et \hat{a} comme défini à la section 2.2. Un avantage considérable des modèles mixtes est qu'il est possible d'utiliser plusieurs méthodes comme le maximum de vraisemblance ou le maximum de vraisemblance restreint et ainsi d'avoir des estimateurs qui ont des propriétés plus intéressantes. On doit par contre supposer une distribution des montants de sinistres, chose qui n'est pas nécessaire avec le modèle de Bühlmann.

4.2.2 Modèle avec poids de Bühlmann-Straub

On suit maintenant les mêmes étapes qu'à la section précédente pour obtenir les résultats du modèle de Bühlmann-Straub. Tout d'abord, la seule différence entre ces deux modèles est la présence de poids distincts w_{it} associés à chacun des y_{it} . La matrice de variance-covariance pour les ε_{it} devient

$$\mathbf{V} = \text{Var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \begin{pmatrix} \mathbf{v}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{v}_n \end{pmatrix},$$

où $\mathbf{v}_i = \text{diag}(1/w_{i1}, \dots, 1/w_{iT})$. En utilisant (4.3), on effectue le produit matriciel pour obtenir le résultat suivant :

$$\hat{\beta} = \frac{\sum_t w_{it} y_{it}}{\sum_t w_{it}} = \bar{y}_{i,w}.$$

Par la suite, avec un peu de travail sur les matrices, on peut trouver un estimé du paramètre aléatoire γ_i avec (4.4) :

$$\hat{\gamma}_i = \zeta_i (\bar{y}_{i,w} - \bar{y}_w),$$

où

$$\zeta_i = \frac{\sum_t w_{it}}{\sum_t w_{it} + \sigma_\varepsilon^2 / \sigma_\gamma^2}$$

et

$$\bar{y}_w = \frac{\sum_i \zeta_i \bar{y}_{i,w}}{\sum_i \zeta_i}.$$

En combinant les paramètres trouvés précédemment, on peut obtenir la prime à la période $T + 1$:

$$W_{BLUP} = \zeta_i \bar{y}_{i,w} + (1 - \zeta_i) \bar{y}_w.$$

4.2.3 Estimation des composantes de variances

On a considéré comme connues les composantes de variance afin de trouver une expression pour la prime de crédibilité. À présent, il reste à estimer σ_γ^2 et σ_ε^2 selon le maximum de vraisemblance ou le maximum de vraisemblance restreint décrit en 3.7.2. Pour le modèle de Bühlmann, on arrive à la forme suivante pour les composantes de la matrice de variance Σ :

$$\hat{\sigma}_{\varepsilon,MLE}^2 = \frac{\sum_i \sum_t (y_{it} - \bar{y}_i)^2}{n(T-1)}$$

et

$$\hat{\sigma}_{\gamma,MLE}^2 = \frac{\sum_i (\bar{y}_i - \bar{y})^2}{n} - \frac{\sum_i \sum_t (y_{it} - \bar{y}_i)^2}{nT(T-1)},$$

où $\bar{y}_i = \sum_t y_{it}$. On arrive au même estimateur de la variance $\hat{\sigma}_\varepsilon^2$ utilisé dans Bühlmann. Par contre, l'estimateur de σ_γ^2 est biaisé. On peut corriger ce biais en utilisant la méthode du maximum de vraisemblance restreint. On obtient alors le même $\hat{\sigma}_\varepsilon^2$, mais $\hat{\sigma}_\gamma^2$ est corrigé et donne

$$\hat{\sigma}_{\gamma,REML}^2 = \frac{\sum_i (\bar{y}_i - \bar{y})^2}{n-1} - \frac{\sum_i \sum_t (y_{it} - \bar{y}_i)^2}{nT(T-1)}.$$

Pour un modèle de crédibilité appliqué sur les montants de sinistres, on a que les estimateurs \hat{s}^2 et \hat{a} sont identiques à ceux trouvés avec un modèle linéaire mixte et une méthode d'estimation pour les composantes de variance basée sur le maximum de vraisemblance restreint.

Chapitre 5

Exemple pratique

Dans cette section, on applique les techniques discutées dans les chapitres précédents en faisant un exemple de tarification automobile. Les données sont celles des immatriculations de la SAAQ pour les véhicules de promenade pour les années 2000 à 2004, connues au 31 mars 2005. Les données de sinistres n'étant pas développées suffisamment, on a dû se contenter des années 2000 à 2003 pour modéliser les montants de sinistres. En effet, pour des sinistres de dommages corporels, les dossiers restent ouverts plus longtemps que des dédommagements matériels, ce qui fait en sorte qu'il est difficile d'estimer avec précision les coûts finaux des sinistres d'une année peu mature. Dans un premier temps, on explique les variables mises à notre disposition. Par la suite, on travaille avec les nombres de sinistres afin de développer un modèle pour la fréquence de sinistres des véhicules de promenade du Québec. On répète ensuite ce travail, mais avec les montants de sinistres pour avoir aussi un modèle de sévérité. Finalement, on combine ces deux modèles pour en arriver à un système de tarification basé sur les GLM. On n'utilise pas dans ce contexte les GLMM, puisqu'on n'a pas un nombre suffisant d'années d'expérience, combiné au fait que la tarification se fait sur le type de véhicule et non sur un individu en particulier. Ce regroupement de véhicules distincts fait en sorte que les effets aléatoires ne sont pas significatifs.

Lorsque pertinent, on introduit les commandes SAS qui s'appliquent à la démarche effectuée. Ces notes ne sont pas assez détaillées pour refaire complètement l'analyse, mais elles aident à comprendre les commandes plus spécifiques de notre démarche.

