UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODÉLISATION DE L'EXPOSITION EN ASSURANCE AUTOMOBILE PAR UN PROCESSUS DE FILES D'ATTENTE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

CONCENTRATION MATHÉMATIQUES ACTUARIELLES ET FINANCIÈRES

PAR

GUILLAUME COUTURE-PICHÉ

JUIN 2015

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.07-2011). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Tout d'abord, je voudrais commencer par remercier mon directeur de mémoire, Jean-Philippe Boucher, pour avoir eu l'idée initiale de ce formidable projet. J'apprécie beaucoup ton aide et les conseils que tu m'as donnés, particulièrement lorsque nous avons eu à mettre en mots les sections un peu plus compliquées. Ces sections ont vraiment nécessité une attention soutenue, mais nous y sommes parvenus. J'ai beaucoup aimé ton style conciliant, c'est d'ailleurs une des raisons que j'ai continué mes études avec une maîtrise. Je vais surtout retenir nos discussions sur Zelda lors de 5 à 7.

Ah Virginie, je suis vraiment heureux de t'avoir dans ma vie. Depuis notre rencontre, tu es toujours là pour moi. Je tiens particulièrement à coeur les moments passés ensemble après que je me sois blessé. Grâce à toi, j'ai réussi cette épreuve dans la vie et j'ai foi que tu seras toujours présente pour les prochaines. Maintenant, c'est à mon tour de t'offrir tout mon support afin que ta thèse de doctorat soit menée à terme.

Quant au collectif Tsarserveur, on se reprend l'année prochaine pour le chalet. Il ne fallait pas qu'il tombe dans les dernières semaines avant la remise du mémoire. Merci, vous êtes véritablement de bons amis, et j'espère qu'on trinquera à Joël encore dans longtemps.

Finalement, je remercie toutes les autres personnes qui m'ont accompagné durant ce projet, notamment les membres de ma famille, qui ont toujours été présents et qui m'ont aidé dans les moments les plus difficiles.

TABLE DES MATIÈRES

LISTE DES TABLEAUX vi

LISTE DES FIGURES vii

RÉSUMÉ viii

INTRODUCTION 1

CHAPITRE I
MODÈLES POUR UN ASSURÉ 2

1.1 Introduction 2

1.1.1 Définition des termes utilisés 3

1.1.2 Données utilisées et notations 4

1.2 Modélisation du nombre de véhicules 7

1.2.1 Ajouts et retraits de véhicule 7

1.2.2 Renouvellement des véhicules 10

1.3 Modèle avec période d'inactivité 12

1.3.1 Modélisation 15

1.3.1.1 Ajout du choc au renouvellement 16

1.4 Inférence statistique 17

1.4.1 Données complètes 17

1.4.1.1 Notations 17

1.4.1.2 Fonction de vraisemblance 19

1.4.2 Données incomplètes 22

1.4.2.1 Inactif vs annulé 22

1.4.3 Applications numériques 27

1.4.3.1 Variables explicatives 28

1.5 Analyse 30

1.5.1 Nombre espéré de voitures assurées 31
1.5.2 Actualisation de l'exposition 34
1.6 Discussion 36
1.7 Conclusion 38
1.8 Annexe 1 : Développements mathématiques 41
1.8.1 Processus de Poisson 41
1.8.2 Naissance et mort 46
1.8.3 Annulation de police 51
CHAPITRE II
MODÈLES POUR UN ASSUREUR 55
2.1 Introduction 55
2.1.1 Définition des termes utilisés 56
2.1.2 Analyse empirique 57
2.2 Modèle du nombre de ménages 58
2.2.1 Nouveaux ménages 59
2.2.2 Anciens ménages 60
2.2.3 Total des ménages 61
2.3 Modélisation du nombre de voitures totales 61
2.3.1 Nouveaux ménages 62
2.3.2 Anciens ménages 64
2.3.3 Total des ménages 65
2.4 Inférence 66
2.4.1 Notations 66
2.4.2 fonction de vraisemblance 68
2.4.3 Inclusion de variables explicatives 71
2.5 Analyse des résultats 75
2.5.1 Espérance et variance du nombre de voitures au temps t . . . 76

2.5.2 Valeur à risque du nombre de voitures au temps t 81
2.5.2.1 Méthode des différences finies 82
2.5.2.2 Algorithme de Lattice-Poisson 83
2.5.3 Valeur actuelle des profits futurs 84
2.6 Conclusion 88
CONCLUSION 90
LEXIQUE 94

LISTE DES TABLEAUX

Tableau		Page
1.1	Valeurs des paramètres (écarts-types)	27
1.2	Variables binaires résumant l'information disponible pour chaque ménage	28
1.3	Valeurs des paramètres de régression avec variables explicatives	29
1.4	Caractéristiques des assurés types choisis	30
1.5	Valeur des paramètres pour chaque profil	31
1.6	Nombre de voitures espérées restantes après 5 ans	33
1.7	Valeurs présentes des profits clients pour 2 voitures initiales	35
1.8	Valeur vie client actualisée à 2%	37
2.1	Valeurs des paramètres (écarts-types)	71
2.2	Variables binaires résumant l'information disponible pour chaque ménage	74
2.3	Valeurs des paramètres de régression avec variables explicatives	74
2.4	Valeurs des paramètres pour chaque type d'assurés	75
2.5	Espérance plus ou moins la l'écart-type du nombre de voitures assurées	78
2.6	Proportion du nombre de voitures selon les type de ménage	79
2.7	Valeur à risque du nombre de voitures assurées des nouveaux ménages	84
2.8	Équité client selon les divers profils d'assurés	87

LISTE DES FIGURES

Figure		Page
1.1	Durée de vie d'une police d'assurance, temps avant l'ajout d'une voiture et temps avant le retrait d'une voiture	5
1.2	Graphe d'un processus de Poisson : chaque cercle représente un état et les flèches indiquent les transitions possibles selon le taux associé	7
1.3	Graphe du modèle $M/M/\infty$	8
1.4	Graphe d'un modèle avec annulation de contrat possible	12
1.5	Graphe d'un modèle avec état absorbant 0*	14
1.6	Exemple du nombre de voitures pour un assuré sur 2 ans	20
1.7	Notation de temps appliquée à un exemple	26
1.8	Espérance du nombre de voitures au temps t	33
2.1	Nombre de nouveaux contrats par date d'arrivée	57
2.2	Comparaison du nombre de voitures lors de l'arrivée d'un ménage	72
2.3	Nombre de voitures assurées dans le passé et espérées dans le futur	80

RÉSUMÉ

Dans ce mémoire, nous proposons en premier lieu un modèle du nombre de voitures assurées pour un ménage, puis pour un portefeuille entier dans la deuxième partie. Nous utilisons la théorie des files d'attente pour construire un nouveau modèle qui nécessite 4 paramètres différents : un qui décrit le taux d'ajout de nouvelles voitures sur le contrat d'assurance, un second qui modélise le taux de retrait de voitures assurées, un troisième paramètre qui modélise le taux d'annulation de la police d'assurance et finalement un dernier paramètre pour le taux de renouvellement de contrat. Pour le second modèle de voitures d'un portefeuille, nous réutilisons ces 4 paramètres, et nous y ajoutons un autre paramètre pour le taux d'arrivées de nouveaux contrats. Des méthodes d'inférence statistique sont ensuite utilisées pour estimer chaque paramètre, même dans les cas où il y a censure. Nous proposons même une généralisation de ces processus de files d'attente avec l'ajout de quelques variables explicatives pour chaque paramètre du modèle. Ceci nous permet de déterminer quels profils de ménages sont plus enclins à ajouter ou retirer des voiture, d'annuler leur contrat ou de renouveler leur police. Les paramètres estimés nous aident à analyser en détail le portefeuille d'assurance parce que les modèles de files d'attente nous permettent de calculer plusieurs statistiques utiles à l'assureur, telles que le nombre espéré de voitures. Un autre résultat très important vient de l'actualisation des profits futurs des voitures. Ce résultat s'appelle valeur vie client lorsqu'il s'agit d'un seul assuré, tandis qu'il s'appelle équité client pour tout le portefeuille d'assurés. En utilisant des données d'assurance automobile, une application de ces modèles sur une compagnie d'assurance canadienne est faite pour supporter cette discussion.

MOTS-CLÉS : Théorie des files d'attente, inférence statistique, valeur vie client, équité client.

INTRODUCTION

Un but qu'ont en commun les entreprises de tout genre est d'attirer une clientèle fidèle et profitable. Ainsi, en ciblant ce type de clients, ceux-ci engendreront des flux monétaires stables à l'entreprise. Dans le domaine de l'assurance automobile, les assurés intéressants correspondent à des clients ayant beaucoup de voitures assurées et qui restent assurés pendant une longue période. En ce sens, il est intéressant de modéliser l'évolution du nombre de voitures pour différents types d'assurés. Dans ce mémoire, nous proposons une nouvelle approche qui utilise la théorie des files d'attente pour modéliser le comportement des assurés.

Dans le premier chapitre, une introduction à la théorie des files d'attente sera présentée dans le cadre de modèles pour un seul assuré. En utilisant une base de données réelle pour faire l'estimation des paramètres, plusieurs résultats intéressants seront montrés pour différents types de ménages. Ainsi, il sera possible de quantifier quels types d'assurés sont plus intéressants du point de vue de l'assureur.

En utilisant une combinaison de deux modèles de files d'attente, un modèle pour un portefeuille entier d'assurés est construit dans le deuxième chapitre. Les résultats montrés au premier chapitre sont donc généralisés. Au lieu de modéliser le mouvement des voitures sur une police d'assurance, nous nous intéressons au mouvement de polices dans un portefeuille d'assurance. Sous quelques hypothèses réalistes, la valeur totale du portefeuille d'un assureur pourra être calculée en fonction de variables explicatives choisies.

CHAPITRE I

MODÈLES POUR UN ASSURÉ

1.1 Introduction

Au cours des dernières années, une variable qui a pris de l'intérêt dans le domaine actuariel est la valeur vie client (traduction de *customer lifetime value*, voir Guillén *et al.* (2012), Guelman *et al.* (2015) ou Farris *et al.* (2010) pour un aperçu général). C'est un terme provenant du domaine marketing et qui permet le ciblage à long terme de clients profitables. Dans le domaine de l'assurance, c'est une valeur que l'assureur peut assigner à chaque assuré et qui serait calculée en actualisant tous les profits futurs de cet assuré. Son application au domaine de l'assurance est particulièrement récente (voir Verhoef et Donkers 2001), possiblement à cause de la complexité du domaine actuariel. Pour calculer la valeur vie client de chaque assuré, ainsi que d'autres statistiques utiles, nous proposons un modèle du nombre de voitures assurées pour chaque ménage. Nous utilisons la théorie des files d'attente (voir Newell 1982 pour un survol) pour construire un nouveau modèle qui nécessite 4 paramètres différents : un qui décrit le taux d'ajout de nouvelles voitures sur le contrat d'assurance, un second qui modélise le taux de retrait de voitures assurées, un troisième paramètre qui modélise le taux d'annulation de la police d'assurance et finalement un paramètre qui décrit le taux de renouvellement de contrat. En utilisant des méthodes de régression, nous pouvons

même identifier les profils assurés qui sont plus intéressants pour les assureurs. En utilisant des données de l'assurance automobile, une illustration numérique d'un portefeuille venant d'un assureur canadien est incluse pour supporter cette discussion.

Dans la deuxième partie de ce chapitre, nous proposons une première approche pour modéliser le nombre de voitures assurées. Ceci inclut un processus pour modéliser l'ajout de nouveaux véhicules et un autre processus pour modéliser le retrait de véhicules assurés d'un contrat d'assurance déjà existant. Un troisième paramètre pour modéliser un processus de renouvellement sera ajouté à ce modèle. Dans la section 3, nous généralisons le modèle de la section 2 pour inclure un processus qui modélise les annulations et les non-renouvellements de contrats. Dans la section 4, nous proposons une méthode pour estimer les paramètres du modèle avec des données complètes et pour des données censurées. Des variables explicatives représentant certaines caractéristiques des ménages seront ensuite ajoutées à chaque paramètre du processus. Dans la section 5, nous appliquons le modèle aux données et calculons plusieurs statistiques utiles, comme le nombre espéré de voitures assurées ou la valeur vie client. Les possibles généralisations du modèle sont discutées à la section 6 et la section 7 conclut ce chapitre.

1.1.1 Définition des termes utilisés

Le terme **ménage** est utilisé pour désigner un seul numéro de client ou assuré. Ce ménage peut d'ailleurs comporter plusieurs membres (ou conducteurs) et plusieurs voitures regroupées sous un même **contrat d'assurance**, qui peut être **renouvelé** chaque année. Ce contrat d'assurance représente le document qui lie l'assureur avec le ménage assuré. Dans ce mémoire, nous nous intéressons particulièrement au nombre de voitures que ce contrat couvre et, par la même occasion, détenues

par le même ménage. Par extension, les **ajouts et retraits de voitures** du contrat d'assurance sont aussi analysés. Finalement, à tout moment durant la couverture d'assurance, un ménage peut décider d'annuler son contrat, signifiant que toutes les voitures assurées sont aussi annulées. Nous appelons cet événement une **rupture de contrat** ou **annulation**.

1.1.2 Données utilisées et notations

L'utilisation de la théorie des files d'attente est basée sur les différents temps d'attente avant le changement dans le nombre de voitures assurées. Nous basons notre recherche sur une analyse empirique venant d'une base de données d'une compagnie d'assurance canadienne. Cette base contient les renseignements généraux d'assurance de chacun des 322 174 ménages pour une période allant de 2003 à 2007. Notons que les données sont censurées puisque la période analysée est courte. Pour chaque ménage, nous avons de l'information sur chacune des voitures assurées. Nous avons également de l'information à propos de contrats nouveaux ou brisés, de renouvellement de contrat et d'ajouts ou retraits de voitures. La section 1.4.3 analyse plus en détail cette base de données et décrit plus précisément les données, en particulier les caractéristiques présentes de chacune des polices.

Une analyse graphique de plusieurs temps d'attente associés dans la modélisation est tout d'abord présentée. Premièrement, dans le premier graphique de la figure 1.1, la distribution de la vie d'une police d'assurance (en années) est présentée. Par la vie d'une police d'assurance, nous voulons dire le temps entre la date effective d'un nouveau contrat et la date de non-renouvellement ou la date de rupture de contrat. Pour éviter de possibles biais, nous n'avons utilisé que les polices générées et annulées dans la période de 2003 à 2007. Cet échantillon nous permet de mieux comprendre les données. De cette figure, nous pouvons remarquer qu'il

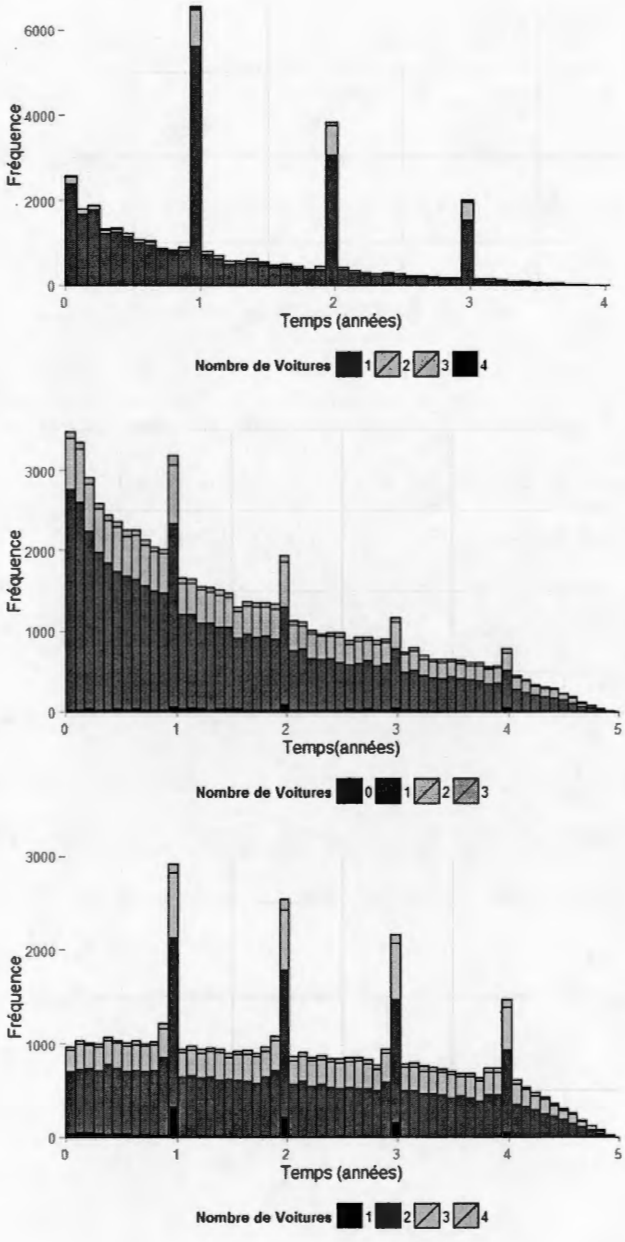

Figure 1.1 Durée de vie d'une police d'assurance, temps avant l'ajout d'une voiture et temps avant le retrait d'une voiture

y a un choc à chaque renouvellement de contrat, c'est-à-dire à tous les ans. Mis à part ce choc, nous pouvons aussi observer une tendance exponentielle décroissante dans les données. Le code de couleur permet d'apercevoir que la grande majorité des départs se font lorsqu'il n'y a qu'une seule voiture assurée sur le contrat. Toujours dans la figure 1.1, le temps avant l'ajout d'une voiture sur un contrat d'assurance déjà existant est présenté. Pour ce graphique, toutes les données ont été utilisées. Le dernier graphique sur la figure 1.1 utilise également toutes les données, et montre le temps avant le retrait d'une voiture sur la police, qui demeure effective malgré le retrait de la voiture. De nouveau, nous pouvons observer une tendance exponentielle et des chocs à chaque renouvellement de contrat, même si ces chocs sont moins importants que ceux observés dans le premier graphique. Le but principal de notre projet est donc de créer un modèle qui serait capable d'approximer ces observations.

Supposons $\mathcal{N}(t)$, une variable aléatoire représentant le nombre d'éléments dans le système de files d'attente au temps t. Dans notre cas, le nombre d'éléments désigne le nombre de voitures assurées pour un ménage spécifique. La fonction de probabilité du nombre de voitures assurées sera exprimée comme $\Pr\{\mathcal{N}(t) = i\} = p_i(t)$qui signifie que la probabilité qu'il y ait i voitures au temps t est $p_i(t)$. Pour résoudre certains problèmes ou pour simplifier la représentation, la fonction génératrice des probabilités (FGP) est utilisée. Dans le cas de la variable aléatoire $N(t)$, elle est définie comme : $P_{\mathcal{N}(t)}(z,t) = \sum_{i=0}^{\infty} p_i(t) \times z^i$ et sa dérivée partielle par rapport à t est :

$$\frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial t}(z,t) = \sum_{i=0}^{\infty} \frac{dp_i(t)}{dt} \times z^i \tag{1.1}$$

Finalement, pour représenter les probabilités conditionnelles, nous avons que :

$$p_{(j,i)}(s,t) = \Pr\{\mathcal{N}(t) = i | \mathcal{N}(s) = j\}.$$

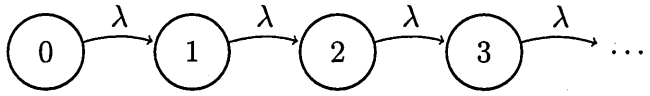

Figure 1.2 Graphe d'un processus de Poisson : chaque cercle représente un état et les flèches indiquent les transitions possibles selon le taux associé

1.2 Modélisation du nombre de véhicules

Dans cette section, nous introduisons comment la théorie des files d'attente, basée sur Newell (1982), peut être utilisée pour modéliser le nombre de voitures assurées. Nous utilisons tout d'abord un processus de Poisson pour modéliser l'ajout de nouveaux véhicules, puis nous y ajoutons un autre processus pour modéliser le retrait de voitures. Cette introduction à la théorie des files d'attente nous permet d'expliquer certains outils qui seront utilisés dans les modèles plus complexes de la troisième section.

1.2.1 Ajouts et retraits de véhicule

Dans un processus de naissance pur, aussi appelé processus de Poisson et illustré à la figure 1.2, il n'y a seulement qu'une composante d'arrivée, définie par un paramètre λ. Il s'agit donc de la modélisation d'une population qui grandit indéfiniment au taux constant λ. Dans notre cas étudié, la population est le nombre de voitures sur une police d'assurance qui s'y ajoute selon ce taux. Les équations de Chapman-Kolmogorov, définies dans notre contexte par :

$$p_i(t) = \sum_{j=0}^{\infty} p_j(s) p_{(j,i)}(s,t),$$

pour $s < t$, peuvent être déduites. Nous interprétons cette équation par le fait qu'une probabilité peut être définie par la somme des parcours différents pour un certain temps intermédiaire. Ces équations nécessitent ainsi de trouver les

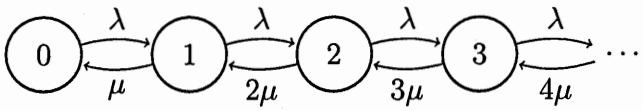

Figure 1.3 Graphe du modèle $M/M/\infty$

probabilités conditionnelles du système soit à un certain niveau i au temps $t+\Delta t$ sachant que le système était au niveau j au temps t. Ceci s'exprime comme :

$$p_{(j,i)}(t, t+\Delta t) = \begin{cases} \lambda\Delta t + o(\Delta t) & \text{quand } j = i-1, \\ 1 - \lambda\Delta t + o(\Delta t) & \text{quand } j = i, \\ o(\Delta t) & \text{quand } j < i-1, \\ 0 & \text{sinon,} \end{cases} \tag{1.2}$$

où $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$. Pour ce modèle introductif, il est montré à l'annexe 1.8.1 que dans ce cas, les probabilités individuelles peuvent s'exprimer comme :

$$p_i(t) = \frac{e^{-\lambda t}(\lambda t)^i}{i!}, \tag{1.3}$$

qui représente la distribution de Poisson classique de paramètre λt.

Afin d'obtenir un modèle qui est plus réaliste, nous ajoutons une composante de service (qu'on peut aussi appeler une composante de mort) au processus de naissance pur. Par simplification, nous supposons l'indépendance entre cette nouvelle composante de mort et le processus de naissance. Ceci est illustré à la figure 1.3 et le modèle résultant est appelé $M/M/\infty$ dans la notation habituelle. Comme le processus de Poisson, le temps entre chaque ajout d'un nouvel élément suit une distribution exponentielle. Cette caractéristique est représentée par le premier M de l'acronyme, signifiant arrivées "markoviennes". Pour ce modèle, chaque élément quitte le système après un certain de temps qui suit lui aussi une distribution exponentielle. Cet ajout dans le processus définit le deuxième M. Finalement, le

symbole ∞ signifie que le processus de départ de chaque voiture dans le ménage peut s'effectuer sans attendre la fin du processus de départ d'un autre véhicule.

En d'autres termes, il s'agit donc de la modélisation dynamique d'une population qui grandit à un taux constant λ, et qui meurt à un taux $i\mu$, avec i représentant le nombre d'éléments de cette population à un temps précis. Dans notre contexte, la population étudiée est le nombre de voitures du ménage et donc, de nouvelles automobiles sont ajoutées à la police selon un taux λ, alors que les véhicules quittent cette même police selon un taux individuel de μ. Pour ce modèle, les probabilités conditionnelles sont :

$$p_{(j,i)}(t, t+\Delta t) = \begin{cases} \lambda\Delta t + o(\Delta t) & \text{quand } j = i-1, \\ 1-(\lambda+i\mu)\Delta t + o(\Delta t) & \text{quand } j = i, \\ (i+1)\mu\Delta t + o(\Delta t) & \text{quand } j = i+1, \\ o(\Delta t) & \text{sinon.} \end{cases}$$

Comparativement au processus de Poisson, certaines différences peuvent être identifiées. Tout d'abord, à un niveau i, chacun des i éléments se fait servir selon un taux μ. Cette nouvelle possibilité peut être insérée à deux endroits. Le premier réduit de $(i\mu\Delta t)$ la probabilité de demeurer au même niveau. La deuxième ajoute $(i+1)\mu\Delta t$ à la probabilité de passer du niveau $(i+1)$ à i. Pour ce modèle bien connu dans la théorie des files d'attente, il est montré à l'annexe 1.8.2 que la FGP au temps t propre à ce processus peut s'exprimer comme :

$$P_{\mathcal{N}(t)}(z,t) \;=\; \left[(z-1)e^{-\mu t}+1\right]^a e^{\frac{\lambda}{\mu}(1-e^{-\mu t})(z-1)}, \tag{1.4}$$

où $\mathcal{N}(0) = a$, c'est-à-dire que le nombre de voitures initial (au temps 0) est a.

