Contribution à l'étude de la prévention en assurance santé

Romain Gauchon

N°d'ordre NNT : 2020LYSE1072

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° accréditation 486
SCIENCES ECONOMIQUES ET DE GESTION

Spécialité de doctorat : Science de Gestion
Discipline : Actuariat

Soutenue publiquement le 05/05/2020, par :
Romain Gauchon

Contribution à l'étude de la prévention en assurance santé

Devant le jury composé de :

Didier Rullière, Maître de conférence à l'université Lyon 1	Président
Meglena Jeleva, Professeure à l'université de Paris Nanterre	Rapporteure
Joël Wagner, Professeur à l'université de Lausanne	Rapporteur
Alexandra Dima, Chercheure à l'université Lyon 1	Examinatrice
Montserrat Guillèn, Professeure à l'université de Barcelone	Examinatrice
Mickaël Schwarzinger, CEO de la société THEN	Examinateur
Stéphane Loisel, Professeur à l'université Lyon 1	Directeur de thèse
Jean-Louis Rullière, professeur à l'université Lyon 1	Co-directeur de thèse
Pierre Arnal, ADDACTIS Groupe,	Invité
Céline Blattner, ADDACTIS France,	Invitée

Remerciements

Tout d'abord, j'aimerai remercier chaleureusement Alexandra Dima, Montserrat Guillén, Didier Rullière et Michaël Schwarzinger d'avoir accepter de faire partie de mon jury. Vous êtes tous des chercheurs que j'estime beaucoup et je suis heureux que vous puissiez utiliser la diversité de vos compétences et de vos parcours dans l'évalutation de mes travaux.

En particulier, je suis très reconnaissant à Meglena Jeleva et Joël Wagner de m'avoir aider à améliorer ce document par la qualité de leurs corrections, suggestions et échanges.

J'exprime toute ma reconnaissance à Stéphane Loisel et Jean-Louis Rullière pour la grande confiance qu'ils m'ont accordée en acceptant de me prendre sous leurs ailes. Plus que des directeurs de thèses, ils ont été pour moi des guides qui m'ont permis d'éviter les méandres et les pièges de la thèse. Leurs connaissances infinies m'ont permis de satisfaire (temporairement !) ma curiosité, que je pensais pourtant insatiable.

J'ai aussi une pensée pour toutes les personnes qui m'ont côtoyé lors de mon séjour à ~~Actuaris~~ addactis. Vous êtes trop nombreux pour être tous cités, mais merci notamment à Audrey, Aymeric[2], Cécile, Céline, Donasian, Guerric, Léo, Mathieu, Nabil, Pierre et Thibault pour tous ces moments échangés ensemble autour de boîtes noires. J'aimerais en particulier complimenter Alexandra pour son management parfait, curieuse, intéressé et toujours à l'écoute de nouvelles idées et à la recherche de nouvelles couleurs, et JP pour m'avoir soutenu dans toutes mes idées plus farfelues.

Un grand merci à tous les doctorants et post-docs dont la bonne humeur et la franchise m'aura été d'un grand réconfort lors des moments difficiles. J'aimerais en particulier remercier Sarah et ses efforts répétés pour tous nous réunir autour d'activités communes, Steve et ses relectures si pertinentes, Morgane de la team expé, Pierre M. pour ses conseils lors de tentatives de dérivées hasardeuses, et Maximilien pour les discussions sur l'IA, même si bien sûr je n'oublie pas Annani, Antoine, Arthur, Baptiste, Bezad, Carine, Claire, Franck, Gilles, Glawdys, Inès, Jean de dieu, Julien, Karim, Marwa, Nesrine, Patrick, Pierre C., Pierrick, Rihem, Saker, Shuwa, Tachfine. Si vous êtes peut-être trop nombreux pour les bureaux de l'ISFA, vous ne l'êtes pas pour mon coeur.

J'en profite pour remercier tous les permanents du laboratoire, pour leur disponibilité, et en particulier Denis Clot, l'homme qui parlait aux ordinateurs, et Christian Robert, pour m'avoir donné l'envie de faire une thèse.

Je remercie aussi chaleureusement les différents chercheurs qui ont eu la gentillesse de partager leur temps avec moi afin de m'aider dans les différents travaux que l'on a menés conjointement. Je pense notamment à Julien Trufin, que je complimente pour sa grande patience et sa réactivité. Je pense aussi à l'ensemble des chercheurs que j'ai croisés grâce à la chaire Prevent'Horizon, à Jean-Yves Lesueur, à Alexandra Dima, à Julie et Frédéric Haesebaert, à Sofia el Ousoul et à Luiza Siqueira do Prado. Un énorme merci à Marine Brocard pour tout le travail administratif qu'elle abat et pour avoir maternés les enfants que nous sommes.

J'aimerais aussi remercier toute ma clique d'amis, grâce à qui je ne suis jamais vraiment tout seul. Du lycée (Etienne, Jeff, Thomas...), de la Martinière (Clément, Lambert, Lucie, Marine, Quentin, Rémy, Ségolène, Willy...), de l'ISFA (Adrien, Alexis, Anthony, Haïtem, Jérôme, Laureline, Nicolas, Paul, Pauline (et son petit mari), Robin, Thibaud, Vincent...). En particulier, j'aimerais remercier Mickaël pour avoir considérablement enrichi ma pensée et Antoine qui m'a appris à me battre pour mes amis. C'est grâce à vous tous que j'ai pu arriver au bout de mon parcours d'étudiant, vous êtes géniaux !

J'aimerais avoir un petit mot pour les diverses personnes que j'ai pu rencontrer au fil de ma vie et qui ont aussi participé à faire de moi qui je suis aujourd'hui : mes compagnons de galère japonaise, l'équipe de Lyon DataScience, JV, la communauté irlandaise lyonnaise, Sinossan, Aryonnar, LowBob, Nima, Cécile, Indra, Mme Mollin, Mme Vessière, M. Lacharmoise et les rôlistes que j'ai eu la chance de côtoyer.

Enfin, je voudrais remercier l'ensemble de ma famille, mes quatres frères et soeurs, mes deux papas / mamans, mes grands-parents et mes oncles / tantes / cousins / cousines pour leur soutien absolu au cours de ses 3 années de thèse. Je profite de ces remerciements pour dire quelque chose que je ne dis pas assez : je vous aime !

落花枝に帰らず
La fleur qui tombe ne peut
retrouver sa branche
Proverbe japonais

Résumé

Cette thèse porte sur la mise en place d'actions de prévention financées par une compagnie d'assurance. Elle est composée de cinq chapitres précédés d'une introduction générale qui a pour vocation de présenter les difficultés liées à la prévention, les outils utilisés et les principaux résultats obtenus.

Le chapitre 1 propose une méthode de classification non supervisée d'assurés santé par groupes de risques homogènes, à partir des prestations versées par un organisme assureur. Cette méthode comporte deux phases : une réduction de dimension des données à l'aide de factorisations de matrice positive (NMF) est d'abord effectuée. La classification est ensuite finalisée à l'aide des cartes de Kohonen. Les tests de la méthode, sont aussi présentés. Les classes finales obtenues sont finalement analysées afin d'étudier si certaines d'entre elles peuvent faire l'objet d'une action de prévention. Une action de prévention sur la psychiatrie est proposée.

Le chapitre 2 est la continuité du chapitre 1 puisqu'il s'attache à comparer la qualité de la réduction de dimension à l'aide des méthodes NMF avec celle obtenue grâce à deux autres méthodes, les méthodes Word2Vec (W2V) et les autoencodeurs débruiteurs empilés marginalisés (mSDA). Notamment, la stabilité des classifications finales est étudiée à l'aide d'une nouvelle mesure de stabilité. Un complément sur la prise en compte de la temporalité avec l'algorithme W2V est également présenté.

Le chapitre 3 propose une étude du risque psychiatrie au sein d'un organisme complémentaire santé, les algorithmes des chapitres précédent ayant permis d'identifier ce risque. A l'aide d'une étude statistique menée sur quatre bases de données, il est notamment montré que les assurés consommant de la psychiatrie coûtent en moyenne deux fois plus cher à l'assureur santé qu'un individu moyen. Quelques actions de prévention potentielles sont suggérées en conclusion.

Le chapitre 4 s'intéresse à la modélisation de la prévention au sein d'une compagnie d'assurance. En intégrant un paramètre de prévention au sein du modèle Poisson composé issu de la théorie de la ruine, il est en effet possible de mesurer l'effet de la prévention sur certains indicateurs, comme la probabilité de ruine. Différentes stratégies optimales de prévention sont proposées, et une analyse de sensibilité est fournie.

Enfin, le chapitre 5 propose d'étendre le modèle considéré dans le chapitre précédent au cas où une compagnie d'assurance est confrontée à un risque léger et à un risque lourd. Dans un tel modèle, la stratégie optimale de prévention dépend des montants de réserves constituées. Des résultats asymptotiques sur les stratégies optimales sont fournis.

Mots-clés : Assurance santé ; Prévention ; Clustering ; Théorie du risque ; Psychiatrie ; Data science

Abstract

In this PhD thesis is studied the implementation of an insurance program by a private insurance company. The manuscript starts with a general introduction, where are presented the difficulties linked to this topic along with the main results and the technical tools used in the other chapters.

In Chapter 1 is described a health policyholders clustering process. This method use the individual health spending in order to create homogeneous risk clusters. The whole process is in two steps : a dimension reduction is first performed using a Nonnegative Matrix Factorization (NMF) algorithm, preceding a clustering made using the Kohonen's map method. The validation process of this method is also presented. It includes tests done on a text corpus database. Final clusters are analysed in order to be allow a better understanding of their makeup.

Chapter 2 completes Chapter 1 by challenging NMF dimension reduction algorithm with two alternative methods : Word2Vec (W2V) and the marginalized Stacked Denoising Autoencoders (mSDA). The final cluster stability is studied using a new stability measure. The ability to tackle the temporality of the Word2Vec algorithm is also considered.

In Chapter 3 is studied the psychiatric risk for an health insurer. A statistical analysis of 4 database is made. It is shown that in France, someone receiving refunds for a psychiatric spending costs overall 2.5 times more than the average policyholder. Potential prevention actions are proposed to conclude this Chapter.

In Chapter 4 is presented a way to model the impact of a prevention program for an insurance company. This model is based on the addition of a prevention parameter into the compound Poisson model. It is then possible to study the prevention effect on some chosen indicators, such a ruin probability. Two optimal prevention strategies are presented, and a sensibility analysis is provided.

In Chapter 5 is proposed an extension of the model presented in Chapter 4. The insurance company is now supposed to face two different risk, a light one and a heavy one. In such a model, the optimal strategy is depending of the initial reserve amounts. Some asymptotic results on optimal strategies are provided.

Keywords: Health insurance; Prevention; Clustering; Risk theory; Psychiatry; Data science

Table des matières

Liste des figures

Introduction Générale

I) Avant-propos

La France possède un système de santé unanimement reconnu dans le monde pour son efficacité. Preuve de cette réussite, en 2017, d'après Eurostat (eurostat, 2019), **l'espérance de vie à la naissance pour les femmes en France était de 85,6 ans,** soit la seconde plus élevée de l'Union européenne.

Le système français repose sur une participation importante de l'Etat, qui prend en charge près de 80% des dépenses de santé en France (Boisguérin et al., 2016), principalement via l'assureur public qu'est la Sécurité Sociale. De plus, le système s'appuie sur des complémentaires santé privées afin de compléter le remboursement des dépenses, essentiellement pour des soins non vitaux, comme les chambres particulières d'hôpital, l'optique ou les prothèses dentaires. En 2015, 13,3% des dépenses santé étaient prises en charge par des mutuelles et autres organismes de complémentaire santé.

L'action commune de l'Etat et de ces organismes complémentaires permet ainsi une bonne prise en charge financière des patients, favorisant l'augmentation de l'espérance de vie (Nixon and Ulmann, 2006). Pourtant, un tel système orienté essentiellement sur la prise en charge des coûts semble avoir favorisé une vision curative de la médecine, au détriment de l'approche préventive [1].

Ainsi, toujours selon Eurostat, **l'espérance de vie en bonne santé** pour les femmes françaises n'est que de **64,9 ans**, soit le dixième rang européen ! À titre de comparaison, les femmes suédoises ont une espérance de vie à la naissance de 84,1 ans, mais une espérance de vie en bonne santé de 71,9 ans, soit une différence beaucoup plus faible.

Les pathologies en France commencent donc tôt et sont prises en charge sur de longues durées, entrainant une explosion des coûts associés. Une étude de la Drees montre que le vieillissement de la population française (et donc la progression de l'espérance de vie) provoque une croissance de près de 1% par an des dépenses liées aux affections longues durées (ALD)

1. En 2017, les dépenses à but préventif atteignaient 6 milliards d'euros, soit seulement 3% des dépenses de soin totales (DREES, 2018).

(Grangier, 2018). Ainsi, la Drees a évalué à 199,3 milliards d'euros la consommation totale de biens et de soins médicaux en France en 2017.

Ces chiffres soulignent que la santé peut être vue comme un bien privé et comme un bien public. En effet, un individu dont la santé se dégrade en est évidemment très affecté : les dépenses liées au traitement d'un cancer peuvent sembler dérisoires par rapport à la souffrance engendrée par cette maladie. Cependant, les coûts associés ne peuvent être complètement négligés et sont en France principalement portés par l'ensemble de la société.

D'un point de vue économique, si l'on considère la santé sous l'angle "bien public", une approche préventive qui tend à limiter les montants des dépenses de soin, menée par les pouvoirs publics ou par des groupements d'individus impactés (mutuelles) est donc cohérente. C'est dans cette optique qu'en 2018 le gouvernement a mis en place le plan "Priorité prévention", qui présente 25 mesures différentes, dont le remboursement des dispositifs de sevrage tabagique ou la création d'une campagne de dépistage du cancer du col de l'utérus.

À ces efforts publics peuvent se joindre les efforts des organismes complémentaires, via la mise en place de remboursement de dépenses préventives, la mise en place d'applications téléphoniques ou de services d'accompagnement des individus et des entreprises. L'objectif est triple : diminuer les montants des prestations versées, améliorer leur image et dégager de la croissance dans un secteur ultra-concurrentiel (en faisant payer cette offre de service sur la prévention). Pourtant, dans ce domaine, les initiatives des organismes complémentaires restent très timides, freinées par un grand nombre de contraintes pratiques.

Cette thèse se propose donc d'étudier certains aspects de la faisabilité de la mise en place d'actions de prévention par une compagnie d'assurance, afin d'aider à résoudre principalement deux difficultés pratiques rencontrées : la modélisation de la prévention et le ciblage d'une action de prévention. Elle est composée de cinq chapitres, chacun ayant fait l'objet d'une soumission pour publication [2], et sont éventuellement complétés d'une annexe en français pour préciser quelques résultats. Bien que certains chapitres de la thèse puissent s'appliquer à tout type de prévention, les travaux présentés ici concernent surtout l'assurance santé.