5.1 Contexte et données

Avant d'expliquer la signification des différentes variables mises à notre disposition, on doit rappeler le contexte dans lequel la SAAQ fait sa tarification pour l'immatriculation. Au Québec, chaque propriétaire d'un véhicule ayant le droit de circuler sur les routes doit être immatriculé et payer une contribution d'assurance à la SAAQ, qui s'élevait à 107 \$ en 2006 pour les véhicules de promenade, pour couvrir les dommages corporels subits ou causés sur les routes du Québec ou subits à l'étranger. On constate alors que cette portion de la prime d'assurance (car un autre montant est prélevé par le biais des frais de permis de conduire) est chargée aux propriétaires de véhicules et ce, même si ce ne sont pas toujours ces derniers qui conduisent leur véhicule. La tarification doit donc se faire sur les véhicules et non sur les conducteurs, ce qui complique grandement le travail. En effet, les variables qui sont généralement utilisées par les assureurs canadiens, telles que le sexe du conducteur, l'âge, le nombre d'années d'expérience ou le dossier de conduite ne peuvent pas être utilisées dans ce contexte. Les principales variables disponibles sont les caractéristiques du véhicule, comme entre autre le poids, le prix, l'année du modèle, la présence de freins ABS, etc.

En plus de ces caractéristiques du véhicule, on a la catégorie d'usage du véhicule, qui distingue les différents usages de véhicule au Québec. Présentement, tous les véhicules de promenade paient le même montant malgré leur différente catégorie d'usage, mais une nouvelle tarification pourrait utiliser cette variable. Tout d'abord, on a la variable qui sera plus tard appelée «A» qui détermine l'utilisation qui est faite du véhicule. Pour les véhicules de promenade dont il est question, on a deux catégories, soit les automobiles utilisées pour un usage personnel (PROME) et les habitations motorisées de 3000 kg ou moins utilisées à des fins personnelles (HABMO). Ensuite, on a la variable qu'on nomme «B» qui décrit le type d'utilisateur. Pour cette variable, on a les conducteurs sans exemption ni privilège (SP) qui inclut la majorité des conducteurs de véhicule de promenade du Québec, les titulaires d'une licence de radioamateur (RA), les corps consulaire et délégation commerciales (CC) et les représentants de l'Organisation civile internationale (CD). Pour ce qui est des deux autres variables qui déterminent la catégorie d'usage, soit le type de véhicule et le lieu où il circule, on n'a que des véhicules de promenade (AU) qui circulent sans lieu spécifique au Québec (SQ), alors on n'utilise pas ces facteurs.

Pour ce qui est des caractéristiques du véhicule, on dispose de dix posi-

tions du numéro d'identification du véhicule (NIV10) qui permettent d'avoir les caractéristiques du véhicule sans toutefois pouvoir identifier un propriétaire en particulier. Les NIV10 des véhicules immatriculés au Québec sont envoyés au Bureau d'Assurance du Canada qui extrait les caractéristiques pour ensuite les envoyer à la SAAQ. On retrouve dans le tableau suivant la liste des variables disponibles :

Année du modèle
Type de carrosserie
Indicateur de traction
Indicateur de décapotable
Freins anti-blocage (ABS)
Système antivol
Type d'essence
Coussins gonflables
Empattement
Poids
Puissance du moteur
Prix

5.1.1 Variables catégoriques

Plusieurs de ces variables sont des variables avec un nombre fini de réponses possibles, c'est-à-dire qu'elles ne sont pas continues. On donne dans cette section les réponses possibles pour les variables catégoriques.

Année du modèle

La base de données contient des véhicules pour les années 1992 à 2005 inclusivement. L'année de modèle 1992 regroupe cependant tous les véhicules des années antérieures, ce qui fait en sorte qu'on est réellement en présence de données censurées. On considère que ce regroupement n'est pas irréaliste et que les voitures de plus de dix ans ont un comportement semblable. Comme on a plusieurs années d'accident, il est préférable de modifier cette variable selon l'âge du véhicule plutôt que selon son année de modèle :

$$\text{age} = \text{ACCYear} - \text{model_yr},$$

où ACCYear est le nom de la variable qui indique l'année d'accident de la base de donnée et model_yr représente l'année du modèle. Plusieurs constructeurs

de voitures commencent à vendre leurs nouveaux véhicules quelques mois avant la nouvelle année, ce qui fait en sorte que certaines immatriculations sont émises sur des voitures dont l'âge est de -1 ou parfois même -2. Comme le nombre de ces véhicules est négligeable, on a préféré les regrouper avec les âges 0, qui sont aussi des voitures de l'année.

Type de carrosserie

Cette variable décrit à quel type général le véhicule appartient. Elle a les niveaux suivants :

2D	2 portes
4D	4 portes
1P	Camionnette 1/2 tonne
2P	Camionnette 3/4 tonne
3P	Camionnette 1 tonne
4P	Camionnette compact
CV	Fourgonnette cargo
MP	Véhicule Utilitaire (VUS)
SS	Sport/ Spécial
SW	Familial

Elle permet de regrouper les différents véhicules selon des groupes généraux déterminés par le BAC.

Indicateur de traction

Il y a deux types de traction, soit deux roues motrices, codé comme 1 et quatre roues motrices, codé comme 2.

Indicateur de décapotable

Les voitures décapotables ont «C» pour la variable appelée «top» et les véhicules avec un toit rigide ont plutôt «H».

Freins anti-blocage

Pour ce qui est de la variable ABS, il y a trois possibilités soit 1 pour aucun freins ABS, 2 pour des freins ABS à l'arrière seulement et 3 pour des freins ABS aux quatre roues.