De cette FGP, il est possible d'en déduire une fonction de probabilité en décomposant le processus, également montrée à l'annexe 1.8.2. Ainsi, la fonction de

probabilité désirée peut être retrouvée par convolution :

$$p_i(t) = \sum_{k=0}^{\min(i,a)} \frac{e^{-(1-e^{-\mu t})\lambda/\mu}\left[(1-e^{-\mu t})\lambda/\mu\right]^{i-k}}{(i-k)!} \times \binom{a}{k} e^{-\mu t k}(1-e^{-\mu t})^{a-k}. \tag{1.5}$$

1.2.2 Renouvellement des véhicules

Il est possible de donner davantage de flexibilité au modèle $M/M/\infty$ par la généralisation de la composante du taux de service. Une telle généralisation signifie que le deuxième M de l'acronyme devrait être remplacé par un G (signifiant distribution générale). Cette généralisation permet d'incorporer un choc à chaque année pour chacune des voitures. En effet, empiriquement et comme nous le montre la figure 1.1, il y a une forte probabilité qu'une voiture quitte le système à son renouvellement annuel.

En se fiant à Benes (1957a), la construction de ce modèle doit être effectuée en séparant en plusieurs composantes le processus modélisant le nombre de voitures. En effet, le processus $\mathcal{N}(t)$ peut se décomposer comme $\mathcal{N}(t) = \mathcal{Z}(t) - \mathcal{Y}(t)$. Dans ce cas, le processus $\mathcal{Z}(t)$ compte le nombre d'ajouts de nouvelles voitures dans le système jusqu'au temps t, alors que le processus $\mathcal{Y}(t)$ compte le nombre de retraits de voitures du contrat d'assurance jusqu'au temps t.

Par le premier M de l'acronyme du système $M/G/\infty$, nous savons que $\mathcal{Z}(t) \sim Poisson(\lambda t)$, indiquant que les arrivées surviennent selon un processus de Poisson pur comme défini plus tôt.

Ensuite, pour construire le modèle, deux variables aléatoires sont utiles pour le développement. Ainsi, supposons I_j, une variable aléatoire modélisant le temps où la voiture j a été ajoutée, et la variable aléatoire Q_j définissant le temps avant le retrait de la voiture j. Nous supposons également que la condition initiale du

processus est $N(t=0)=0$.

À l'aide de ces nouvelles variables aléatoires, conditionnellement à la survenance d'une seule arrivée de voiture, la probabilité que cette voiture soit toujours présente dans le ménage peut être calculée à l'aide du théorème des probabilités totales :

$$\Pr(\mathcal{N}(t)=i)=\sum_{j=i}^{\infty}\Pr(\mathcal{N}(t)=i|\mathcal{Z}(t)=j)\times\Pr(\mathcal{Z}(t)=j). \tag{1.6}$$

Ainsi, pour $i=j=1$, nous avons :

$$\begin{aligned}\Pr(\mathcal{N}(t)=1|\mathcal{Z}(t)=1) &= \int_0^t \Pr(Q_1>t-x|I_1=x)\times\Pr(I_1=x)dx\\ &= \int_0^t \Pr(Q_1>t-x|I_1=x)\frac{1}{t}dx\\ &= \int_0^t \frac{S(t-x)}{t}dx\\ &= \int_0^t \frac{S(x)}{t}dx\\ &\equiv q_t,\end{aligned}$$

où $S(\cdot)$ est la fonction de survie du temps de service, qui correspond ici à la période assurée de la voiture. Par les propriétés de la distribution de Poisson, le moment où l'occurrence I_j se produit dans un intervalle de temps t, conditionnellement au fait que cet événement se produira, est uniformément distribué sur $(0,t)$.

En utilisant la distribution binomiale et l'indépendance entre les arrivées, nous pouvons généraliser ce résultat pour i arrivées de sorte que : $\mathcal{N}(t)|\mathcal{Z}(t)=j\sim$ Binomiale(j,q_t). Ainsi, nous pouvons finalement résoudre l'équation (1.6) :

$$\begin{aligned}\Pr(\mathcal{N}(t)=i) &= \sum_{j=i}^{\infty}\binom{j}{i}q_t^i(1-q_t)^{j-i}\frac{e^{-\lambda t}(\lambda t)^j}{j!}\\ &= \frac{(\lambda t q_t)^i e^{-\lambda t}}{i!}\sum_{j=i}^{\infty}\frac{[\lambda t(1-q_t)]^{j-i}}{(j-i)!}\\ &= \frac{(\lambda t q_t)^i e^{-q_t\lambda t}}{i!}.\end{aligned} \tag{1.7}$$

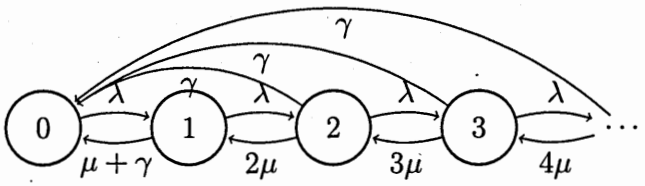

Figure 1.4 Graphe d'un modèle avec annulation de contrat possible

Cette fonction de probabilité peut être reconnue comme étant celle d'une distribution de Poisson de paramètre $\lambda t q_t$. Similairement au développement de l'équation (1.5), il est aussi possible de trouver la fonction de probabilité d'un ménage ayant initialement a véhicules. À noter qu'il est possible de généraliser la fonction de survie de plusieurs manières. Par exemple, nous pourrions inclure un choc qui se produirait à tous les anniversaires de l'ajout de la voiture en utilisant la fonction de survie suivante :

$$S(t) = e^{-\mu t} p^{\lfloor t \rfloor}, \tag{1.8}$$

avec p représentant la probabilité de renouvellement de la voiture et où $\lfloor \cdot \rfloor$ est la fonction "partie entière". Ainsi, lorsque cette fonction de survie est choisie, le paramètre q_t est égal à :

$$q_t = \left(1 - e^{-\gamma}\right) \frac{1 - \left(pe^{-\gamma}\right)^{\lfloor t \rfloor}}{\gamma t \left(1 - pe^{-\gamma}\right)} + \frac{p^{\lfloor t \rfloor} \left(e^{-\gamma \lfloor t \rfloor} - e^{-\gamma t}\right)}{\gamma t}. \tag{1.9}$$

Il est à noter qu'en utilisant cette équation avec un taux de renouvellement p égal à 1, nous retournons à un modèle $M/M/\infty$.

1.3 Modèle avec période d'inactivité

Il serait intéressant de généraliser le modèle pour inclure une possible annulation de contrat, où l'ensemble des véhicules assurés quitte la compagnie simultanément. Ce modèle serait construit par l'ajout d'un processus poissonnien permettant de

passer des niveaux $n \geq 1$ à 0 selon un taux γ, recréant ainsi l'effet d'une annulation de police. La figure 1.4 illustre ce modèle, et le système d'équations est développé à l'annexe 1.8.3. En étudiant ce nouveau modèle, nous voyons toutefois que ce modèle engendre une situation particulière qui pose problème. En effet, ce modèle permet la possibilité qu'un nouveau véhicule soit ajouté à un contrat d'assurance, même si ce contrat n'a pas eu de voitures assurées (état 0) pendant un certain temps. Cette situation est contre-intuitive : il nous semble plus logique de croire qu'une police ayant un nombre de véhicules à 0 est simplement annulée. En ce sens, l'état 0 du modèle devrait être absorbant.

Toutefois, une analyse empirique plus approfondie des données nous a révélé une situation un peu plus complexe. En effet, il arrive qu'une police d'assurance retirant son dernier véhicule ne soit pas annulée. Cela signifie que parfois un nouveau véhicule est ajouté à un contrat d'assurance déjà existant sans voitures assurées. Dans nos données, ceci est survenu 1340 fois, et le temps moyen passé dans cet état 0 est presque d'une demi-année (0.4789).

Ceci représente une forme d'inactivité du contrat, qui permet la réactivation du contrat quand un nouveau véhicule est ajouté au contrat. Nous pensons à des situations comme l'entreposage du dernier véhicule de la police ou la présence d'un autre type d'assurance (habitation ou autres) associée à la police automobile prévenant l'annulation complète de la police. Il peut d'ailleurs avoir plusieurs autres explications spécifiques générant cette situation. En ce sens, empiriquement, nous ne pouvons pas considérer l'état 0 du modèle comme absorbant. Toutefois, la présence d'un état absorbant désignant le départ final et définitif de l'assuré doit être considérée. Cet état n'est simplement pas celui correspondant à un contrat sans aucune voiture assurée, mais un état légèrement différent.

Nous proposons ainsi un modèle ayant une forme illustrée à la figure 1.5. Dans

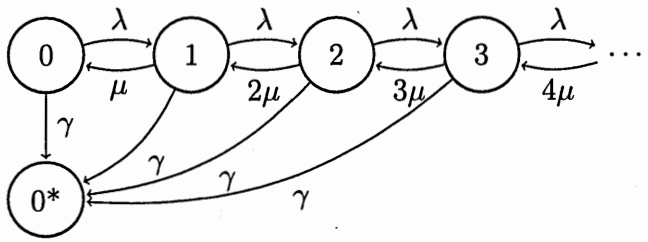

Figure 1.5 Graphe d'un modèle avec état absorbant 0*

ce nouveau modèle, un nouvel état est créé et est identifié par 0*. Ce nouveau niveau est un état absorbant ne permettant plus au ménage d'assurer de nouveaux véhicules[1]. En d'autres mots, l'état 0* n'est accessible que dans la situation où le contrat est annulé. Il est à noter que ce passage vers 0* s'effectue à un taux γ pour tous les niveaux. Bien que nous explorerons cette situation plus en détail à la section 1.4, nous aimerions mettre en évidence une difficulté conceptuelle de ce nouveau modèle. Dans le cas d'une police n'ayant qu'un véhicule, empiriquement, l'annulation de ce dernier véhicule est similaire à l'annulation du contrat. En ce sens, il n'est pas évident de savoir si la police vient d'être annulée (et passe à l'état absorbant 0*), ou si la police ne se situe que dans une période d'inactivité (état 0). Nous proposerons une solution à ce problème plus tard.

La construction de ce nouveau modèle est complexe et nécessite l'ajout de notations. Ainsi, en plus de $\mathcal{N}(t)$, correspondant au nombre de voitures assurés au temps t, la variable aléatoire binaire $\mathcal{M}(t)$ est introduite. Cette variable définit la vie ou mort du processus au temps t. Dans notre contexte, un processus est en vie au temps t, ou $\mathcal{M}(t) = 1$, si le contrat d'assurance est actif, ou en d'autres mots, si le contrat n'a pas été brisé. Un processus en vie n'ayant aucun véhicule assuré est simplement considéré inactif, et à l'état 0. À l'inverse, un processus mort, ou

1. Si l'assuré brise son contrat et veut plus tard retourner s'assurer, les assureurs vont souvent créer un nouveau contrat d'assurance.

$\mathcal{M}(t) = 0$, signifie que le ménage a annulé son contrat d'assurance et que l'état absorbant 0^* a été atteint.

1.3.1 Modélisation

Pour ce modèle les équations de Chapman-Kolmogorov auraient pu être utilisées pour trouver les probabilités de transition, desquelles les équations de différence-différentielles auraient pu être trouvées. Nous prendrons à la place un développement mathématique plus simple. Nous supposons premièrement que le processus temporel relié à la variable aléatoire $\mathcal{M}(t)$ est distribuée exponentiellement et indépendant du processus relié à $\mathcal{N}(t)$. En d'autres mots, nous faisons l'hypothèse que le temps entre le début du contrat et la fin du contrat est $U \sim \text{Exponentiel}(\gamma)$. Conséquemment, nous avons $\mathcal{M}(t) = 0$ lorsque $t > U$ et :

$$\Pr(\mathcal{M}(t) = 1) = e^{-\gamma t}. \tag{1.10}$$

Nous pouvons définir le nombre de voitures assurées pour les deux situations :

1. Le cas $\mathcal{M}(t) = 1$: Tant que le ménage est assuré, les processus d'ajout et de retrait de véhicules sont actifs. Ainsi, nous pouvons modéliser le nombre de voitures assurées $\mathcal{N}(t)$ avec le modèle de files d'attente $M/M/\infty$ défini vu précédemment.
2. Le cas $\mathcal{M}(t) = 0$: Quand l'annulation de contrat survient, le nombre de voitures assurées est 0. Nous avons donc $\Pr(\mathcal{N}(t) = 0 | M(t) = 0) = 1$.

Ainsi, la FGP jointe de $\mathcal{N}(t)$ et $\mathcal{M}(t)$ peut facilement être développée :

$$\begin{aligned}
P_{\mathcal{N}(t),\mathcal{M}(t)}(z,y,t) &= \sum_{m=0}^{1}\sum_{i=0}^{\infty}\Pr(\mathcal{M}(t)=m,\mathcal{N}(t)=i)y^m z^i \\
&= \sum_{m=0}^{1}\sum_{i=0}^{\infty}\Pr(\mathcal{N}(t)=i|\mathcal{M}(t)=m) \\
&\quad \times \Pr(\mathcal{M}(t)=m)y^m z^i \\
&= \Pr(\mathcal{M}(t)=0)+\sum_{i=0}^{\infty}\Pr(\mathcal{N}(t)=i|\mathcal{M}(t)=1) \\
&\quad \times \Pr(\mathcal{M}(t)=1)yz^i \\
&= 1-e^{-\gamma t}+e^{-\gamma t}yP_{\mathcal{N}(t)}(z,t) \\
&= 1-e^{-\gamma t}+e^{-\gamma t}y\left((z-1)e^{-\mu t}+1\right)^a \\
&\quad \times e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}. \qquad (1.11)
\end{aligned}$$

(1.12)

La fonction génératrice des probabilités du modèle $M/M/\infty$ peut être reconnue de la partie droite de l'équation, de laquelle l'équation 1.10 a été ajoutée.

1.3.1.1 Ajout du choc au renouvellement

À chaque anniversaire de la police, le ménage assuré doit renouveler son contrat d'assurance. Il est possible de modifier la fonction génératrice de probabilité pour y ajouter une épreuve de Bernoulli qui reflète ce comportement. Pour ce faire, l'équation (1.10) est modifiée pour y insérer une probabilité de renouvellement p :

$$\Pr(\mathcal{M}(t)=1)=e^{-\gamma t}p^{\lfloor t \rfloor}, \qquad (1.13)$$

où la variable t représente le temps en années.

À chaque anniversaire de la police, pour $p < 1$, cette fonction engendre une augmentation de la probabilité d'annulation. Il est facile de réécrire l'équation

1.11 avec cette nouvelle définition pour avoir le résultat :

$$\begin{aligned} P_{\mathcal{N}(t),\mathcal{M}(t)}(z,y,t) &= 1 - e^{-\gamma t} p^{\lfloor t \rfloor} + e^{-\gamma t} p^{\lfloor t \rfloor} \\ &\quad \times y \left((z-1)e^{-\mu t} + 1 \right)^{a} e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}, \qquad (1.14) \\ & \qquad (1.15) \end{aligned}$$

qui représente le processus final que nous proposons pour modéliser le nombre de voitures d'un seul ménage.

1.4 Inférence statistique

Il est possible d'estimer les paramètres λ, μ, γ et p du modèle défini par l'équation (1.14) en utilisant le maximum de vraisemblance. Cette technique d'estimation nécessite de trouver la densité de probabilités jointes et de l'ajuster en fonction des observations de la base de données. Certaines hypothèses doivent être posées afin de pouvoir utiliser les informations disponibles de la base de données. Nous résumerons tout d'abord les étapes d'estimation des paramètres d'un modèle simplifié où il est possible de distinguer chacun des événements. Ensuite, en introduisant de la censure et des variables explicatives, le modèle sera adapté et généralisé.

1.4.1 Données complètes

1.4.1.1 Notations

Pour estimer tous les paramètres propres du nouveau modèle, une liste des variables utilisées dans la fonction de vraisemblance doit être présentée. Dans un premier groupe, tous les événements observés pendant la vie de la police affectant directement le nombre de voitures assurés seront notés comme :

1. Événement de type 1, qui représente l'ajout d'une voiture au contrat d'as-

surance. La variable aléatoire E représente le nombre total des événements de ce type (excluant celles présentes au départ du contrat) ;

2. Événement de type 2, qui représente le retrait d'une voiture d'un contrat d'assurance. La variable aléatoire S représente le nombre total d'événements de ce type ;
3. Événement de type 3, qui représente l'annulation de la police d'assurance à un moment différent de l'anniversaire de la police. La variable aléatoire A, est une variable indicatrice pour ce type d'événement ;
4. Événement de type 4, qui représente l'annulation de la police d'assurance à l'anniversaire de la police. La variable aléatoire Q est une variable indicatrice pour ce type d'événement. En d'autres mots, $Q = 1$ s'il n'y a pas eu renouvellement de contrat.

Ainsi, nous obtenons que l'ensemble des événements K est égal à la somme des éléments précédents, telle que $K = E + A + Q + S$. Nous pouvons aussi noter les variables révélant les informations sur le temps du contrat d'assurance comme :

1. t_j, le temps (en années) de survenance du j^{e} événement affectant le nombre de voitures ;
2. $\widetilde{t_j}$ la durée de temps (en années) entre le $(j-1)^{\text{e}}$ et le j^{e} événement tel que $\widetilde{t_j} = t_j - t_{j-1}$, avec $\widetilde{t_1} = t_1$;
3. T, le nombre d'années où le processus était en vie ou inactif, tel que $T = \sum_{j=1}^{K} \widetilde{t_j}$ et donc $T = t_K$;
4. Finalement, définissons respectivement par Ψ_1, Ψ_2, Ψ_3 et Ψ_4 les variables qui définissent le temps avant la survenance d'un événement de type 1, 2, 3 ou 4.

Aussi, nous avons :

1. N_j, une variable qui compte le nombre de voitures immédiatement avant le j^e événement et donc N_1 est le nombre de voitures au début ;
2. V, la somme du temps d'exposition (en années) de toutes les voitures d'un ménage de sorte que $V = \int_0^T N(t)dt = \sum^K \widetilde{t}_j N_j$.

1.4.1.2 Fonction de vraisemblance

Les fonctions de vraisemblance sont développées à partir des travaux de Benes (1957b) sur le modèle $M/M/\infty$. Pour estimer les paramètres, nous construirons la fonction de vraisemblance conditionnelle à l'état de départ. En ce sens, nous ne nous intéressons pas au nombre de voitures déjà assurées dans le portefeuille (nous travaillons avec cette hypothèse initiale), mais plutôt aux processus qui affectent le nombre de voitures dans le futur. Selon les hypothèses supposées dans la conception du système, ces variables ont été définies telles que $\Psi_1 \sim$ Exponentielle$(\lambda), \Psi_2 \sim$ Exponentielle$(\mu), \Psi_3 \sim$ Exponentielle(γ) et $\Psi_4 \sim$ Géométrique(p).

Pour définir la distribution jointe de tous les événements d'une police, il suffit d'analyser tous les cas possibles. Tous les processus sont supposés indépendants, ce qui facilite la décomposition du modèle. Par exemple, sachant que le premier événement est un ajout de véhicule observé au temps $\widetilde{t}_1$, nous obtenons :

$$\Pr(\Psi_1 = \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1) = \lambda e^{-\lambda \widetilde{t}_1} e^{-N_1 \mu \widetilde{t}_1} e^{-\gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)}, \qquad (1.16)$$

où $h(x)$ est une fonction qui compte le nombre de fois que le ménage s'est retrouvé dans une possibilité de renouvellement dans l'intervalle de temps x. L'équation a été développée en utilisant les propriétés du premier événement d'une distribution conjointe d'exponentielles. Ensuite, par un cheminement analogue, il est possible de trouver les probabilités que le premier événement soit respectivement une sortie

Figure 1.6 Exemple du nombre de voitures pour un assuré sur 2 ans

de véhicule, un arrêt de contrat ou un choc au renouvellement [2] :

$$\Pr(\Psi_1 > \tilde{t}_1, \Psi_2 = \tilde{t}_1, \Psi_3 > \tilde{t}_1, \Psi_4 > \tilde{t}_1) = e^{-\lambda\tilde{t}_1} N_1 \mu e^{-N_1\mu\tilde{t}_1} e^{-\gamma\tilde{t}_1} p^{h(\tilde{t}_1)} \quad (1.17)$$

$$\Pr(\Psi_1 > \tilde{t}_1, \Psi_2 > \tilde{t}_1, \Psi_3 = \tilde{t}_1, \Psi_4 > \tilde{t}_1) = e^{-\lambda\tilde{t}_1} e^{-N_1\mu\tilde{t}_1} \gamma e^{-\gamma\tilde{t}_1} p^{h(\tilde{t}_1)} \quad (1.18)$$

$$\Pr(\Psi_1 > \tilde{t}_1, \Psi_2 > \tilde{t}_1, \Psi_3 > \tilde{t}_1, \Psi_4 = \tilde{t}_1) = e^{-\lambda\tilde{t}_1} e^{-N_1\mu\tilde{t}_1} e^{-\gamma\tilde{t}_1} p^{h(\tilde{t}_1)-1}(1-p) \quad (1.19)$$

Par la suite, sachant que les processus exponentiels utilisés sont sans mémoire, il est possible de calculer le produit des K événements pour obtenir la distribution jointe des événements d'un seul assuré, ou fonction de vraisemblance :

$$\begin{aligned} & f(E, A, S, Q, T, V | \lambda, \gamma, \mu, p) \\ = \; & \lambda^E e^{-\lambda \sum^K \tilde{t}_j} \left(\prod^S N_s \right) \mu^S e^{-\mu \sum^K N_j \tilde{t}_j} \gamma^A e^{-\gamma \sum^K \tilde{t}_j} (1-p)^Q p^{\sum^K h(\tilde{t}_j) - Q} \\ \propto \; & \lambda^E \gamma^A \mu^S (1-p)^Q e^{-(\lambda+\gamma)T} e^{-\mu V} p^{\lfloor T \rfloor - Q}, \end{aligned} \quad (1.20)$$

où les termes constants ont été enlevés de la seconde équation puisqu'ils ne sont pas utilisés dans le calcul du maximum de vraisemblance.

---

2. Cette technique est légèrement biaisée par le fait qu'elle ne considère pas la possibilité d'annulation à la journée d'anniversaire de la police. Néanmoins, l'impact est minimal puisqu'il correspond approximativement à 1/365.

Pour visualiser un exemple de la fonction de vraisemblance sur $T = 2$ ans, supposons un ménage assuré qui passe de 3 à 4 voitures après 9 mois de couverture, puis ne renouvelle pas son contrat à la fin de sa 2e année d'assurance. Il est possible de visualiser cet exemple sur la figure 1.6.

Pour le premier événement, il s'agit de compter un ajout de voiture tandis que pour le deuxième événement, il s'agit d'un non-renouvellement. Ces deux phénomènes peuvent être vus séparément et être développés comme :

$$\begin{aligned} & \mathcal{L}(E=0, A=0, S=1, Q=1, T=2, V=7.25|\lambda, \gamma, \mu, p) \\ \propto\ & \underbrace{\left(\lambda e^{-0.75(\lambda+\gamma)} e^{-3\cdot 0.75\mu}\right)}_{\text{Premier événement}} \times \underbrace{\left((1-p) e^{-1.25(\lambda+\gamma)} e^{-4\cdot 1.25\mu} p\right)}_{\text{Deuxième événement}} \\ \propto\ & \lambda(1-p) e^{-2(\lambda+\gamma)} e^{-7.25\mu} p. \end{aligned}$$

Maintenant, la fonction de logvraisemblance partielle du total des ξ ménages de la base de données peut être exprimée comme :

$$\ell = \sum_{i=1}^{\xi} E_i \ln\lambda + A_i \ln\gamma + S_i \ln\mu + Q_i \ln(1-p) + (\lfloor T_i \rfloor - Q_i) \ln p - (\lambda+\gamma) T_i - \mu V_i, \tag{1.21}$$

où l'indice i identifie chacun des ménages. Finalement, en maximisant la fonction de logvraisemblance pour chacun des paramètres, il est possible d'arriver aux estimateurs suivants :

$$\hat{\lambda} = \frac{\sum^{\xi} E_i}{\sum^{\xi} T_i}, \tag{1.22}$$

$$\hat{\gamma} = \frac{\sum^{\xi} A_i}{\sum^{\xi} T_i}, \tag{1.23}$$

$$\hat{\mu} = \frac{\sum^{\xi} S_i}{\sum^{\xi} V_i}, \tag{1.24}$$

$$\hat{p} = \frac{\sum^{\xi} \lfloor T_i \rfloor - Q_i}{\sum^{\xi} \lfloor T_i \rfloor}. \tag{1.25}$$

1.4.2 Données incomplètes

Puisque l'horizon de temps de l'échantillon n'est que de 5 ans, il n'est toujours possible d'observer la vie entière d'une police. Donc, les données avec lesquelles nous travaillons sont nécessairement censurées et il n'est pas possible d'utiliser les techniques d'estimation basées sur les données complètes. Pour estimer les paramètres correctement, d'importantes hypothèses doivent être faites. En ce sens, notons une nouvelle variable Ω, qui représente la date d'observation de la base de données. Par exemple, dans le cas de notre base de données, cette date est le 31 décembre 2007.