Cette introduction est constituée de quatre parties. La première partie vise à introduire plus en détail le contexte de la prévention en assurance. Après avoir défini le sujet, elle expose les difficultés rencontrées en pratique par un assureur. Celles-ci ont servi de motivation pour la réalisation des différents chapitres de la thèse. L'aspect médical de la prévention sera très peu abordé, à l'exception notable du cas de la psychiatrie, présenté en fin de cette première section. En effet, la psychiatrie semble être un excellent champ d'application de la prévention en assurance, et sera considérée dans le chapitre 3 de la thèse.

2. Le chapitre 4 a été accepté pour publication à la revue IME, et les chapitres 1 et 3 sont en cours de révision.

L'une des difficultés rencontrées étant le manque d'outil de modélisation, les chapitres 4 et 5 de cette thèse ont pour but d'adapter les modèles classiques de théorie du risque afin qu'ils puissent prendre en compte la prévention. La seconde partie de l'introduction s'attache donc à présenter les différents outils utilisés dans ce manuscrit, en partant des modèles de Poisson composés, puis en étendant le modèle à des cadres plus compliqués (prise en compte des dividendes, de la taille des queues de distribution,...) .

Une fois les outils théoriques présentés, la troisième partie de l'introduction présentera les algorithmes d'apprentissage statistique utilisés en pratique dans les chapitres 1 et 2 de cette thèse. Leur emploi a pour objectif de répondre à une seconde problématique rencontrée par les assureurs : le manque de données médicales qui leur sont disponibles afin, par exemple, de cibler une action de prévention. Afin de répondre à cette problématique, les chapitres 1 et 2 de la thèse présentent la mise en place de procédures de classification des assurés en assurance santé, afin d'améliorer la connaissance du portefeuille. Cette troisième partie de l'introduction commencera donc par présenter les algorithmes de clustering nécessaires à la compréhension des autres chapitres de thèse. Elle présentera ensuite diverses méthodes de réduction de dimension utilisées afin d'améliorer les résultats. Enfin, elle présentera rapidement le domaine du text mining, qui a inspiré les différentes architectures de classification présentées dans les deux premiers chapitres de thèse.

Enfin une dernière partie présentera plus en détail les chapitres et les apports de la thèse.

II) La prévention dans un contexte d'assurance

Cette partie vise à présenter la prévention, et sa perception en économie, notamment les liens avec l'activité assurantielle, afin de donner une idée générale du contexte de cette thèse. Ainsi, seuls des résultats principaux et quelques exemples d'études sont présentés ici. Le lecteur pourra se référer à Courbage et al. pour une revue plus détaillée de la littérature économique sur la prévention (Courbage et al., 2013).

Comme dit précédemment, cette partie commence par définir l'objet d'étude de la thèse : la prévention. Elle présente ensuite les freins pratiques rencontrés par les assureurs lors de la mise en oeuvre d'une action de prévention. Ces freins sont tous des raisons évoquées par les praticiens pour expliquer le manque d'action ambitieuse initiée par les assureurs. Du fait de l'ampleur du sujet, le contexte médical de la prévention dans son ensemble ne peut être évoqué ici. En revanche, un focus sur la psychiatrie est effectué en fin de cette partie de l'introduction, puisque celle-ci semble particulièrement pertinente dans le cadre de l'assurance et qu'elle fait l'objet du chapitre 3 de la thèse.

II-A) Définition(s) de la prévention

Afin de bien appréhender le sujet d'étude de cette thèse, il convient tout d'abord de le définir précisément. Le dictionnaire Larousse propose la définition suivante du verbe prévenir :

"**Prévenir** : Prendre les mesures nécessaires pour éviter un mal, un danger."

Ainsi, la prévention serait l'ensemble des actions permettant d'éviter un mal, un danger, autrement dit permettant de limiter un risque. Déjà, cette définition très simple fait ressortir l'intérêt de la prévention pour un organisme assureur : elle pourrait même mot pour mot faire partie de la description de la gestion des risques. Néanmoins, la définition proposée par le Larousse est large. Le milieu médical propose des définitions plus précises de la prévention. Il distingue généralement quatre types de prévention : les trois premiers, semble-t-il proposés par Gruenberg en 1953 (Gruenberg, 1953) sont définis comme suit par l'OMS :

"**Prévention primaire :** mesures applicables à une maladie ou groupe de maladies pour en bloquer les causes avant qu'elles n'agissent sur l'homme ; en d'autres termes, pour empêcher la survenue de la maladie". Il s'agit donc de l'ensemble des actions se pratiquant avant la survenance d'un risque. L'exemple typique de prévention primaire est le vaccin contre la grippe.

"**Prévention secondaire :** mesures destinées à interrompre un processus morbide en cours pour prévenir de futures complications et séquelles, limiter les incapacités et éviter le décès". Cette fois, il s'agit de détecter le plus tôt possible la survenance d'un risque avant qu'il n'empire. Les actions de dépistage d'une maladie illustrent bien cette catégorie.

"**Prévention tertiaire :** ensemble de mesures visant à permettre aux personnes handicapées de recouvrer leurs fonctions initiales ou d'utiliser au maximum les capacités qui leur restent ; la réadaptation comprend à la fois des interventions individuelles et des actions sur l'environnement." Souvent, la prévention tertiaire est vue de manière plus globale, comme l'ensemble des actions visant à diminuer les conséquences et les rechutes d'une maladie. Par exemple, la Haute Autorité de Santé dit de la prévention tertiaire qu'elle "agit sur les complications et les risques de récidive". Au sens le plus large, elle comprend donc l'ensemble des activités curatives.

À ces trois définitions traditionnelles s'ajoute celle de la prévention quaternaire proposée dans les années 1990 notamment par Jamoulle et Roland (Jamoulle and Roland, 1995). La prévention quaternaire a pour objectif de limiter la surmédicalisation des patients.

Il est important de noter que, suite aux suggestions d'Ehrlich et Becker, les économistes utilisent des définitions de la prévention différentes (Ehrlich and Becker, 1972). Ils distinguent deux types de prévention :

L'auto-protection : Il s'agit de l'ensemble des actions de prévention visant à diminuer **la probabilité** de survenance d'un risque.

L'auto-assurance : Il s'agit de l'ensemble des actions de prévention visant à diminuer **le coût** de survenance d'un risque.

Les liens précis entre les concepts économiques et médicaux ont été pointés par Dervaux et Eeckhoudt (Dervaux and Eeckhoudt, 2004) et Courbage et Rey (Courbage and Rey, 2016), mais on peut globalement considérer que l'auto-protection correspond à la prévention primaire et l'auto-assurance aux préventions secondaire et tertiaire.

Les chapitres 4 et 5 traiteront uniquement d'auto-protection, alors que les autres chapitres sont plus généraux.

II-B) Les difficultés de la prévention en assurance

Comme souligné lors de la partie précédente, la prévention semble être un outil naturel à la disposition des compagnies d'assurances. Pourtant, le mot prévention est rarement associé au domaine de l'assurance par le grand public, et les initiatives dans ce domaine de la part de l'ensemble des acteurs de l'assurance santé sont longtemps restées timides, notamment à cause des difficultés techniques détaillées ci-dessous[3].

Quels programmes de prévention pour les sociétés d'assurances ?

Le premier défi que rencontre un assureur santé voulant mettre en place une action de prévention est de choisir quel type d'action mettre en place. En effet, il est nécessaire de se demander sur quel risque il est le plus pertinent d'agir, et quels programmes de prévention peuvent être efficaces pour cela.

Les organismes assureurs peuvent pour ce faire compter sur une littérature médicale très développée afin de les aider dans cette tâche. De nombreux programmes ont ainsi été évalués et comparés par des médecins. Certains sites internet spécialisés, comme la plateforme "Health evidence" animée par l'université de Mac Master au Canada, centralisent les différentes ressources disponibles afin de guider un éventuel financeur d'une action de prévention dans ce choix.

Les possibilités de programme sont alors multiples et les sociétés d'assurance peuvent tester des actions de prévention dans des domaines divers. AG2R a ainsi mené une campagne d'information et de dépistage des pathologies dentaires auprès des apprentis boulangers. IPECA

3. Il est néanmoins nécessaire de noter que de plus en plus de projets sur la prévention portés par des organismes assureurs voient le jour. En effet, dans un secteur comme la santé ou les marges sont extrêmement petites, les acteurs se tournent petit à petit vers la proposition de service, notamment liés à la prévention, afin d'arriver à dégager un nouvel axe de croissance.

a ouvert un espace de prévention, composé d'une cabine de téléconsultation, d'un centre d'ophtalmologie et d'un centre d'ostéopathie offert aux salariés d'Airbus à Blagnac.

Plusieurs organismes ont axé leurs efforts sur la facilitation de l'accès au sport. C'est ainsi que la mutuelle Just, la MGEN, le groupe Pasteur mutualité ou Harmonie mutuelle ont tous mis en place, sous diverses formes, la possibilité de rembourser une partie des dépenses sportives.

De plus, de nombreux acteurs développent des applications santé en lien avec la prévention. Ainsi, Eovie mutuelle a créé un partenariat avec médecin direct afin de faciliter l'accès aux soins, et donc de limiter le renoncement à ces mêmes soins. Generali, en collaboration avec Discovery, a lancé le programme Vitality, qui permet de gagner des points synonymes de bons d'achats pour les personnes adoptant un comportement sain. Le programme Emprunteur 625 de Afi Esca propose une application qui mesure l'activité sportive et entraîne des diminutions des primes d'assurance.

Enfin, comme souligné dans la quatrième édition de l'étude BVA-Réhalto "Comprendre pour agir", l'accompagnement au retour au travail des personnes arrêtées depuis longtemps est très efficace. Ainsi, de nombreuses sociétés proposant des assurances collectives offrent ce type de service.

Cette problématique n'est pas l'objet principale de cette thèse. Néanmoins, une action concrète de prévention du risque psychiatrique est proposée en ouverture à la fin du chapitre 1

Le choix d'un programme de prévention approprié est la première difficulté rencontrée par un assureur. La seconde est l'estimation de l'impact de ce programme.

Il est difficile de modéliser et d'évaluer l'impact de prévention au cours du temps

Une fois un programme de prévention déterminé, il est nécessaire d'évaluer s'il sera assez efficient pour être mis en place. L'estimation d'un retour sur investissement peut ainsi convaincre les différents décideurs de l'organisme assureur d'investir dans la mise en place de l'action de prévention. L'utilisation de modèles appropriés peut permettre de répondre à cette problématique. Cependant, il existe très peu de modèles liant prévention et assurance.

Le modèle le plus connu de ce type est celui proposé par Ehrlich et Becker (Ehrlich and Becker, 1972). Ce modèle s'intéresse aux décisions d'un individu effectuant ses choix à l'aide d'une fonction d'utilité u de type Von Neumann et Morgenstern. Cet agent possède une richesse initiale R.

Les deux auteurs supposent que l'individu a la possibilité de fournir un effort e de prévention. Dans le cas de l'auto-protection, ils supposent que l'agent est confronté à un risque ayant une probabilité $p(e)$ de se réaliser, p étant une fonction décroissante de l'effort de prévention, et qui lui coûtera L en cas de réalisation.

Ils supposent aussi que l'agent a la possibilité d'acheter une quantité α d'assurance proportionnelle, à un prix de $\alpha(1+\lambda)p(e)L$, λ désignant le taux de chargement fixé par la compagnie d'assurance. La prime d'assurance dépendant de e, ils supposent donc que l'effort de prévention est observable. Le modèle est en deux périodes : dans la première, l'agent décide combien il investit en assurance et en d'auto-protection (ou auto-assurance), et dans la seconde le risque se réalise ou non. Finalement, l'individu cherche à calculer

$$\underset{\alpha,e}{\text{Argmax}}[(1-p(e))u(R-e-\alpha(1+\lambda)p(e)L)+p(e)u(R-e-\alpha(1+\lambda)p(e)L-(1-\alpha)L)]. \quad (1)$$

Dans le cas de l'auto-assurance, les auteurs considèrent que l'individu devra payer un montant $L(e)$ en cas de réalisation du risque. Cependant, ils supposent désormais la prévention inobservable. Le problème de l'agent devient alors de trouver

$$\underset{\alpha,e}{\text{Argmax}}[(1-p)u(R-e-\alpha(1+\lambda)pL)+pu(R-e-\alpha(1+\lambda)pL-(1-\alpha)L(e))]. \quad (2)$$

Dans les deux cas, il est alors possible de calculer les dérivées partielles par rapport à α et e, puis d'utiliser le théorème des fonctions implicites afin d'étudier les liens entre assurance et auto-assurance ou auto-protection. Ehrlich et Becker montrent ainsi que l'assurance et l'auto-assurance sont substituables, mais que l'assurance et l'auto-protection sont complémentaires, à condition que le taux de chargement λ soit suffisamment proche de 0.

Ce modèle simple souffre cependant de plusieurs limitations. Tout d'abord, Ehrich et Becker font deux hypothèses difficilement compatibles dans leur article. En effet, ils supposent dans le cas de l'auto-protection que l'effort de prévention est connu de l'assureur (c'est pourquoi la prime dépend de e ci-dessus). En revanche, ils supposent que ce n'est pas le cas pour l'effort d'auto-assurance. De plus, le modèle ne considère que deux périodes temporelles, et le montant de sinistre (en cas de sinistre) est entièrement connu de l'individu avant qu'il ne prenne sa décision, ce qui n'est généralement pas vrai en pratique.

De nombreuses extensions ont été proposées. Par exemple, Courbage et ses coauteurs ont analysé le cas d'un individu soumis à plusieurs risques au lieu d'un seul (Courbage et al., 2017). Cela revient à considérer que l'assuré risque non seulement de perdre un montant L avec une probabilité p, mais aussi de perdre un montant G avec probabilité q. La corrélation entre les deux risques est modélisée par un paramètre k. Cependant, ils ne s'intéressent qu'au cas de l'auto-protection et ne considèrent pas la possibilité de s'assurer.

Lee propose d'étudier une action de prévention qui est à la fois de l'auto-assurance et de l'auto-protection (Lee, 1998), ce qu'il appelle "Self-insurance-cum-protection". Il étudie l'impact de l'aversion au risque sur l'effort de prévention, mais le modèle proposé ne prend pas en compte la possibilité de s'assurer.

Briys et al. supposent que la prévention a un effet incertain (par exemple, le détecteur de fumée ne fonctionnant pas) (Briys et al., 1991). Ils considèrent les cas de l'auto-assurance et auto-protection, mais ils n'autorisent l'individu à acheter une police d'assurance que dans le cas de l'auto-assurance.

Enfin, Konrad et Skaperdas proposent de se placer dans le cadre de l'utilité dépendante au rang (*rank dependent utility*) et de regarder l'impact de l'aversion au risque (Konrad and Skaperdas, 1993). La possibilité de s'assurer n'est en revanche pas prise en compte dans leur modèle.

La plupart des modèles de prévention sont dérivés de celui d'Erlich et Becker. Ils ne peuvent cependant pas être utilisés directement par une compagnie d'assurance pour évaluer l'impact d'une action de prévention puisqu'il s'agit de modèles s'intéressant à la demande et non à l'offre de prévention et d'assurance.

Il est ainsi difficile d'estimer à priori le retour sur investissement d'une action de prévention, ce qui a limité le nombre de programmes préventifs mis en place. En conséquence, il existe aussi très peu de travaux empiriques calculant l'impact d'une action de prévention sur une société d'assurance.