Système antivol

La valeur 1 indique que le véhicule n'a pas de système antivol approuvé par le BAC, alors que les voitures ayant un système approuvé ont 2 comme valeur. Il est préférable d'utiliser des variables ayant une correspondance dans les systèmes de la SAAQ et la présence d'un système antivol ne semble pas significative pour des blessures corporelles ; on n'utilisera donc pas cette variable.

Type d'essence et coussins gonflables

Les voitures fonctionnant au diesel sont codées à 2 alors que les autres types d'essence sont à 1. Pour ce qui est des coussins gonflables, la présence de coussins gonflables est identifiée par la valeur 2 alors que l'absence est indiquée par la valeur 1.

5.1.2 Variables continues

Les variables continues sont plus difficiles à utiliser, puisqu'on doit les regrouper si on veut les utiliser pour la tarification. En effet, on ne peut pas avoir des primes différentes pour toutes les combinaisons possibles de prix, poids, puissance du moteur et empattement, puisque pratiquement tous les véhicules auraient une prime différente, ce qui n'est pas le but du modèle. Il faut donc décider du nombre de groupes optimal, ainsi que des divisions à faire, ce qui est un processus long qui repose en grande partie sur le jugement de l'actuaire qui les fait.

Pour prendre ces décisions, on a d'abord tracé des graphiques afin de voir s'il y a des différences évidentes à certains seuils. Par la suite, avec l'aide des graphiques, on a calculé des espérances pour des tranches données, afin de vérifier si les seuils séparaient vraiment des niveaux différents. Finalement, une fois la variable insérée dans le modèle linéaire généralisé, on a vérifié que les niveaux étaient significatifs. On a alors ajusté les seuils pour avoir des β significativement différents les uns des autres et ainsi d'avoir des classes qui représentent bien les données. La technique adoptée a été de conserver la majorité des véhicules plus communs dans une ou deux classes au centre de la distribution et de séparer les extrémités avec deux nouvelles classes qui se comportent généralement différemment. Suite à ces étapes, on a laissé tomber la variable *empattement*, puisqu'elle n'est pas significative pour la fréquence

ni pour la sévérité. On peut penser que cela résulte du fait que la variable *poids* est aussi dans le modèle, et qu'en général une voiture plus spacieuse aura également un poids plus élevé. On confirme aussi ce résultat en faisant une régression linéaire entre ces variables. 70% de la variable empattement est expliquée par le poids pour la fréquence et 65% pour la sévérité.

Prix

Pour ce qui est de la variable prix, on a séparé les véhicules en quatre groupes : les voitures de moins de 15 000 \$, celles entre 15 000 et 45 000, celles entre 45 000 et 75 000 et celles de plus de 75 000 dollars. On a utilisé une variable indexée pour le prix. Le BAC a utilisé des taux d'inflations cohérents selon les années de modèle combinées au prix de vente d'origine pour obtenir des dollars constants. On doit utiliser un prix en dollars constants afin d'obtenir des données comparables.

Poids

On a aussi choisi de séparer les voitures en quatre classes de poids, soit les moins de 750 kg, celles entre 750 et 1250, celles entre 1250 et 2250 et celles de plus de 2250 kg. Comme on peut le constater, le but est de séparer les voitures qui ont des poids extrêmes (soit très petits, soit très élevés) des autres véhicules et de finalement séparer les véhicules de poids moyens en deux groupes.

Puissance du moteur

La dernière variable continue est la puissance du moteur (variable «*hpower*»). On a regroupé ensemble les moteurs à faible puissance à moins de 80 chevaux vapeur, puis les moteurs très puissants de plus de 250 chevaux vapeur et finalement, pour les véhicules restants, on les a séparés en deux groupes avec 150 comme point de séparation.

Ratio de la puissance du moteur sur le poids

On a aussi introduit une nouvelle variable, soit la puissance du moteur divisée par le poids du véhicule. En effet, plus un véhicule est lourd, plus il a besoin d'un moteur puissant. Le ratio puissance/poids est donc un autre indicateur de la puissance, puisqu'il élimine l'effet du poids du véhicule. Le

domaine de cette variable s'étend approximativement de 0,05 jusqu'à 0,30. On a également séparé les extrémités de la distribution en deux classes distinctes, soit les ratios inférieurs à 0,06 et ceux supérieurs à 0,15. Pour les ratios moyens, on les a aussi séparés en deux groupes, avec la valeur 0,11 utilisée comme seuil.

5.2 Modèles linéaires généralisés appliqués aux données

Dans le contexte décrit à la section précédente, il a été mis en évidence que le but premier d'utiliser les GLM est d'établir de nouvelles primes d'assurance pour les dommages corporels avec les variables disponibles. On aurait pu la calculer directement avec un seul modèle et des données présentées comme plusieurs montants de sinistres de zéro, et quelques montants non nuls. Cependant, il est très difficile de trouver un modèle adéquat pour les données, puisque aucune distribution continue disponible dans SAS n'a une masse considérable à zéro. C'est pour cette raison qu'on a choisi de faire deux modèles, soit un pour la fréquence et un pour la sévérité. De plus, les variables peuvent s'avérer significatives pour un modèle et non pour l'autre. Il est donc intéressant de les considérer séparément pour bien mesurer l'impact des variables sur les deux composantes de la prime. On débute dans un premier temps par modéliser la fréquence, on poursuit ensuite avec la sévérité et on combine les deux primes pour terminer. Pour combiner les deux modèles, on se base sur le fait que la prime pure est égale à la fréquence multipliée par la sévérité :

$$\begin{aligned} PP &= \frac{S}{E} = \frac{S}{N} \frac{N}{E} \\ &= \text{Sev} * \text{Freq}, \end{aligned}$$

où S représente les sinistres totaux, E l'exposition et N les nombres de sinistres.