Ainsi, nous n'avons pas les informations d'une police qui n'avait pas été annulée avant cette date. Peut-être que cette police a été annulée en 2008, ou est encore active en 2014. Pour ces contrats, nous devons ajouter la fonction de vraisemblance pour la période de temps entre le dernier événement observé K et la date d'observation Ω, c'est-à-dire une fonction de probabilité qui indique qu'il n'y a pas eu d'événement dans cet intervalle de temps. Ceci correspond à la probabilité jointe de survie de 4 événements. Il est donc permis de dire, par indépendance des variables aléatoires, que :

$$\begin{aligned}
&\Pr(\Psi_1 > \Omega - t_K, \Psi_2 > \Omega - t_K, \Psi_3 > \Omega - t_K, \rho > \Psi_4 - t_K) \\
&= \Pr(\Psi_1 > \Omega - t_K) \Pr(\Psi_2 > \Omega - t_K) \Pr(\Psi_3 > \Omega - t_K) \Pr(\Psi_4 > \Omega - t_K) \\
&= e^{-\lambda(\Omega - t_K)} e^{-N_K \mu(\Omega - t_K)} e^{-\gamma(\Omega - t_K)} p^{h(\Omega - t_K)}. \qquad (1.26)
\end{aligned}$$

1.4.2.1 Inactif vs annulé

Lorsqu'il y a censure, il faut distinguer les cas d'inactivité du contrat d'assurance (état 0) et les cas d'annulation de police (état 0*). En effet, dans les cas où un seul véhicule était assuré au temps t_K, c'est-à-dire si $N_K = 1$, il est impossible de distinguer un événement 2 d'un événement 3. Identifions cette situation par la

variable $AS(\Omega - t_{K^*}) = AS = 1$, signifiant que cet état inconnu a une durée de $\Omega - t_K$. Puisque nous travaillons avec des variables aléatoires exponentielles, il est possible d'additionner les équations (1.17) et (1.18). Nous obtenons la probabilité que le ménage entre dans cet état inconnu au temps t_K. Cette probabilité s'exprime comme $e^{-(\lambda+\mu+\gamma)\tilde{t}_K} p^{h(\tilde{t}_K)}(\mu + \gamma)$, que l'on peut incorporer dans la fonction de vraisemblance.

Avec les propriétés des distributions exponentielles, les probabilités que l'assuré soit à l'état 0* (événement de type 3) ou à l'état 0 (événement de type 2)sont respectivement de :

$$\Pr(\mathcal{M}(t_K) = 0 | AS = 1) = \frac{\gamma}{\gamma + \mu} \tag{1.27}$$

$$\Pr(\mathcal{M}(t_K) = 1, \mathcal{N}(t_K) = 0 | AS = 1) = \frac{\mu}{\gamma + \mu}. \tag{1.28}$$

Il est possible d'inclure le temps $\Omega - t_K$, la période passée à l'état inconnu, dans la fonction de vraisemblance afin de savoir quel état entre 2 et 3 est le plus probable.

Dans une première situation où aucune possibilité de renouvellement n'est possible entre t_K et Ω, 3 situations sont possibles au temps t_K :

1. Un événement de type 3 est survenu à t_K, ce qui implique que la police est annulée ;
2. Un événement de type 2 est survenu à t_K et aucun autre événement n'est survenu jusqu'à Ω. Ceci implique que la police d'assurance est inactive ;
3. Un événement 4 est survenu à t_K et un événement de type 3 survient pendant la période (t_K, Ω), annulant ainsi la police.

Au lieu de calculer chacune de ces probabilités, il est plus simple de calculer la probabilité complémentaire. Cette probabilité correspond à un seul événement : l'occurrence d'un événement de type 2 au temps t_K suivie d'un ajout de voiture (type 1) durant la période (t_K, Ω). Cette probabilité peut être calculée de façon

suivante :

$$\begin{aligned} & \Pr(\mathcal{M}(t_K) = 1, \mathcal{N}(t_K) = 0 | AS = 1) \times \Pr(\Psi_1 < \min(\Psi_3, \Omega - t_K)) \\ = \; & \frac{\mu}{\gamma + \mu} \times \left(1 - e^{-(\gamma+\lambda)(\Omega - t_K)}\right) \frac{\lambda}{\lambda + \gamma}. \end{aligned} \tag{1.29}$$

Cette équation correspond au produit de trois probabilités : la probabilité qu'un événement de type 2 survienne au temps t_K, la probabilité qu'un événement de type 1 ou 3 survienne avant Ω et la probabilité que ce dernier événement soit spécifiquement un ajout de voiture (événement de type 1). Il est important de noter que cette équation ne s'utilise que dans le cas où il n'y a pas de renouvellement possible dans la période inconnue. Il nous faut donc généraliser cette équation afin d'inclure chacun des renouvellements de contrat dans le calcul de la probabilité de rester dans l'état inconnu. Pour ce faire, il est nécessaire de séparer les probabilités d'une arrivée de véhicule selon chacune des années.

Mais avant, plusieurs éléments doivent être expliqués en détail. D'une part, il faut noter que le temps t_K, correspondant au moment où le statut de l'assuré devient inconnu, ne correspond pas au moment du renouvellement de police. Cela signifie aussi que la première année avant le premier renouvellement ; ainsi que la dernière année avant Ω (si elles sont différentes), ne sont que partielles, et donc inférieures à un an.

Ainsi, s'il y a au moins un renouvellement lors de la période inconnue $\Omega - t_K$, un moyen d'écrire la partie de la première année inconnue est $1 - \{t_K\}$, où $\{\cdot\}$ est la fonction de partie décimale. Similairement, la dernière année avant Ω peut être écrite comme $\{\Omega\}$ et le nombre renouvellements effectués dans cette période comme $\lfloor \Omega \rfloor - \lfloor t_K \rfloor = R$. Nous avons alors :

$$\begin{aligned} & \underbrace{\Pr(\Psi_1 < \min(\Psi_3, 1 - \{t_K\}))}_{\text{Avant le premier renouvellement après } t_K} \\ + \; & \underbrace{\Pr(\min(\Psi_1, \Psi_3, \Psi_4) > 1 - \{t_K\}) \times \Pr(\Psi_1 < \min(\Psi_3, 2 - \{t_K\}) | \min(\Psi_1, \Psi_3, \Psi_4) > 1 - \{t_K\})}_{\text{Entre le premier et le second renouvellement après } t_K} \\ + \; & \ldots \end{aligned} \tag{1.30}$$

Cette équation peut être expliquée comme la somme des différentes possibilités d'occurrence entre chaque renouvellement. Le premier élément est simplement la probabilité qu'un événement de type 1 survienne avant le premier renouvellement après t_K. Le second élément de la somme est aussi la probabilité qu'un événement de type 1 survienne, mais entre le premier et le deuxième renouvellement après t_K. Dans ce cas, la probabilité doit aussi considérer que la police d'assurance a du être renouvelée au premier renouvellement après t_K. Toutes les autres situations entre le renouvellement potentiel j et $j+1$ doivent aussi être calculées pour $j = 3, \ldots,$ de sorte que :

$$\begin{aligned}
& \Pr(\Psi_1 < \min(\Psi_3, \Psi_4, \Omega - t_K) | R \geq 1) \\
= \; & \frac{\lambda}{\gamma+\lambda}\left(1 - e^{-(\gamma+\lambda)(1-\{t_K\})}\right) + \frac{\lambda}{\gamma+\lambda} e^{-(\gamma+\lambda)(1-\{t_K\})} p \left(1 - e^{-(\gamma+\lambda)}\right) + \ldots \\
& + \frac{\lambda}{\gamma+\lambda} e^{-(\gamma+\lambda)(R-\{t_K\})} p^R \left(1 - e^{-(\gamma+\lambda)\{\Omega\}}\right) \\
= \; & \frac{\lambda}{\lambda+\gamma}\left[\left(1 - e^{-(\gamma+\lambda)(1-\{t_K\})}\right)\right. \\
& + e^{-(\gamma+\lambda)(1-\{t_K\})} p \left(1 - e^{-(\gamma+\lambda)}\right) \frac{1 - \left(pe^{-(\gamma+\lambda)}\right)^{R-1}}{1 - pe^{-(\gamma+\lambda)}} \\
& \left. + e^{-(\gamma+\lambda)(R-\{t_K\})} p^R \left(1 - e^{-(\gamma+\lambda)\{\Omega\}}\right)\right],
\end{aligned}$$

où la simplification s'effectue grâce aux propriétés de la série géométrique. Les cas pour $R = 0$ et $R \geq 1$ peuvent être rassemblés pour obtenir :

$$\begin{aligned}
& \Pr(\mathcal{M}(t_K) = 1, \mathcal{N}(t_K) = 0 | AS = 1) \times \Pr(\Psi_1 < \min(\Psi_3, \Psi_4, \Omega - t_K)) \\
= \; & \frac{\mu}{\mu+\gamma}\frac{\lambda}{\gamma+\lambda}\left[\mathbb{1}(R=0)\left(1 - e^{-(\gamma+\lambda)(\Omega - t_K)}\right) + \mathbb{1}(R \geq 1)\left[\left(1 - e^{-(\gamma+\lambda)(1-\{t_K\})}\right)\right.\right. \\
& + e^{-(\gamma+\lambda)(1-\{t_K\})} p \left(1 - e^{-(\gamma+\lambda)}\right) \frac{1 - \left(pe^{-(\gamma+\lambda)}\right)^{R-1}}{1 - pe^{-(\gamma+\lambda)}} \\
& \left.\left. + e^{-(\gamma+\lambda)(R-\{t_K\})} p^R \left(1 - e^{-(\gamma+\lambda)\{\Omega\}}\right)\right]\right] \equiv G(t_K, \Omega). \qquad (1.31)
\end{aligned}$$

Le cas où $p = 1$ retourne le résultat de l'équation (1.29).

Cette situation peut être visualisée par un exemple simple, illustré à la figure 1.7.

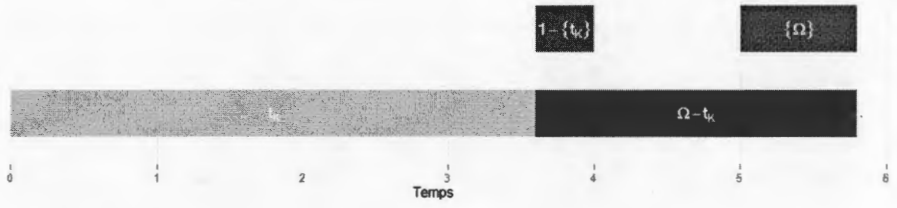

Figure 1.7 Notation de temps appliquée à un exemple

Dans cet exemple, l'état inconnu du ménage débute avant le 4^e renouvellement, et se poursuit jusqu'à après le 5^e renouvellement potentiel. Le ménage n'atteint toutefois pas le temps où un 6^e renouvellement aurait pu être effectué. Dans ce cas, le terme G se calculerait avec un $t_K \approx 3.6$ et $\Omega \approx 5.8$, et serait ainsi égal à :

$$\begin{aligned} G(t_K = 3.6, \Omega = 5.8) &= \left[\left(1 - e^{-(\gamma+\lambda)(1-0.6)}\right) + e^{-(\gamma+\lambda)(1-0.6)} p \left(1 - e^{-(\gamma+\lambda)}\right)\right. \\ &\quad \left. + e^{-(\gamma+\lambda)(2-0.6)} p^2 \left(1 - e^{-(\gamma+\lambda)0.8}\right)\right] \frac{\lambda}{\gamma+\lambda} \frac{\mu}{\mu+\gamma}. \end{aligned}$$

Finalement, par souci de simplification et comme nous l'avons déjà noté, nous avons calculé la probabilité complémentaire.

À la suite de la correction du modèle afin d'inclure les situations où il est difficile de distinguer les cas inactifs des cas où la police serait annulée, la contribution d'un assuré à la fonction de vraisemblance (1.20) peut être ajustée en ajoutant le terme $1 - G(t_K, \Omega)$:

$$\begin{aligned} &\mathcal{L}(\lambda, \gamma, \mu, p | E, A, S, Q, T, V, \Omega) \\ &\propto e^{-(\lambda+\gamma)T} e^{-\mu V} \lambda^E \mu^S \gamma^A (1-p)^Q p^{\lfloor T \rfloor - Q} \left[(\mu+\gamma)(1-G)\right]^{A^*}, \end{aligned} \tag{1.32}$$

où $A^* = 1$ si l'assuré est effectivement passé au niveau inconnu. En passant au logarithme de cette fonction et en sommant pour l'ensemble des ménages d'un portefeuille donné, il devient donc possible d'estimer les 4 paramètres du modèle, λ, γ, μ et p.

Ajout de voitures λ	Retrait de voitures μ	Arrêt de contrat γ	Renouvellement p
0,0730 (0,0003)	0,0828 (0,0003)	0,0400 (0,0003)	0,9179 (0,0003)

Tableau 1.1 Valeurs des paramètres (écarts-types)

1.4.3 Applications numériques

La fonction de vraisemblance finale a été appliqué aux données d'assurance décrites dans la section 1.1.2. Les calculs ont été effectués avec SAS et son complément de langage matriciel, SAS/IML. Pour l'estimation des 4 paramètres requis, un algorithme de Newton-Raphson a été en mesure de converger et d'obtenir un résultat valide. Puisqu'il s'agit d'une imposante base de données, la fonction de première dérivée a été fournie à l'algorithme afin d'en accélérer les calculs. Il aurait été également possible de fournir la matrice hessienne pour avoir un algorithme encore plus rapide. Cette matrice hessienne a été estimée par un algorithme de différences finies et puis a été inversée pour en approximer les écarts-types propres à chacun des paramètres. Les résultats apparaissent dans la table 1.1.

Cette valeur de $\widehat{\lambda}$ peut être interprétée dans le sens qu'il y a, pour un contrat actif, un ajout de voiture à tous les $0.0730^{-1} = 13.69$ ans. Le paramètre $\widehat{\mu}$ signifie que chaque voiture reste en moyenne $0.0828^{-1} = 12.08$ ans avec le ménage assuré. Il est à noter que le taux d'arrivée de voitures ne suffit pas à compenser pour le taux de départ de voitures puisque $\widehat{\lambda} < \widehat{\mu}$. Notons aussi que le taux de renouvellement annuel est d'environ 92% et que la probabilité d'annulation est approximativement égale à 4% par an. Pour l'assureur analysé, l'arrivée de nouveaux assurés serait requise pour garantir une profitabilité à long terme.

$\boldsymbol{X}_i$	Description
x_1	égale 1 si le ménage vient du grand marché[3]
x_2	égale 1 si le ménage a au moins une voiture louée
x_3	égale 1 si les assurés du ménage ne sont pas mariés
x_4	égale 1 si le ménage est avec l'assureur depuis 9 ans ou moins
x_5	égale 1 si la date effective du contrat est entre janvier et juin
x_6	égale 1 si la date effective du contrat est au mois de juillet
x_7	égale 1 si la date effective du contrat est le premier du mois

Tableau 1.2 Variables binaires résumant l'information disponible pour chaque ménage

1.4.3.1 Variables explicatives

Intuitivement, nous savons que certains profils d'assurés sont plus enclins à ajouter ou retirer des voitures à leur contrat. Similairement, nous pouvons croire que certains profils annulent plus rapidement que d'autres ou qu'ils ont un taux de renouvellement plus bas. Ainsi, l'ajout de variables explicatives dans les paramètres du modèle semble justifié. Les variables choisies pour définir le vecteur $\boldsymbol{X}_i$ apparaissent à la table 1.2. Les dates effectives du contrat d'assurance ont été utilisées pour montrer la stabilité des ménages assurés le premier juillet. Même si les caractéristiques d'un ménage changent au cours des années, pour simplifier, seulement celles de la première année sont considérées. Nous croyons que l'effet est minimal, puisque l'horizon de temps est court et que la majorité des ménages garde les mêmes caractéristiques au cours des années. Cependant, les recherches futures pourraient considérer d'améliorer la modélisation.

3. À l'opposé de l'assurance de groupe.

Paramètre	β_λ	β_μ	β_γ	β_p
β_0	-2,494 (0,015)	-2,726 (0,014)	-3,709 (0,039)	2,769 (0,019)
β_1	-0,264 (0,009)	0,122 (0,009)	-0,070 (0,019)	-0,152 (0,010)
β_2	-0,142 (0,011)	-0,048 (0,011)	·	0,092 (0,013)
β_3	-0,454 (0,009)	0,410 (0,008	0,214 (0,018))	-0,246 (0,009)
β_4	0,201 (0,014)	0,130 (0,013)	0,677 (0,038)	-0,399 (0,017)
β_5	·	·	·	0,137 (0,009)
β_6	-0,098 (0,012)	-0,034 (0,012)	-0,181 (0,029)	0,400 (0,016)
β_7	-0,129 (0,009)	-0,129 (0,009)	-0,647 (0,021)	0,089 (0,010)

Tableau 1.3 Valeurs des paramètres de régression avec variables explicatives

Une certaine fonction de lien $g(\boldsymbol{X}_i\boldsymbol{\beta})$ est associé à chacun des paramètres, où $\boldsymbol{\beta}$ est un vecteur de paramètres à être estimés. Dans notre modèle, les paramètres satisfont $\lambda, \gamma, \mu \in \mathbb{R}^+$, et donc une fonction de lien logarithmique est choisie : $\exp(\boldsymbol{X}_i\boldsymbol{\beta})$. Cette fonction de lien permet aux paramètres d'être toujours positifs. Aussi, puisque le paramètre modélisant le taux de renouvellement doit satisfaire $p \in [0, 1]$, il faut donc choisir une fonction de lien logistique pour ce paramètre : $p_i = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_p)}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_p)}$. Comme pour l'estimation des paramètres sans variables explicatives, un algorithme de Newton-Raphson a été utilisé pour l'estimation des paramètres et la matrice hessienne a été inversée pour le calcul des écarts-types. La valeur estimée des éléments du vecteur $\boldsymbol{\beta}$ est disponible à la table 1.3.

L'objectif d'un assureur devrait être de maximiser le nombre de voitures assurées au temps t. Dans ce cas, nous cherchons un λ et un p ayant une valeur élevée tandis que μ et γ devraient avoir une valeur faible. Nous pouvons remarquer certaines caractéristiques ayant des effets assez importants sur les paramètres comme le statut civil ou la date effective du contrat au mois de juillet. Il faut aussi

Paramètre	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Ménage A	1	0	1	1	0	0	0
Ménage B	1	1	1	0	0	0	0
Ménage C	0	0	1	0	0	0	1
Ménage D	1	1	0	0	0	0	0
Ménage E	0	1	0	0	0	1	1

Tableau 1.4 Caractéristiques des assurés types choisis

remarquer que les assurés renouvelant les premiers du mois offrent également une stabilité accrue. De plus, la variable x_4, qui identifie les ménages étant assurés depuis moins de 9 ans avec l'entreprise, montre que les familles assurées depuis longtemps montrent une loyauté accrue envers leur assureur. Pour tester si les variables explicatives sont statistiquement significatives, un test de Wald a été effectué selon un intervalle de confiance de 95%. Ainsi, β_5 n'a donc pas été inclus dans β_λ, β_γ et β_μ alors que β_2 n'a pas été inclus dans β_γ.

1.5 Analyse

Dans cette section, des applications sont présentées qui utilisent les valeurs trouvées par régression dans la section précédente et montrées au tableau 1.3. Ainsi, par exemple, même si nous travaillons avec $\widehat{\lambda}$, $\widehat{\gamma}$, $\widehat{\mu}$ ou $\widehat{p}$, par simplicité, nous noterons ces paramètres par λ, γ, μ et p.

5 profils d'assurés ont été choisis pour représenter l'impact de la segmentation de marché. Puisqu'il y a 96 possibilités de types d'assurés, seulement certains assurés typiques seront montrés pour représenter les résultats. Le premier profil sélectionné représente le meilleur type de ménage (E) en terme du nombre de voitures espéré, tandis que le pire type correspond au ménage (A). À titre d'illustration,

Paramètre	λ	γ	μ	p
Ménage A	0,0492	0,0557	0,1270	0,87795
Ménage B	0,0349	0,0283	0,1063	0,92154
Ménage C	0,0460	0,0159	0,0868	0,93163
Ménage D	0,0550	0,0228	0,0705	0,93756
Ménage E	0,0570	0,0107	0,0530	0,96608

Tableau 1.5 Valeur des paramètres pour chaque profil

nous avons aussi utilisé 3 types moyens (B,C,D). La table 1.4 exprime chacun des profils en terme de ses variables explicatives.

Nous pouvons voir que la seule différence entre le ménage de type B et D réside dans l'état civil. Mais comme il est possible de l'entrevoir dans la table 1.3, cette variable explicative a des effets significatifs sur chaque paramètre du modèle. Les valeurs des paramètres λ, γ, μ et p de chacun des ménages sont présentées dans le tableau 1.5.

1.5.1 Nombre espéré de voitures assurées

De la fonction génératrice de probabilité du processus complet présenté à la section 1.3.1.1, d'intéressantes propriétés peuvent être extraites. Pour simplifier la notation et les calculs, nous utilisons la variable $H(t)$ comme le nombre de voitures assurées d'un assuré actif ou inactif. Sachant que $\Pr(X = n) = \frac{d^n P(z=0)}{dz^n}$, nous pouvons calculer l'espérance du nombre de voitures assurées $H(t)$ comme :

$$\mathbb{E}[\mathcal{H}(t)] = \frac{\partial P_{\mathcal{N}(t),\mathcal{M}(t)}(z=1,y=1,t)}{\partial z} = \sum_{i=0}^{\infty} i p_i^{(1)} 1^{i-1} 1.$$

En utilisant l'équation (1.14) du processus final, nous commençons par le calcul

de la dérivée selon z :

$$\begin{aligned}\frac{\partial P_{\mathcal{N}(t),\mathcal{M}(t)}(z,y,t)}{\partial z} &= \frac{\partial\left[1-e^{-\gamma t}p^{\lfloor t\rfloor}\left(1-((z-1)e^{-\mu t}+1)^{a}\,ye^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\right)\right]}{\partial z}\\ &= e^{-\gamma t}p^{\lfloor t\rfloor}\left((z-1)e^{-\mu t}+1\right)^{a-1}ye^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\\ &\quad\times\left[ae^{-\mu t}+\left((z-1)e^{-\mu t}+1\right)\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right]. \qquad (1.33)\end{aligned}$$

Puis, en mettant $z=1$ et $y=1$, l'espérance peut être retrouvée :

$$\mathbb{E}[\mathcal{H}(t)] = e^{-\gamma t}p^{\lfloor t\rfloor}\left(ae^{-\mu t}+\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right). \qquad (1.34)$$

À titre d'illustration, nous avons supposé la valeur $c=0$, signifiant que les assurés sont au tout début de leur contrat et que le prochain renouvellement se fera dans un an. Avec de calculs similaires, les autres moments du nombre de voitures peuvent être trouvés, tels que la variance. Ces calculs peuvent également être appliqués à toutes les fonctions génératrices trouvées précédemment.

Pour chaque profil, le nombre de voitures assurées a été calculé. Les fonctions résultantes peuvent être visualisées sur la figure 1.8. Dans la figure, nous pouvons clairement voir le choc au renouvellement de contrat, modélisé par le paramètre p. De plus, une grande amplitude est visible entre chaque profil, où par exemple, le ménage E semble beaucoup plus avantageux pour l'assureur qu'un de type A. Les valeurs numériques apparaissent à la table 1.6, où le nombre espéré de voitures après 5 ans (immédiatement après le renouvellement) est montré. Nous voyons que le nombre initial de voitures assurées accroît linéairement l'espérance.