En revanche, plusieurs études évaluent le retour sur investissement d'un programme de prévention mis en place pour un employeur. Les effets de nombreux programmes sur la baisse des dépenses de santé, l'absentéisme et l'augmentation de la productivité ont ainsi été chiffrés (e.g. (Parks and Steelman, 2008), (Baicker et al., 2010), (Gubler et al., 2017)). Une étude plus large, menée par l'Agence Internationale de la Sécurité Sociale a évalué à 2,2 euros gagnés par euro investi le retour sur investissement de la prévention (Hans-Horst, 2011). De tels programmes de prévention pourraient être mis en place par un organisme assureur en collaboration avec les entreprises qu'elle assure. Cependant, le contexte international de ces études empêche généralement de transposer directement ces résultats en France.

La seconde étape que doit franchir l'assureur pour mettre en place une action de prévention est donc de se doter d'outils d'estimation de l'impact de la prévention. Cette problématique fera l'objet des chapitres 4 et 5 de cette thèse. Une fois le modèle mis en place, l'organisme assureur doit essayer de maximiser l'efficacité médicale et économique de son programme de prévention. Pour ce faire, il doit proposer la participation au programme uniquement aux assurés qui en bénéficieront le plus.

Il est difficile en France de savoir à qui proposer une action de prévention

Une fois un programme de prévention fixé et validé, il convient de se demander à quels assurés il faut le proposer. Si dans certains pays, les organismes assureurs peuvent facilement accéder à la donnée qui leur permettra de répondre à cette question (par exemple, aux USA, ils ont

le droit de demander un certificat médical avant d'activer le remboursement d'un dépistage), ce n'est pas le cas en France. Ainsi, une entreprise relevant du code de la mutualité ne peut pas poser un questionnaire médical lors de la souscription d'un assuré.

Pourtant, la prévention doit être ciblée afin d'être efficace, que ce soit par âge, par sexe, par milieu social, ou par antécédents médicaux (Brisson et al., 2016). Le ciblage permet à la fois de limiter les coûts (Yamin et al., 2016), les risques de surmédicalisation (Courbage and Rey, 2016) et d'avoir une gestion plus efficace des ressources disponibles (Mbah et al., 2014).

Aux anciens dispositifs sur la protection des données santé s'est de plus rajoutée récemment la nouvelle réglementation sur les données personnelles. Le Règlement Général de Protection des Données est en effet entré en application le 25 mai 2018. Il mentionne spécifiquement le ciblage à titre commercial, et impose des contraintes fortes sur les données santé, réduisant encore la marge de manœuvre des sociétés d'assurance sur le sujet.

Mettre en place une stratégie de prévention avec un accès à des données aussi limitées est donc un vrai défi. Afin d'y faire face, il est nécessaire d'améliorer la connaissance du risque représenté pas les assurés et de le décrire du mieux possible. Les chapitres 1 et 2 abordent cette problématique.

Proposer le programme de prévention aux personnes les plus pertinentes est donc un vrai enjeu pour les assureurs. Mais cela ne suffi pas : il faut aussi qu'elles y participent sérieusement.

Les résultats des programmes de prévention peuvent être sous optimaux, car les assurés n'y participent pas ou n'y adhèrent pas suffisamment.

Une fois que l'action de prévention est proposée à des assurés, il est encore nécessaire qu'ils acceptent d'y participer et qu'ils y adhèrent. Or, il est difficile de la part d'un assureur de convaincre ses assurés, notamment en raison d'un manque de proximité entre ces deux acteurs. Une trop faible participation a ainsi été à l'origine de l'échec de programmes de prévention d'origine privée, comme celui d'AXA utilisant les bracelets connectés Withings.

Ce manque de participation peut trouver sa source dans le manque de confiance des Français envers les sociétés d'assurance[4]. Mais cela ne semble pas suffisant pour comprendre la totalité du phénomène, puisque des actions d'origine médicale peuvent aussi être confrontées à une faible participation ou une faible adhésion (Newby et al., 2006).

La seconde raison pouvant expliquer cette difficulté est d'ordre cognitif : les individus n'arrivent pas à apprécier à leur juste valeur les bienfaits d'une action de prévention à cause de biais cognitifs. En reprenant la définition de Buss (Buss, 2015), les biais cognitifs sont des

4. Un sondage réalisé par YouGov pour LeComparateurAssurance.com en avril 2019 a ainsi estimé que 50% des Français ne faisaient pas confiance aux compagnies d'assurance.

situations dans lesquels la réflexion humaine conduit à une perception du monde différente de la réalité vis-à-vis de certains aspects [5].

Par exemple, le biais de préférence pour le présent décrit la propension qu'a l'être humain à surévaluer excessivement un bienfait immédiat par rapport à un bienfait futur. Les efforts préventifs sont souvent perçus comme pénibles (faire du sport, surveiller son alimentation, aller chez le médecin). C'est pourquoi l'individu aura tendance à préférer son confort immédiat (ne pas participer au programme de prévention) au confort futur incertain (ne pas être malade).

De plus, le cadre de vie et l'état d'esprit d'un individu peuvent évoluer avec le temps, obligeant les programmes de prévention à être eux aussi dynamiques, ce qui n'est pas toujours le cas (Etner and Jeleva, 2013).

Le biais de surconfiance peut lui aussi expliquer une baisse de la participation à un programme de prévention. Il décrit la tendance des individus à être trop confiants dans leur jugement ou leurs capacités. En matière de prévention, il s'exprime souvent via des raisonnements du type "cela n'arrive qu'aux autres" et peut expliquer une chute de participation au cours du temps (Malmendier et al., 2002). De manière générale, les individus ont du mal à évaluer les probabilités, en particulier sur des risques rares (voir par exemple Andersson and Lundborg (2007)).

Il est aussi très difficile à l'individu d'exploiter rationnellement ces probabilités. Les choix des individus dans un contexte incertain est un sujet de recherche actif depuis la proposition du modèle d'espérance d'utilité de Von Neuman et Morgenstern Morgenstern and Von Neumann (1953), qui supposait que l'individu faisait ses choix en pondérant l'utilité des états possibles par leur probabilité d'occurance. Depuis, les travaux de Kahneman et Tversky Kahneman and Tversky (1972) confirmés par des travaux expérimentaux comme ceux de Wu et gonzales Wu and Gonzalez (1996) ont montré que les individus évaluaient mal les probabilités et ne faisaient pas leurs choix en fonction d'une espérance. Différente théorie, dont la rank dependant utility (voir par exemple les travaux de Quiggin Quiggin (1991)). Ces mauvaises estimations des probabilités peuvent aussi expliquer une faible participation à un programme de prévention.

Pour résumer, afin que le programme de prévention puisse être un succès, il est donc nécessaire de choisir le programme adequat, d'estimer son impact, d'optimiser son effet en le proposant aux personnes les plus appropriées et de s'assurer que celles-ci sont bien motivées. Mais il reste une dernière contrainte afin que l'assureur observe une baisse effective de son risque : il faut qu'il assure encore les personnes ayant participées.

La mobilité des assurés peut rendre une action de prévention moins efficiente

5. Citation originale : "By cognitive bias, we mean cases in which human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality."

Une fois l'action de prévention mise en place, il reste donc un dernier défi à relever pour l'assureur : conserver les assurés qui ont participé.

En effet, la fluidité du marché de l'assurance est favorisée par les gouvernements successifs puisqu'elle fait partie des hypothèses majeures nécessaires à la perfection de la concurrence. C'est une des raisons qui ont poussé le parlement français à voter en mai 2019 la possibilité pour les assurés de résilier un contrat individuel de complémentaire santé à tout moment. Cependant, une mobilité importante des assurés a pour effet secondaire de freiner l'investissement préventif : pourquoi investir pour réduire les dépenses à long terme d'un assuré quand celui-ci peut quitter le portefeuille avant que l'entreprise n'ait pu récupérer les bénéfices de l'investissement ?

En France, il existe deux marchés de l'assurance : l'assurance collective, la plus importante en termes de chiffre d'affaires, et l'assurance individuelle. En assurance collective, la mobilité des assurés est plus élevée : l'assureur est soumis à la fois au risque qu'une entreprise quitte son portefeuille, mais les individus peuvent aussi partir du portefeuille lorsqu'ils changent d'entreprise. Ainsi, Cebul et al. ont estimé à 20% le turnover des assurés d'une société d'assurance santé aux USA (Cebul et al., 2011). Fang et Gavazza (Fang and Gavazza, 2011) et Herring (Herring, 2010) ont montré que cette mobilité entrainait un investissement sous optimal de la part des entreprises et des sociétés d'assurance, notamment en termes de prévention.

De plus, si le départ des assurés est un frein potentiel à la mise en place d'actions de prévention, l'arrivée de nouveaux assurés suite à de potentiels effets d'anti-sélection en est un autre. L'anti-sélection (ou sélection adverse) est un phénomène introduit par Rothschild et Stiglitz en 1978 (Rothschild and Stiglitz, 1978) et très étudié dans le domaine de l'assurance (voir par exemple Chiappori et Salaniè pour une étude théorique (Chiappori and Salanié, 2000), et Buchmueller pour une étude empirique en France (Buchmueller et al., 2004)). Il décrit le fait que certains assurés peuvent profiter d'informations sur eux-mêmes non connues par l'assureur, afin de souscrire un contrat moins cher que ce qu'il aurait été si la compagnie d'assurance avait eu connaissance de ces informations.

La mise en place d'actions de prévention ciblées dans le cadre d'un contrat est propice à l'apparition de ce phénomène : ceux qui souscrivent au contrat sont ceux qui vont utiliser ces services. Lorsqu'il s'agit de prévention tertiaire, cela peut nuire fortement au retour sur investissement de l'action de prévention. Ainsi, Beaulieu et al. ont observé qu'un programme de prévention sur le diabète pouvait avoir un retour sur investissement négatif au global alors qu'il est très positif au niveau individuel (Beaulieu et al., 2006).

Ainsi, l'organisme assureur rencontre des difficultés pratiques à la plupart des étapes de la mise en place d'une action de prévention. Cela a beaucoup freiné l'essor de la prévention en assurance. Si les organismes de complémentaires santé ont expérimenté dans quelques

domaines (notamment le domaine du sport), il reste beaucoup de secteurs ou tout reste encore à construire en matière de prévention. Par exemple, la santé mentale est identifiée dans le chapitre 3 comme un secteur qui pourrait se révéler comme propice à la mise en place d'un programme préventif, mais n'est que rarement considérée comme tel en pratique. La partie suivante a pour but de présenter le contexte de la psychiatrie en France et les coûts qui y sont associés.

II-C) Un champ possible d'application de la prévention en assurance : la psychiatrie

"Dans la cartographie médicalisée des dépenses de santé, le poids de la « santé mentale » est chaque année particulièrement marquant, tant en termes d'effectifs de patients concernés que de dépenses. Ainsi, en 2016, plus de 7 millions de personnes, soit plus d'un français sur dix, ont eu recours à des soins ou des prestations que l'on peut rattacher soit à une maladie psychiatrique soit à un traitement chronique par psychotrope. Les dépenses correspondantes se sont élevées à près de 20 milliards d'euros pour les bénéficiaires du régime général, une somme extrapolable à 23 milliards d'euros pour correspondre aux dépenses de tous les régimes, soit 14% des dépenses d'Assurance Maladie."

C'est en ces termes que la Sécurité Sociale introduit le chapitre entier concernant la psychiatrie dans ses "Propositions pour l'Assurance Maladie en 2019" (Assurance Maladie, 2018). Avec leur coût annuel de 20,6 milliards d'euros, les maladies psychiatriques sont en effet les pathologies les plus coûteuses pour l'Assurance Maladie, loin devant les maladies cardiovasculaires (16,6 milliards) ou les cancers (16,5 milliards) - et ces dépenses ne concernent que les coûts liés aux dépenses de santé. C'est pourquoi ce domaine représente un enjeu majeur pour l'assurance santé française. Cette partie vise à présenter le monde de la psychiatrie en général, avec un focus particulier sur le poids financier qu'elle représente pour la société et qui est notamment supporté par les assureurs publics et privés. L'ensemble des informations évoquées ici ont pour objectif d'exposer le contexte général du chapitre 3.

La psychiatrie : qu'est ce que c'est ?

En France, la psychiatrie fait l'objet d'une sorte de tabou, si bien que de nombreuses erreurs sont souvent commises à ce sujet. Notamment, plus de 40% des Français associent à tort la psychiatrie à la folie (IPSOS and Klesia, 2014). Il est donc nécessaire de bien définir le sujet afin de bien en comprendre les enjeux. Le dictionnaire médical de l'académie de médecine définit la psychiatrie ainsi :

"Discipline médicale destinée à l'étude, à la prévention, au traitement des maladies mentales et à la réadaptation des patients."

Elle définit aussi une maladie mentale comme une :

"Altération des capacités psychiques d'une personne. *Une maladie mentale peut altérer différentes fonctions, l'intelligence, la mémoire, le jugement, l'orientation temporospatiale, etc., isolées ou associées. La gravité peut varier depuis de légers troubles du comportement, des névroses, jusqu'à des psychoses graves rendant la personne dangereuse pour autrui et pour elle-même (agressivité, tendances suicidaires) et nécessitant l'hospitalisation sans internement.*"

La psychiatrie est donc l'étude de tout ce qui affecte la psyché, c'est-à-dire l'esprit. Les maladies psychiatriques se distinguent des maladies neurologiques, qui affectent physiquement le cerveau, de par leur dimension quasi immatérielle : elles sont souvent indiscernables par imagerie (Sandu et al., 2017). Les maladies psychiatriques comprennent généralement :

— Les troubles névrotiques et troubles graves du comportement (troubles anxieux, addictions graves, anorexie...)
— Les troubles de l'humeur (dépression, bipolarité...)
— Les psychoses (schizophrénies, paranoïa...)
— Les démences (Alzeihmer)
— Les retards mentaux

Les prévalences en Europe de chacun de ces troubles ont été estimées par Wittchen et al. (Wittchen et al., 2011). D'après cette étude, les dépressions, à elles seules, concerneraient environ 7% de la population, et les troubles névrotiques plus de 14%. La prévalence de ces maladies est généralement plus élevée chez les femmes et chez les personnes âgées. Lorsque la guérison est envisageable, elle reste aléatoire, les taux de rechutes étant très importants (Cruwys et al., 2013).

Si les maladies mentales concernent une part très élevée de la population, beaucoup ne sont jamais détectées. En fonction des études, le non-recours aux soins concerne ainsi entre 40% et 66% de la population (Demailly, 2011). La principale raison de ce non-recours aux soins provient de l'image associée à la psychiatrie dans l'imaginaire collectif. En effet, si les psychiatres sont considérés comme les bons interlocuteurs pour soigner les maladies mentales, celles-ci sont perçues négativement, étant souvent associées à tort à la violence, au viol ou à l'inceste par la population générale (Roelandt et al., 2010).