La dernière chose à faire avant de décrire les deux modèles qui définissent la prime est d'expliquer la variable de l'année d'accident («ACCYear») qui est incluse dans les modèles sans but de tarification. Cette variable est cependant nécessaire pour absorber les effets externes aux risques des véhicules, comme par exemple l'inflation pour la sévérité. S'il y a une tendance évidente dans

cette variable, on peut l'utiliser pour projeter la prime pure de l'année où le modèle se termine jusqu'à celle où la tarification sera appliquée. Autrement, on peut projeter la prime pure avec des données externes et simplement ignorer cette variable pour la tarification. Avec les données de la SAAQ, on préfère ignorer cette variable et faire la tarification pour l'année 2004. Il faudra donc indexer les nouvelles primes au moment de leur mise en vigueur.

5.2.1 Distribution de fréquence

On cherche à trouver la meilleure distribution possible pour le nombre de sinistres subits entre 2000 et 2004 par les 21 813 502 mois d'exposition des véhicules de la base de données. Pour ce faire, on commence avec un modèle de Poisson, un lien logarithmique et toutes les variables explicatives décrites précédemment. On utilise aussi dans le modèle un terme d'offset égal au logarithme de l'exposition de chacune des cellules.

Le modèle a un facteur de surdispersion $\hat{\phi} = 1,1376$, ce qui est raisonnable pour le nombre élevé de données qu'on a. Étant donné qu'on veut garder une structure simple et facile à expliquer, on conserve la distribution de Poisson. Cependant, pour s'assurer de la justesse des inférences, on ajuste les variances des paramètres avec le facteur de dispersion, comme expliqué à la section 3.5.1. Pour ce faire, on doit multiplier les variances des bêta du modèle par l'estimé du paramètre de surdispersion. En effet, cela revient à supposer que la vraie distribution n'est pas une Poisson et que les données suivent plutôt une binomiale négative¹.

Maintenant qu'on s'est assuré que les inférences sont valides, on peut retirer du modèle les variables qui ne sont pas significatives. Pour ce faire, on utilise des tests du rapport des vraisemblances. On ne peut pas utiliser les statistiques t de Student puisque les variables sont catégoriques et possèdent plusieurs niveaux. Dû à ce fait, on doit tester si tous les paramètres liés à une variable sont simultanément différents de zéro. On doit donc retirer la variable du modèle, trouver la valeur de la fonction de vraisemblance et calculer la valeur p^2 .

Pour le modèle, on retire d'abord la variable «essence» qui a une valeur p de 0,2535. Une fois cette variable retirée, il s'avère que toutes les autres ont une importance significative dans le modèle. On aurait pu retirer la trac-

¹En SAS, on n'a qu'à ajouter l'option PSCALE à la ligne MODEL.

²On peut cependant effectuer directement ce travail avec l'option TYPE3 de SAS.

tion qui a une valeur p de 0,06 alors que toutes les autres ont des valeurs inférieures à 1%. Cependant, cette variable est importante pour la sévérité et on a préféré la conserver. Il est normal que la majorité des variables restent significatives, car aucune des informations qu'on possède n'explique suffisamment le nombre de sinistres pour sortir les autres du modèle. Par exemple, si on avait accès à l'âge du conducteur, au sexe et au dossier de conduite, qui sont des variables reconnues pour bien expliquer le nombre de sinistres, probablement que plusieurs des variables utilisées deviendraient non significatives.

À présent, on choisi de regrouper la variable de l'âge du véhicule pour lisser la tarification. En effet, en conservant des primes différentes pour les voitures d'âge différent, le prime change tous les ans pour les propriétaires. Il est alors préférable de regrouper les âges de véhicules pour stabiliser les primes. De plus, on a quelques exemples d'écarts significatifs entre deux classes d'âge. En effet, les voitures âgées de quatre ans se voient accorder un rabais de 5%, rabais qui diminue à 2,2 % pour les véhicules de cinq ans, alors qu'il revient à 5% pour ceux de six ans. En regroupant les âges pour avoir moins de groupes, on se trouve à stabiliser les rabais et surprimes. Pour faire ces regroupements, on a dû regarder simultanément la fréquence et la sévérité. En effet, si deux âges de véhicules sont très semblables du côté de la fréquence et très différents du côté de la sévérité, on ne peut pas les regrouper puisqu'on se trouve à biaiser la sévérité et le paramètre âge ne sera plus significatif pour la sévérité. On doit donc regrouper de façon à perdre le moins d'informations possibles sur la fréquence et sur la sévérité. En regardant l'allure générale des $\hat{\beta}$, on a choisi de définir des primes différentes à chaque deux années d'âge de véhicules. Par la suite, on s'assure que les regroupements sont cohérents en testant $\beta_i = \beta_j$ pour i et j deux années consécutives. Il s'agit d'une autre forme de statistique t de Student³. Avec ce test, on s'aperçoit qu'il est préférable de laisser les voitures de l'année (âge 0) séparées des autres, puisqu'elles se comportent très différemment des automobiles âgées d'un an.

On arrive enfin à notre modèle final pour la fréquence avec les variables catégoriques et leurs valeurs données dans le tableau 5.1. Pour ce qui est des variables continues qu'on a regroupées, on a pris un point dans l'intervalle qui sert de référence à la catégorie.

Bien que ce modèle n'explique pas parfaitement les données, il est satisfaisant considérant les contraintes de la SAAQ. Les relativités obtenues sont

³Avec SAS, on a recours à la ligne ESTIMATE.