Les résultats montrent une différence importante entre les profils où, par exemple, le ménage E semble beaucoup plus avantageux pour les assureurs que le ménage A. Ainsi, nous croyons qu'un assureur aurait intérêt à adapter ses efforts marketing pour viser certains types d'assurés. Dans le cas ici montré, il s'agit de ménages de type E au lieu de ménages de type A. Effectivement, les assureurs seraient

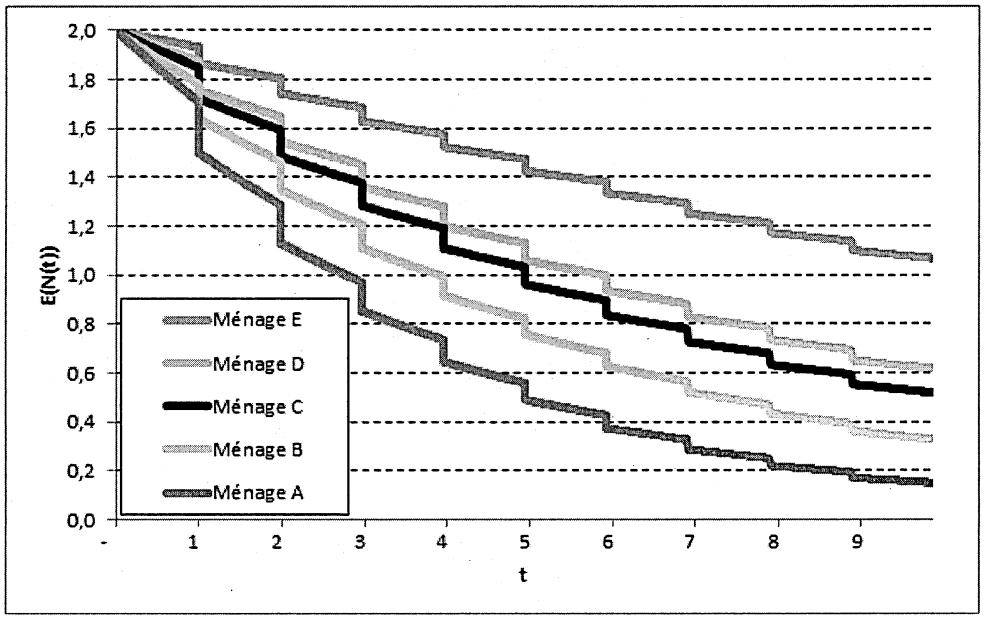

Figure 1.8 Espérance du nombre de voitures au temps t

Nombre de véhicules de départ (a)	1	2	3	4
Ménage A	0,281	0,490	0,700	0,909
Ménage B	0,417	0,756	1,096	1,435
Ménage C	0,541	0,961	1,381	1,801
Ménage D	0,604	1,058	1,513	1,967
Ménage E	0,812	1,424	2,036	2,648

Tableau 1.6 Nombre de voitures espérées restantes après 5 ans

capables d'offrir des escomptes à certains profils d'assurés qui génèrent des coûts administratifs plus faibles puisqu'ils restent plus longtemps dans la compagnie.

1.5.2 Actualisation de l'exposition

La valeur présente monétaire d'un ménage peut être retrouvée en calculant la valeur présente de flux monétaires sur un certain horizon de temps futur. Dans le jargon marketing, ce résultat est appelé valeur vie client (traduction de *customer lifetime value* de l'anglais). Cette valeur est d'ailleurs analysée dans plusieurs articles dont Donkers *et al.* (2007) qui compare différentes méthodes de calcul, Guillén *et al.* (2012) et Guelman *et al.* (2015). Dans notre cas, nous faisons l'hypothèse que la compagnie d'assurance fait 1 $ de profit sur l'exposition annuelle de chaque voiture. Cette hypothèse pourrait facilement être adoptée selon les données du marché pour être plus réaliste. Nous utilisons aussi un taux d'actualisation instantané de δ. Donc, pour calculer la valeur vie client, il est nécessaire de calculer la valeur actualisée de l'exposition future jusqu'à un certain temps T en utilisant un taux d'actualisation continu δ. Notée par ω_T, la valeur vie client se calcule en intégrant la variable $\mathcal{H}(t)$ de sorte que :

$$\begin{aligned} \omega_T &= \int_0^T \mathcal{H}(t) e^{-\delta t} dt \\ &= \lim_{n\to\infty} \frac{T}{n} \sum_{i=1}^{n} \mathcal{H}\left(\frac{iT}{n}\right) e^{-\frac{\delta i T}{n}} \end{aligned} \tag{1.35}$$

en utilisant les sommes de Riemann. L'espérance peut facilement être calculée, de telle sorte que :

$$\begin{aligned} \mathbb{E}(\omega_T) &= \lim_{n\to\infty} \frac{T}{n} \sum_{i=1}^{n} \mathbb{E}\left(\mathcal{H}\left(\frac{iT}{n}\right)\right) e^{-\frac{\delta i T}{n}} \\ &= \int_0^T \mathbb{E}(\mathcal{H}(t)) e^{-\delta t} dt. \end{aligned} \tag{1.36}$$

Pour 2 voitures de départ, les résultats apparaissent dans le tableau 1.7 pour

Nombre d'années (T)	1	5	10	20
Ménage A	1,833	5,468	6,720	7,122
Ménage B	1,869	6,359	8,581	9,731
Ménage C	1,904	7,003	10,104	12,237
Ménage D	1,917	7,286	10,823	13,506
Ménage E	1,946	8,218	13,553	19,433

Tableau 1.7 Valeurs présentes des profits clients pour 2 voitures initiales

chaque profil. Il faut remarquer qu'à long terme, les différences deviennent apparentes de sorte qu'un ménage de type E est en moyenne 2.5 fois plus profitable après 20 ans pour l'assureur qu'un de type A.

Bien que classiquement dans la littérature marketing, il est rare de calculer la valeur vie client sur un horizon de temps très grand. Toutefois, nous trouvons qu'il est utile de faire ce calcul puisqu'il permet d'avoir de simples équations. Nous pouvons trouver le profit total en tendant T vers l'infini de l'équation (1.35). Définissons cette valeur telle que $\omega \equiv \lim_{t\to\infty} \omega_t$. Son espérance peut alors être

calculée par l'équation :

$$\begin{aligned}
\mathbb{E}(\omega) &= \int_0^\infty \mathbb{E}(\mathcal{H}(t))e^{-\delta t}dt \\
&= \int_0^\infty e^{-(\gamma+\delta)t}p^{\lfloor t \rfloor}\left(ae^{-\mu t} + \left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right)dt \\
&= \sum_{i=0}^{\infty}\int_i^{i+1} e^{-(\gamma+\delta)t}p^i\left(ae^{-\mu t} + \left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right)dt \\
&= \sum_{i=0}^{\infty} a\frac{\left(e^{-i(\gamma+\delta+\mu)} - e^{-(i+1)(\gamma+\delta+\mu)}\right)}{\gamma+\delta+\mu}p^i + \frac{\left(e^{-i\gamma+\delta} - e^{-(i+1)(\gamma+\delta)}\right)\lambda}{(\gamma+\delta)\mu}p^i \\
&\qquad -\frac{\left(e^{-i(\gamma+\delta+\mu)} - e^{-(i+1)(\gamma+\delta+\mu)}\right)\lambda}{(\gamma+\delta+\mu)\mu}p^i \\
&= \frac{\left(1-e^{-(\gamma+\delta+\mu)}\right)\left(a-\frac{\lambda}{\mu}\right)}{(\gamma+\delta+\mu)\left(1-e^{-(\gamma+\delta+\mu)}p\right)} + \frac{(1-e^{-(\gamma+\delta)})\lambda}{(\gamma+\delta)\mu(1-e^{-(\gamma+\delta)}p)}, \qquad (1.37)
\end{aligned}$$

où a est le nombre de véhicules de départ. Les résultats, qui apparaissent dans le tableau 1.8[4], montrent une grande disparité entre les meilleurs et pires résultats. Dans notre cas, une entreprise a donc grand intérêt à cibler les assurés ayant les caractéristiques des assurés E. Cette analyse est basée sur un profit d'un dollar par voiture. Ceci peut être généralisé en un profit proportionnel à la prime d'assurance. En outre, en utilisant ce modèle et un modèle économique incorporant l'élasticité du prix, la prime peut être définie comme une fonction des profits futurs.

1.6 Discussion

Comme mentionné précédemment, les résultats numériques de notre étude empirique suggère que l'assureur devrait porter plus attention à certains ménages qu'à d'autres. Nous voulions introduire à la littérature actuarielle des processus

4. Nous avons fait une approximation avec la variable (x_4) en la gardant fixe, tandis qu'elle aurait dû être changée après un certain temps. Cette particularité n'est vraie que pour le type de ménage A.

Nombre de véhicules de départ (a)	1	2	3	4
Ménage A	3,964	7,161	10,358	13,555
Ménage B	5,588	9,992	14,395	18,799
Ménage C	7,661	13,011	18,361	23,711
Ménage D	8,782	14,586	20,390	26,194
Ménage E	16,130	24,735	33,340	41,945

Tableau 1.8 Valeur vie client actualisée à 2%

de files d'attente qui sont réalistes et directement applicables à l'exposition au risque. Mais, il est important de comprendre que ces conclusions sont basées sur certaines hypothèses de notre modèle.

Dans la dernière section du papier, nous avons mentionné que l'hypothèse du profit sur la prime devrait être généralisé et plus réaliste. D'autres changements sont possibles comme rendre les variables explicatives davantage dynamiques afin de refléter le changement des caractéristiques du ménage à travers le temps. Par exemple, la variable modélisant le temps assuré avec la compagnie, (x_4), devrait logiquement changer avec le temps. Aussi, l'ajout de la sinistralité pourrait améliorer l'ajustement du modèle, principalement dans le paramètre p représentant le taux de renouvellement. En effet, il est intuitif de croire que le comportement de l'assuré ne sera pas le même s'il y a eu des réclamations dans l'année. Ce changement de comportement est probablement lié à la modification de prime qui lui sera proposée à son renouvellement. En ce sens, dans une telle généralisation du modèle, un système de tarification par expérience pourrait être introduit. Une introduction au phénomène de soif du bonus (voir Lemaire 1977 ou Boucher *et al.* 2009) pourrait aussi être ajoutée, donnant ainsi un modèle plus utile et qui a une approche plus réaliste.

D'autres généralisations sont possibles. Une dépendance entre les 4 processus utilisés dans le modèle pourrait être ajouté dans une modélisation future. Par exemple, une dépendance entre le taux d'annulation et le taux de renouvellement semble logique. Effectivement, un assuré qui est au courant des prix du marché devrait avoir une probabilité plus grande de changer d'assureur durant la couverture et au renouvellement. Similairement, il est logique de croire qu'un assuré qui ajoute une voiture sur son contrat est plus susceptible d'en enlever un autre par la site. Plus généralement, un assuré qui ajoute plusieurs nouvelles voitures sur son contrat devrait aussi enlever plusieurs automobiles. D'autres types de dépendances entre les processus sont possibles. Toutefois, même si la structure de dépendance semble évidente, l'utilisation de variables explicatives devrait diminuer l'impact de cette dépendance.

Même si l'inclusion de dépendances entre les processus est une généralisation intéressante du modèle, nous pensons que des améliorations plus simples seraient davantage utiles. Par exemple, l'utilisation de d'autres distributions de temps d'attente (gamma au lieu d'une distribution exponentielle, par exemple) pour modéliser le processus devrait être envisagée. D'autres utilisations du modèle proposé (comme l'assurance maison ou l'assurance de groupe) pourrait être d'intéressantes avenues. Le mélange de plusieurs lignes d'affaire peut aussi être considéré.

1.7 Conclusion

Nous voulions un modèle du nombre de voitures assurées pour chaque ménage. En partant d'un simple processus de Poisson, il est possible de le modifier afin d'obtenir différents modèles de files d'attente. Celui qui a été conçu pour la modélisation d'un contrat d'assurance automobile peut être vu comme une généralisation du processus $M/M/\infty$. Nous avons ajouté un nouvel état de "mort" du processus qui

permet la possibilité d'annulation ou de non-renouvellement de contrat. Justifiés par les données empiriques, nous avons aussi proposé une distinction entre une annulation de contrat et un état d'inactivité.

Ce nouveau modèle proposé est ainsi construit avec 4 paramètres : soit le taux d'ajout de nouveaux véhicules au contrat d'assurance, le taux de retrait de véhicules, le taux d'annulation de la police d'assurance et finalement le taux de renouvellement de la police. Des méthodes d'inférence statistique nous ont permis d'estimer chacun de ces paramètres, en utilisant souvent les propriétés de la loi exponentielle, et en conditionnant sur chacun des événements possibles. Finalement, puisque nous travaillons empiriquement avec des données censurées, nous avons développé une manière d'estimer les paramètres dans le cas où il n'est pas possible de distinguer entre les états inactifs et annulés d'un contrat d'assurance.

Nous avons proposé de généraliser le modèle en ajoutant quelques variables explicatives pour chacun des 4 paramètres du modèle. Il nous a été ainsi possible de segmenter le portefeuille et de déterminer ainsi certains profils d'assurés ayant davantage de disposition à ajouter ou enlever des véhicules, à annuler en cours de terme, ou à renouveler annuellement leur contrat d'assurance. Les paramètres obtenus ont pu mener à divers types de résultats, comme le nombre espéré de voitures assurées ou la valeur vie client, mesure qui calcule les profits futurs d'un assuré.

Nous croyons que ce modèle offre une bonne approximation des données empiriques et propose une intéressante première étape dans l'analyse du temps d'exposition en assurance. Il semble approprié d'améliorer le modèle proposé dans le futur. Plusieurs généralisations possibles ont été mentionnées dans le chapitre.

Finalement, notons que dans cette analyse, seul le développement des ménages déjà présents dans le portefeuille est considéré. Ainsi, nous n'incluons par l'arrivée

de nouvelles polices dans le modèle et le portefeuille assuré ne peut ainsi que diminuer à travers les années. Sachant que l'objectif du travail était d'analyser le comportement des assurés et l'impact des caractéristiques des ménages, l'analyse du modèle ne souffre pas de cet oubli. Toutefois, si un assureur voulait analyser l'évolution de son portefeuille en entier, il faudrait prendre en compte l'arrivée de nouveaux ménages. Ceci permettrait de calculer la valeur monétaire du portefeuille (aussi appelée équité client), de manière analogue au calcul menant à la valeur vie client.

1.8 Annexe 1 : Développements mathématiques

Pour la représentation matricielle, il est nécessaire de définir le vecteur de longueur infini : $\boldsymbol{p}(t) = [p_0(t), p_1(t), p_2(t), \ldots]$, ainsi que le vecteur de dérivée première : $\boldsymbol{p}'(t) = [\frac{dp_0(t)}{dt}, \frac{dp_1(t)}{dt}, \frac{dp_2(t)}{dt}, \ldots]$.

1.8.1 Processus de Poisson

Avec les probabilités conditionnelles (1.2), il est possible de retrouver l'équation de Chapman-Kolmogorov. Dans le cas où $i \geq 1$, elle peut être développée pour obtenir :

$$\begin{aligned} p_i(t+\Delta t) &= \sum_{j=0}^{\infty} p_j(t) \cdot p_{(j,i)}(t, t+\Delta t) \\ &= \sum_{j=0}^{i-2} p_j(t) \cdot o(\Delta t) + p_{i-1}(t) \cdot (\lambda \Delta t + o(\Delta t)) + p_i(t) \cdot (1 - \lambda \Delta t + o(\Delta t)) \\ &= o(\Delta t) \sum_{j=0}^{i} p_j(t) + \lambda \Delta t \cdot p_{i-1}(t) + (1 - \lambda \Delta t) \cdot p_i(t). \end{aligned}$$

Ensuite, en effectuant quelques opérations et en tendant Δt vers 0, nous arrivons à une équation différentielle :

$$\begin{aligned} \frac{p_i(t+\Delta t) - p_i(t)}{\Delta t} &= \lambda p_{i-1}(t) - \lambda p_i(t) + \frac{o(\Delta t)}{\Delta t} \sum_{j=0}^{i} p_j(t) \\ \Leftrightarrow \lim_{\Delta t \to 0} \frac{p_i(t+\Delta t) - p_i(t)}{\Delta t} &= \lambda p_{i-1}(t) - \lambda p_i(t) + \sum_{j=0}^{i} p_j(t) \cdot \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} \\ \Leftrightarrow \frac{dp_i(t)}{dt} &= \lambda p_{i-1}(t) - \lambda p_i(t). \end{aligned} \tag{1.38}$$

Ce genre d'équation est appelée de différence-différentielle, puisqu'elle est définie selon des niveaux relatifs. Dans ce cas, la dérivée de la probabilité d'être au niveau i est fonction de la probabilité d'être au niveau i et $i-1$.

Pour un niveau $i = 0$, qui signifie dans notre contexte 0 voiture, un développement équivalent à celui générant (1.38) nous mène à :

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t). \tag{1.39}$$

Pour faire le parallèle avec la notation matricielle et les chaînes de Markov à temps continu, il est possible de refaire le cheminement précédent en utilisant vecteurs et matrices.

Tout d'abord, l'équation (1.2) peut être réécrite sous forme de matrice de transition :

$$\boldsymbol{p}(t, t+\Delta) = \begin{pmatrix} 1-\lambda\Delta t + o(\Delta t) & \lambda\Delta t + o(\Delta t) & o(\Delta t) & o(\Delta t) & \dots \\ 0 & 1-\lambda\Delta t + o(\Delta t) & \lambda\Delta t + o(\Delta t) & o(\Delta t) & \dots \\ 0 & 0 & 1-\lambda\Delta t + o(\Delta t) & \lambda\Delta t + o(\Delta t) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Ensuite, il faut utiliser l'équation de Chapman-Kolmogorov pour recréer le cheminement menant aux résultats (1.38) et (1.39). Pour se faire, il faut séparer $\boldsymbol{p}(t, t+\Delta)$ en sommes de matrices de telle sorte que :

$$\begin{aligned} \boldsymbol{p}(t+\Delta t) &= \boldsymbol{p}(t)\boldsymbol{p}(t, t+\Delta) \\ &= \boldsymbol{p}(t)\left[\boldsymbol{I}_\infty + \Delta t\boldsymbol{Q} + o(\Delta t)\boldsymbol{U}\right], \end{aligned} \tag{1.40}$$

où $\boldsymbol{I}_\infty$ est une matrice identité d'ordre ∞. Les autres matrices sont :

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

alors que $\boldsymbol{U}$ est une matrice triangulaire supérieure de forme :

$$\boldsymbol{U} = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots \\ 0 & 1 & 1 & 1 & \dots \\ 0 & 0 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Les équations sont ensuite manipulées pour obtenir le vecteur de dérivée première $\boldsymbol{p'(t)}$ de sorte que :

$$\begin{aligned} \lim_{\Delta t \to 0} \frac{\boldsymbol{p}(t+\Delta t) - \boldsymbol{p}(t)\boldsymbol{I_\infty}}{\Delta t} &= \boldsymbol{p}(t)\left[\frac{\Delta t}{\Delta t}\boldsymbol{Q} + \frac{o(\Delta t)}{\Delta t}\boldsymbol{U}\right] \\ \boldsymbol{p'}(t) &= \boldsymbol{p}(t)\boldsymbol{Q}. \end{aligned} \tag{1.41}$$

Dans une équation comme celle-ci, la matrice $\boldsymbol{Q}$ est surnommée matrice des taux de transition dans la littérature.

Ainsi, l'équation (1.41) résume les équations (1.38) et (1.39). Pour la suite, les FGP peuvent être utilisées pour résoudre les équations (1.38) et (1.39). Il faut commencer par les multiplier par z^i où i est le nombre de voitures et ensuite les additionner. La fonction (1.1) peut être recréée du côté gauche de l'équation. La nouvelle formule ainsi générée est alors :

$$\sum_{i=0}^{\infty} \frac{dp_i(t)}{dt} z^i = -\lambda \sum_{i=0}^{\infty} p_i(t) z^i + \lambda z \sum_{i=0}^{\infty} p_i(t) z^i,$$

qui peut être réécrite comme :

$$\frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial t} = \lambda(z-1) P_{\mathcal{N}(t)}(z,t). \tag{1.42}$$

Cette équation aux dérivées partielles a comme solution :

$$\ln(P_{\mathcal{N}(t)}(z,t)) = \lambda t(z-1) + C,$$

où C est une certaine constante déterminée par la condition initiale. En prenant la condition initiale $\mathcal{N}(t=0) = 0$, c'est-à-dire que le nombre de voitures initial est 0, la solution $P(z, t=0) = 1$ est trouvée, et ainsi, $C = 0$. Finalement, la solution de ces équations est :

$$P_{\mathcal{N}(t)}(z,t) = e^{(z-1)\lambda t}, \tag{1.43}$$

ce qui correspond à la FGP d'une distribution de Poisson de paramètre λt. Sachant cela, il peut être dit que les probabilités individuelles sont :

$$p_i(t) = \frac{e^{-\lambda t}(\lambda t)^i}{i!}. \tag{1.44}$$

Encore une fois, il est possible d'utiliser la notation matricielle dans le développement du modèle. De façon analogue à la version scalaire, l'équation (1.41) a comme solution :

$$\boldsymbol{p(t)} = \boldsymbol{p(0)}e^{\boldsymbol{Q}t}, \tag{1.45}$$

où l'exponentielle de la matrice est définie comme :

$$\begin{aligned} e^{\boldsymbol{Q}t} &= \sum_{i=0}^{\infty} \frac{\boldsymbol{Q}^i t^i}{i!} \\ &= \boldsymbol{I}_\infty + t\boldsymbol{Q} + \frac{t^2}{2}\boldsymbol{Q} \times \boldsymbol{Q} + \dots \quad . \end{aligned} \tag{1.46}$$

Résoudre cette équation requiert de trouver une représentation générale pour $\boldsymbol{Q}^i$. Pour ce faire, la matrice $\boldsymbol{Q}$ peut être séparée en une addition telle que :

$$\boldsymbol{Q} = -\lambda \boldsymbol{I}_\infty + \lambda \boldsymbol{Q_1}, \tag{1.47}$$

où :

$$\boldsymbol{Q_1} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Les puissances de la matrice $\boldsymbol{Q_1}$ ont la particularité d'avoir la diagonale transposée d'une unité vers la droite chaque fois que l'exposant augmente de 1. Par exemple,

pour la puissance de 2, nous avons :

$$\boldsymbol{Q}_1^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

et ainsi de suite. Sachant cela, l'exponentielle de l'équation (1.45) peut être développée en utilisant (1.46) de sorte que :

$$\begin{aligned} e^{\boldsymbol{Q}t} &= e^{-\lambda t \boldsymbol{I}_\infty + \lambda t \boldsymbol{Q}_1} \\ &= \left[\sum_{i=0}^{\infty} \frac{\boldsymbol{I}_\infty{}^i(-\lambda t)^i}{i!}\right]\left[\sum_{i=0}^{\infty} \frac{\boldsymbol{Q}_1{}^i(\lambda t)^i}{i!}\right] \\ &= e^{-\lambda t}\boldsymbol{I}_\infty \begin{pmatrix} 1 & \lambda t & \lambda^2 t^2/2! & \lambda^3 t^3/3! & \dots \\ 0 & 1 & \lambda t & \lambda^2 t^2/2! & \dots \\ 0 & 0 & 1 & \lambda t & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \end{aligned}$$

Notons que pour que la décomposition $e^{\boldsymbol{X}+\boldsymbol{Y}} = e^{\boldsymbol{X}}e^{\boldsymbol{Y}}$ fonctionne, les matrices doivent satisfaire la condition $\boldsymbol{XY} = \boldsymbol{YX}$, ce qui est le cas ici. Finalement, en incorporant le résultat dans l'équation (1.45) et en choisissant 0 comme état initial avec $\boldsymbol{p}(0) = [1, 0, 0, \dots]$, il est retrouvé que :

$$\boldsymbol{p}(t) = e^{-\lambda t}\left[1, \lambda t, \frac{\lambda^2 t^2}{2!}, \dots\right], \tag{1.48}$$

c'est-à-dire une représentation vectorielle de (1.3). Pour les prochains modèles, la méthode utilisant les FGP sera utilisée permettant ainsi d'avoir les informations nécessaires au temps t. Mais, puisque ces modèles sont plus complexes, il ne sera pas toujours possible de trouver de solutions simples pour les probabilités individuelles. Aussi, utiliser la solution impliquant l'exponentielle d'une matrice n'est pas toujours une possibilité puisque la matrice des taux de transition ne pourra pas être séparée comme il a été effectué ici.