Les maladies mentales sont souvent soignées par des psychiatres (si la réponse adéquate est médicamenteuse) ou par des psychologues (via la psychothérapie). Les troubles les moins lourds peuvent tout simplement être suivis par un médecin généraliste (Debreu and Catteau, 2003). Ces maladies peuvent entrainer une hospitalisation. Les établissements concernés peuvent être les hôpitaux psychiatriques, mais aussi les hôpitaux classiques, les établissements de Soins de Suite et de Réadaptation, ou les hospitalisations à domicile.

Si ces maladies sont difficiles à guérir, il est en revanche possible de les prévenir. Plusieurs axes peuvent être adoptés afin de réduire la prévalence de ces pathologies (voir (Muñoz et al., 2012)

pour une revue de littérature). Ces actions de prévention, telles les actions de dépistage, sont souvent efficaces tant économiquement que médicalement (Mihalopoulos et al., 2011), (Knapp et al., 2011).

La psychiatrie en assurance santé

Comme beaucoup de dépenses de soins en France, le coût de la psychiatrie est essentiellement supporté par la Sécurité Sociale. La majorité des maladies psychiatriques[6] sont classées en Affection Longue Durée, permettant aux malades de bénéficier de remboursements accrus de la part de l'Assurance Maladie. Aux coûts directs des soins s'ajoutent des coûts indirects, liés aux arrêts de travail ou à la baisse de productivité. Ainsi, le coût total de la psychiatrie pour la société française s'élèverait à 109 milliards d'euros en 2012 (Chevreul et al., 2013).

Ces coûts élevés ont poussé la Sécurité Sociale à mener nombre d'études sur le sujet, et notamment celle présentée dans les Propositions pour l'Assurance Maladie de 2019 (Assurance Maladie, 2019). On y apprend que 7,2 millions de personnes seraient suivies pour une maladie psychiatrique, un chiffre important, mais nettement en deçà de ceux avancés dans la littérature médicale, certainement à cause du non-recours aux soins, mais aussi à cause des dépressions qui ne sont pas reconnues comme ALD. Un français pris en charge pour une maladie psychiatrique se voyait rembourser pour l'année 2016 en moyenne 6686 euros pour ses dépenses de soin contre 1126 euros pour un français moyen.

De plus, les maladies psychiatriques ont un risque de comorbidité très fort. À titre d'illustration, toujours selon le même rapport, la probabilité d'avoir une embolie pulmonaire ou de faire un AVC est deux fois supérieure, lorsque la personne à des antécédents psychiatriques. Bien sûr, de nombreuses raisons peuvent expliquer ces corrélations : les personnes atteintes d'une maladie psychiatrique peuvent adopter des comportements plus à risque (alcool, cigarette...), subir un stress excessif, ou être dans un état d'immunodéficience. Des risques liés aux effets secondaires des traitements peuvent également exister. À l'inverse, des facteurs de risque externe, comme la prise de drogue ou une maladie peuvent déclencher une pathologie psychique.

Les personnes souffrant d'au moins une maladie psychiatrique sont ainsi sujettes à une surmortalité très forte. Cela peut être dû aux comorbidités ou à un risque accru de suicide. 200 000 tentatives de suicide (dont 10 000 ont entrainé le décès) ont ainsi été enregistrées en France en 2014 (Ministère des Affaires sociales de la Santé et des Droits des femmes, 2014).

Si le coût des maladies psychiatriques pour la Sécurité Sociale est très bien documenté, il n'en est pas de même pour les complémentaires santé françaises. Rien n'a été publié sur le sujet. C'est ce qui a motivé le chapitre 3. Il y est montré que la psychiatrie représente un

6. À l'exception notable des dépressions, qui ne sont considérées comme Affection Longue Durée que si le malade a déjà fait deux dépressions par le passé.

surcoût important pour un assureur santé, malgrès le statut d'ALD dont bénéficient souvent les personnes concernées. En effet, un assuré qui a consulté un psychiatre au moins une fois dans l'année demande deux fois plus de remboursement que l'assuré moyen.

Les personnes souffrant de pathologie psychiatrique sont donc dispendieuses, globalement faciles à identifier, et des actions de prévention pertinentes pouvant être mises en place facilement existent tout en étant potentiellement efficaces économiquement. Pour les compagnies d'assurance, la psychiatrie apparaît donc comme un bon angle de prévention, même si des études supplémentaires sont nécessaires pour pouvoir conclure définitivement sur le sujet.

La psychiatrie peut donc être un élément de réponse parmi d'autres à la problématique "quelle action de prévention mettre en place". Cependant, cela ne permet pas de lever les autres freins que sont le manque de modèle prenant en compte la prévention, la difficulté de cibler une action pour un assureur, la nécessité d'avoir une adhésion aux programmes de prévention ou encore la possibilité que les personnes bénéficiant de la prévention partent à la concurrence. La prochaine section a pour objectif d'introduire des éléments de théorie du risque, qui permettront dans les chapitres 4 et 5 d'apporter des éléments de réponse à la problématique de la modélisation de la prévention.

III) Théorie de la ruine et liens potentiels avec la prévention

La théorie de la ruine a pour objet l'étude de l'évolution des réserves d'une société d'assurance au cours du temps. Si la norme Solvabilité 2 l'a mise sous les projecteurs en la plaçant au cœur du dispositif réglementaire du calcul des provisions, cette théorie a vu le jour bien avant, au tout début du vingtième siècle. Elle a été l'objet d'un nombre très important de recherches, mais l'accent est ici mis sur les outils majeurs utilisés dans les chapitres 4 et 5. Pour approfondir, le lecteur pourra se référer au livre d'Asmussen et Albrecher (Asmussen and Albrecher, 2010).

Cette introduction à la théorie de la ruine commence par la présentation du modèle le plus standard : celui de Poisson composé. Une modification de ce modèle afin de prendre en compte l'auto-protection, tel que détaillé dans les chapitres 4 et 5 est aussi proposée. Une extension de ces modèles afin d'introduire le versement de dividende sera ensuite proposée, puisqu'elle permet de mesurer les gains issus de la prévention. Deux autres extensions du modèle portant sur les lois des sinistres seront aussi introduites : les modèles multi-risques, qui permettent de considérer une entreprise confrontée à diverses sources de risque dont elle n'en prévient que certains, et les modèles de réassurance, puisqu'ils permettent par analogie de modéliser l'auto-assurance. Enfin, les notions de queues de distribution et d'ordres stochastiques seront présentées : ces deux notions permettent de comparer et de mieux caractériser les risques, et seront donc utiles dans le chapitre 5 pour étudier l'impact de la prévention.

III-A) Le modèle de Poisson composé

Le modèle de Poisson composé est le plus simple, et le plus ancien, des modèles étudiés en théorie de la ruine. Il a été d'abord proposé par Lundberg en 1903 (Lundberg, 1903), puis développé notamment par Cramer en 1930 (Cramér, 1930). C'est ce modèle que nous avons choisi de modifier afin de prendre en compte la prévention.

Ce modèle résume une compagnie d'assurance à quatre éléments :

- Un montant de réserves initiales $u \in \mathbb{R}_+$.
- Un montant de prime gagnée par unité de temps c.
- Des montants de sinistres $(X_i)_{i\in\mathbb{N}^*}$, supposés i.i.d. et suivant la même loi qu'une variable aléatoire positive X admettant pour fonction de répartition F_X. Il est de plus supposé que le risque est assurable, c'est-à-dire que $\mathbb{E}(X_i) = \mu < \infty$.
- Un processus de survenance des sinistres modélisé par un processus de Poisson N_t de paramètre $\lambda \in \mathbb{R}_+^*$. Ce processus est supposé indépendant des montants de sinistre X_i.

Dans ce modèle, les réserves disponibles pour l'assureur varient dans le temps, suivant le processus

$$R_t = u + ct - \sum_{i=1}^{N_t} X_i, \tag{3}$$

avec la convention $\sum_{i=1}^{N_t} X_i = 0$ si $i = 0$. Ce processus R_t est l'objet d'étude de la théorie de la ruine.

En particulier, la théorie va régulièrement s'intéresser à la probabilité que l'entreprise soit un jour ruinée, appelée probabilité de ruine ψ (et à la probabilité qu'elle ne soit jamais ruinée φ) de la compagnie d'assurance : $\psi(u) = 1 - \varphi(u) = \mathbb{P}(\exists t > 0 \ tel \ que \ R_t < 0 | R_0 = u)$. Evidemment, cet indicateur de ruine est à prendre au sens large et peut par exemple modéliser la probabilité que la compagnie passe en dessous des seuils réglementaires, poussant l'Autorité de Contrôle Prudentiel et de Résolution à la mettre sous tutelle.

Afin de pouvoir espérer ne jamais être ruinée, l'entreprise a besoin d'être rentable. Ainsi, si le rapport S/P est plus grand que 1, la ruine est certaine : si $\lambda\mu > c$, alors, $\forall u \geq 0,\ \varphi(u) = 0$. Il est donc nécessaire de supposer que

$$\lambda\mu < c. \tag{4}$$

Cette condition est appelée la condition du **chargement de sécurité**. Il est alors possible de calculer explicitement les probabilités de ruine : La probabilité de non-ruine sans réserve initiale vaut

$$\varphi(0) = 1 - \frac{\lambda\mu}{c}. \tag{5}$$

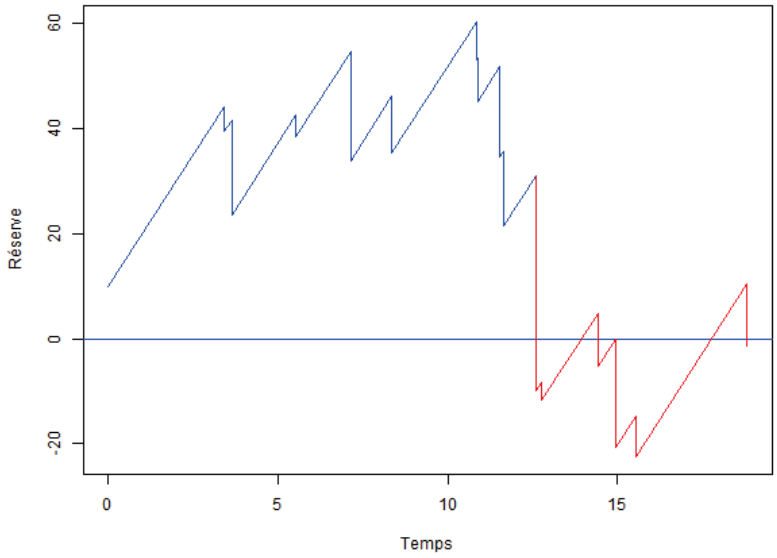

FIGURE 1 – Exemple d'un processus des réserves. Chaque sinistre agit comme un saut du processus.

La probabilité de non-ruine lorsque le montant de réserves initiales vaut $u > 0$ est donnée par

$$\varphi(u) = \varphi(0) \sum_{i=0}^{\infty} (1 - \varphi(0))^n F_{e,X}^{*n}(u), \tag{6}$$

avec

$$F_{e,X} = \frac{1}{\mu} \int_0^u 1 - F_X(x) dx \tag{7}$$

la **fonction de répartition d'équilibre** de X. L'opérateur $*n$ désigne la convolution n-ième pour une fonction de répartition. Par exemple, $F_X^{*2} = F_{X_1+X_2}$. La formule 6 est appelée **formule de Pollaczeck-Khinchine**.

La simplicité apparente de ce modèle lui permet d'être facilement modifié afin d'étudier l'impact de différents paramètres assurantiels, comme les dividendes, la réassurance, le multi-risque, ou encore la dangerosité de la loi de X.

Ainsi, les chapitres 4 et 5 proposent une modification afin de prendre en compte la prévention primaire. Ils considèrent un investissement monétaire p en prévention, et supposent que le processus de Poisson qui modélise la survenu des sinistres est de paramètre $\lambda(p)$, avec λ une fonction décroissante positive convexe. Un tel modèle peut permettre par exemple de modéliser la mise en place de campagne annuelle de vaccination. Le modèle devient alors

$$R_t(p) = u + (c-p)t - \sum_{i=1}^{N_t(p)} X_i. \tag{8}$$

À p fixé, l'ensemble des résultats présentés plus haut sur le modèle de Poisson composé restent vrais. L'objectif est alors de déterminer le montant optimal à investir en prévention, c'est-à-dire le montant qui optimise une métrique de mesure arbitraire du risque. Le chapitre 4 propose ainsi de considérer différentes métriques, afin d'étudier les montants optimaux de prévention. La Proposition 1 de ce chapitre établit par exemple que le montant optimal de prévention minimisant la probabilité de ruine ne dépend pas du montant de réserves initiales u.

III-B) Intégration de versements de dividendes

Que ce soit aux actionnaires (compagnies d'assurance) ou aux adhérents (mutuelles), les sociétés d'assurance reversent généralement une partie de leurs bénéfices. Or, le versement de tels dividendes ralentit le rythme de constitution des réserves, affaiblissant généralement la pérennité de l'entreprise. Un objectif possible de la prévention est alors de maximiser les dividendes versés au cours de la vie de l'entreprise. Cette métrique sera considérée dans le chapitre 4 de la thèse.

De nombreux chercheurs ont étendu le modèle de Poisson composé pour prendre en compte le phénomène de redistribution des profits, prolongeant ainsi des travaux initiés par De Finetti (De Finetti, 1957). Diverses stratégies de versement de dividende ont été considérées dans la littérature, mais Asmussen et Taksar ont montré que la stratégie consistant à verser en dividende la totalité des bénéfices lorsque les réserves ont atteint un niveau cible K pouvait être optimale, dès lors que certaines hypothèses sont respectées (Asmussen and Taksar, 1997). Cette stratégie est appelée barrière de dividende constante.

Dans un tel modèle, la ruine est désormais certaine : en effet, lorsque les réserves atteignent la barrière K, elles stagnent jusqu'au prochain sinistre. Lorsque celui-ci arrive, il est possible de considérer un nouveau processus de montant de réserve initiale $K - x$, x étant le montant de ce sinistre. Cet enchaînement d'événements va se produire jusqu'à ce que la ruine survienne.