Variables	Niveaux
ACCYear	2000, 2001, 2002, 2003, 2004
A	HABMO, PROME
ABS	1, 2, 3
Age	0, 1, 3, 5, 7, 9, 11
B	CC, CD, RA, SP
Body_style	1P, 2D, 2P, 3P, 4D, 4P, CV, MP, PV, SS, SW
Coussin	1, 2
Hpower	75, 100, 200, 250
HPPoids	0,05, 0,1, 0,12, 0,2
Poids	750, 1250, 2000, 2500
Prix	10 000, 30 000, 50 000, 100 000
Top	C, H
Traction	1, 2

TAB. 5.1 – Niveaux possibles des variables utilisé

présentées avec celles du modèle pour la sévérité à la section 5.2.3.

5.2.2 Distribution pour la sévérité

Pour ce qui est de la sévérité, on doit d'abord se choisir une distribution du montant d'un sinistre. Les montants sont positifs et continus, alors on doit opter pour une distribution comme la gamma, l'inverse gaussienne, la lognormale, la Pareto, etc. On doit éliminer les distributions discrètes qui ne sont pas appropriées, et la loi normale, puisqu'elle permet les observations négatives. On a fait plusieurs tests qui semblaient appropriés à la situation, mais on a eu quelques difficultés à faire converger le modèle. On a donc conservé deux distributions à tester qui sont présentes dans le logiciel SAS, soit la gamma et l'inverse gaussienne. En régression linéaire, pour comparer deux modèles, on prendrait simplement le R_a^2 qui mesure le degré d'ajustement en tenant compte du nombre de paramètres. Dans la situation présente, la distribution des deux modèles est différente et SAS ne calcule aucun R^2 . On ne peut pas faire non plus un test du maximum de vraisemblance, puisque le premier modèle n'est pas imbriqué dans le second. On a donc calculé la somme des erreurs carrées pondérées pour les modèles optimaux des deux distributions.

5.2. MODÈLES LINÉAIRES GÉNÉRALISÉS APPLIQUÉS AUX DONNÉES81

Pour calculer les modèles optimaux de sévérité, le cheminement est semblable à celui utilisé pour la fréquence. Par contre, on s'intéresse moins à la surdispersion que dans le cas de la Poisson où une surdispersion évidente conduisait à la binomiale négative. De plus, comme on a une observation par sinistre, on n'a pas besoin de terme d'offset. Finalement, comme on voulait avoir une prime de forme multiplicative, on a conservé le lien logarithmique même si les liens canoniques des lois gamma et inverse gaussienne sont respectivement le lien inverse et le lien inverse au carré. On a donc utilisé l'option TYPE3 et retiré une à une les variables non significatives des modèles.

En utilisant la distribution gamma, les facteurs suivants et leurs niveaux sont demeurés dans le modèle :

Variabiles	Niveaux
ACCYear	2000, 2001, 2002, 2003
Age	0, 1, 3, 5, 7, 9, 11
B	CC, CD, RA, SP
Body_style	1P, 2D, 2P, 3P, 4D, 4P, CV, MP, PV, SS, SW
HPPoids	0,05, 0,1, 0,12, 0,2
Poids	750, 1250, 2000, 2500
Prix	10 000, 30 000, 50 000, 100 000
Traction	1, 2

Dans le cas de la distribution inverse gaussienne, on conserve seulement trois variables de tarification :

Variabiles	Niveaux
ACCYear	2000, 2001, 2002, 2003, 2004
Age	0, 1, 3, 5, 7, 9, 11
Body_style	1P, 2D, 2P, 3P, 4D, 4P, CV, MP, PV, SS, SW
Prix	10 000, 30 000, 50 000, 100 000

Avec ces deux modèles, on a calculé les fréquences et les montants totaux observés pour toutes les primes possibles sur les quatre années d'accident. Selon les relativités déterminées par le GLM, on a déduit les montants totaux de sinistres espérés pour ces primes. Avec ces valeurs, on a pu trouver la somme pondérée des résidus mis au carré :

$$SS = \frac{\sum_{t=1}^4 \sum_{i=1}^{N_t} w_{it} (y_{it} - \hat{y}_{it})^2}{\sum_{t=1}^4 \sum_{i=1}^{N_t} w_{it}},$$

où N_t est le nombre de classes dans la base de données pour chacune des années. Les w_{it} représentent les nombres de sinistres pour chacune des nouvelles classes, y_{it} est le montant total des sinistres observés pour la classe i et l'année t , alors que \hat{y}_{it} est le montant total des sinistres prédits par le modèle pour la classe i au temps t . On a pondéré nos erreurs parce que les sinistres sont regroupés et qu'on veut donner la même importance à chacun des sinistres individuels. On obtient les résultats suivants :

$$SS_{gamma} = 2\ 332\ 801\ 609\ 672$$

$$SS_{IG} = 6\ 315\ 877\ 875\ 800.$$

Comme le modèle inverse gaussien a une somme des résidus carrés environ trois fois plus élevée que la gamma, on conclut que la gamma est plus appropriée pour modéliser le montant d'un sinistre dans notre contexte.