1.8.2 Naissance et mort

Pour développer la fonction génératrice de probabilité de ce système, il faut passer par une écriture des équations de Chapman-Kolmogorov, qui équivaut, pour $i > 0$, à :

$$\begin{aligned} p_i(t+\Delta t) &= \sum_{j=0}^{\infty} p_j(t) \cdot p_{(j,i)}(t, t+\Delta t) \\ &= p_{i-1}(t) \cdot [\lambda \Delta t + o(\Delta t)] + p_i(t) \cdot [1 - (\lambda + i\mu)\Delta t + o(\Delta t)] \\ &\quad + p_{i+1}(t) \cdot [(i+1)\mu \Delta t + o(\Delta t)] + \sum_{j \neq i-1, i, i+1} p_j(t) o(\Delta t) \\ &= \lambda \Delta t \cdot p_{i-1}(t) + [1 - (\lambda + i\mu)\Delta t] \cdot p_i(t) \\ &\quad + (i+1)\mu \Delta t \cdot p_{i+1}(t) + \sum_{j=0}^{\infty} p_j(t) o(\Delta t). \end{aligned}$$

Maintenant, le raisonnement est continué pour retrouver la formule du calcul différentiel :

$$\begin{aligned} p_i(t+\Delta t) - p_i(t) &= \lambda \Delta t \cdot p_{i-1}(t) - (\lambda + i\mu)\Delta t \cdot p_i(t) \\ &\quad + (i+1)\mu \Delta t \cdot p_{i+1}(t) + o(\Delta t) \\ \Leftrightarrow \lim_{\Delta t \to 0} \frac{p_i(t+\Delta t) - p_i(t)}{\Delta t} &= \lambda p_{i-1}(t) - (\lambda + i\mu) p_i(t) + (i+1)\mu p_{i+1}(t) + \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} \\ \Leftrightarrow \frac{dp_i(t)}{dt} &= \lambda p_{i-1}(t) - (\lambda + i\mu) p_i(t) + (i+1)\mu p_{i+1}(t). \end{aligned} \tag{1.49}$$

Pour $i = 0$, il est possible de retrouver par calculs similaires que :

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t). \tag{1.50}$$

Comme pour le processus de Poisson, il est possible de voir ce modèle selon les chaînes de Markov à temps continu. Les équations (1.49) et (1.50) peuvent être reproduites en utilisant la matrice $\boldsymbol{Q}$ telle que

$$\boldsymbol{p}'(t) = \boldsymbol{p}(t)\boldsymbol{Q}. \tag{1.51}$$

Cette matrice des taux de transition $\boldsymbol{Q}$ est égale à :

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \dots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Bien que la solution à l'équation différentielle soit $\boldsymbol{p}(t) = \boldsymbol{p}(0)e^{\boldsymbol{Q}}$, comme le processus de Poisson, il est difficile de trouver une combinaison linéaire pour séparer la matrice $\boldsymbol{Q}$.

Puisque dans les données étudiées, il n'y a qu'au maximum 4 voitures, cela veut dire que seulement les premières lignes et colonnes de la matrice sont utiles. Nous pouvons donc penser à un modèle ajusté où le processus de naissance s'arrête quand le nombre de voitures $\mathcal{N}(t)$ est à 4. Malheureusement, les équations associées à ce modèle réduit n'offrent pas de formes analytiques simples pour exprimer la fonction génératrice de probabilités $P_{\mathcal{N}(t)}(z,t)$ ou des probabilités de niveaux $p_i(t)$.

Maintenant, pour résoudre les équations (1.49) et (1.50), il faut passer par la fonction génératrice de probabilités qui permettent un résultat fermé. Comme pour le processus de Poisson, il faut commencer par multiplier les équations par z^i pour ensuite faire la sommation :

$$\sum_{i=0}^{\infty} \frac{dp_i(t)}{dt} z^i = \sum_{i=1}^{\infty} p_{i-1}(t) \cdot z^i \lambda - \sum_{i=0}^{\infty} p_i(t) \cdot (\lambda + i\mu) z^i + \sum_{i=0}^{\infty} p_i(t) \cdot (i+1)\mu z^i. \quad (1.52)$$

Ensuite, en réécrivant cette équation de façon à en faire ressortir les termes de FGP et ses dérivées, il s'ensuit que :

$$\sum_{i=0}^{\infty} \frac{dp_i(t)}{dt} z^i = \lambda z \sum_{i=0}^{\infty} p_i(t) z^i - \lambda \sum_{i=0}^{\infty} p_i(t) z^i - \mu z \sum_{i=0}^{\infty} i p_i(t) z^{i-1} + \mu \sum_{i=0}^{\infty} i p_i(t) z^{i-1}.$$

Cela peut être réécrit comme étant :

$$\begin{aligned}\frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial t} &= \lambda z P_{\mathcal{N}(t)}(z,t) - \lambda P_{\mathcal{N}(t)}(z,t) - \mu z \frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial z} + \mu \frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial z} \\ &= (z-1)\lambda P_{\mathcal{N}(t)}(z,t) - (z-1)\mu \frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial z}, \end{aligned} \tag{1.53}$$

qui utilise la dérivée de la FGP :

$$\frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial z} = \sum_{i=0}^{\infty} i p_i(t) z^{i-1}. \tag{1.54}$$

Il s'agit alors d'une équation aux dérivées partielles à résoudre. Nous pouvons résoudre cette équation par une dérivation en chaîne avec une nouvelle variable r, telle que z et t sont fonctions de r. Cette technique est appelée la méthode des caractéristiques. Nous avons alors que la dérivée de notre fonction $P(z,t)$ peut être écrite de la façon suivante :

$$\frac{dP}{dr} = \frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial t}\frac{dt}{dr} + \frac{\partial \mathcal{P}_{N(t)}(z,t)}{\partial z}\frac{dz}{dr}. \tag{1.55}$$

En mettant les éléments de l'équation (1.53) de façon à reprendre la forme de l'équation (1.55), nous obtenons :

$$P_{\mathcal{N}(t)}(z,t)\lambda(z-1) = \frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial t} + \frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial z}\mu(z-1). \tag{1.56}$$

Nous avons alors 3 équations suivantes à résoudre :

$$\frac{dt}{dr} = 1; \tag{1.57}$$

$$\frac{dz}{dr} = \mu(z-1); \tag{1.58}$$

$$\frac{dP}{dr} = P\lambda(z-1). \tag{1.59}$$

En supposant $t(r=0)=0$, la première équation a comme solution :

$$t = r. \tag{1.60}$$

Pour la deuxième équation, cela mène à la solution

$$\ln(z-1) = \mu r + C,$$

pour une certaine constante d'intégration C. Avec la condition initiale $z(r=0) = z_0$, nous avons l'égalité $C = \ln(z_0 - 1)$. Nous pouvons maintenant dire que :

$$\begin{aligned} \ln(z-1) &= \mu r + \ln(z_0 - 1) \\ \Leftrightarrow z - 1 &= e^{-\mu r}(z_0 - 1). \end{aligned} \tag{1.61}$$

Finalement, pour la troisième équation, nous avons que :

$$\begin{aligned} \frac{dP}{dr} &= P_{\mathcal{N}(t)}(z,t)\lambda(z-1) \\ \frac{dP}{dr} &= P\lambda e^{-\mu r}(z_0 - 1), \end{aligned}$$

en remplaçant certains termes par les résultats (1.60) et (1.61). Ensuite, puisque la dernière fonction n'utilise que r et P, la solution est facilement obtenue :

$$\ln P = \frac{\lambda}{\mu} e^{\mu r}(z_0 - 1) + C,$$

où C est encore une fois une constante d'intégration. Avec comme condition initiale que le système est à un niveau a, c'est-à-dire qu'il y a a voitures au temps 0, nous avons que la fonction génératrice de probabilité à ce temps est :

$$\begin{aligned} P_{\mathcal{N}(t)}(z, t=0) &= \sum_{n=0}^{\infty} p_n(t=0) z^n \\ &= z^a, \end{aligned}$$

et donc $P(r=0) = z_0^a$ avec le changement de l'équation (1.61). Nous pouvons alors trouver la valeur de la constante C de sorte que :

$$\begin{aligned} \ln(z_0^a) &= \frac{\lambda}{\mu}(z_0 - 1) + C \\ \Leftrightarrow z_0^a &= e^{\frac{\lambda}{\mu}(z_0-1)} e^C \\ \Leftrightarrow e^C &= z_0^a e^{-\frac{\lambda}{\mu}(z_0-1)}. \end{aligned}$$

Maintenant, nous pouvons introduire cette constante dans l'équation originale pour trouver :

$$\begin{aligned} P &= e^{\frac{\lambda}{\mu}e^{\mu r}(z_0-1)} z_0^a e^{-\frac{\lambda}{\mu}(z_0-1)} \\ \Leftrightarrow P_{\mathcal{N}(t)}(z,t) &= \left[(z-1)e^{-\mu t} + 1\right]^a e^{\frac{\lambda}{\mu}(1-e^{-\mu t})(z-1)}. \end{aligned} \tag{1.62}$$

Tout d'abord, il peut être intéressant de noter que lorsque $\mu \to 0$, nous retrouvons le résultat (1.43) du processus de Poisson. Ce résultat peut être interprété par le fait que dans ce cas, les voitures ne quittent jamais le système. Dans ce cas, seuls les ajouts de voitures sont nécessaires pour trouver le nombre de voitures dans le système.

Notons aussi que pour une valeur de 0 voiture au temps initial, donc $a = 0$, le processus se simplifie comme :

$$P_{\mathcal{N}(t)}(z,t) = e^{\frac{\lambda}{\mu}(1-e^{-\mu t})(z-1)}. \tag{1.63}$$

Il peut être remarqué qu'il s'agit de la FGP d'une distribution de Poisson de paramètre $\frac{\lambda}{\mu}(1-e^{-\mu t})$. Définissons par $\mathcal{A}(t)$ cette variable liée à cette équation génératrice (1.63) et représentant le nombre de voitures arrivées après le temps 0 et étant encore présentes au temps à t. Définissons aussi par $\mathcal{B}(t)$ un processus comptant les voitures qui étaient déjà présentes au temps 0 et qui y sont demeurées jusqu'à temps t. Ainsi, $N(t)$, le nombre de voitures au temps t, peut être défini par la relation $\mathcal{N}(t) = \mathcal{A}(t) + \mathcal{B}(t)$.

Puisque les processus $\mathcal{A}(t)$ et $\mathcal{B}(t)$ sont indépendants, par la propriété de l'addition des variables aléatoires sur leurs FGP, nous avons que :

$$P_{\mathcal{N}(t)}(z,t) = P_{\mathcal{A}(t)}(z,t) \times P_{\mathcal{B}(t)}(z,t).$$

Ainsi, en utilisant les résultats (1.4) et (1.63), nous trouvons que la FGP de $\mathcal{B}(t)$ est donc :

$$P_{\mathcal{B}(t)}(z,t) = \left[(z-1)e^{-\mu t}+1\right]^a, \tag{1.64}$$

de laquelle une distribution binomiale de paramètre a et $e^{-\mu t}$ peut être reconnue. Il est aussi possible de dire que la variable $\mathcal{B}(t)$ suit un processus de mort pur par

le fait qu'il n'y ait pas d'ajouts de voitures, c'est-à-dire que $\lambda = 0$. Ainsi, il n'y a que le taux individuel de retrait de voitures de la police μ qui intervient.

Finalement, la fonction de probabilité désirée pour modéliser le nombre de voitures peut être retrouvée par convolution en utilisant les distributions de chacune des variables

$$\begin{aligned} p_i(t) &= \Pr\{\mathcal{N}(t) = i\} \\ &= \Pr\{\mathcal{A}(t) + \mathcal{B}(t) = i\} \\ &= \sum_{k=0}^{\min(i,a)} \Pr\{\mathcal{A}(t) = i-k\} \cdot \Pr\{\mathcal{B}(t) = k\} \\ &= \sum_{k=0}^{\min(i,a)} \frac{e^{-(1-e^{-\mu t})\lambda/\mu}\left[(1-e^{-\mu t})\lambda/\mu\right]^{i-k}}{(i-k)!} \\ &\quad \times \binom{a}{k} e^{-\mu t k}(1-e^{-\mu t})^{a-k}. \end{aligned} \tag{1.65}$$

1.8.3 Annulation de police

Pour modéliser les départs de ménages de la compagnie d'assurance, un modèle avec catastrophes totales peut être développé. Se basant sur les modèles appliqués à des populations, il ajoute un processus poissonnien passant des niveaux $n \geq 1$ à 0 selon un taux γ, recréant ainsi l'effet d'une catastrophe qui anéantirait une population à zéro. Cet événement, appliqué dans notre contexte à un ménage assuré, peut recréer la rupture de contrat et fait en sorte que le nombre de voitures assurées pour ce ménage passe à 0 d'un seul coup. En partant du modèle $M/M/\infty$, nous avons que la probabilité conditionnelle d'être à i voitures au temps $t + \Delta t$,

sachant qu'il y avait j voitures au temps t, est :

$$p_{(j,i)}(t, t+\Delta t) = \begin{cases} \lambda\Delta t + o(\Delta t) & \text{quand } j = i-1, \\ 1-(\lambda+\gamma+i\mu)\Delta t + o(\Delta t) & \text{quand } j = i, \\ (i+1)\mu\Delta t + o(\Delta t) & \text{quand } j = i+1, \\ o(\Delta t) & \text{sinon,} \end{cases} \tag{1.66}$$

pour $i \geq 1$. Pour $i = 0$, nous avons :

$$p_{(j,0)}(t, t+\Delta t) = \begin{cases} 1-\lambda\Delta t + o(\Delta t) & \text{quand } j = 0, \\ (\gamma+\mu)\Delta t + o(\Delta t) & \text{quand } j = 1, \\ \gamma\Delta t + o(\Delta t) & \text{quand } j > 1. \end{cases} \tag{1.67}$$

Nous pouvons alors utiliser les équations de Chapman-Kolmogorov dans un développement similaire aux sections précédentes pour trouver les équations suivantes :

$$\begin{aligned} \frac{dp_i(t)}{dt} &= p_{i-1}(t)\lambda - p_i(t)(\lambda+\gamma+i\mu) + (i+1)\mu p_{i+1}(t) \quad \text{pour } i \geq 1 \\ \frac{dp_0(t)}{dt} &= -p_0(t) + \sum_{k=1}^{\infty} p_k(t)\gamma + p_1(t)\mu, \end{aligned}$$

où nous pouvons remplacer $\sum_{k=1}^{\infty} p_k(t)$ dans la dernière égalité par $1 - p_0(t)$. Encore une fois, il est possible de résumer les dernières équations par l'équation matricielle $\boldsymbol{p'(t)} = \boldsymbol{p(t)Q}$ où la matrice $\boldsymbol{Q}$ est cette fois-ci égale à :

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \gamma+\mu & -(\lambda+\gamma+\mu) & \lambda & 0 & \dots \\ \gamma & 2\mu & -(\lambda+\gamma+2\mu) & \lambda & \dots \\ \gamma & 0 & 3\mu & -(\lambda+\gamma+3\mu) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Nous pouvons alors calculer la fonction génératrice de probabilité en multipliant les deux côtés de l'équation par z^i pour le niveau i approprié et en sommant sur

tous les niveaux :

$$\begin{aligned}\sum_{i=0}^{\infty}\frac{dp_i(t)}{dt}z^i &= \sum_{i=1}^{\infty}p_{i-1}(t)z^i\lambda-\sum_{i=0}^{\infty}p_i(t)z^i\lambda+\gamma-\sum_{i=0}^{\infty}p_i(t)z^i\gamma\\ &\quad-\sum_{i=0}^{\infty}p_i(t)i\mu z^i+\sum_{i=0}^{\infty}p_i(t)(i+1)\mu z^i\\ &= \lambda(z-1)\sum_{i=0}^{\infty}p_i(t)z^i+\gamma-\gamma\sum_{i=0}^{\infty}p_i(t)z^i-\mu(z-1)\sum_{i=0}^{\infty}p_i(t)iz^i.\end{aligned}$$

Avec les mêmes relations qu'à la section précédente, nous pouvons trouver l'équation différentielle partielle suivante à résoudre :

$$\frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial t}=(\lambda(z-1)-\gamma)P_{\mathcal{N}(t)}(z,t)+\gamma-(z-1)\mu\frac{\partial P_{\mathcal{N}(t)}(z,t)}{\partial z}.$$

En utilisant encore une fois la méthode des caractéristiques pour recréer l'équation (1.55), nous avons les 3 équations suivantes à résoudre :

$$\begin{aligned}\frac{dt}{dr} &= 1;\\ \frac{dz}{dr} &= (z-1)\lambda;\\ \frac{dP}{dr} &= P((z-1)\lambda-\gamma)+\gamma.\end{aligned}$$

Avec les solutions déjà trouvées dans la section 1.8.2, nous avons les réponses aux deux premières équations. Nous pouvons les utiliser pour remplacer les termes en z et t de la troisième équations qui peut donc être résolue comme :

$$P=e^{(z_0-1)e^{\mu r}\frac{\lambda}{\mu}-\gamma r}\int e^{-(z_0-1)e^{\mu r}\frac{\lambda}{\mu}-\gamma r}\gamma dr+Ce^{(z_0-1)e^{\mu r}\frac{\lambda}{\mu}-\gamma r},$$

pour une certaine constante C. L'intégrale peut être résolue en développant en série de Taylor une partie des termes exponentiés :

$$\begin{aligned}
\int e^{-(z_0-1)e^{\mu r}\frac{\lambda}{\mu}-\gamma r}\gamma dr &= \int \sum_{k=0}^{\infty} \frac{\left[-(z_0-1)e^{\mu r}\lambda/\mu\right]^k}{k!}\gamma e^{-\gamma r}dr \\
&= \int \sum_{k=0}^{\infty} \frac{\left[-(z_0-1)\lambda/\mu\right]^k}{k!}\gamma e^{r(\mu k-\gamma)}dr \\
&= \sum_{k=0}^{\infty} \frac{\left[-(z_0-1)\lambda/\mu\right]^k}{k!}\gamma \int e^{r(\mu k-\gamma)}dr \\
&= \sum_{k=0}^{\infty} \frac{\left[-(z_0-1)\lambda/\mu\right]^k}{k!(\mu k-\gamma)}\gamma e^{r(\mu k-\gamma)}.
\end{aligned}$$

Avec la condition initiale $P_{\mathcal{N}(t)}(z, t=0) = z^a$, que le ménage assure a voitures au temps 0, nous pouvons finalement résoudre l'équation et arriver à :

$$\begin{aligned}
P_{\mathcal{N}(t)}(z,t) &= e^{(z-1)\lambda/\mu-\gamma t} + \left(\sum_{k=0}^{\infty} \frac{\left[(z-1)\lambda/\mu\right]^k\left(1-e^{-\mu k t}\right)}{k!(\mu k+\gamma)}\gamma \right. \\
&\quad \left. + \left((z-1)e^{-\mu t}\lambda/\mu\right)^a e^{-(z-1)e^{-\mu t}\lambda/\mu}\right). \qquad (1.68)
\end{aligned}$$

En imposant la contrainte $\gamma = \mu$, nous simplifions le résultat (1.68) précédent permettant ainsi d'éviter les sommations pour trouver :

$$P_{\mathcal{N}(t)}(z,t) = -\frac{\mu}{\lambda(z-1)} + \left[e^{-\mu t}((z-1)e^{-\mu t}+1)^a + \frac{\mu}{\lambda(z-1)}\right] e^{(z-1)(1-e^{-\mu t})\lambda/\mu}. \qquad (1.69)$$

CHAPITRE II

MODÈLES POUR UN ASSUREUR

2.1 Introduction

Dans ce chapitre, il sera question de généraliser le modèle de files d'attente développé au chapitre précédent qui estime le nombre de voitures assurées sur un contrat d'assurance. Une importante nouveauté de ce chapitre vient de l'inclusion d'un processus d'arrivées de nouveaux ménages. Puisque ce nouveau modèle inclut les ménages déjà assurés, de même que les nouveaux ménages qui pourraient s'ajouter au portefeuille, l'approche proposée ne permet non seulement de déterminer la valeur vie client(voir Guillén *et al.* 2012,Guelman *et al.* 2014 ou Guelman *et al.* 2015), mais aussi la valeur globale d'un portefeuille d'assurés en utilisant le concept d'équité client (traduction de *customer equity*, voir Rust *et al.* 2004).

Nous proposons une modélisation du nombre de ménages qui incorporera (1) un processus d'arrivées de nouveaux ménages, (2) un processus d'annulation de contrat et finalement (3) un processus de renouvellement de contrat. Pour le nombre de voitures par ménage, le modèle sera composé (4) d'un processus d'ajout de voitures ainsi (5) qu'un processus de retrait de voitures. Ensuite, puisque les modèles mathématiques sont sensiblement similaires, les modèles du nombre de voitures par ménage et du nombre de ménages par compagnie d'assurance pour-

ront aisément s'imbriquer pour former un modèle complet capable d'émuler le nombre total de voitures assurées par la compagnie d'assurance.

Comme au chapitre 1, nous nous intéressons aux fonctions génératrices de probabilités (FGP) puisqu'il s'agit d'une façon efficace de contenir toute l'information du modèle dans une seule équation simple. Ainsi, à la section 2, la FGP du nombre de ménages sera construite. À la section 3, la FGCP du nombre total de voitures sera développée. Par la suite, à la section 4, les paramètres nécessaires au modèles seront estimés par maximum de vraisemblance selon les hypothèses propres à ce chapitre. Ensuite, en utilisant ces estimations de paramètres, des statistiques potentiellement intéressantes et utiles seront calculées à la section 5, telles que le nombre espéré de voitures assurées au temps t ou la valeur actualisée des profits futurs. Finalement, la section 6 conclut ce chapitre.

2.1.1 Définition des termes utilisés

Le terme **ménage** est utilisé pour désigner un seul numéro de client ou assuré. Ce ménage peut d'ailleurs comporter plusieurs membres (ou conducteurs) et plusieurs voitures regroupées sous un même **contrat d'assurance**, qui peut être **renouvelé** chaque année. Ce contrat d'assurance représente le document qui lie l'assureur avec le ménage assuré. Dans ce mémoire, nous nous intéressons particulièrement au nombre de voitures que ce contrat couvre et, par la même occasion, détenues par le même ménage. Par extension, les **ajouts et retraits de voitures** du contrat d'assurance sont aussi analysés. Finalement, à tout moment durant la couverture d'assurance, un ménage peut décider d'annuler son contrat, signifiant que toutes les voitures assurées sont aussi annulées. Nous appelons cet événement une **rupture de contrat** ou **annulation**.

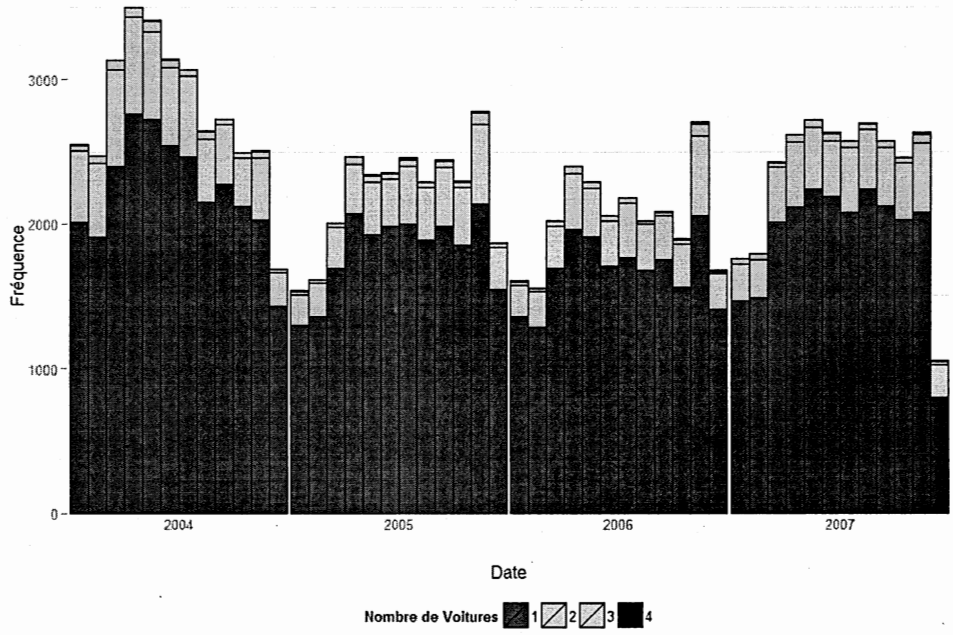

Figure 2.1 Nombre de nouveaux contrats par date d'arrivée

2.1.2 Analyse empirique

Nous nous baserons encore une fois sur les observations empiriques de notre base de données décrites à la section 1.1.2 pour construire un modèle dynamique du nombre de ménages. Ainsi, nous avons besoin des informations concernant les arrivées de nouveaux ménages dans l'entreprise. Les arrivées de contrats selon leur date d'entrée dans le portefeuille d'assurance, et selon le nombre de voitures assurés sur le contrat peuvent être observées sur la figure 2.1.

Nous remarquons qu'en fonction de la date de calendrier, le taux d'arrivée varie grandement et affiche une saisonnalité apparente où l'arrivée de nouveaux contrats survient davantage en été qu'en hiver. Bien entendu, il faut remarquer que le

nombre de voitures assurées lors de la souscription d'un nouveau contrat n'est jamais égal à 0, ce qui nécessitera certains ajustements des équations utilisées au chapitre 1.

2.2 Modèle du nombre de ménages

Dans le chapitre précédent, nous avons fait certaines hypothèses sur les comportements des ménages concernant le temps avant leur départ éventuel de l'entreprise. En premier lieu, nous avons fait l'hypothèse que la distribution du temps avant le bris du contrat venait d'une distribution exponentielle de paramètre γ. Cette modélisation provient de l'hypothèse que le taux de départ est constant dans le temps. Pour une meilleure approximation du comportement des assurés, nous avons ajouté au modèle le paramètre p correspondant au taux de renouvellement annuel de contrat.