La formalisation du modèle est un peu plus compliquée. Afin d'obtenir le processus de réserve en présence d'une barrière de dividende $R_t(K)$, il est nécessaire de partir d'un processus de réserve R_t comme défini dans la partie précédente. Soit P_t le montant de perte à l'instant t :

$$P_t = \sum_{n=1}^{N_t} X_i - ct. \tag{9}$$

Soit L_t le montant de dividende total versé jusqu'à la date t :

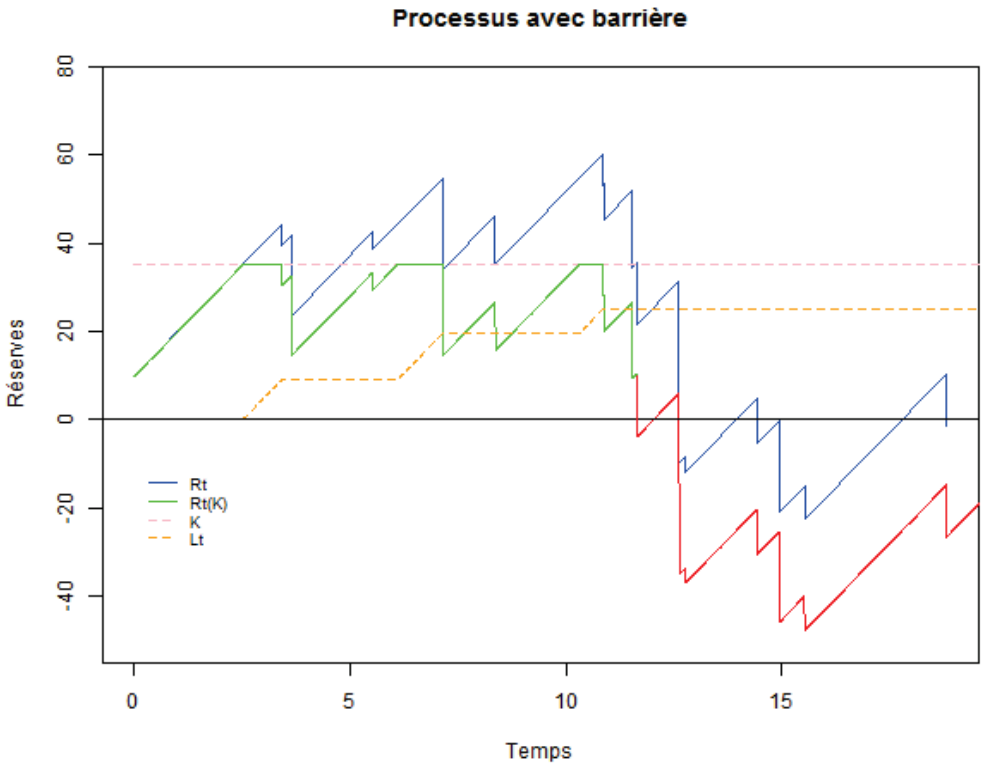

FIGURE 2 – Il s'agit ici du même processus que pour le graphe 1, mais avec un versement de dividende lorsque les réserves atteignent 35 (courbe verte). On voit que la ruine arrive un peu plus tôt. La courbe orange représente le montant de dividende versé par l'entreprise depuis le début.

$$L_t = -\inf_{0\leq s\leq t} min(K - u + P_s; 0). \tag{10}$$

En effet, si le processus R_t n'a pas encore atteint K, pour tout $s < t$, $min(K-u+P_s; 0) = 0$. Si R_t a atteint la barrière, $L_t = -\inf_{0\leq s\leq t}(K - u + P_s)$. Or $-\inf_{0\leq s\leq t}(-u + P_s)$ représente le plus haut point atteint par le processus R_t au cours de son histoire. À chaque fois que de nouveaux dividendes sont versés, le processus original R_t est à un niveau plus grand qu'il n'a jamais été.

On a alors

$$R_t(K) = u - P_t - L_t. \tag{11}$$

Considérons désormais $\mathbb{E}(L_T)$. Pour verser des dividendes, il faut dans un premier temps que le processus des réserves atteigne la barrière. Soit $\xi(u, K)$ la probabilité d'être ruiné avant d'avoir

versé le moindre dividende. Segerdhal (Segerdahl, 1970) a montré que $\lim_{K \mapsto \infty} \xi(u, K) = \psi(u)$, et que, pour tout $u, K \in \mathbb{R}_+$,

$$\xi(u, K) = \frac{\psi(u) - \psi(K)}{1 - \psi(K)}. \tag{12}$$

Une fois que la barrière est atteinte, le processus des réserves va en moyenne stagner pendant un temps λ avant de subir un nouveau saut.

Notons T_1 la date du premier sinistre. Posons

$$r = \mathbb{P}(T > \inf\{t > T_1, \quad R_t(K) = K\}) \tag{13}$$

la probabilité que les réserves atteignent à nouveau la barrière avant que la ruine ne survienne. Le nombre de fois où les réserves vont désormais atteindre le niveau K va donc suivre une loi géométrique de paramètre $1-r$. Mélanger ces trois éléments permet d'obtenir que le montant moyen de dividende versé avant la ruine vaut :

$$\mathbb{E}(L_T) = \frac{(1 - \xi(u, K))c}{\lambda(1 - r)}. \tag{14}$$

Il peut sembler raisonnable d'essayer d'utiliser la prévention afin de maximiser l'espérance des dividendes versés avant la ruine. Le corollaire 3 du chapitre 4 de cette thèse montre que dans le modèle à un risque, optimiser à l'aide de la prévention les dividendes versés jusqu'à la ruine dans un modèle à barrière revient à minimiser la probabilité de ruine dans le modèle de Poisson composé standard.

Les modèles à dividende modifient le modèle de Poisson composé en y ajoutant une nouvelle dimension. Ce n'est pas le cas de toutes les extensions : certaines ont été proposées afin d'étudier une dimension déjà présente dans le modèle, comme la loi des sinistres.

III-C) Extensions du modèle de Poisson composé portant sur la loi des sinistres

Le modèle de Poisson composé, qui ne s'intéresse qu'à un seul risque connu, indépendant des autres variables et des actions de l'assureur est assez restrictif. De nombreux chercheurs ont essayé de relâcher ces hypothèses afin d'assouplir le cadre de travail. Deux grandes catégories de travaux seront évoquées dans cette partie : les travaux étendant le modèle de Poisson composé à des entreprises qui couvrent deux risques différents, et ceux portant sur les politiques de réassurance.

Il est très rare qu'une société d'assurance ne soit confrontée qu'à un seul type de risque. Par exemple, dans le domaine de la santé, les contrats proposent très souvent des garanties couplées santé et prévoyance. Le risque santé est lui-même divisé en divers risques non homogènes, notamment des risques de fréquence (comme les soins courants, liés à des maladies comme les grippes) et des risques plus lourds (comme les prothèses dentaires ou la psychiatrie). C'est dans ce cadre de travail que ce placera le chapitre 5 de cette thèse. Nous supposerons ici que les divers risques auxquels est exposée l'entreprise sont décorrélés.

Cela revient à considérer deux risques $X^{(1)}$ et $X^{(2)}$ indépendants, survenant selon des processus de Poisson N_1 de paramètre λ_1 et N_2 de paramètre λ_2 [7]. Ces modèles ont d'abord été étudiés par Cramer (Cramér, 1954). Ce cadre a par exemple été considéré dans le cas où $X^{(1)}$ et $X^{(2)}$ suivent des lois exponentielles par Gerber et al. (Gerber et al., 1987) et Dickson et Gray (Dickson and Gray, 1984).

Le processus des réserves peut alors s'écrire

$$R_t^{multi-risque} = u + ct - \sum_{i=1}^{N_{1,t}} X_i^{(1)} - \sum_{i=1}^{N_{2,t}} X_i^{(2)}. \tag{15}$$

Ces modèles permettent d'obtenir un certain nombre de résultats intéressants. Par exemple il est possible, dans le cas où les sinistres suivent des lois exponentielles de paramètres différents, d'obtenir une formule fermée pour la probabilité de ruine : elle est de la forme $\varphi(u) = 1 + c_1 e^{r_1(u)} + c_2 e^{r_2 u}$, avec r_1 le coefficient d'ajustement du modèle (voir partie "Autres outils actuariels utilisés").

Dans un tel modèle, le nombre total de sinistres $N_t = N_{1,t} + N_{2,t}$ rencontrés à la date t suit un processus de Poisson de paramètre $\lambda_1 + \lambda_2$. De plus, chaque sinistre rencontré suit la même loi que $X^{(1)}$ avec probabilité $\frac{\lambda_1}{\lambda_1+\lambda_2}$ et la même que $X^{(2)}$ avec probabilité $\frac{\lambda_2}{\lambda_1+\lambda_2}$. Une autre manière d'écrire R_t est donc la suivante :

$$R_t^{multi-risque} = u + ct - \sum_{i=1}^{N_t} X_i, \tag{16}$$

avec les X_i i.i.d admettant pour fonction de répartition

$$F_X = \frac{\lambda_1}{\lambda_1 + \lambda_2} F_{X^{(1)}} + \frac{\lambda_2}{\lambda_1 + \lambda_2} F_{X^{(2)}}. \tag{17}$$

Il est alors possible de se ramener du modèle multirisque au modèle à un risque en utilisant une loi mélange. Cela permet de conserver des résultats importants, comme la formule de Pollaczeck Khinchine (6). C'est ce type de modèle qui sera considéré lors du chapitre 5 de cette thèse. Il y sera montré que l'introduction d'un second risque modifie certaines propriétés

7. L'ensemble des résultats peut être facilement généralisé à n risques.

constatées dans le chapitre 4. Par exemple, la Proposition 14 montre que dans un tel modèle, le montant de prévention qui minimise la probabilité de ruine dépend du niveau de réserve initiale u, ce qui n'est pas le cas dans un modèle à un risque.

Une seconde extension classique modifiant le processus des sinistres du modèle de Poisson composé est l'intégration de paramètres de réassurance. Supposons qu'une société d'assurance paie un montant r par unité de temps pour acheter une couverture de réassurance. On appellera programme de réassurance la fonction $h : (x, r) \in \mathbb{R}^2 \mapsto h(x, r)$. On peut alors écrire le processus des réserves

$$R_t^{reass} = u + (c - r)t - \sum_{i=1}^{N_t} h(X_i, r). \tag{18}$$

Le programme de réassurance modifie la loi des sinistres, puisque l'assureur verra une partie de ses charges couverte par le réassureur.

L'intégration d'un paramètre de réassurance soulève deux questions :

— Quel programme de réassurance choisir ? En effet, la forme de la fonction h peut être fixée par l'assureur suite à des négociations avec les réassureurs. Diverses formes, comme une réassurance proportionnelle, peuvent être envisagées. Cependant, Gerber aurait montré que les contrats minimisant le risque de ruine seraient les contrats de réassurance de type XS (Gerber, 1979) [8].

— Combien investir en réassurance une fois le programme de réassurance choisi ? Cette problématique a par exemple été résolue par Centeno quand des contrats de réassurance proportionnels et excess of loss sont disponibles (Centeno, 1985). Les résultats ont d'ailleurs été étendus au cas où l'assureur est exposé à plusieurs risques (Centeno and Simões, 1991).

Les modèles de Poisson composé les plus généraux prenant en compte la réassurance sont très souples puisqu'ils ne spécifient pas la forme de h. Cela permet de les adapter afin de modéliser l'auto-assurance. Cette dernière consiste en effet à réduire les montants de sinistre en investissant en prévention. Le programme de réassurance peut donc jouer le rôle de l'efficacité de la prévention, et la prime de réassurance joue le rôle de l'investissement en prévention. La proximité entre auto-assurance et réassurance nous a donc poussé à plutôt étudier l'auto-prévention dans les chapitres 4 et 5, pour lequel il n'existait pas d'équivalent naturel en théorie du risque.

Cependant, il est nécessaire de préciser que la majorité des travaux sur la réassurance porte sur les contrats de réassurance proportionnelle et XS. Il est difficile de transposer ces contrats directement en prévention. En effet, la réassurance proportionnelle supposerait une efficacité

8. Référence faite d'après (Hesselager, 1990) ; le document de Gerber n'est malheureusement pas facilement consultable et cela n'a pas pu être vérifié.

constante de la prévention, alors qu'il semble raisonnable de devoir supposer que la fonction h est convexe en r. De plus, ces contrats donneraient la possibilité à l'assureur de complètement prévenir un risque, ce qui n'est pas forcément réaliste. Les contrats XS pourraient correspondre à une politique de souscription consistant à rompre les contrats des assurés trop coûteux, ce qui est interdit en France. Les deux formes de h issues de ces deux contrats particuliers ne peuvent donc pas être appliquées telles qu'elles à la prévention.

III-D) Autres outils actuariels utilisés

La théorie de la ruine s'inscrit généralement dans un cadre plus large appelé théorie du risque. Celui-ci fournit un certain nombre d'outils permettant de mieux modéliser le monde réel. Deux d'entre eux seront présentés ici : les queues de distribution et les ordres stochastiques.

Les queues de distribution Les queues de distribution décrivent le comportement des lois de distribution dans les zones extrêmes. On distingue généralement deux types de queues : les queues légères et les queues lourdes. Les réalisations de ces dernières sont souvent situées dans des zones extrêmes, ce qui rend les risques décrits par ce type de loi particulièrement dangereux pour une compagnie d'assurance.

Formellement, une variable aléatoire X suit une loi dite à queue légère (ou sur-exponentielle) s'il existe $s>0$ tel que la fonction génératrice de X $M_X(s) = \mathbb{E}(e^{sX})$ existe pour au moins une valeur de s. Sont notamment concernées les lois exponentielles et les lois gamma.

À l'inverse, les lois à queue lourde sont les lois dont la queue de distribution n'est pas exponentiellement bornée. Plusieurs classes de loi à queue lourde se distinguent. Parmi elles, les lois dites sous-exponentielles sont régulièrement étudiées. Une variable aléatoire X suit une loi dite sous-exponentielle si, pour toute suite de variable aléatoire (X_i) i.i.d de même loi que X, alors pour tout $n \in \mathbb{N}^*$, $\lim\limits_{x\to\infty} \frac{\mathbb{P}(\sum_{i=1}^n X_i > x)}{\mathbb{P}(X>x)} = n$. Sont notamment concernées les lois de Pareto et les lois log-normales. Les livres de Asmussen (Asmussen, 2008) et de Adler et al. (Adler et al., 1998) s'attardent plus longuement sur ces notions.

Dans le cadre de la théorie de la ruine, supposer que les montants de sinistre suivent des lois à queue légère permet de majorer la probabilité de ruine en fonction des montants de réserve. En effet, dans le modèle à un risque exposé précédemment, si les X_i suivent une loi à queue légère, alors l'équation d'inconnu s

$$\lambda + cs = \lambda M_{X_1(s)} \tag{19}$$

admet une solution κ appelée **le coefficient d'ajustement**. Dans un tel cas, il est alors possible d'écrire

$$\psi(u) \leq e^{-\kappa u}. \tag{20}$$

La probabilité de ruine décroit donc exponentiellement avec le montant de réserves initiales. La vitesse de cette diminution dépend fortement de κ, qui peut donc être vu comme un indicateur de l'efficacité des réserves : plus κ est grand, plus augmenter les réserves initiales agira efficacement sur la probabilité de ruine.

L'existence du coefficient d'ajustement permet aussi d'étudier le comportement de la probabilité de ruine lorsque les réserves deviennent grandes. En effet, $\psi(u) \sim_{u\to\infty} \frac{c-\lambda\mu}{\lambda M_X'(\kappa)-c} e^{-\kappa u}$.

Lorsque les lois considérées ne sont pas à queue légère, le coefficient d'ajustement peut ne pas exister. Dans certains cas, il est cependant toujours possible d'obtenir une estimation de la vitesse de décroissance de la probabilité de ruine lorsque les réserves deviennent grandes. Par exemple, lorsque la variable d'équilibre (voir 7) suit une loi sous exponentielle, $\psi(u) \sim_{u\to\infty} \frac{\lambda\mu}{c-\lambda\mu}(1 - F_e(u))$.