5.2.3 Modèle final

Une fois que la fréquence et la sévérité ont été modélisées, il faut combiner les deux modèles pour obtenir la prime pure. Pour la sévérité, on a un bon nombre de variables qui ne sont pas significatives. Étant donné qu'on souhaite les conserver pour la fréquence, on va aussi les inclure dans le modèle de sévérité. En effet, les divisions existant déjà pour la fréquence, il est maintenant préférable de conserver les variables pour la sévérité et ainsi augmenter la quantité d'information disponible. On diminue l'erreur carrée moyenne de moitié. Pour la gamma, elle est de $SS_{gamma} = 1\ 227\ 873\ 853\ 954$ et pour l'inverse gaussienne, de $1\ 212\ 407\ 970\ 066$. Les deux valeurs sont très rapprochées et on préfère la gamma puisque les paramètres sont plus significatifs et que cette loi est plus utilisée en actuariat, donc la compréhension est plus facile. Pour avoir une expression de la prime, on multiplie la fréquence par la sévérité. Étant donné qu'on a utilisé un lien logarithmique pour les deux modèles, cela revient à additionner les $\hat{\beta}$

$$PP = \exp(\hat{\beta}_F + \hat{\beta}_S + r),$$

où $\hat{\beta}_F$ représente les paramètres de la distribution de fréquence, $\hat{\beta}_S$ sont ceux pour la sévérité et r représente la projection entre la période sur laquelle on a basé nos observations et celle où les taux seront utilisés.

On arrive aux coefficients et aux relativités donnés dans le tableau 5.2. Pour déterminer quelles étaient les classes de base, on a choisi celles avec les plus grandes expositions.

5.2. MODÈLES LINÉAIRES GÉNÉRALISÉS APPLIQUÉS AUX DONNÉES83

Paramètres	Classes	$\hat{\beta}_F$	$\hat{\beta}_S$	Relativités
intercept		-7,2555	10,0653	16,6066
A	HABMO	-0,9017	-1,3951	0,1006
	PROME	0	0	1
B	CC	-0,3575	1,5359	3,2492
	CD	-0,8717	1,9087	2,8207
	RA	-0,2635	0,1523	0,8948
	SP	0	0	1
Traction	2	0,066	0,0635	1,1383
	1	0	0	1
ABS	2	0,0006	0,017	1,0178
	3	-0,0359	0,0299	0,9940
	1	0	0	1
Coussin	1	-0,046	-0,0102	0,9454
	2	0	0	1
Prix	12000	0,0183	-0,0911	0,9298
	50000	-0,2652	-0,0799	0,7081
	100000	-0,6501	0,043	0,5449
	30000	0	0	1
Poids	750	-0,032	0,3401	1,3608
	2000	-0,1632	0,0302	0,8755
	2500	-0,1058	0,2271	1,1290
	1250	0	0	1
HPower	75	-0,0642	-0,0281	0,9118
	200	-0,0799	-0,027	0,8986
	250	-0,1418	0,0057	0,8728
	100	0	0	1
Top	C	-0,2177	0,0553	0,8501
	H	0	0	1

On constate qu'il y a un très grand nombre de variables et que le nombre de primes est considérable. Ceci est principalement dû au manque d'une variable de groupe plus détaillée. Si la variable de type de carrosserie («body_style») était subdivisée en quelques sous-groupe, on pourrait probablement éliminer plusieurs variables comme le poids, le prix et les chevaux-vapeur, puisque ces petits groupes en tiendraient compte indirectement. Cependant, avec le NIV10 d'un véhicule et les caractéristiques physiques mises à notre disposition, le modèle construit permet de calculer une prime pour tous les véhicules sous

Paramètres	Classes	$\hat{\beta}_F$	$\hat{\beta}_S$	Relativités
Age	0	0,0565	-0,2027	0,8640
	1	-0,093	-0,2469	0,7118
	3	-0,0864	-0,2122	0,7419
	5	-0,0598	-0,1838	0,7838
	7	-0,0362	-0,1093	0,8646
	11	-0,0281	0,0549	1,0272
	9	0	0	1
Body_style	1P	-0,3833	0,3373	0,9550
	2D	0,1978	0,103	1,3509
	2P	-0,3281	0,5178	1,2089
	3P	-0,5244	0,229	0,7442
	4P	-0,1656	0,2381	1,0752
	CV	-0,3254	0,3768	1,0527
	MP	-0,0322	0,2347	1,2245
	PV	-0,1241	0,0686	0,9460
	SS	0,2036	0,4563	1,9346
	SW	-0,2005	0,0678	0,8757
	4D	0	0	1
HPPoids	0,05	0,061	0,1249	1,2043
	0,12	0,0284	0,0462	1,0775
	0,2	-0,0574	0,2372	1,1970
	0,1	0	0	1

TAB. 5.2 – Relativités selon les différents paramètres

étude.

En observant le tableau 5.2, on remarque qu'il y a un nombre impressionnant de combinaisons possibles entre les facteurs utilisés dans le modèle. Il est bien de s'assurer que l'ordre de grandeur pour les primes est raisonnable avant d'appliquer le modèle. Pour certaines classes ayant peu d'exposition, les primes peuvent être plus variables, et il est pertinent dans ces cas d'utiliser un facteur de crédibilité ou de considérer une méthode alternative de tarification. On s'est donc d'abord assuré que les primes des véhicules de promenade («PROME») ayant les classes de conducteurs les plus courantes, soit «sans privilège ni exemption» et «radioamateurs» («SP» et «RA») étaient comprises dans un domaine raisonnable. La figure 3.2 illustre bien l'ordre de