Le modèle $M/G/\infty$ sera utile dans cette section pour modéliser le nombre de ménages. Ce modèle a déjà été analysé au chapitre 1 à la section 1.2.2, où la fonction de probabilité du nombre de voitures assurées au temps t est exprimée par l'équation (1.7). Afin de pouvoir utiliser ce modèle dans notre nouveau contexte de modélisation du mouvement des ménages, les paramètres de l'équation (1.7) doivent être ajustés de manière à refléter le fait que nous modélisons un nombre de ménages et non un nombre de voitures. Ainsi, par notation, nous devons remplacer :

1. L'ancien taux d'arrivée de voitures λ par un taux d'arrivée de ménages noté τ ;
2. L'ancien taux de retrait de véhicules μ par un taux d'annulation de police γ. Nous pouvons remarquer que ce taux d'annulation de police a le même rôle que celui du chapitre précédent (voir la Figure 1.5)

La correspondance entre le modèle du nombre de voitures assurées et du nombre de ménages assurés est possible puisqu'ils sont tous deux bâtis sous des hypothèses de service similaires. Puisque le modèle exprimé par l'équation (1.7) doit être utilisé en supposant que le nombre initial de ménages est de 0, nous séparons la modélisation du nombre de ménage en deux composantes :

1. Les nouveaux assurés qui incorporent le portefeuille que nous appelerons **nouveaux ménages** ;
2. Les assurés déjà présents dans le portefeuille lorsqu'il est analysé que nous appelerons **anciens ménages** ;

2.2.1 Nouveaux ménages

Nous notons par $\mathcal{K}(t)$, la variables aléatoire représentant le nombre de nouveaux ménages qui sont toujours présents dans le portefeuille au temps t. Nous supposons que cette variable aléatoire suit un processus $M/G/\infty$ avec 0 ménage initial. Ainsi, sa FGP est celle d'une distribution de Poisson telle que :

$$\begin{aligned} P_{\mathcal{K}(t)}(z,t) &= \sum_{i=0}^{\infty} \frac{z^i e^{-\tau q_t t}(\tau q_t t)^i}{i!} \\ &= e^{(z-1)\tau q_t t}. \end{aligned} \tag{2.1}$$

Pour un tel modèle, les ajouts de ménages dans le portefeuille se font selon un processus de Poisson de paramètre τ. Les départs de ménages, quant à eux, sont modélisés par le paramètre q_t, qui peut être calculé en utilisant la fonction de survie :

$$S(t) = e^{-\gamma t} p^{\lfloor t \rfloor} \tag{2.2}$$

où p représente la probabilité de renouvellement du contrat d'assurance et $\lfloor t \rfloor$ est la fonction fonction plancher, et permet au paramètre p d'affecter la fonction de survie à chaque renouvellement de contrat. Dans ce cas, tous les ménages ont un

$c = 0$. Ainsi, nous avons que :

$$\begin{aligned} q_t &= \int_0^t \frac{e^{-\gamma t} p^{\lfloor t \rfloor}}{t} dx \\ &= \left(1 - e^{-\gamma}\right) \frac{1 - \left(pe^{-\gamma}\right)^{\lfloor t \rfloor}}{\gamma t \left(1 - pe^{-\gamma}\right)} + \frac{p^{\lfloor t \rfloor} \left(e^{-\gamma \lfloor t \rfloor} - e^{-\gamma t}\right)}{\gamma t}, \end{aligned} \tag{2.3}$$

qui ressemble à l'équation homologue (1.9).

Même si les données apparaissant à la figure 2.1 ne vont pas dans le même sens, nous supposons que les arrivées de nouveaux ménages sont modélisées par un processus de Poisson de paramètre fixe τ. Garder ce paramètre fixe fait en sorte que le nombre de ménages au temps t suit un modèle de files d'attente simple de type $M/G/\infty$, duquel nous connaissons déjà la fonction génératrice résultante.

2.2.2 Anciens ménages

Définissons par $\mathcal{R}(t)$ le nombre de ces ménages qui étaient présents initialement dans le portefeuille et qui sont encore assurés au temps t. Contrairement à $\mathcal{K}(t)$, la variable aléatoire $\mathcal{R}(t)$ n'est pas égale à 0 au temps 0. Notons $R_0 = \mathcal{R}(0)$, le nombre initial de ménages. Cette variable aléatoire ne concerne que le processus de départ d'assuré puisque l'arrivée des nouveaux ménages est modélisée par $\mathcal{K}(t)$. Ainsi, cette variable aléatoire n'est fonction que du taux de renouvellement annuel du contrat (modélisé par le paramètre p), et le taux d'annulation de contrat (modélisé par le paramètre γ).

Notons par $\mathcal{M}_i(t)$, la variable aléatoire indicatrice qui indique si l'assuré i est encore assuré au temps t, dans quel cas $\mathcal{M}_i(t) = 1$. Il est ainsi possible d'utiliser la relation suivante : $\mathcal{R}(t) = \sum_{i=1}^{R_0} \mathcal{M}_i(t)$ pour modéliser le nombre d'anciens ménages au temps t. Pour $\mathcal{M}_i(t)$, une distribution de Bernoulli peut être utilisée qui a comme fonction génératrice de probabilités :

$$P_{\mathcal{M}(t)}(z) = e^{-\gamma t} p^{\lfloor t+c \rfloor}(z - 1) + 1, \tag{2.4}$$

avec comme paramètre $p^{\lfloor t+c \rfloor}$. La constante c permet de placer le ménage au temps adéquat. Par exemple, certains ménages sont analysés alors que leur renouvellement de contrat surviendra dans 3 mois. Dans un tel cas, c serait égal à 9/12, tandis qu'un autre ménage avec $C = 1/12$ aura son renouvellement dans 11 mois.

La FGP de $\mathcal{R}(t)$ peut être trouvée comme le produit des fonctions génératrices :

$$\begin{aligned} P_{\mathcal{R}(t)}(z,t) &= \prod_{i=1}^{R_0} P_{\mathcal{M}_i(t)}(z,t) \\ &= \prod_{i=1}^{R_0} e^{-\gamma t} p^{\lfloor t+c_i \rfloor}(z-1)+1. \end{aligned} \tag{2.5}$$

2.2.3 Total des ménages

Définissons par $\mathcal{W}(t)$ le nombre total de ménages au temps t, c.-à-d. le nombre total d'anciens ménages et de nouveaux ménages encore assurés au temps t, ou encore $\mathcal{W}(t) = \mathcal{K}(t) + \mathcal{R}(t)$. Puisque les variables aléatoires sont indépendantes, nous avons ainsi la relation $P_{\mathcal{W}(t)}(z,t) = P_{\mathcal{K}(t)}(z,t) \times P_{\mathcal{R}(t)}(z,t)$, de laquelle nous obtenons la fonction génératrice de $\mathcal{W}(t)$:

$$P_{\mathcal{W}(t)}(z,t) = \prod_{i=1}^{R_0} \left[e^{-\gamma t} p^{\lfloor t+c_i \rfloor}(z-1)+1\right] \times e^{\tau t q_t (z-1)}. \tag{2.6}$$

Cette FGP de $\mathcal{W}(t)$ sera incorporée par la suite dans le modèle du nombre total de voitures assurés.

2.3 Modélisation du nombre de voitures totales

Dans le premier chapitre, nous avons supposé un profit moyen par voiture assuré, ce qui nous a permis de calculer une valeur pour chaque client. Pour faire un calcul similaire afin de donner une valeur au portefeuille d'assurés, il est nécessaire de connaître le nombre total de voitures assurées de ce portefeuille.

Pour avoir un tel modèle, nous utilisons une combinaison des modèles du nombre de ménages et du nombre de voitures par ménage. Si la variable aléatoire déjà définie $\mathcal{W}(t)$ équivalant au nombre de ménages est utilisée, le nombre de voitures assurées $\mathcal{J}(t)$ peut être calculé grâce à la relation :

$$\mathcal{J}(t) = \sum_{j=1}^{\mathcal{W}(t)} \mathcal{N}_j(t), \tag{2.7}$$

où $\mathcal{N}_j(t)$ est le nombre de voitures au temps t du ménage assuré j et où $\mathcal{W}(t) = 0$ signifie que $\mathcal{J}(t) = 0$.

Avant de continuer dans le développement du modèle, il est important de comprendre la signification du paramètre t. Il doit être vu comme un indicateur du temps calendaire, et est donc partagé par tous les ménages. Nous supposons $t = 0$ comme le temps où le portefeuille d'assurance est analysé. Ainsi, le processus $\mathcal{N}_j(t)$ dépend du temps calendaire et non pas de l'âge du contrat. Le nombre de voitures $\mathcal{N}_j(t)$ du ménage j change donc dans le temps t, même si le ménage n'est pas assuré chez l'assureur que nous sommes en train d'analyser.

Encore une fois, le nombre total de voitures assurées provenant de ménages déjà assurés ($\mathcal{J}_A(t)$) et le nombre total de voitures assurées provenant de futurs ménages ($(\mathcal{J}_N(t))$) seront analysés séparément. Cela permet de traiter le problème spécifiquement pour chacun.

2.3.1 Nouveaux ménages

Pour modéliser $\mathcal{J}_N(t)$, le nombre de voitures assurées provenant des nouveaux ménages, nous utilisons l'équation suivante :

$$\mathcal{J}_N(t) = \sum_{j=1}^{\mathcal{K}(t)} \mathcal{N}_j(t), \tag{2.8}$$

où la variable $\mathcal{N}_j(t)$ suit un processus $M/M/\infty$. Tel que trouvé précédemment, la FGP de $\mathcal{K}(t)$ peut être exprimée par l'équation (2.1). Pour rappel, $\mathcal{K}(t)$ est un modèle $M/G/\infty$ qui ne suppose aucun ménage assuré initialement. LA FGP de $\mathcal{K}(t)$ est exprimée dans l'équation 2.1. Par propriété de composition des fonctions génératrices, nous pouvons développer l'équation :

$$\begin{aligned} P_{\mathcal{J}_N(t)}(z,t) &= P_{\mathcal{K}(t)}(P_{\mathcal{N}(t)}(z,t),t) \\ &= \exp\left[\tau t q_t \left[\left((z-1)e^{-\mu t}+1\right)^a e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}} - 1\right]\right]. \end{aligned} \quad (2.9)$$

Nous voyons que la FGP nécessite de donner une valeur pour a, le nombre de voitures initiales lorsque le portefeuille d'assurance a été analysé. Pour les anciens ménages, le nombre de voitures initiales d'un ménage, $\mathcal{N}(0) = a$, est toujours considéré comme connu, et cette information est disponible dans la base de données analysée. Dans le cas d'un ménage arrivant plus tard dans le portefeuille, il faut donc faire une hypothèse sur la distribution du nombre de voitures présentes sur le contrat à l'arrivée du ménage.

Une solution possible serait de mettre une valeur fixe pour a, mais il est évident que le nombre de voitures par ménage au temps $t = 0$ n'est pas le même pour tous. Une autre possibilité pour déterminer le nombre de voitures par police consiste à prendre la distribution stationnaire de $\mathcal{N}(t)$, c.-à-d. la distribution de $\mathcal{N}(t)$ lorsque $t \to \infty$. La FGP résultante de cette opération est simplement :

$$\lim_{t\to\infty} P_{\mathcal{N}(t)}(z,t) = e^{\frac{\lambda}{\mu}(z-1)}, \quad (2.10)$$

qui correspond à la FGP d'une distribution de Poisson de paramètre $\frac{\lambda}{\mu}$. Ainsi, nous pourrions supposer que le nombre de voitures d'un nouveau ménage suit une telle distribution. Ce choix pour la distribution est avantageux, car il a pour effet de simplifier les équations du modèle puisqu'il s'agit d'une distribution qui n'est pas fonction du temps calendaire t.

Un problème causé par cette distribution de Poisson est qu'il devient possible qu'un assuré intègre le portefeuille d'assurance sans avoir de voitures à assurer. Pour régler ce problème, nous modifions le support de la variable du nombre de voitures par ménage pour que le minimum soit de 1. Ceci suggère une transformation où 1 est ajouté à une variable suivant une distribution de Poisson telle que le nombre de voitures par ménage $\mathcal{N}(t)$ est égal à :

$$\mathcal{N}(t) = \mathcal{N}^*(t) + 1, \tag{2.11}$$

où $\mathcal{N}^*(t) \sim$ Poisson(λ/μ). Dans un cas comme celui-ci, toutefois, il est important de noter que l'interprétation du paramètre μ change légèrement. Ainsi, nous obtenons la FGP suivante pour la nouvelle variable aléatoire $\mathcal{N}(t)$:

$$\begin{aligned} P_{\mathcal{N}(t)}(z,t) &= P_{\mathcal{N}(t)}(z,t) \times z \\ &= ze^{\frac{\lambda}{\mu}(z-1)}, \end{aligned} \tag{2.12}$$

$$\tag{2.13}$$

qui signifie que nous avons :

$$e^{\tau t q_t (e^{\frac{\lambda}{\mu}(z-1)} z - 1)}. \tag{2.14}$$

2.3.2 Anciens ménages

Pour les R_0 ménages déjà assurés, nous pouvons nous baser sur la modélisation de la variable aléatoire $\mathcal{M}_i(t)$, qui a comme FGP de l'équation (2.4). Cette variable représente le nombre de voitures assurées au temps t du ménage i et pour elle, nous avons la relation suivante :

$$\mathcal{H}_j(t) = \mathcal{M}_j(t) \times \mathcal{N}_j(t),$$

où la variable $\mathcal{N}_j(t)$ est la même que celle exprimée dans la relation 2.11. Le j-ème ménage verra son nombre de voitures modélisé par la variable aléatoire

$\mathcal{H}_j(t)$ en ayant comme paramètres son nombre de voitures a_j et la constante de temps c_j précédemment introduite. Ainsi, nous retrouvons l'équation suivante, par composition de fonctions génératrices :

$$\begin{aligned} P_{\mathcal{H}_j(t)}(z,t) &= P_{\mathcal{M}_j(t)}(P_{\mathcal{N}_j(t)}(z,t),t) \\ &= P_{\mathcal{M}_j(t)}(P_{\mathcal{N}_j^*(t)}(z,t) \times z, t) \\ &= 1 - e^{-\gamma t}p^{\lfloor t \rfloor} + e^{-\gamma t}p^{\lfloor t \rfloor}\left((z-1)e^{-\mu t}+1\right)^a e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}} \end{aligned} \quad (2.15)$$

où nous voyons un changement dû à la relation $\mathcal{N}(0) = a$ qui implique $\mathcal{N}^*(0) = a - 1$.

Pour modéliser $\mathcal{J}_A(t)$, le nombre de voitures assurées provenant des anciens ménages, nous avons donc :

$$\mathcal{J}_A(t) = \sum_{j=1}^{R_0} \mathcal{H}_j(t),$$

qui correspond à la somme des voitures des R_0 anciens ménages. Par la somme de fonctions génératrices, nous avons le résultat :

$$P_{\mathcal{J}_A(t)}(z,t) = \prod_{i=1}^{R_0}\left[1 - e^{-\gamma t}p^{\lfloor t+c_i \rfloor}\left(1 - \left((z-1)e^{-\mu t}+1\right)^{a_i-1} z e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\right)\right], \quad (2.16)$$

exprimant la FGP de $\mathcal{J}_A(t)$.

2.3.3 Total des ménages

Pour modéliser le nombre total de voitures assurées venant de nouveaux et anciens ménages, nous devons développer la fonction génératrice des probabilités de la variable :

$$\mathcal{J}(t) = \mathcal{J}_A(t) + \mathcal{J}_N(t). \quad (2.17)$$

Finalement, la FGP du modèle complet selon les hypothèses choisies est :

$$\begin{aligned} P_{\mathcal{J}(t)}(z,t) &= P_{\mathcal{J}_A(t)}(z,t) \times P_{\mathcal{J}_N(t)}(z,t) \\ &= \prod_{i=1}^{R_0} \left[1 - e^{-\gamma t} p^{\lfloor t+c_i \rfloor} \left(1 - \left((z-1)e^{-\mu t} + 1\right)^{a_i - 1} z e^{(z-1)(1-e^{-\mu t})\frac{\lambda}{\mu}}\right)\right] \\ &\quad \times e^{\tau t q_t (e^{\frac{\lambda}{\mu}(z-1)} z - 1)}, \end{aligned} \tag{2.18}$$

où les informations du ménage i sont intégrées avec c_i et a_i.

2.4 Inférence

La fonction génératrice (2.18) utilise plusieurs paramètres que nous pouvons estimer en utilisant des données réelles. En effet, nous devons estimer τ, le taux d'arrivée de nouveaux ménages dans le portefeuille, γ, le taux de départ d'assuré, p, la probabilité de renouvellement de contrat, λ, le taux d'ajout de nouvelles voitures sur un contrat et μ, le taux de retrait de voitures d'un contrat. Pour en estimer leurs valeurs, nous allons effectuer des opérations similaires aux calculs de la section 1.4.1.2. Comme mentionné plus tôt, une importante différence existe entre les deux procédures par le fait que l'inférence de ce chapitre se fait seulement sur une ligne de temps calendaire, tandis que l'inférence du chapitre précédent supposait une ligne de temps par assuré.

2.4.1 Notations

Pour estimer tous les paramètres propres du nouveau modèle, une liste des variables utilisées dans la fonction de vraisemblance doit être présentée. Tous les événements observés affectant le nombre de polices ou le nombre de voitures sont notés comme :

1. Événement de type 1, qui représente l'ajout d'une voiture au contrat d'as-

surance. La variable aléatoire E représente le nombre total des événements de ce type (excluant celles présentes au départ du contrat) ;

2. Événement de type 2, qui représente le retrait d'une voiture d'un contrat d'assurance. La variable aléatoire S représente le nombre total d'événements de ce type ;
3. Événement de type 3, qui représente l'annulation de la police d'assurance à un moment différent de l'anniversaire de la police. La variable aléatoire A, est une variable indicatrice pour ce type d'événement ;
4. Événement de type 4, qui représente l'annulation de la police d'assurance à l'anniversaire de la police. La variable aléatoire Q est une variable indicatrice pour ce type d'événement. En d'autres mots, $Q = 1$ s'il n'y a pas eu renouvellement de contrat ;
5. Événement de type 5, qui représente l'arrivée d'un nouveau ménage. La variable aléatoire B représente le nombre total d'événements de ce type.

Ainsi, nous obtenons que l'ensemble des événements K est égal à la somme des éléments précédents, telle que $K = E + A + Q + S + B$. De plus, nous définissons respectivement par $\Psi_1, \Psi_2, \Psi_3, \Psi_4$ et Ψ_5 les variables aléatoires qui définissent le temps avant l'occurence d'un événement de type 1, 2, 3, 4 ou 5. Nous proposons aussi les notations suivantes :

1. t_j, le temps (en années) de survenance du j^{e} événement affectant le nombre de voitures ;
2. $\widetilde{t_j}$ la durée de temps (en années) entre le $(j-1)^{\text{e}}$ et le j^{e} événement tel que $\widetilde{t_j} = t_j - t_{j-1}$, avec $\widetilde{t_1} = t_1$;
3. ξ, le nombre total d'assurés dans la base de données.

Aussi, nous avons :

1. N_j, une variable qui compte le nombre de voitures immédiatement avant le j^e événement et donc N_1 est le nombre de voitures au début ;
2. T_j, le nombre d'années où l'assuré j était assuré ;
3. V_j^*, la somme du temps d'exposition (en années) de toutes les voitures du ménage j. Notons qu'à cause de la transformation $\mathcal{N}^*(t) = \mathcal{N}(t) - 1$, nous avons $V_j = V_j^* - T_j$.

2.4.2 fonction de vraisemblance

Notons tout d'abord que le temps avant l'arrivée de nouveaux ménages vient d'une distribution exponentielle. Selon l'hypothèse markovienne, ces arrivées sont indépendantes des ajouts/retraits de voiture et de rupture de contrat. Ainsi, il est possible de calculer la fonction de vraisemblance en ajustant la fonction de vraisemblance (1.21) pour qu'elle incorpore le nouveau paramètre τ.

De façon analogue à la fonction (1.26), nous pouvons utiliser les propriétés des distributions conjointes exponentielles pour calculer la probabilité que le premier événement soit l'ajout d'une voiture. Nous avons que la probabilité que le premier événement soit un ajout de voiture pour un des ménages est :

$$\Pr(\Psi_1 = \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) = W_1 \lambda e^{-\lambda W_1 \widetilde{t}_1} e^{-J_1 \mu \widetilde{t}_1} e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)} e^{-\tau \widetilde{t}_1},$$

où W_j est le nombre de ménages assurés et J_j est le nombre de voitures totales assurés juste avant l'évenement j, alors que $h(\widetilde{t}_j)$ est le nombre de renouvellement de contrats effectués durant la période $\widetilde{t}_j$. Pour les autres réalisations, les probabilités sont :

$$\begin{aligned}
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 = \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} J_1 \mu e^{-J_1 \mu \widetilde{t}_1} e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)} e^{-\tau \widetilde{t}_1} \\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 = \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} e^{-J_1 \mu \widetilde{t}_1} W_1 \gamma e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)} e^{-\tau \widetilde{t}_1} \\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 = \widetilde{t}_1, \Psi_5 > \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} e^{-J_1 \mu \widetilde{t}_1} e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)-1} (1-p) e^{-\tau \widetilde{t}_1} \\
\Pr(\Psi_1 > \widetilde{t}_1, \Psi_2 > \widetilde{t}_1, \Psi_3 > \widetilde{t}_1, \Psi_4 > \widetilde{t}_1, \Psi_5 = \widetilde{t}_1) &= e^{-\lambda W_1 \widetilde{t}_1} e^{-J_1 \mu \widetilde{t}_1} e^{-W_1 \gamma \widetilde{t}_1} p^{h(\widetilde{t}_1)} \tau e^{-\tau \widetilde{t}_1}
\end{aligned}$$

Puisque les processus exponentiels sont sans mémoire, la fonction de vraisem-

blance peut être calculée par le produit des K événements s'étant produits dans un certain intervalle de temps U pour tous les assurés de telle sorte que :

$$\begin{aligned}
&\mathcal{L}(\tau,\gamma,p,\lambda,\mu|U,T,V,E,A,Q,S,B)\\
&= e^{-\sum_{j=1}^{K}\tilde{t}_j(\tau+W_j(\lambda+\gamma)+J_j\mu)}p^{\sum^{K}h(\tilde{t}_j)}\tau^{\nu}\prod^{E}(\lambda W_j)\prod^{S}(\mu J_j)\prod^{A}(\gamma W_j)(1-p)^{Q}\\
&\propto e^{-(U\tau+\sum_{j=1}^{\xi}T_j(\lambda+\gamma)+\sum_{j=1}^{\xi}V_j\mu)}p^{\sum_{j=1}^{\xi}\lfloor T_j\rfloor-Q}\tau^{B}\lambda^{E}\mu^{S}\gamma^{A}(1-p)^{Q}, \qquad (2.19)
\end{aligned}$$

où les constantes ont été enlevées parce qu'elles ne servent pas au calcul du maximum de vraisemblance. Nous remarquons la similitude entre l'équation (2.19) et l'équation (1.21) du premier chapitre.

Néanmoins, nous devons adapter l'équation de vraisemblance afin d'introduire la distribution du nombre de voitures que le ménage détient lorsqu'il intègre le portefeuille d'assurance. Cette distribution est fonction de λ et μ. Telle que définie à la section par l'équation (2.10), la distribution du nombre de voitures est telle que :

$$\begin{aligned}
\Pr(\mathcal{N}(0)=a) &= \Pr(\mathcal{N}^{*}(0)=a-1)\\
&= \frac{e^{-\frac{\lambda}{\mu}}}{(a-1)!}\left(\frac{\lambda}{\mu}\right)^{a-1}, \qquad (2.20)
\end{aligned}$$

que nous multiplions à la fonction de vraisemblance (2.19). Ainsi, la fonction de vraisemblance pour les 5 paramètres est :

$$\begin{aligned}
\mathcal{L}(\tau,\gamma,p,\lambda,\mu) \propto\ & e^{-(U\tau+\sum_{j=1}^{\xi}T_j(\lambda+\gamma)+\sum_{j=1}^{\xi}V_j\mu+\xi\frac{\lambda}{\mu})}p^{\sum_{j=1}^{\xi}\lfloor T_j\rfloor-Q}\\
&\times\tau^{\nu}\lambda^{E+\sum_{j=1}^{\xi}(a_j-1)}\mu^{S-\sum_{j=1}^{\xi}(a_j-1)}\gamma^{A}(1-p)^{Q}, \qquad (2.21)
\end{aligned}$$

où a_j est le nombre de voitures de départ pour l'assuré j.