Considérer une loi à queue lourde ou à queue légère permet donc d'obtenir des comportements asymptotiques différents. C'est pourquoi dans le chapitre 5, ces deux cas seront distingués, permettant d'obtenir deux résultats distincts. Dans le cas de loi à queues légères, la Proposition 15 établit ainsi de montrer que les montants de prévention qui minimisent la probabilité de ruine tendent, lorsque le montant de réserve initiale tend vers l'infini, vers le montant de prévention qui optimise le coefficient d'ajustement. Dans le cas des lois à queue sous exponentielles, la Proposition 16 fournie également la limite vers lequel convergent les montants optimaux de prévention lorsque les réserves initiales deviennent grandes.

Les ordres stochastiques Les ordres stochastiques sont une notion importante lorsqu'il s'agit d'analyser des risques. Ils permettent en effet de comparer des variables aléatoires entre elles : on appelle ordre stochastique toute relation d'ordre sur l'ensemble des lois de probabilités sur $\mathbb{R}$.

Ces outils ont un vrai intérêt en micro-économie. En effet, pouvoir comparer des lois de variable aléatoire permet de mieux comprendre les décisions des individus lorsqu'ils sont en présence d'un aléa. Rotschild a par exemple proposé des interprétations économiques de ces ordres (Rothschild and Stiglitz, 1970). Une revue de littérature peut être trouvée dans le livre de Shaked et Shanthikumar(Shaked and Shanthikumar, 2007).

Les ordres stochastiques sont nombreux. Le plus connu, et le plus intuitif, est le "premier ordre". Si X et Y sont deux variables aléatoires, on dira que X domine au premier ordre Y (noté $Y \preceq_{\text{st}} X$) si et seulement si pour toute fonction croissante u, $\mathbb{E}(u(X)) > \mathbb{E}(u(Y))$. Autrement dit, tout individu rationnel préfère la variable aléatoire X à la variable aléatoire Y, quelle que soit la fonction d'utilité u qu'il utilise pour effectuer ses choix.

Un autre ordre très utilisé est l'ordre convexe. Si X et Y sont deux variables aléatoires de fonction de répartition F_X et F_Y, on dit que X domine à l'ordre convexe Y (noté $Y \preceq_{\text{cv}} X$) si et seulement si pour toute fonction u convexe, $\mathbb{E}(u(Y)) \leq \mathbb{E}(u(X))$. Cet ordre s'interprète donc comme le premier ordre, sauf que la fonction d'utilité u est supposée convexe. A noter que, si $\mathbb{E}(Y) = \mathbb{E}(X)$, alors $Y \preceq_{\text{cv}} X \Leftrightarrow \forall t \geq 0,\ \int_t^\infty 1-F_Y(p)dp \leq \int_t^\infty 1-F_X(p)dp$. Autrement dit, lorsque les espérances sont égales, l'ordre convexe revient à comparer toutes les queues de distributions possibles des deux lois.

Cependant, l'hypothèse d'égalité des espérances est très restrictive. L'ordre HMRL (harmonic mean residual life order) est un autre ordre courant qui permet lui aussi de comparer des variables aléatoires à l'aide des queues de distribution, sans cette hypothèse. En effet, si X et Y sont deux variables aléatoires positives de fonction de répartition F_X et F_Y, on dit que X domine à l'ordre HMRL Y (noté $Y \preceq_{\text{hmrl}} X$) si et seulement si pour tout $t \geq 0$,

$$\frac{\int_t^\infty 1 - F_X(p)dp}{\int_t^\infty 1 - F_Y(p)dp} \geq \frac{\mathbb{E}(X)}{\mathbb{E}(Y)}. \tag{21}$$

De manière équivalente, $Y \preceq_{\text{hmrl}} X$ si et seulement si pour tout $t \geq 0$, $F_{e,X}(t) \leq F_{e,y}(t)$, avec F_e désignant les fonctions d'équilibre des variables aléatoires, comme définies dans l'équation 7 (voir théorème 2.B.2 dans (Shaked and Shanthikumar, 2007) pour une démonstration).

En réécrivant l'équation 21, on obtient, pour tout $t > 0$, $\frac{\int_t^\infty 1-F_X(p)dp}{\mathbb{E}(X)} \geq \frac{\int_t^\infty 1-F_Y(p)dp}{\mathbb{E}(Y)}$. Le paramètre t peut être vu comme un seuil de gravité. Alors, quel que soit ce seuil, dire que $Y \preceq_{\text{hmrl}} X$ signifie que, si un grand nombre de sinistres i.i.d était observé, alors la proportion de sinistres qui dépassent le seuil t sera plus grande si les sinistres suivent la même loi que X plutôt que la même loi que Y. De tels ratios apparaissent parfois naturellement dans les calculs. C'est pourquoi cet ordre sera utilisé dans le chapitre 5. En effet, dans le cas d'un modèle à deux risques, la Proposition 13 montre qu'il suffit que la prévention soit efficace lorsqu'il n'y a pas de réserves initiales (u=0) et que le risque prévenu domine à l'ordre HMRL le second risque pour que il existe un montant de prévention optimal quelque soit le montant de réserves initiales.

La théorie du risque dispose donc de nombreux outils facilitant la modélisation des problématiques actuarielles. Cela permettra de proposer des modèles intégrant la prévention dans les chapitres 4 et 5 de cette thèse. En revanche, comme souligné dans la partie II-B) de cette introduction, si il est nécessaire de savoir modéliser la prévention, cela ne suffit pas : la prévention pose de nombreux problèmes pratiques. La difficulté pour un assureur de savoir quand et à qui proposer un service de prévention en fait partie. Le manque de variables à disposition des assureurs, ainsi que la réglementation obligent ces entreprises à adopter de nouvelles approches. Ainsi, la Section suivante présente divers algorithmes qui seront utilisés dans les chapitres 1 et 2, qui présenteront une méthode de classification d'assurés santé inspiré du text mining et pouvant servir dans le cadre du ciblage d'un programme de prévention.

IV) L'apprentissage automatique au service de la classification

L'augmentation de la puissance de calcul disponible a permis aux actuaires de compléter leurs techniques avec de nouveaux outils d'apprentissage statistique. Ces avancées technologiques ont été accompagnées d'un débat presque philosophique : l'assurance doit-elle être solidaire ou personnalisée ?

Historiquement, l'assurance a une vocation solidaire, basée sur la mutualisation des risques. Au sein d'un même portefeuille, les bons et les mauvais risques se mélangeaient et se compensaient, permettant à chacun d'être indemnisé correctement moyennant la même prime d'assurance. Ce système, profondément égalitaire, a permis l'émergence des mutuelles, sociétés d'assurance à but non lucratif. Mais, outre un principe de solidarité apparent, la mutualisation devait aussi son existence à des raisons plus pragmatiques : les variables observées par les assureurs étaient peu nombreuses et peu fiables ; et les calculs, effectués manuellement, se devaient d'être simples et se basaient sur la loi des grands nombres.

Néanmoins, un traitement égalitaire n'est pas forcément équitable. Avec la vision mutualiste coexiste une vision individualiste de l'assurance, où chacun doit payer conformément au risque qu'il représente. Cette philosophie a notamment donné naissance aux systèmes de bonus/malus présents en assurance automobile. Or, avec le développement des technologies informatiques, une puissance de calcul nouvelle est apparue, accompagnée d'une explosion du nombre de variables observables. Rêve pour certains, cauchemar pour d'autres, ces avancées ont rendu tangible la réalisation d'un vieux fantasme assurantiel : la connaissance parfaite du risque représenté par chacun.

Tarifer de manière aussi précise présenterait cependant certains effets pervers. Alors que ce sont les plus mauvais risques qui en ont le plus besoin, l'explosion des primes d'assurance pourrait être telle qu'ils ne pourraient plus s'assurer. De plus, cette vision peut aussi se révéler injuste, lorsque le risque représenté par un assuré ne dépend pas de ses choix. Ainsi, dans le domaine de la santé, le risque est partiellement expliqué par le patrimoine génétique dont a hérité l'assuré, ce qu'il ne peut contrôler. A l'inverse, un modèle économique basé sur une mutualisation parfaite n'est pas compatible avec un marché concurrentiel : les bons risques vont quitter le portefeuille d'une compagnie qui mutualise fortement ses tarifs, attirés par les primes plus faibles proposées par une compagnie qui segmente plus. L'assurance mutualisant les risques se verra alors obligée d'augmenter ses tarifs, initiant un cercle vicieux. En conséquence, la réponse adéquate semble se trouver entre les deux visions : naturellement, les sociétés d'assurance essayent de grouper les assurés homogènes dans un nombre de classes approprié afin de proposer des tarifs à la fois personnalisés et mutualisés.

L'objectif de cette Section est de présenter l'ensemble des méthodes utilisées dans les chapitres 1 et 2, dans lesquels sont proposées des méthodes de classification non supervisée

d'assurés. Dans un premier temps, ce sont donc les méthodes de clusterings qui seront exposées. Celles-ci souffrant généralement d'une perte de performance lorsque la dimension des données augmente, des méthodes de réduction de dimension sont ensuite présentées. Enfin, le domaine du text mining sera introduit, puisque celui-ci a fait l'objet de nombreux travaux de classification et a inspiré les process présentés dans les chapitres 1 et 2 de la thèse.

IV-A) Classification non supervisée

La classification désigne la tâche informatique consistant à partitionner des données en groupes homogènes. Celle-ci peut être supervisée (*classification* en anglais) ou non supervisée (*clustering* en anglais).

Les algorithmes de classification non supervisée sont capables d'analyser un grand nombre de données afin de détecter des points communs entre elles. Cependant, dans un contexte non supervisé, il n'est pas possible de considérer objectivement qu'un individu est mal classé. Il n'est donc pas possible d'obtenir une mesure d'erreur pour vérifier la qualité des résultats. Aussi, avant de réaliser une telle tâche, il est nécessaire de s'interroger sur deux points essentiels :

— Comment déterminer si deux individus sont similaires ? Pour résoudre ce problème, il est nécessaire de s'interroger sur la distance (ou la similarité) utilisée pour comparer des individus. Comme montré par Huang, ce choix peut grandement influencer les résultats finaux (Huang, 2008).
— Combien de classes constituer ? En effet, constituer trop de classes rendra le résultat moins exploitable, quand constituer trop peu de classes aboutira à des groupes de mauvaise qualité.

Les réponses à ces questions peuvent permettre de déterminer quel algorithme choisir et comment le paramétrer. De nombreuses méthodes ont en effet été développées (voir (Xu and Wunsch, 2005) pour une revue de la littérature à ce sujet). Seules les méthodes utilisées dans cette thèse seront détaillées dans cette partie.

Nous supposerons avoir à disposition une base de données B constituée de n individus représentés par m variables numériques. Nous souhaitons construire une classification en k classes de cette base de données. Formellement, B est un sous ensemble fini $(x_1, ..., x_n)$ de $\mathbb{R}^m$. L'objectif est de construire une partition de B en k sous-ensembles. On supposera de plus disposer d'une fonction de similarité sim.

La Classification Ascendante Hiérarchique

La Classification Ascendante Hiérarchique (CAH) est à la fois la plus simple et l'une des plus efficaces des méthodes de classification non supervisée.

Lors de l'initialisation, chaque individu est supposé constituer une classe à part entière. Puis, à chaque itération, l'algorithme va chercher les deux classes $C^{(1)}$ et $C^{(2)}$ les plus similaires

entre elles. Une fois cette tâche terminée, il regroupe ces deux classes dans une même nouvelle classe. Il commence alors une nouvelle itération.

Cet algorithme suppose donc d'être capable de calculer une similarité entre deux classes C_1 et C_2 (potentiellement constituées de plusieurs individus). Pour ce faire, plusieurs méthodes sont possibles (liste non exhaustive) :

- L'algorithme calcule l'inertie intragroupe de la classe issue de la fusion de C_1 et C_2 (méthode de Ward).
- L'algorithme considère les deux individus les moins similaires de chacune des deux classes : $sim(C_1, C_2) = \min_{x \in C_1, y \in C_2} sim(x, y)$ (stratégie dite du saut maximum).
- L'algorithme considère les deux individus les plus similaires de chacune des deux classes : $sim(C_1, C_2) = \max_{x \in C_1, y \in C_2} sim(x, y)$ (stratégie dite du saut minimum).

Cette méthode présente néanmoins le défaut de n'être utilisable que sur de petites bases de données, à cause de temps de calcul trop importants.

Cette méthode à été utilisé dans des domaines de recherches variés, comme le text mining (e.g. pour classifier des documents similaires Premalatha and Natarajan (2010)), la santé (e.g. pour faire des groupes de symptomes Fan et al. (2007)), ou l'IoT (e.g. pour les moteurs de recherche, Chen et al. (2015), Beeferman and Berger (2000)). Elle sera utilisée dans les chapitres 1 et 2 afin de regrouper des classes entre elles et réduire le nombre de classes total créés par le processus complet.

Les k-means

La méthode des k-means est certainement l'algorithme de classification non supervisée le plus connu. Il a été introduit par Steinhaus en 1956 (Steinhaus, 1956)). Afin d'utiliser cet algorithme, il faut disposer d'une partition initiale de B en k classes (cette partition peut être constituée aléatoirement). Les deux étapes suivantes sont alors répétées jusqu'à ce que l'algorithme converge :

- Le barycentre de chacune des classes est calculé.
- chaque observation est allouée à la classe dont le barycentre est le plus proche.

Cette méthode, bien que facile à utiliser et efficace, comporte cependant un certain nombre de limites :

- Le nombre de classes est à déterminer en amont
- Les performances de l'algorithme dépendent grandement de la distance et de l'initialisation choisie
- Elle ne garantit pas d'obtenir la meilleure classification possible
- Elle n'est pas efficace si le nombre de variables est trop important (Beyer et al., 1999)

Il existe de nombreuses variantes, permettant de combler certaines de ces faiblesses. Cette méthode est très populaire, et a été utilisé dans divers secteurs, comme l'analyse de gènes

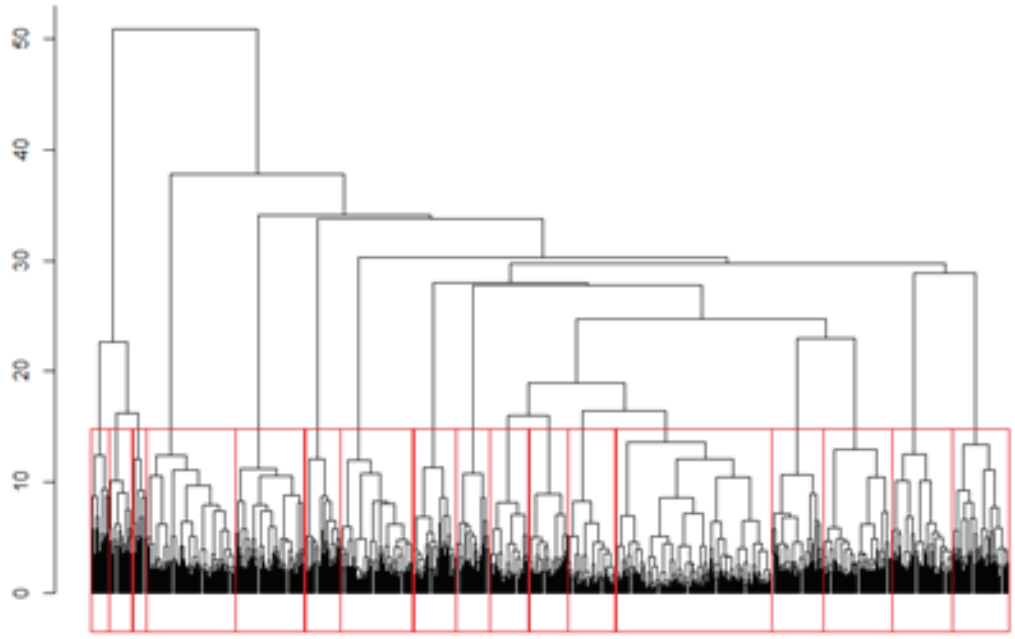

FIGURE 3 – Exemple de CAH. Chaque trait horizontal illustre la fusion de deux classes. En ordonnée est marquée la dissimilarité entre deux classes au moment de la fusion des classes. À titre d'illustration, les 17 dernières classes restantes sont distinguées par les séparations rouges. Un tel graphe est appelé un dendrogramme.