5.2. MODÈLES LINÉAIRES GÉNÉRALISÉS APPLIQUÉS AUX DONNÉES85

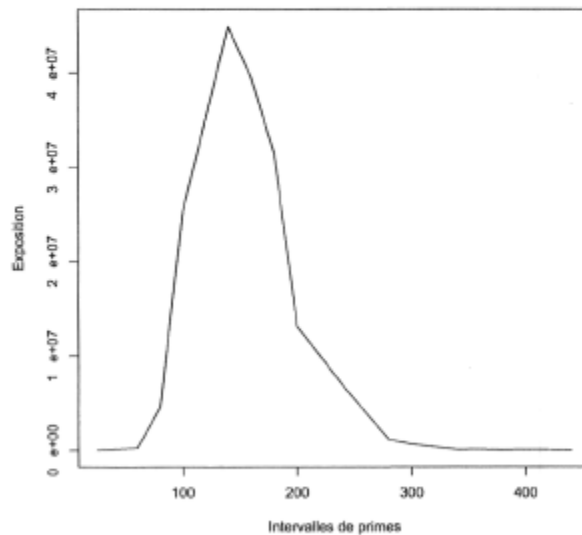


FIG. 5.1 – Distribution de l'exposition selon le niveau des primes

grandeur des primes des véhicules. On peut aussi calculer entre quelles deux valeurs se situe 95 % des primes, et on obtient entre 100 \$ et 260 \$.

En somme, la tarification a été faite dans le contexte particulier de la SAAQ. En tant qu'organisme gouvernemental, il est dans son obligation d'offrir à tous un produit d'assurance abordable. De cette façon, des règles d'équité doivent s'appliquer et certaines variables descriptives du conducteur utilisées dans le privé ne peuvent être utilisées par la SAAQ pour discriminer les risques. Le processus a donc été plus complexe, mais on considère que le résultat est statistiquement significatif. Il serait ainsi intéressant d'obtenir plus d'information sur le conducteur des véhicules et de pouvoir l'utiliser à des fins de tarification. On obtiendrait alors un modèle qui explique encore plus l'expérience de la SAAQ. Finalement, mentionnons que l'opinion de l'auteure n'engage en rien le BAC ou la SAAQ.

Conclusion

Dans ce mémoire, on a établi des liens entre les modèles de régression utilisés en statistiques et deux contextes actuariels, soit la tarification et la crédibilité. On a d'abord présenté des méthodes de tarification traditionnelles, soit l'analyse univariée et la procédure du biais minimum. L'analyse univariée est une des premières méthodes de tarification présentée dans les articles actuarielles. Bien qu'elle soit simple et qu'elle ne puisse plus vraiment être appliquée dans les compagnies d'assurance étant donné les techniques de tarification maintenant plus raffinées et spécialisées, elle donne un point de référence intéressant pour comparer les nouvelles méthodes.

Une de ces nouvelles méthodes, le biais minimum, est ensuite décrite et comparée à l'analyse univariée. On constate que, comparativement à l'analyse univariée, la procédure du biais minimum réussit à éliminer la dépendance entre les facteurs ce qui est essentiel si on veut bien mesurer le risque de chacune des classes sans en surestimer une ou plusieurs. Cependant, avec la procédure du biais minimum, tout comme avec l'analyse univariée, on ne peut pas choisir les facteurs utilisés dans la tarification selon qu'ils soient significatifs ou non. En effet, on doit prendre notre décision a priori sur une base intuitive, car la méthode ne permet pas de savoir si la variable est importante dans le modèle.

Après avoir présenté la théorie de la tarification, on a élaboré les principes de crédibilité de base. Avec le modèle de Bühlmann, on trouve quelle est la meilleure prime linéaire qui minimise l'erreur quadratique moyenne. On obtient que la prime optimale au sens des moindres carrés est une pondération entre l'expérience individuelle (ou du groupe) et l'expérience collective. Une limitation du modèle de Bühlmann est qu'il considère que l'exposition de chacun des assurés pour chacune des période est la même, alors que dans la pratique, il arrive régulièrement que les assurés sont de taille différente. Avec le modèle de Bühlmann-Straub, on introduit des poids permettant de

considérer la taille de l'assuré pour trouver son facteur de crédibilité.

La crédibilité est très importante dans le processus de tarification utilisé en assurance dommages. En effet, il ne suffit pas pour les assureurs de trouver les primes appropriées pour chacune des classes. Ils doivent aussi déterminer dans quelle mesure l'expérience de la classe est crédible, c'est-à-dire à quel point la variabilité de la classe est petite. Pour une classe qui n'a pas beaucoup d'expérience, il est préférable d'utiliser une pondération entre l'expérience du groupe et celle de l'ensemble du portefeuille, comme le démontrent les modèles de Bühlmann et Bühlmann-Straub.

Par la suite, on présente la théorie statistique sur les modèles linéaires généralisés en débutant par le modèle de régression linéaire. On explique les hypothèses de ce modèle et comment introduire les variables explicatives dans le modèle. Étant donné que les contextes qui sont intéressants en actuariat ne peuvent généralement pas être expliqués par des modèles linéaires, on présente la famille exponentielle, qui regroupe plusieurs distributions comme la gamma, la Poisson, la binomiale négative, etc. Avec toutes les distributions de cette famille, on peut construire des modèles de régression pour modéliser une variable, comme le nombre de sinistres ou les montants de sinistres. Finalement, on peut estimer les paramètres de ces modèles en se basant sur leur fonction de vraisemblance lorsque la fonction de densité de la variable endogène est connue, et on peut corriger les inférences dans les cas où la fonction de densité est erronée.

Lorsque les observations proviennent de plusieurs individus qui sont suivis dans le temps, on peut ajouter au modèle un paramètre pour chacune de ces personnes afin de quantifier leur comportement individuel qui n'est pas expliqué par les autres variables. On peut ainsi faire une tarification plus personnalisée et mieux prédire le comportement futur. Pour la majorité des distributions de la famille exponentielle, on doit linéariser le modèle à l'aide du développement en séries de Taylor afin d'estimer les paramètres. On obtient alors ce qu'on appelle un pseudo-modèle, qu'on peut traiter comme les modèles linéaires mixtes.