À noter que l'ajout de la distribution du nombre de voitures que le ménage détient lorsqu'il intègre le portefeuille d'assurance a comme conséquence de rendre plus complexe l'équation d'estimation des paramètres μ et λ. Ainsi, ces paramètres ne

pourront plus être estimés par une formule explicite comme ils l'étaient au premier chapitre grâce aux équations (1.22) et (1.24).

Néanmoins, en maximisant l'équation (2.21), nous retrouvons les estimateurs suivants pour les autres paramètres :

$$\widehat{\tau} = \frac{\nu}{U}, \tag{2.22}$$

$$\widehat{\gamma} = \frac{A}{\sum_{j=1}^{\xi} T_j}, \tag{2.23}$$

$$\widehat{p} = \frac{\sum_{j=1}^{\xi} \lfloor T_j \rfloor - Q}{\sum_{j=1}^{\xi} \lfloor T_j \rfloor}. \tag{2.24}$$

en plus du nouvel estimateur $\widehat{\tau}$, qui correspond au nombre annuel moyen de nouveaux ménages.

Les calculs ont encore une fois été effectués avec SAS et son complément de langage matriciel, SAS/IML. Pour l'estimation des 4 paramètres requis, un algorithme de Newton-Raphson a été utilisé pour maximiser la fonction. Puisqu'il s'agit d'une imposante base de données, la fonction de première dérivée a été fournie à l'algorithme afin d'en accélérer les calculs. Il aurait été également possible de fournir la matrice hessienne pour avoir un algorithme encore plus rapide. Cette matrice hessienne a été estimée par différences finies et puis a été inversée pour en approximer les écarts-types de chacun des paramètres. Les résultats se trouvent à la table 2.1. Nous pouvons voir que les valeurs trouvées sont similaires à celle du tableau 1.1 du premier chapitre.

Ainsi, avec $\widehat{\tau} = 28014$, il y aurait environ $\frac{28014}{365.25} = 76.7$ nouveaux assurés par jour pour le portefeuille de l'assureur. Les autres estimateurs peuvent être comparés à ceux de la table 1.1 du premier chapitre. La valeur de $\widehat{\lambda}$ peut être interprétée dans le sens qu'il y a, pour un contrat actif, un ajout de voiture à tous les $0.0624^{-1} = 16.02$ ans. Le paramètre $\widehat{\mu}$ signifie que chaque voiture reste en moyenne $0.2315^{-1} =$

Ajout de voitures $\widehat{\lambda}$	Retrait de voiture $\widehat{\mu}$	Arrêt de contrat $\widehat{\gamma}$	Renouvellement de contrat $\widehat{p}$	Nouveau contrat $\widehat{\tau}$
0,0624 (0,0002)	0,2315 (0,0007)	0,0918 (0,0003)	0,9188 (0,0003)	28014 (83,68)

Tableau 2.1 Valeurs des paramètres (écarts-types)

4.32 ans avec le ménage assuré. Notons aussi que le taux de renouvellement annuel est d'environ 92% et que la probabilité d'annulation est approximativement égale à 9% par an.

Le principal changement à remarquer de cette estimation est dans la valeur du taux de départ de voitures $\widehat{\mu}$. La différence provient de la transformation $\mathcal{N}^*$ et fait en sorte que la distribution du nombre de voitures à l'arrivée du ménage dans la compagnie est plus près de ce qui est observé dans la réalité, comme l'indique la figure 2.2. Il est intéressant de voir que si nous n'avions pas modifié le modèle afin que le nombre de voitures d'une nouvelle police d'assurance soit toujours plus grand que 0, le nombre de nouveaux ménages sans voitures assurées aurait été de 40%, ce qui est évidemment irréaliste. La transformation $\mathcal{N}^*$ est aussi responsable du faible changement dans le paramètre λ, tandis que le paramètre de renouvellement p est demeuré très similaire. Puisque le modèle de ce chapitre ne suppose d'état inactif 0*, le passage de 1 à 0 voiture assurée pour un ménage a simplement été considéré comme une annulation de contrat. Ainsi, pour le paramètre γ, sas valeur est surévaluée par rapport au premier chapitre.

2.4.3 Inclusion de variables explicatives

Intuitivement, nous savons que certains profils d'assurés sont plus enclins à ajouter ou retirer des voitures à leur contrat. Similairement, nous pouvons croire que

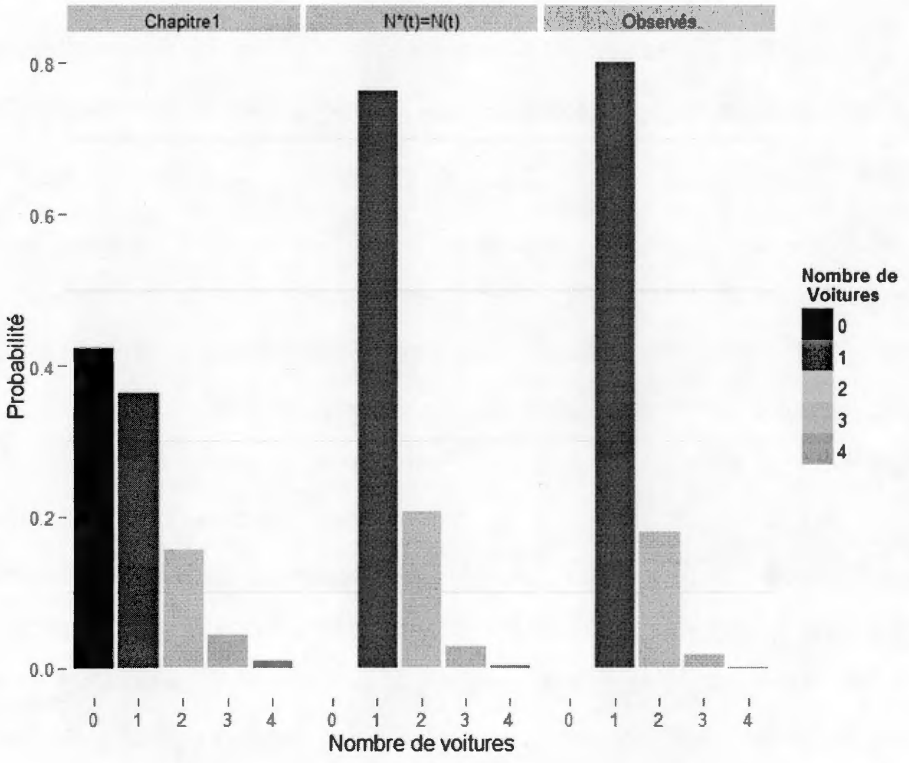

Figure 2.2 Comparaison du nombre de voitures lors de l'arrivée d'un ménage

certains profils annulent plus ou moins rapidement que d'autres ou qu'ils ont un taux de renouvellement plus ou moins haut. Ainsi, l'ajout de variables explicatives dans les paramètres du modèle semble justifié.

Les variables choisies pour définir le vecteur $\boldsymbol{X}_i$ apparaissent ainsi à la table 2.2. Parmi les régresseurs utilisés dans le chapitre précédent, certains avaient un lien avec la date effective, tel que le mois ou la journée de l'entrée en vigueur du contrat d'assurance. L'inclusion de ces régresseurs dans le présent modèle nécessiterait des modifications dans les hypothèse de base du modèle $M/G/\infty$. Ainsi, seules 3 variables explicatives sont utilisées dans le modèle du nombre de ménages. Cela crée donc 8 différents types de profils puisqu'il y a trois variables explicatives binaires, et donc $2^3 = 8$ types d'assurés.

Comme précédemment, une certaine fonction de lien $g(\boldsymbol{X}_i\boldsymbol{\beta})$ est associé à chacun des paramètres, où $\boldsymbol{\beta}$ est un vecteur de paramètres à être estimés. Dans notre modèle, les paramètres satisfont $\lambda, \gamma, \mu \in \mathbb{R}^+$, et donc une fonction de lien logarithmique est choisie : $\exp(\boldsymbol{X}_i\boldsymbol{\beta})$. Cette fonction de lien permet aux paramètres d'être toujours positifs. Aussi, puisque le paramètre modélisant le taux de renouvellement doit satisfaire $p \in [0, 1]$, il faut donc choisir une fonction de lien logistique pour ce paramètre. Alors, $p_i = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}_p)}{1+\exp(\boldsymbol{X}_i\boldsymbol{\beta}_p)}$. Les résultats des paramètres de régression se trouvent à la table 2.3. Ici encore, un algorithme de Newton-Raphson a été utilisé pour maximiser la fonction et La matrice hessienne a été inversée pour en approximer les écarts-types propres à chacun des paramètres.

L'utilisation de variables explicatives à inclure dans le paramètre τ, correspondant au nombre de nouveaux ménages, est plus difficile. En effet, le paramètre τ mesure le temps entre deux arrivés de nouveau ménage. Ainsi, il n'est pas aisé d'introduire des variables explicatives dans ce paramètre sans en changer sa nature ou l'interprétation du processus. Pour contourner ce problème, puisque notre

X_i	Description
x_1	égale 1 si le ménage vient du grand marché
x_2	égale 1 si le ménage a au moins une voiture louée
x_3	égale 1 si les assurés du ménage ne sont pas mariés

Tableau 2.2 Variables binaires résumant l'information disponible pour chaque ménage

Paramètre	β_λ	β_μ	β_γ	β_p
β_0	-2,5788 (0,0045)	-1,7056 (0,0043)	-2,7137 (0,0058)	2,5989(0,0066)
β_1	-0,0868 (0,0074)	0,1562 (0,0071)	0,2492 (0,0074)	-0,2541(0,0090)
β_2	-0,2403 (0,0106)	0,3735 (0,0103)	0,0717 (0,0100)	0,0784(0,0127)
β_3	-0,4222 (0,0076)	0,6885 (0,0071)	0,5453 (0,0071)	-0,2628(0,0087)

Tableau 2.3 Valeurs des paramètres de régression avec variables explicatives

Variables			Paramètres				
x_1	x_2	x_3	λ	μ	γ	p	τ
1	1	1	0,0359	0,6142	0,1576	0,8966	133
1	1	0	0,0547	0,3085	0,0914	0,9186	111
1	0	1	0,0456	0,4228	0,1467	0,8891	2 060
1	0	0	0,0696	0,2124	0,0851	0,9125	985
0	1	1	0,0391	0,5254	0,1229	0,9179	1 187
0	1	0	0,0597	0,2639	0,0712	0,9357	1 880
0	0	1	0,0497	0,3617	0,1144	0,9118	9 596
0	0	0	0,0759	0,1817	0,0663	0,9308	12 098

Tableau 2.4 Valeurs des paramètres pour chaque type d'assurés

exemple n'utilise que très peu de variables et ainsi peu de types de ménages, nous utilisons un estimateur $\hat{\tau_j}$ différent pour chacun des types de ménage. Cela signifie que nous supposons 8 processus d'arrivées différents pour le nombre de nouveaux ménages. Pour estimer les paramètres, il faut prendre l'équation de vraisemblance (2.21) comme fonction à maximiser pour chaque type de ménage. Nous obtenons ainsi une fonction de vraisemblance pour le paramètre τ_j de laquelle nous trouvons l'estimateur : $\hat{\tau_j} = \frac{B_j}{U}$, où B_j est le nombre d'arrivée de ménage de type $j = 1, \ldots, 8$.

Ainsi, les paramètres propres à chaque type d'assurés apparaissent à la table 2.4. Il est apparent que le taux d'arrivée diffère grandement selon le type de ménage.

2.5 Analyse des résultats

Grâce aux paramètres des tableaux 2.3 et 2.4, nous pouvons extraire différents résultats en utilisant les fonctions génératrices trouvées précédemment. D'ailleurs,

même si nous travaillons avec $\widehat{\lambda}$, $\widehat{\gamma}$, $\widehat{\mu}$, $\widehat{p}$ ou $\widehat{\tau}$, par simplicité, nous noterons ces paramètres par λ, γ, μ, p et τ. Comme pour le premier chapitre, la première analyse sera de calculer le nombre de voitures au temps t, mais cette fois-ci pour le portefeuille entier d'assurés de la compagnie d'assurance. Ensuite, nous actualiserons ce nombre de voitures pour obtenir une valeur actualisée des ménages présents et futurs, ce qui nous permet de donner une valeur au portefeuille.

2.5.1 Espérance et variance du nombre de voitures au temps t

Nous commençons par le calcul de l'espérance du nombre de voitures assurées par la compagnie au temps t. Bien qu'il soit possible de trouver ce résultat par dérivation de la FGP de l'équation (2.18), l'utilisation de l'espérance conditionnelle mène au même résultat de façon plus simple. En commençant par l'espérance de l'équation(2.17), nous développons :

$$\begin{aligned}
& \mathbb{E}(\mathcal{J}(t)) \\
= \; & \mathbb{E}\left[\sum_{i=1}^{R_0}\mathcal{H}_i(t)+\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)\right] \\
= \; & \mathbb{E}\left[\sum_{i=1}^{R_0}\mathcal{H}_i(t)\right]+\mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)|\mathcal{K}(t)\right]\right] \\
= \; & \sum_{i=1}^{R_0}\mathbb{E}(\mathcal{H}_i(t))+\mathbb{E}(\mathcal{K}(t)\mathbb{E}(\mathcal{N}(t))) \\
= \; & \sum_{i=1}^{R_0}\mathbb{E}(\mathcal{H}_i(t))+\frac{\lambda}{\mu}\mathbb{E}(\mathcal{K}(t)) \\
= \; & \sum_{i=1}^{R_0}e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left((a_i-1)e^{-\mu t}+1+\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right)+\tau t q_t\left(\frac{\lambda}{\mu}+1\right) \quad (2.25)
\end{aligned}$$

La variance ou écart-type peut être calculé par un cheminement similaire, duquel

nous obtenons :

$$
\begin{aligned}
& \mathbb{V}\mathrm{ar}(\mathcal{J}(t)) \\
= \; & \mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{R_0}\mathcal{H}_i(t)+\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)\right] \qquad (2.26)\\
= \; & \sum_{i=1}^{R_0}\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\sum_{j=1}^{\mathcal{M}_i(t)}\mathcal{N}_i(t,a_i)\middle|\mathcal{M}(t)\right]\right]+\sum_{i=1}^{R_0}\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\sum_{j=1}^{\mathcal{M}_i(t)}\mathcal{N}_i(t,a_i)\middle|\mathcal{M}(t)\right]\right] \\
& +\mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)\middle|\mathcal{K}(t)\right]\right]+\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{\mathcal{K}(t)}\mathcal{N}_i(t)\middle|\mathcal{K}(t)\right]\right] \\
= \; & \sum_{i=1}^{R_0}\mathbb{V}\mathrm{ar}\left[\mathcal{M}_i(t)\mathbb{E}[\mathcal{N}_i(t,a_i)]\right]+\sum_{i=1}^{R_0}\mathbb{E}\left[\mathcal{M}_i(t)\mathbb{V}\mathrm{ar}[\mathcal{N}_i(t,a_i)]\right] \\
& +\mathbb{V}\mathrm{ar}\left[\mathcal{K}(t)\mathbb{E}[\mathcal{N}(t)]\right]+\mathbb{E}\left[\mathcal{K}(t)\mathbb{V}\mathrm{ar}[\mathcal{N}(t)]\right] \\
= \; & \sum_{i=1}^{R_0}\mathbb{V}\mathrm{ar}\left[\mathcal{M}_i(t)\left[(a_i-1)e^{-\mu t}+1+\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\right]\right] \\
& +\sum_{i=1}^{R_0}\mathbb{E}\left[\mathcal{M}_i(t)\left[(a_i-1)e^{-\mu t}\left(1-e^{-\mu t}\right)+\frac{\lambda}{\mu}\left(1-e^{-\mu t}\right)\right]\right] \\
& +\mathbb{V}\mathrm{ar}\left[\mathcal{K}(t)\left(\frac{\lambda}{\mu}+1\right)\right]+\mathbb{E}\left[\mathcal{K}(t)\frac{\lambda}{\mu}\right] \\
= \; & \sum_{i=1}^{R_0}e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left(1-e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\right)\left((a_i-1)e^{-\gamma t}+1+\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right)^2 \\
& +\sum_{i=1}^{R_0}e^{-\gamma t}p^{\lfloor t+c_i\rfloor}\left((a_i-1)e^{-\mu t}\left(1-e^{-\mu t}\right)+\left(1-e^{-\mu t}\right)\frac{\lambda}{\mu}\right) \\
& +\left(\frac{\lambda}{\mu}+1\right)^2\tau t q_t+\frac{\lambda}{\mu}\tau t q_t. \qquad (2.27)
\end{aligned}
$$

Les résultats appliqués à la base de données de l'assureur étudié après 1 an, 5 ans et 10 ans apparaissent au tableau 2.5. En plus de l'espérance, nous pouvons y voir un variation de plus ou moins l'écart-type du nombre de voitures assurées, montrant ainsi la variabilité possible des résultats.

En prenant la limite de l'espérance ou en utilisant la fonction génératrice lorsque

t	1	5	10
$\mathbb{E}(\mathcal{J}(t)) + \sqrt{\mathbb{V}\text{ar}(\mathcal{J}(t))}$	240 414	228 915	225 932
$\mathbb{E}(\mathcal{J}(t))$	240 075	228 379	225 355
$\mathbb{E}(\mathcal{J}(t)) - \sqrt{\mathbb{V}\text{ar}(\mathcal{J}(t))}$	239 737	227 843	224 779

Tableau 2.5 Espérance plus ou moins la l'écart-type du nombre de voitures assurées

$t \to \infty$, nous pourrions voir que le nombre de ménages espérés converge vers une certaine valeur quand t augmente. De plus, nous verrions qu'aucune saisonnalité n'est observée. Un modèle plus près de la réalité aurait eu un taux de nouvelles arrivées qui est fonction du temps. Malheureusement, comme nous le mentionnions plus tôt, un taux d'entrée comme celui-ci aurait enlevé certaines propriétés markoviennes du modèle et aurait ainsi compliqué les équations puisqu'il s'agirait d'un modèle de type $G/G/\infty$.

Même si le nombre espéré de nouveaux ménages par année est élevé, l'écart-type sur le nombre de voitures est faible. Visuellement, nous pouvons observer le nombre de voitures attendu par année à la figure 2.3, où il peut être remarqué que l'écart-type sur le nombre de véhicules n'augmente que très peu au fil du temps. L'ajout de l'erreur d'estimation dans les prédictions serait un élément à analyser dans de futures recherches.

Le second graphique de cette figure montre le nombre espéré de voitures provenant de nouveaux assurés ($\mathcal{J}_N$) et le nombre espéré de voitures des anciens assurés ($\mathcal{J}_A$). Évidemment, la somme des deux variables nous donne le nombre total de voitures assurées, comme l'indique l'équation $\mathcal{J}(t) = \mathcal{J}_N(t) + \mathcal{J}_A(t)$. Dans le cas de cet assureur, le nombre de voitures total demeure sensiblement le même, où nous voyons que les anciens ménages sont rapidement remplacés par des nouveaux.

Variables			Voitures au 31/12/2007		Voitures à l'état stationnaire	
x_1	x_2	x_3	Nombre	Pourcentage	Nombre	Pourcentage
1	1	1	1 929	0,8%	556	0,2%
1	1	0	4 432	1,7%	773	0,3%
1	0	1	17 977	7,6%	9 141	4,0%
1	0	0	30 040	12,3%	7 741	3,4%
0	1	1	6 579	2,6%	6 375	2,8%
0	1	0	17 657	6,8%	17 302	7,7%
0	0	1	48 605	20,0%	55 221	24,5%
0	0	0	119 445	48,2%	128 735	57,0%
Total			246 664	100,0%	225 845	100,0%

Tableau 2.6 Proportion du nombre de voitures selon les type de ménage

Ainsi, après seulement 5 ans, la moitié des voitures espérées provient de nouveaux ménages, ce qui montre l'importance pour l'assureur de continuer ses campagnes de promotion pour attirer de nouveaux clients.

Conséquemment, il peut être intéressant d'analyser si ce comportement est le même pour tous les types d'assurés. Dans un premier temps, le tableau 2.6 offre un aperçu de $\mathcal{J}(t)$ pour $t = 0$ et $t \to \infty$. La figure 2.3 montre l'évolution du nombre de voitures assurées par profil de risque.

Nous pouvons remarquer que la composition du portefeuille changera avec le temps, et ce à cause de grands changements pour certains profils. Il est permis de croire que le taux d'arrivée de certains ménages a été historiquement différent dans le passé comparativement à ce qu'il est selon nos analyses. Ainsi, si ces assurés sont profitables, une analyse marketing pourrait permettre de cerner quels changements aux politiques de l'entreprise ont générés une telle différence.

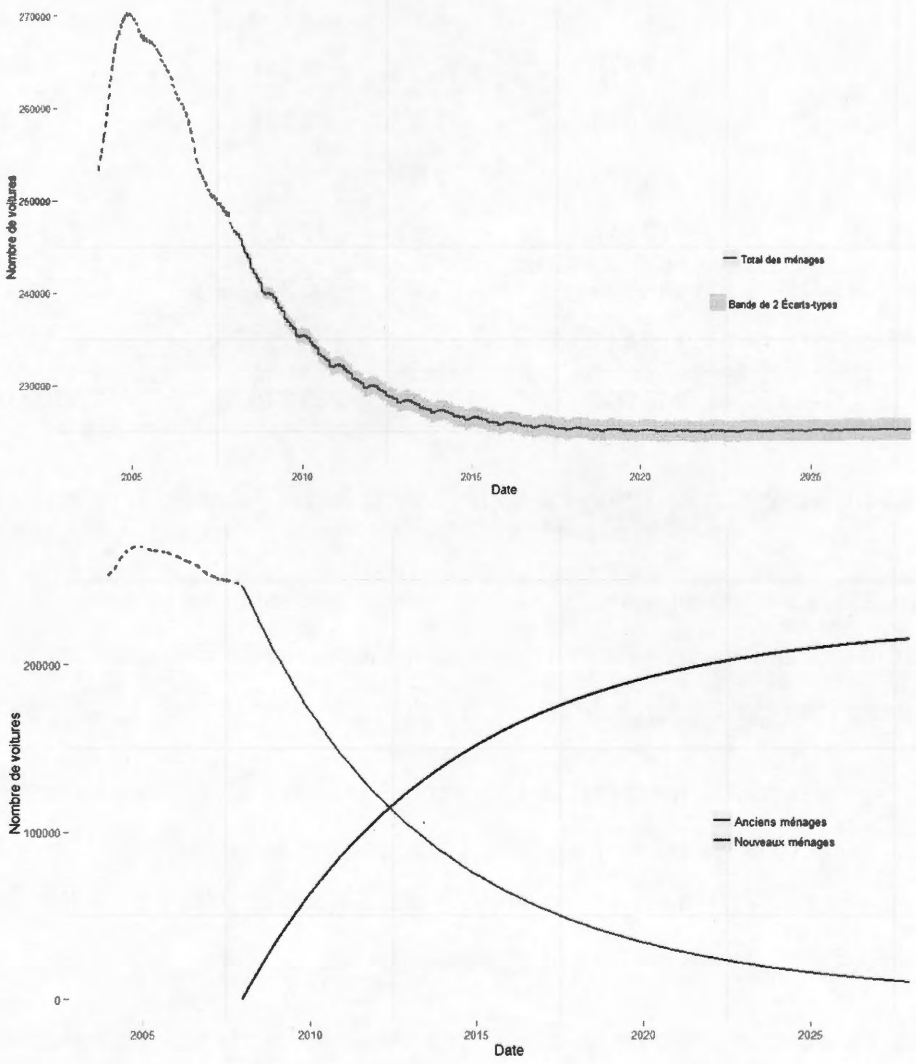

Figure 2.3 Nombre de voitures assurées dans le passé et espérées dans le futur

2.5.2 Valeur à risque du nombre de voitures au temps t

Un autre résultat pouvant être extrait des fonctions génératrices est la valeur à risque (VaR). Si nous voulons calculer la VaR ou percentiles, nous pouvons utiliser les probabilités cumulatives telles que : $\Pr(\mathcal{J}(t) \leq l)$. Il est aisé de transformer les fonctions génératrices de probabilités trouvées précédemment pour obtenir les fonctions de répartition. Supposons que la série génératrice de la probabilité cumulative de la variable $\mathcal{J}(t)$ est $\Upsilon_{\mathcal{J}(t)}(z,t)$. La relation avec la fonction génératrice $P_{\mathcal{J}(t)}(z,t)$ est trouvée en effectuant ces opérations :

$$\begin{aligned}
\Upsilon_{\mathcal{J}(t)}(z,t) &= \sum_{l=0}^{\infty} \Pr(\mathcal{J}(t) \leq l) z^l \\
&= \sum_{l=0}^{\infty} \sum_{j=0}^{l} \Pr(\mathcal{J}(t) = j) z^l \\
&= \sum_{j=0}^{\infty} \sum_{l=j}^{\infty} Pr(\mathcal{J}(t) = j) z^l \\
&= \sum_{j=0}^{\infty} \frac{\Pr(\mathcal{J}(t) = j) z^j}{1-z} \\
&= \frac{P_{\mathcal{J}(t)}(z,t)}{1-z}.
\end{aligned} \tag{2.28}$$

Sous la condition que $z < 1$, cette relation tient pour toutes les FGP trouvées précédemment. Elle permet d'aisément calculer les probabilités cumulatives par ses dérivées, où :

$$\Pr(\mathcal{J}(t) \leq l) = \frac{\partial^l \Upsilon_{\mathcal{J}(t)}(z=0,t)}{\partial z^l}. \tag{2.29}$$

Puisque les FGP trouvées sont parfois complexes, des méthodes numériques sont nécessaires, et doivent être développées.