Yano and Kotani (2003), la sociologie (e.g. prédiction des résultats scolaire, Oyelade et al. (2010)) ou encore en gestion (e.g. recherche d'entreprise au profil similaires Vagizova et al. (2014)). Cette méthode a été utilisée afin de challenger les cartes de Kohonen utilisées dans les chapitres 1 et 2, mais les résultats étaient décevants, principalement à cause de la difficulté à choisir les barycentres initiaux, et ne sont pas présentés ici.

Les fuzzy c-means

Les k-means appartiennent à la catégorie des algorithmes mono-labellisés, c'est-à-dire que chaque individu est supposé n'appartenir qu'à une seule classe. Cette hypothèse peut cependant être limitante : il est parfois impossible d'attribuer une étiquette unique à chaque input.

Les classifications floues (ou multilabélisées) sont apparues afin de contourner ce problème. Elles autorisent en effet l'appartenance d'un individu à plusieurs classes simultanément. Ainsi, en 1973, Dunn a proposé une variante floue des k-means : les fuzzy c-means (Dunn, 1973).

A la différence des k-means, l'algorithme des fuzzy c-means suppose qu'un individu x est caractérisé non pas par l'appartenance à une seule classe, mais par un degré d'appartenance $w_j(x)$ à chacune des classes, avec $\sum_{j=1}^{k} w_j(x) = 1$.

À l'initialisation de l'algorithme, chaque individu se voit attribuer des poids d'appartenance à chaque classe (aléatoirement par exemple). Puis, à chaque itération, les étapes suivantes sont répétées :

— Les barycentres pondérés c_i de chaque classe i sont calculés selon la formule $c_i = \frac{\sum_x w_i^\omega(x)x}{\sum_x w_i^\omega(x)}$. Le paramètre $\omega \geq 1$ est fixé à l'avance et permet de contrôler si la classification finale doit être plus ou moins floue.
— Les poids sont mis à jour afin de minimiser $\sum_x \sum_{i=1}^k w_i^\omega(x)||x - c_i||^2$.

Depuis les premiers travaux de Dunn, la méthode à été notamment très utilisé en analyse d'image (e.g. Chuang et al. (2006)), mais a été utilisé dans de nombreux autres domaines, et notamment l'assurance (Verrall and Yakoubov (1999)).

Si cet algorithme a l'avantage de proposer une classification floue, il possède en revanche les autres défauts de l'algorithme des k-means : les résultats sont très sensibles à l'initialisation, et la qualité de la classification est sensible à la dimension du problème. Il a été testé d'utiliser cette méthode en initialisant l'algorithme à l'aide des barycentres des classes obtenus par l'algorithme des cartes de Kohonen afin de rendre les résultats des chapitres 1 et 2 flous. Si cette manière de faire fonctionne, les résultats sont très sensibles au paramètre ω, et ne sont pas présentés dans cette thèse.

Les cartes de Kohonen

Les cartes de Kohonen (ou cartes auto-organisatrices de Kohonen, ou som) sont des algorithmes de clustering basés sur les réseaux de neurones. Kohonen, en 1982, a ainsi proposé d'organiser les neurones d'une couche selon une topologie fixée (cf. (Kohonen, 1982) pour l'article d'introduction et (Kohonen, 1990) pour un aperçu plus global). Ainsi disposés, les neurones vont former une carte dont les zones géographiques vont attirer les individus similaires.

Les résultats dépendent de la topologie choisie. Généralement, les neurones sont disposés en deux dimensions, afin de faciliter la visualisation des résultats. Ils sont le plus souvent disposés de manière hexagonale (voir par exemple la Figure 4). Chaque neurone se voit attribuer des coordonnées : le neurone en bas à gauche possède pour coordonnées $(1, 1)$, celui à sa droite $(1, 2)$, etc. Choisir une topologie suppose aussi de fixer une distance $||\ ||$ qui permet de mesurer si deux neurones sont loin l'un de l'autre. La distance euclidienne suffit souvent à obtenir de bons résultats. Il est enfin nécessaire de choisir un nombre total N de neurones.

Visuellement, chaque neurone se retrouve entouré d'un certain nombre de neurones voisins. Pour modéliser ceci, Kohonen a introduit le concept de fonction de voisinage h : celle-ci est chargée de calculer le degré de voisinage entre deux neurones. Un choix classique consiste à prendre une fonction de voisinage gaussienne : si i et j sont deux neurones de coordonnées c_i et c_j, alors $h(c_i, c_j, t) = \alpha(t)exp(-\frac{||c_i - c_j||}{2\sigma^2(t)})$, avec t désignant la période d'apprentissage de

l'algorithme et α et σ étant deux fonctions décroissantes au cours des itérations. Ainsi, plus le nombre d'itérations effectuées par l'algorithme est important, moins deux neurones seront voisins.

Au début de l'algorithme, chaque neurone i se voit attribuer aléatoirement un vecteur de poids $w_i(0)$ de même taille que le nombre de variables m. Ces poids vont permettre de mesurer l'adéquation entre un neurone et un individu. L'algorithme est ensuite itératif ; pour une période t :

1. Un individu x est choisi
2. Pour chaque neurone i, la distance $||x - w_i(t)||$ est calculée. Cela permet de trouver le neurone I qui a les poids les plus similaires aux caractéristiques de l'individu x : $I = argmin_{i \in [\![1,N]\!]} ||x - w_i(t)||$.
3. Les poids du neurone I sont ajustés afin qu'il soit encore plus représentatif de l'individu x. De même, les poids des neurones voisins du neurone I sont ajustés selon la formule $w_i(t) = w_i(t-1) + h(c_I, c_i, t)(x - w_i(t-1))$. De cette manière, des zones d'attraction d'individus similaires sont créées.
4. Retour au point 1.

Lorsque tous les individus ont été choisis un certain nombre de fois (fixé à l'avance par l'utilisateur), l'algorithme se termine. Chaque individu se voit alors attribuer définitivement le neurone qui a les poids les plus similaires à ses caractéristiques. Chaque neurone devient une classe différente.

Les cartes de Kohonen souffrent elles aussi d'une perte de performance importante lorsque le nombre de variables augmente. De même, leurs performances diminuent lorsque trop de classes sont à réaliser (Mingoti and Lima, 2006).

Ce dernier point peut contraindre l'utilisateur à diminuer sensiblement le nombre de neurones de la carte et peut diminuer l'intérêt de la représentation sous-jacente. Afin de contourner cette contrainte, il est possible de choisir un grand nombre de neurones, puis de réduire le nombre de classes en regroupant les neurones similaires au sein d'une même classe grâce à une classification ascendante hiérarchique (Murtagh, 1995). Les grandes classes créées ainsi sont alors plus stables que les classes directement issues des neurones.

Des exemples d'utilisation de ces cartes dans la littérature sont par exemple l'industrie (e.g. prédiction des durées de vie des roulements à bille, Huang et al. (2007)), l'écologie (e.g. gestion des ressources d'eau Kalteh et al. (2008)) ou l'IoT (classification de pages webs Smith and Ng (2003)). Les cartes de Kohonen sont la méthode qui a été retenue afin de procéder aux classifications dans les chapitres 1 et 2. Elles permettent en effet d'obtenir et de visualiser des classes de risque cohérentes.

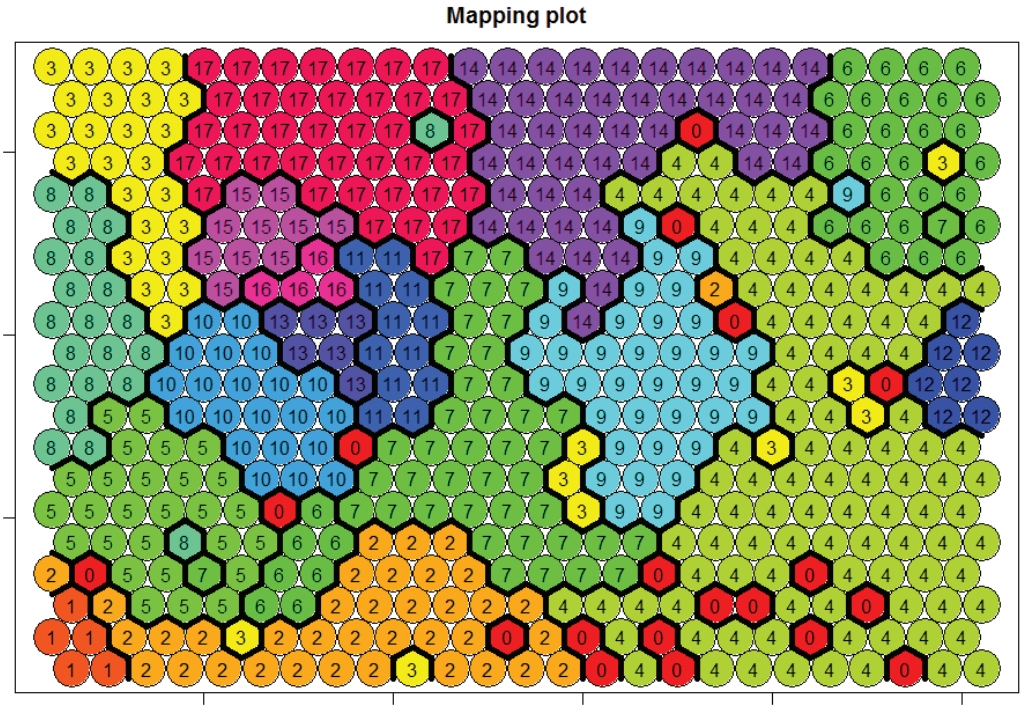

FIGURE 4 – Exemple de carte de Kohonen obtenue en classifiant des assurés à partir de leurs prestations santé. Chaque cercle est un neurone, et chaque groupe de même couleur est une classe finale différente, obtenue en regroupant des neurones par CAH.

Si les méthodes présentées jusqu'ici peuvent être utilisées seules, il est souvent nécessaire de les combiner avec une méthode de réduction de dimension afin d'améliorer leurs performances en grande dimension.

IV-B) Réduction de dimension

Ce n'est pas un hasard si la quasi-totalité des méthodes évoquées lors de la partie précédente souffre d'une baisse de performance lorsque de nombreuses variables sont disponibles. Il s'agit de l'illustration d'un phénomène statistique appelé par Bellman le Fléau de la dimension (Bellman, 2015).

En proposant ce terme, Bellman illustre le fait qu'au fur et à mesure que la dimension grandit, il est nécessaire d'avoir de plus en plus d'individus dans la base de données étudiée afin de pouvoir correctement modéliser les structures sous-jacentes à ces données. A titre d'illustration, la méthode des k-means éprouve des difficultés à gérer plus de 15 variables différentes (Beyer et al., 1999).

Le fléau de la dimension est un problème pour la totalité des méthodes d'apprentissage automatique. Il existe cependant deux moyens de le contrer : la suppression de variables peu

informatives et la création de nouvelles variables résumant l'information des anciennes. Il s'agit des méthodes de réduction de dimension. Cette section présentera quelques algorithmes utilisant la seconde manière de faire. Le lecteur pourra se référer à Sorzano et al. pour avoir un aperçu plus complet de l'ensemble des méthodes existantes (Sorzano et al., 2014).

Pour cette partie, nous supposerons disposer d'un nuage de n points $(x_1, ..., x_n) \in \mathbb{R}^m$. Nous supposerons vouloir réduire la dimension initiale m en créant un espace de dimension $k < m$.

Analyse en Composante Principale

L'Analyse en Composante Principale (ACP) est une des plus connues et des plus anciennes méthodes de réduction de dimension (voir Wall et al. pour une revue de littérature Wall et al. (2003)). Formalisée par Hotelling dans les années 1930 (Hotelling, 1933), elle consiste à projeter les données dans un nouvel espace de dimension plus faible. Cet espace est choisi de manière à maximiser la dispersion des données : ainsi, la perte d'information est minimisée. La base de cet espace est composée de nouvelles variables créées par combinaison linéaire des précédentes.

Maximiser ainsi la dispersion des données revient à maximiser l'inertie dans le nouvel espace des données étudiées. Autrement dit, cela revient à trouver une base qui maximise la variance des données. Pour ce faire, il est possible de calculer la matrice de variance - covariance de nos variables initiales. Cette matrice étant symétrique, elle est diagonalisable. Il est alors possible de former une base avec les vecteurs propres associés. Ce sont les k vecteurs propres associés aux k plus grandes valeurs propres qui forment la base de l'espace de dimension réduite.

L'ACP présente plusieurs avantages : elle est simple, les variables obtenues sont décorrélées, et elle permet de sélectionner la dimension finale a posteriori à l'aide de méthodes simples.

En revanche, la dimension de la matrice de covariance étant égale à celle du problème, sa diagonalisation peut être très coûteuse en temps pour certains problèmes, comme ceux issus du text mining. De plus, l'ACP étant une méthode linéaire, elle peut offrir des résultats de moindre qualité sur certains jeux de données non linéaires. En particulier, l'ACP est très sensible à la présence de bruit et d'outlier Wu et al. (2018). Enfin, les résultats de l'ACP dépendent de la distance euclidienne (lors de la projection des données), qui n'est pas forcément appropriée.

L'ACP a été comparé avec la méthode NMF dans le chapitre 1, mais les performances n'étaient pas satisfaisantes. Ce manque de performance provient certainement du bruit important contenu dans les données santé.

Décomposition en Valeurs Singulières

La seconde méthode classique est la décomposition en valeurs singulières (SVD), dont le premier algorithme a été proposé par Golub et Kahan, puis amélioré par Golub et Reinsch

Métrique	NMF	SVD
R^2 moyen	24,7%	24,3%
R^2 maximum	37,3%	25,8%

FIGURE 5 – 100 cartes de Kohonen ont été construites à partir d'une réduction de dimension réalisée à l'aide d'un algorithme NMF et 100 autres à partir d'une réduction de dimension réalisée à l'aide de l'algorithme SVD. Pour chaque carte, le R^2 (i.e. l'inertie intra-groupe divisée par l'inertie totale) a été calculé. Ce tableau montre les R^2 moyens et maximums obtenus à l'aide de cette méthodologie. Selon cette métrique, l'algorithme SVD est moins efficace que l'algorithme NMF.