Il est intéressant de lier les modèles statistiques présentés aux techniques actuarielles utilisées dans le processus de tarification. On trouve alors que plusieurs processus de biais minimum nous donnent les mêmes résultats que si on utilise des modèles linéaires généralisés. Par exemple, les résultats du modèle de biais minimum avec une prime de forme additive et une fonction de biais de déviance moyenne peuvent être reproduits avec un modèle de régression linéaire. L'avantage d'utiliser les modèles de régression est qu'ils

5.2. MODÈLES LINÉAIRES GÉNÉRALISÉS APPLIQUÉS AUX DONNÉES89

permettent de choisir les variables selon leur pertinence et ainsi de mieux justifier leur présence dans le modèle. Finalement, on exprime les modèles de Bühlmann et Bühlmann-Straub avec des modèles linéaires mixtes.

La dernière section de ce travail est consacrée à un exemple concret où les GLM ont été utilisés dans le but d'effectuer une tarification basée sur l'expérience. On utilise les données de fréquence et de sévérité fournies par la SAAQ et on construit deux modèles, soit un pour la fréquence et un pour la sévérité. On termine en combinant ces deux modèles et en s'assurant que les primes obtenues avec les modèles sont vraisemblables. Cette étude a été très pertinente entre autre pour déterminer quelles variables avaient une grande influence sur l'expérience des sinistres. Il serait très intéressant dans le futur de répéter l'expérience pour une compagnie d'assurance privée où un plus grand nombre de variables est disponible.

Bibliographie

- Bailey, R. A. & Simon, L. J. (1960), 'Two studies in automobile insurance ratemaking', *Proceedings of the Casualty Actuarial Society XLVII*, 1–19.
- Bühlmann, H. & Straub, E. (1970), 'Glaubwürdigkeit für Schadensätze', *Bulletin Suisse des Actuaires* **70**(1), 111–133.
- Brown, R. L. (1988), 'Minimum bias with generalized linear models', *Proceedings of the Casualty Actuarial Society LXXV*, 187–217.
- Bühlmann, H. (1967), 'Experience rating and credibility', *Astin Bull.* **4**(3), 199–207.
- Bühlmann, H. & Gisler, A. (2005), *A course in credibility theory and its applications*, Universitext, Springer-Verlag, Berlin.
- Cameron, A. C. & Trivedi, P. K. (1998), *Regression analysis of count data*, Vol. 30 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.
- Cossette, H. (1993), Modèles récents de crédibilité et estimation, Master's thesis, Université Laval.
- Dannenburg, D. R., Kaas, R. & Goovaerts, M. J. (1996), *Practical actuarial credibility models*, Institute of Actuarial Science and Econometrics.
- Denuit, M. & Charpentier, A. (2004), *Mathématiques de l'Assurance Non-Vie. Tome I : Principes Fondamentaux de Théorie du Risque*, Collection Économie et Statistiques Avancées, Economica, Paris.
- Denuit, M. & Charpentier, A. (2005), *Mathématiques de l'Assurance Non-Vie. Tome II : Tarification et Provisionnement*, Collection Économie et Statistiques Avancées, Economica, Paris.
- Diggle, P. J., Heagerty, P. J., Liang, K.-Y. & Zeger, S. L. (2002), *Analysis of longitudinal data*, Vol. 25 of *Oxford Statistical Science Series*, second edn, Oxford University Press, Oxford.

- Feldblum, S. & Brosius, E. J. (2002), 'The minimum bias procedures : A practitioner's guide', *Forums of the Casualty Actuarial Society* **Fall**, 591–684.
- Frees, E. W., Young, V. R. & Luo, Y. (1999), 'A longitudinal data analysis interpretation of credibility models', *Insurance Math. Econom.* **24**(3), 229–247.
- Frees, E. W., Young, V. R. & Luo, Y. (2001), 'Case studies using panel data models', *N. Am. Actuar. J.* **5**(4), 24–42.
- Goovaerts, M. J. & Hoogstad, W. J. (1987), *Credibility theory*, Nationale-Nederlanden N.V., Rotterdam.
- Goulet, V. (1994), *Théorie de la crédibilité : histoire, principes et applications*, Master's thesis, Université Laval.
- Gourieroux, C. (1999), *Statistique de l'Assurance*, Collection Économie et Statistiques Avancées, Economica, Paris.
- Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2004), *Loss models : From data to decisions*, Wiley Series in Probability and Statistics, second edn, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
- McCullagh, P. & Nelder, J. A. (1983), *Generalized linear models*, Monographs on Statistics and Applied Probability, Chapman & Hall, London.
- McNeil, A. J. (1997), 'Estimating the tails of loss severity distributions using extreme value theory', *Astin Bull.* **27**(1), 117–137.
- Mildenhall, S. (1999), 'A systematic relationship between minimum bias and generalized linear models', *Proceedings of the Casualty Actuarial Society* **LXXXVI**, 393–487.
- Mullahy, J. (1997), 'Instrumental-variable estimation of count data models : Applications to models of cigarette smoking behavior', *The Review of Economics and Statistics* **79**(4), 586–593.
- Shaked, M. (1980), 'On mixtures from exponential families', *J. Roy. Statist. Soc. Ser. B* **42**(2), 192–198.
- Straub, E. (1988), *Non-Life Insurance Mathematics*, Springer-Verlag, New-York.
- Winkelmann, R. (2000), *Econometric analysis of count data*, Springer-Verlag, Berlin.