2.5.2.1 Méthode des différences finies

Pour les probabilités d'être à un faible niveau, la méthode des différences finies est avantageuse, étant donné sa simplicité. Elle se base sur le quotient de différence de Newton qui mène au calcul différentiel, telle que la dérivée de $\Upsilon(z,t)$ par rapport à z est égale à

$$\frac{\partial \Upsilon_{\mathcal{J}(t)}(z,t)}{\partial z} = \lim_{h\to 0} \frac{\Upsilon_{\mathcal{J}(t)}(z+h,t) - \Upsilon_{\mathcal{J}(t)}(z,t)}{h}. \tag{2.30}$$

Maintenant, un algorithme numérique peut faire l'approximation de cette formule en choisissant un h assez petit. Pour les dérivées d'ordre supérieur, la même approche peut être effectuée. Par exemple, pour une dérivée d'ordre 2, nous avons :

$$\begin{aligned}
\Upsilon''_{\mathcal{J}(t)}(z,t) &\approx \frac{\Upsilon'_{\mathcal{J}(t)}(z+h,t) - \Upsilon'_{\mathcal{J}(t)}(z,t)}{h} \\
&\approx \frac{\Upsilon'(z+2h,t) - \Upsilon'_{\mathcal{J}(t)}(z+h,t)}{h^2} - \frac{\Upsilon'_{\mathcal{J}(t)}(z+h,t) - \Upsilon'_{\mathcal{J}(t)}(z,t)}{h^2} \\
&= \frac{\Upsilon_{\mathcal{J}(t)}(z+2h,t) - 2\Upsilon_{\mathcal{J}(t)}(z+h,t) + \Upsilon_{\mathcal{J}(t)}(z,t)}{h^2}.
\end{aligned} \tag{2.31}$$

Ensuite, pour la formule générale, nous avons que la dérivée d'ordre n par le même développement donne :

$$\Upsilon^{(n)}_{\mathcal{J}(t)}(z) \approx \sum_{j=0}^{n} \frac{\Upsilon_{\mathcal{J}(t)}(z+jh)}{h^n} \binom{n}{j} (-1)^{n-j}. \tag{2.32}$$

Il y a quelques versions de la méthode des différences finies qui existent. Celle-ci se nomme "décentrée en amont" (traduction libre de *forward*), puisque les fonctions sont évaluées au point à dériver (z) et à un certain point plus à droite $(z+h)$.

Une autre méthode possible pour des fonctions dérivables dans un certain intervalle $(z-\epsilon, z+\epsilon)$ autour du point z consiste à évaluer la fonction à dériver en deux points $(z-h, z+h)$. L'approximation de la n-ème dérivée selon cette méthode, nommée "approximation centrée", devient :

$$\Upsilon^{(n)}_{\mathcal{J}(t)}(z) \approx \sum_{j=0}^{n} \frac{\Upsilon_{\mathcal{J}(t)}\left(z + \left(\frac{n}{2} - j\right)h, t\right)}{h^n} \binom{n}{j} (-1)^{j}. \tag{2.33}$$

Pour ces deux façons de calculer, il faut faire une division avec, au dénominateur, un terme très petit pour obtenir une précision considérable. Pour des valeurs de n assez élevées, il faut que la valeur de h^n demeure assez élevée afin de demeurer dans les nombres pouvant être représentés par le programme informatique utilisé. Dans le contexte de compagnie d'assurance, cette méthode s'applique difficilement pour le calcul de probabilités puisque nous sommes intéressés aux valeurs de $n > 10000$.

2.5.2.2 Algorithme de Lattice-Poisson

L'algorithme de Lattice-Poisson se base sur l'inversion de la FGP grâce à une méthode qui s'apparente aux transformées de Fourier. Cette méthode est d'ailleurs décrite par (Abate et Whitt, 1992). Pour pouvoir extraire les probabilités de niveau élevé, cette méthode excelle par sa rapidité. Elle permet également de préalablement définir la précision voulue. Comme d'autres méthodes de type transformées de Fourier, il faut utiliser les nombres complexes lorsque les calculs sont effectués. Ceci peut d'ailleurs poser problème pour les logiciels n'ayant pas de méthodes d'évaluations pour ces valeurs. Lorsque $l \geq 1$, l'algorithme est le suivant :

$$\Pr(\mathcal{J}(t) \leq l) = \frac{1}{2lr^l} \sum_{k=1}^{2l} (-1)^k \Re\left(\Upsilon_{\mathcal{J}(t)}(z = re^{\frac{\pi k i}{l}}, t)\right) + \epsilon, \tag{2.34}$$

où $\Re(\cdot)$ est la fonction qui retourne la partie réelle et ϵ est l'erreur commise. Avec cet algorithme, l'erreur peut être bornée de telle sorte que :

$$\epsilon \leq \frac{r^{2l}}{1 - r^{2l}} \approx r^{2l}. \tag{2.35}$$

Pour que cette borne soit valide, il faut que la série génératrice n'aie que des termes inférieurs ou égaux à 1, comme c'est le cas pour les équations génératrices $P(z,t)$ et $\Upsilon(z,t)$. Pour des petites valeurs de r, le dénominateur est près de 0, ce qui explique l'approximation. Un r est choisi de sorte qu'il apporte une précision voulue.

t	1	2	5
$VaR_{0,05}(\mathcal{J}(t))$	33 565	62 053	123 806
$VaR_{0,95}(\mathcal{J}(t))$	34 306	63 061	125 232

Tableau 2.7 Valeur à risque du nombre de voitures assurées des nouveaux ménages

L'inclusion des ménages déjà présents se fait difficilement à cause de la lenteur des calculs. En effet, une multiplication de toutes leurs fonctions génératrices doit se faire avant d'effectuer la somme. Ainsi, seulement les résultats du nombre de voitures assurées des nouveaux ménages $\mathcal{J}_N(t)$ apparaissent à la table 2.7. Dans ce cas, les résultats confirment la faible variabilité du modèle due aux relations avec la distribution de Poisson.

2.5.3 Valeur actuelle des profits futurs

Dans la section 1.5.2, nous avons fait le calcul de la valeur vie client. Pour obtenir ce résultat, nous avons additionné les potentiels profits futurs (supposés de 1$ par voiture) qui ont ensuite été actualisés. Maintenant, si nous appliquons cette méthode sur tous les clients qui sont présentement assurés et sur ceux qui le seront dans le futur, nous obtenons une valeur totale du portefeuille. Cette valeur porte parfois le nom d'équité client et permet d'avoir une opinion sur la santé du portefeuille d'assurés de l'entreprise.

En notant par Φ cette valeur de portefeuille, par un développement similaire à

l'équation(1.37), nous pouvons en retrouver la valeur espérée. Ainsi :

$$\begin{aligned}
\mathbb{E}(\Phi) &= \int_0^\infty \mathbb{E}(\mathcal{J}(t))e^{-\delta t}dt \\
&= \int_0^\infty \sum_{i=1}^{R_0} \mathbb{E}(H_i(t))e^{-\delta t} + \left(\frac{\lambda}{\mu}+1\right)\tau t q_t e^{-\delta t}dt \\
&= \sum_{i=1}^{R_0}\int_0^\infty \mathbb{E}(H_i(t))e^{-\delta t}dt + \left(\frac{\lambda}{\mu}+1\right)\int_0^\infty \tau t q_t e^{-\delta t}dt \qquad (2.36)
\end{aligned}$$

Bien que la première intégrale, $\int_0^\infty \mathbb{E}(H_i(t))e^{-\delta t}dt$ avait déjà été calculée avec l'équation (1.37), ce résultat doit être ajusté selon les hypothèses de ce chapitre, c.-à-d. qu'il ne peut pas y avoir 0 voiture assurée chez un ménage. Nous obtenons alors :

$$\begin{aligned}
&\int_0^\infty \sum_{i=1}^{R_0} \mathbb{E}(H_i(t))e^{-\delta t}dt \\
&= \sum_{i=1}^{R_0}\left(a_i - 1 - \frac{\lambda}{\mu}\right) e^{-(1-c_i)(\gamma+\delta+\mu)} \frac{pe^{-(\gamma+\delta+\mu)}\left(1-e^{-(\gamma+\delta+\mu)}\right)}{\left(1-pe^{-(\gamma+\delta+\mu)}\right)(\gamma+\delta+\mu)} \\
&\quad + \left(\frac{\lambda}{\mu}+1\right) e^{-(1-c_i)(\gamma+\delta)} \frac{pe^{-(\gamma+\delta)}\left(1-e^{-(\gamma+\delta)}\right)}{\left(1-pe^{-(\gamma+\delta)}\right)(\gamma+\delta)} \\
&\quad + \left(a_i - 1 - \frac{\lambda}{\mu}\right) \frac{1-e^{-(1-c_i)(\gamma+\delta+\mu)}}{\gamma+\delta+\mu} \\
&\quad + \left(\frac{\lambda}{\mu}+1\right) \frac{1-e^{-(1-c_i)(\gamma+\delta)}}{\gamma+\delta}. \qquad (2.37)
\end{aligned}$$

Pour ce qui est de la seconde intégrale, nous avons :

$$\begin{aligned}
&\int_0^\infty \tau t q_t e^{-\delta t} dt \\
&= \int_0^\infty \frac{\tau}{\gamma}\left[\left(1-e^{-\gamma}\right)\frac{1-\left(pe^{-\gamma}\right)^{\lfloor t\rfloor}}{1-pe^{-\gamma}} + p^{\lfloor t\rfloor}\left(e^{-\gamma\lfloor t\rfloor} - e^{-\gamma t}\right)\right] e^{-\delta t} dt \\
&= \sum_{i=0}^{\infty}\int_i^{i+1} \frac{\tau}{\gamma}\left[\left(1-e^{-\gamma}\right)\frac{1-\left(pe^{-\gamma}\right)^{i}}{1-pe^{-\gamma}} + p^{i}\left(e^{-\gamma i} - e^{-\gamma t}\right)\right] e^{-\delta t} dt \\
&= \sum_{i=0}^{\infty}\frac{\tau}{\gamma}\left[\left(1-e^{-\gamma}\right)\frac{1-\left(pe^{-\gamma}\right)^{i}}{1-pe^{-\gamma}}\frac{e^{-\delta i}\left(1-e^{-\delta}\right)}{\delta}\right. \\
&\qquad \left. + \left(pe^{-(\gamma+\delta)}\right)^{i}\left(\frac{1-e^{-\delta}}{\delta} - \frac{1-e^{-(\delta+\gamma)}}{\delta+\gamma}\right)\right] \\
&= \frac{\tau}{\gamma}\left[\frac{1-e^{-\gamma}}{\delta\left(1-pe^{-\gamma}\right)}\left(1 - \frac{1-e^{-\delta}}{1-pe^{-(\gamma+\delta)}}\right) + \frac{1-e^{-\delta}}{\delta\left(1-pe^{-(\gamma+\delta)}\right)}\right. \\
&\qquad \left. - \frac{1-e^{-(\delta+\gamma)}}{(\gamma+\delta)\left(1-pe^{-(\gamma+\delta)}\right)}\right]
\end{aligned} \tag{2.38}$$

En utilisant un δ égal à 0.02, les résultats apparaissent dans le tableau 2.8 pour chacun des types de ménages. Afin d'avoir un aperçu de la provenance des profits, ils ont été séparés selon les nouveaux et les anciens assurés.

De la table 2.8, nous observons que la valeur totale du portefeuille d'assurance est approximativement 11.3 millions. De ce nombre, seulement 1.3 vient des anciens assurés, signifiant que seulement une petite proportion des assurés présentement actifs participe aux profits à long terme de l'entreprise. Ainsi, notre analyse peut être utilisée pour justifier que plus d'efforts doivent être faits pour attirer de nouveaux assurés, comparativement aux efforts pour garder les assurés actuels.

Toutefois, cette analyse ne spécifie pas ce qui considéré comme un nouveau ménage assuré. Par exemple, un assuré annulant son contrat pourrait vouloir être assuré encore par la même compagnie quelques années plus tard. La base de données utilisée dans notre analyse ne permet pas de différencier les nouveaux clients des clients passés qui sont revenus. Donc, les efforts pour garder les clients présents

Variables			Anciens ménages		Nouveaux ménages		Total	
x_1	x_2	x_3	Valeur	%	Valeur	%	Valeur	%
1	1	1	6 481	0,5%	25 867	0,3%	32 349	0,3%
1	1	0	22 156	1,7%	35 630	0,4%	57 786	0,5%
1	0	1	62 223	4,7%	426 630	4,3%	488 852	4,3%
1	0	0	149 871	11,4%	342 266	3,4%	492 137	4,3%
0	1	1	27 754	2,1%	299 241	3,0%	326 995	2,9%
0	1	0	105 349	8,0%	757 722	7,6%	863 071	7,6%
0	0	1	210 379	16,0%	2 486 015	24,8%	2 696 394	23,8%
0	0	0	730 249	55,6%	5 644 101	56,3%	6 374 350	56,3%
Total			1 314 462	100%	10 017 472	100%	11 331 934	100%

Tableau 2.8 Équité client selon les divers profils d'assurés

ne devraient pas être minimisés suite à notre analyse.

2.6 Conclusion

Dans ce deuxième chapitre, il a été question de généraliser le modèle du premier chapitre. Notre méthode proposée modélise le nombre de ménages pour la totalité du portefeuille d'un assureur. En se basant encore une fois sur un modèle de files d'attente, ce modèle suppose que chaque assuré suit un processus homologue à celui du premier chapitre, mais permet également l'ajout de nouveaux ménages assurés selon un processus de Poisson. Finalement, le modèle final a été fait en imbriquant deux modèles différents de files d'attente, ce qui a permis la réutilisation de théories développées plus tôt.

Une première ébauche du modèle avait le problème de permettre aux ménages de s'assurer une première fois sans voitures. Ceci aurait pu être interprété de façon symbolique, mais il a été préférable de changer le support de la variable aléatoire. Cette modification a été bénéfique car la distribution du nombre de voitures des nouveaux assurés est beaucoup plus près des observations empiriques.

Ce modèle complet nécessite 5 paramètres pour être utilisé : 4 paramètres ayant une fonction analogue à ceux du premier chapitre et un cinquième et nouveau paramètre correspondant au taux d'arrivée de nouveaux ménages. Bien que ces 4 paramètres aient la même utilité qu'avant, l'estimation de ceux-ci a dû être faite de nouveau suivant les hypothèses propres à ce chapitre. Ainsi, une nouvelle fonction de vraisemblance a pu être construite, qui utilise entres autres l'information contenue dans le nombre de voitures initial.

Le modèle a encore une fois été généralisé pour permettre d'utiliser des variables explicatives dans chacun des paramètres du modèle. Toutefois, pour le taux d'arrivée, nous avons utilisé les caractéristiques du risque de manière différente. Les

paramètres ainsi calculés ont été utilisés pour générer divers statistiques pertinentes pour une compagnie d'assurance. Le concept d'équité client a été exploré. Cela nous a permis de remarquer que certains types d'assurés, bien que correspond à une grande partie du portefeuille actuel, ne seront la cause que d'une très petite part des profits futurs de l'entreprise. Un assureur pourrait donc orienter ses politiques envers ces assurés afin d'analyser cette perte de profits potentiels.

CONCLUSION

Les différents modèles avaient comme but principal de recréer les comportements possibles des assurés ayant un effet sur le nombre de voitures assurées. Nous avons commencé par des modèles généraux pour un seul assuré, en allant progressivement vers un modèle complet du nombre de voitures assurées par une compagnie d'assurance. Dans le premier chapitre, la théorie des files d'attente a modélisé le mouvement des voitures sur un contrat d'assurance. Dans le second chapitre, cette même théorie a été utilisée pour modéliser le mouvement des ménages pour un portefeuille d'assurés chez un assureur. À chaque étape, la fonction génératrice des probabilités a été calculée. Les paramètres ont été calculés par maximum de vraisemblance pour chacun des modèles.

Bien entendu, les modèles de ce mémoire reposent sur d'importantes hypothèses et simplifications. Cela a permis d'obtenir des résultats qui étaient simples et analytiques. Pour aller vers des modèles plus complets et réalistes, une première supposition pouvant être libérée est celle des distributions des temps d'attente. Par exemple, en remplaçant l'hypothèse de distribution exponentielle par une Erlang, nous permettons une plus grande flexibilité de la variance tout en gardant certaines propriétés markoviennes. Une autre hypothèse exponentielle a été utilisée pour la distribution du temps entre l'arrivée de deux nouveaux ménages. Celle-ci ne prend pas en compte ni les variations saisonnières, ni la croissance de l'entreprise et est difficile à justifier en analysant les résultats à long terme.

Un autre point qui aurait pu être étudié est la relation entre la survenance des sinistres et le taux de départ chez l'assuré. Les modèles de tarification supposent

qu'un sinistre occasionne nécessairement une augmentation de la prime. Cette augmentation pourrait ainsi motivé l'assuré à changer d'assureur afin d'avoir une prime plus basse. L'information à propos de la sinistralité de chaque assuré est déjà présente dans les bases de données pour ce type d'analyse et pourraient donc facilement être incluses dans un modèle plus complet. Finalement, les variables explicatives choisies ne forment qu'une petite partie de toutes celles possibles. Bien que certaines variables choisies semblaient importantes, celles qui n'ont pas été choisies pourraient peut-être réveler des informations importantes.

RÉFÉRENCES

Abate, J. et Whitt, W. (1992). Numerical inversion of probability generating functions. *Operations Research Letters*, *12*, 245 – 251.

Benes, V. E. (1957a). Fluctuations of telephone traffic. *Bell System Technical Journal*, *36*, 965–973.

Benes, V. E. (1957b). A sufficient set of statistics for a simple telephone exchange model. *Bell System Technical Journal*, *36*, 939–964.

Boucher, J.-P., Denuit, M. et Guillén, M. (2009). Number of accidents or number of claims ? an approach with zero-inflated poisson models for panel data. *Journal of Risk and Insurance*, *76*(4), 821–846.

Donkers, B., Verhoef, P. et Jong, M. (2007). Modeling clv : A test of competing models in the insurance industry. *Quantitative Marketing and Economics*, *5*(2), 163–190.

Farris, P. W., Bendle, N. T., Pfeifer, P. E. et Reibstein, D. J. (2010). *Marketing Metrics : The Definitive Guide to Measuring Marketing Performance (2nd Edition)* (2 éd.). Pearson Prentice Hall.

Guelman, L., Guillén, M. et Pérez-Marín, A. M. (2014). A survey of personalized treatment models for pricing strategies in insurance. *Insurance : Mathematics and Economics*, *58*(0), 68 – 76.

Guelman, L., Guillén, M. et Pérez-Marín, A. M. (2015). Uplift random forests. *Cybernetics and Systems*, *46*(3-4), 230–248.

Guillén, M., Nielsen, J. P., Scheike, T. H. et Pérez-Marín, A. M. (2012). Time-varying effects in the analysis of customer loyalty : A case study in insurance. *Expert Systems with Applications*, *39*(3), 3551 – 3558.

Lemaire, J. (1977). La soif du bonus. *Astin Bulletin*, *9*(181), 90.

Newell, G. (1982). *Applications of queueing theory.* Monographs on statistics and applied probability. Chapman and Hall.

Rust, R. T., Lemon, K. N. et Zeithaml, V. A. (2004). Return on marketing : using customer equity to focus marketing strategy. *Journal of marketing*, *68*(1), 109–127.

Verhoef, P. C. et Donkers, B. (2001). Predicting customer potential value an application in the insurance industry. *Decision Support Systems*, *32*(2), 189 – 199.

LEXIQUE

Cette liste contient les symboles couramment utilisés dans le mémoire. Ils ont été placés en ordre alphabétique, incluant les symboles grecs, qui ont été mis selon leur nom français.

A : Nombre d'annulation de contrats

A^* : Variable indiquant si le ménage est entré dans l'état inconnu

a : Nombre de voitures initiales pour un ménage

AS : Indicatrice de l'état inconnu

β_θ : Vecteur de réponse du paramètre θ

B : Nombre de nouveaux ménages

C : Constante d'intégration

c : Constante permettant à la fonction $\lfloor t+c \rfloor$ de changer à chaque renouvellement de contrat

E : Nombre de voitures ajoutées au contrat

$G(T, \Omega)$: Probabilité qu'il y ait un ajout de voiture durant l'état inconnu

γ : Taux de bris de contrat

$\mathcal{H}(t)$: Nombre de voitures assurées au temps t pour un ménage

J_j Nombre total de voitures assurées avant l'événement j

$\mathcal{J}(t)$: Nombre total de voitures assurées au temps t par l'assureur

$\mathcal{J}_A(t)$: Nombre total de voitures assurées au temps t par l'assureur venant d'anciens ménages

$\mathcal{J}_N(t)$: Nombre total de voitures assurées au temps t par l'assureur venant de nouveaux ménages

K : Nombre d'événements affectant le nombre de voitures

K^* : Nombre d'événements observés affectant le nombre de voitures

$\mathcal{K}(t)$: Nombre de ménages arrivés après l'état initial qui sont toujours présents chez l'assureur au temps t

κ : Temps avant l'arrivée d'un nouveau ménage

λ : Taux d'ajout de voitures

$\mathcal{M}(t)$: Variable du statut du ménage indiquant si le ménage est assuré (1) ou non (0) au temps t

μ : Taux de retrait d'une voiture pour un ménage assuré

$\mathcal{N}(t)$: Nombre de voitures au temps t d'un ménage

N_j : Nombre de voitures d'un ménage immédiatement avant l'événement j

$o(\Delta t)$: Valeur négligeable par rapport à Δt

Ω : Date d'observation de la base de données

ω : Valeur vie client avec un horizon de temps infini

ω_t : Valeur vie client avec un horizon de temps t

$P_{\mathcal{N}(t)}(z,t)$: Fonction génératrice des probabilités de la variable $\mathcal{N}(t)$

$P_{\mathcal{N}(t),\mathcal{M}(t)}(z,y,t)$: Fonction génératrice des probabilités des variables $\mathcal{N}(t)$ et $\mathcal{M}(t)$

p : Probabilité de renouvellement de contrat

$p_i(t)$: Probabilité qu'il y ait i voitures au temps t pour un ménage

$p_{(j,i)}(s,t)$: Probabilité qu'il y ait i voitures au temps t, sachant qu'il y en avait j assurées au temps s

$\boldsymbol{p(t)}$: Vecteur des probabilités de nombre de voitures au temps t

Φ : Valeur du portefeuille

Ψ_j : Temps avant un événement de type j

Q : Nombre de renouvellements de contrat

$\boldsymbol{Q}$: Matrice des taux de transition

q_t : Composante de service dans les modèles de file d'attente $M/G/\infty$

R : Nombre de renouvellements dans l'état inconnu

$\mathcal{R}(t)$: Nombre de ménages présents initialement chez l'assureur qui sont toujours présents au temps t

R_0 : Nombre de ménages présents initialement chez l'assureur

S : Nombre de voitures retirées du contrat

$S(t)$: Fonction de survie du temps de service

T : Durée de temps observé d'un ménage

t_j : Temps d'occurrence de l'événement j

$\widetilde{t}_j$: Durée de temps entre l'événement $j-1$ et l'événement j

τ : Taux d'arrivée de nouveaux ménages

U : Intervalle de temps analysé dans la base de données

$\Upsilon_{\mathcal{J}(t)}(z,t)$: Série génératrice de la fonction de répartition de $\mathcal{J}(t)$

V : Somme des temps d'exposition des voitures pour un ménage

W_j : Nombre de ménages assurés avant l'événement j

$\mathcal{W}(t)$: Nombre de ménages présents chez l'assureur au temps t

$\boldsymbol{X}$: Vecteur des variables explicatives

ξ : Nombre de ménages analysés dans la base de données

$\mathcal{Y}(t)$: Nombre de voitures retirées du ménage du temps 0 jusqu'au temps t

$\mathcal{Z}(t)$: Nombre de voitures ajoutées au ménages du temps 0 jusqu'au temps t