(Golub and Kahan, 1965), (Golub and Reinsch, 1971). Elle se base sur un résultat, dû à Beltrami et Jordan (Stewart, 1993), qui dit que si V est une matrice réelle de taille $n * m$, alors il existe U une matrice unitaire de taille $n * n$, Σ une matrice de taille $n * m$ dont les coefficients sur la diagonale sont positifs (et les autres coefficients sont nuls), et W une matrice unitaire de taille $m * m$ et de matrice transposée W^t tel que $V = U\Sigma W^t$.

Cette décomposition est appelée décomposition en valeurs singulières (les coefficients de la matrice Σ étant appelés les valeurs singulières de la matrice V). Usuellement, les coefficients de la matrice Σ sont triés par ordre décroissant, tel que la plus grande valeur singulière corresponde avec la première valeur de la diagonale. Ceci permet d'assurer l'unicité de la matrice Σ. Contrairement à la décomposition en valeurs propres, qui ne peut être utilisée que sur une matrice carrée (c'est pourquoi l'ACP se base sur la matrice de variance - covariance des variables, et non sur les données elles-mêmes), cette décomposition peut être utilisée pour n'importe quelle matrice réelle.

La décomposition en valeur singulière est exacte, et il est possible de tronquer la matrice Σ afin de ne conserver que les k valeurs singulières les plus grandes. La matrice W tronquée de $m-k$ colonnes contient alors une base de k vecteurs d'un nouvel espace de dimension réduite. La matrice U tronquée de ses $m-k$ dernières colonnes contient elle la projection des données initiales dans l'espace de dimension réduite. La réduction de dimension est alors finalisée.

Si la méthode SVD est très efficace et très utilisée (voir Wall et al. pour une revue de littérature Wall et al. (2003)), elle peut souffrir de certaines limites lorsque la dimension est très grande et que les matrices décomposées sont très creuses (Yang et al., 2011). Elle a aussi été utilisée pour challenger la méthode NMF lors de l'étude présentée dans le chapitre 1. Si les résultats sont acceptables, ils sont légèrement moins bons que ceux des NMFs (voir par exemple la Figure 5), c'est pourquoi la méthode n'a pas été retenue et que les résulats issus de cette méthode ne sont pas présentés dans les chapitres de thèse.

Décomposition en matrices positives

Les méthodes ACP et SVD sont des méthodes très populaires, mais comportent certaines limitations. Par exemple, le fait que les coefficients des combinaisons linéaires calculés afin de créer les bases des espaces de dimension réduite puissent être négatifs peut, pour certains types de données (image), rendre l'interprétation des résultats délicate (Welling and Weber, 2001).

Certaines méthodes spécialisées ont alors vu le jour afin de pallier les différentes limites de ces méthodes. Parmi elles, les méthodes de décomposition en matrices positives (NMF) ont connu un certain succès. Introduites par Paatero et Tapper (Paatero and Tapper, 1994), puis popularisées par Lee et Seung (Lee and Seung, 1999), elles ont pour but d'approximer une matrice positive par le produit de deux matrices positives de taille fixée. En particulier, si V est la matrice de données considérée de taille $n * m$, et étant donné une norme $||\ ||$, les méthodes NMF cherchent à trouver une matrice W de taille $n * k$ et une matrice H de taille $k * m$ de manière à minimiser

$$d(W, H) = ||V - WH||. \tag{22}$$

Bien entendu, la fonction d dépend de la norme utilisée. Des algorithmes, comme celui proposé par Lee et Seung, se basent sur la distance Euclidienne, quand d'autres, comme celui proposé par Brunet et al. (Brunet et al., 2004), se basent sur l'entropie.

Certains auteurs essayent de minimiser des fonctions légèrement différentes de la fonction (22). Par exemple, Kim et Park proposent de considérer la fonction

$$d'(W, H) = \frac{1}{2}[\|V - WH\|_2^2 + \alpha \|H\|_2^2 + \beta \sum_{i=1}^{n} \|W(i, .)\|_1^2], \tag{23}$$

avec α et β des coefficients choisis en amont par l'utilisateur permettant de contrôler la forme de H et W, et $\|\|_1$ et $\|\|_2$ les normes L_1 et L_2 (Kim and Park, 2007).

Si les fonctions présentées ci-dessus sont convexes lorsqu'un des arguments est fixé, elles ne le sont plus lorsque W et H varient en même temps. Les méthodes d'optimisation de ces fonctions étant similaires à la méthode de descente de gradient, elles n'aboutissent donc qu'à des minima locaux. Leurs résultats dépendent donc de l'initialisation de la méthode. Il existe plusieurs méthodes fixant les coefficients initiaux des matrices W et H, mais celle consistant à initialiser aléatoirement ces deux matrices un grand nombre de fois, puis à comparer les différentes décompositions obtenues afin de sélectionner la plus performante est très efficace, bien que peu efficiente (Gaujoux and Seoighe, 2010).

La méthode NMF semble mieux performer que l'ACP sur des données de type texte (Chikhi et al., 2007), mais moins bien que le SVD (Kumar, 2009). Cependant, aucun de ces articles ne présente la méthodologie employée afin d'initialiser l'algorithme.

Il s'agit de la méthode de réduction utilisée dans le chapitre 1. Le fait qu'elle permette d'obtenir deux matrices simultanés permet d'interpréter facilement les résultats : utiliser la méthode NMF sur des assurés santé permet d'une part de faire ressortir les liens entre acte et d'obtenir différent groupes d'actes, et d'autre part d'obtenir directement les assurés projetés dans un espace de dimension réduite, ce qui permet d'améliorer leur classification. Elle est challengée avec deux méthodes plus complexes dans le chapitre 2, les méthodes mSDA et Word2Vec (voir après).

Les auto-encodeurs

L'amélioration de la puissance de calcul disponible informatiquement a largement favorisé le développement d'approches basées sur les réseaux de neurones. De nombreuses architectures neuronales sont alors apparues, répondant à diverses problématiques. Parmi elles, les auto-encodeurs constituent une méthode de réduction de dimension visant à compresser puis reconstituer des données.

Les premiers auto-encodeurs décrits par Ackley et al. en 1985 (Ackley et al., 1985) ne comptaient qu'une centaine de neurones maximum et une seule couche cachée. Depuis, les auto-encodeurs ont été utilisés dans de nombreux domaines, comme l'étude d'image (Wang et al., 2016a) ou l'étude de texte (Liou et al., 2014), pour des tâches allant de la compression au débruitage de données (Zhai et al., 2018).

Un auto-encodeur est un réseau de neurones composé au minimum d'une couche d'entrée et d'une couche de sortie comptant chacune un neurone par variable de la base de données étudiée, ainsi que d'un certain nombre de couches cachées comprenant moins de neurones que les couches d'entrée et de sortie. Un auto-encodeur a généralement pour mission de reconstituer sur sa couche de sortie les données qu'il a reçues en entrée. Les couches cachées étant de dimension plus petite que les couches visibles, le réseau va être obligé de projeter les données dans des espaces de dimension plus faible (phase d'encodage), et de comprendre la structure de ces données afin de pouvoir les décompresser correctement (phase de décodage, voir la Figure 6). Les poids des couches intermédiaires sont alors estimés grâce à l'algorithme de backpropagation (Rétropropagation du gradient en français).

Une architecture rencontrée régulièrement est celle des auto-encodeurs débruiteurs. Ces auto-encodeurs ont la même architecture que les auto-encodeurs classiques. En revanche, les données sont bruitées (par exemple, en forçant aléatoirement certaines variables à être nulles) avant d'entrer dans le réseau. L'objectif de l'auto-encodeur n'est alors plus de reconstituer les données d'entrée, mais de retrouver les données non bruitées (Vincent et al., 2008).

Généralement, plus le nombre de données en entrée est élevé, plus les résultats offerts par les auto-encodeurs sont satisfaisants. Dans le cas d'auto-encodeurs débruiteurs, il est possible d'utiliser plusieurs fois le même individu, mais de le bruiter à chaque fois différemment, afin

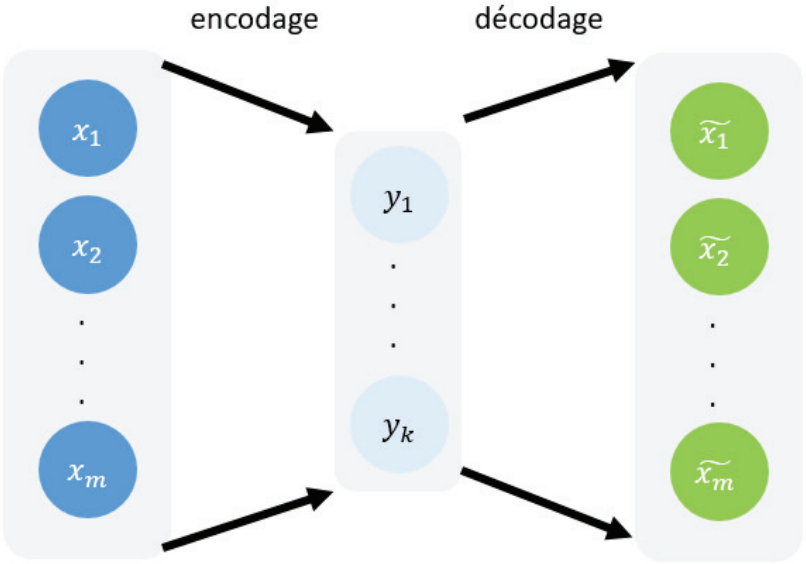

FIGURE 6 – Schéma d'un auto-encodeur classique

d'augmenter le nombre de données disponibles. Chen et al. ont remarqué que dans le cas où le bruit consiste en l'annulation aléatoire de certaines caractéristiques, il est possible d'utiliser la loi des grands nombres afin de simuler un nombre infini de bruitages (Chen et al., 2012). L'algorithme de backpropagation devient alors inutile, car les poids optimaux peuvent être explicitement calculés. On parle alors d'auto-encodeurs débruiteurs marginalisés.

Enfin, il est possible de disposer plusieurs auto-encodeurs à la suite. La couche de sortie d'un auto-encodeur devient alors la couche d'entrée de l'auto-encodeur suivant. On parle alors d'auto-encodeurs empilés. C'est ainsi que Chen et al. ont créé les auto-encodeurs débruiteurs marginalisés empilés (marginalized Stacked Denoising Autoencoder, ou mSDA).

Les méthodes d'auto-encodeur ont souvent été utilisées pour des tâches de réduction de dimension. Elles ont prouvé être relativement efficaces, offrant de meilleures performances lorsque la dimension est élevée (Wang et al., 2016b), et performent mieux que de nombreux algorithmes (Van Der Maaten et al., 2009).

Les auto-encodeurs, et en particulier l'architecture mSDA, ont été utilisés dans le chapitre 2 afin de challenger la méthode NMF. Ils permettent d'obtenir des groupes d'actes très cohérent, et notamment de grouper les actes de maternité qui echappent aux algorithmes NMF sur les bases de données utilisées.

La dernière méthode (l'algorithme Word2Vec) utilisée dans le chapitre 2 exploite une analogie entre les données d'assurance santé et les textes. La partie suivante à donc pour but de

présenter le domaine du text mining, afin d'introduire l'algorithme Word2Vec. De plus, des tests de méthodes (dont les résultats sont présentés en complément) ont été effectués dans le cadre du chapitre 1 sur une base de données issues du text mining.

IV-C) Etude des textes et du langage

Les méthodes évoquées plus haut trouvent leurs applications dans un grand nombre de domaines. Chaque secteur possède des données différentes, obligeant les chercheurs à spécialiser les algorithmes afin de contourner les difficultés liées à ces spécificités. En particulier, l'étude automatique des textes et des langages présente des données peu similaires avec les données numériques, si bien qu'elle a longtemps été un domaine où peu d'algorithmes étaient exploitables à grande échelle.

Que ce soit pour étudier des textes (text mining) ou pour étudier les langues (Native Language Processing, NLP), cette discipline a pour but d'analyser automatiquement les mots, les phrases et les textes, afin d'offrir des applications aussi multiples que la traduction automatique, l'analyse de sentiments, la synthèse de textes, offrant à terme la perspective de créer une intelligence artificielle capable de communiquer avec l'homme.

Les données issues de textes présentent quelques spécificités. Tout d'abord, il est difficile d'assimiler les mots à des données numériques. Les espaces dans lesquels vivent les textes sont ainsi très différents des espaces normés usuels. Il est donc nécessaire d'utiliser des métriques plus adaptées à ces données, comme la similarité cosinus (Huang, 2008).

De plus, tous les mots n'ont pas la même importance dans une phrase. Des mots comme "artifice" ou "football" vont apporter beaucoup plus d'informations que des mots comme "être" ou "la", trop fréquents pour être discriminants. Des méthodes, comme la méthode tf-idf (term frequency - inverse document frequency) sont alors nécessaires afin de faire ressortir les mots les moins fréquents, qui sont souvent les mots les plus caractérisants.

Les mots présentent une difficulté supplémentaire liée aux évolutions historiques du langage : ils peuvent posséder plusieurs sens. Sans contexte, il est impossible de savoir si le terme "constitution" fait référence à la constitution d'un état, à la constitution physique, ou à la création d'une entité. L'analyse des mots proches d'un terme dans une phrase permet normalement d'en identifier le sens précis, mais constitue un obstacle informatique important.

Du fait du grand nombre de difficultés associées, le text mining et le NLP ont fait l'objet d'un nombre impressionnant de recherches. Seules deux techniques classiques - l'indexation sémantique latente, ou Latent Semantic indexing (LSI) et l'embedding de mots, ou word embedding - seront évoquées ici, mais une revue de littérature recensant les avancées effectuées avant 2007 a été effectuée par Feldman et Sanger (Feldman and Sanger, 2007).

Latent Semantic Analysis

L'approche informatique d'un texte suit généralement un procédé standard. Chaque texte est d'abord découpé en morceaux, appelé mots, et la ponctuation est supprimée. Cette opération est assez simple, mais ne doit pas forcément être traitée trop vite : des mots comme "aujourd" et "hui", ou "Victor" et "Hugo" doivent-ils être considérés comme deux mots différents ou comme un seul lorsqu'ils se suivent ? Cette opération est appelée segmentation (ou Tokenization en anglais).

Les mots sont ensuite retraités afin de fusionner certains mots qui contiennent la même information. Par exemple "jouent" et "jouer" font référence à la même action et peuvent être régroupés au sein d'un même mot, "jou". Cette opération est appelée racinisation (ou stemming en anglais).

Le procédé classique consiste ensuite à construire une matrice de fréquence. Chaque texte représente une ligne, chaque mot représente une variable différente, et la matrice contient le nombre de fois que chaque mot apparait dans le texte.

Cette matrice est ensuite prétraitée, par exemple avec la méthode tf-idf. Soit N le nombre de textes présents dans un corpus. Soit i un mot présent dans le corpus, alors, en notant N_i le nombre de textes qui contiennent au moins une fois le mot i, on définit le coefficient

$$idf(i) = log(\frac{N}{n_i}). \tag{24}$$

Ensuite, si un texte j contient au total n_j mots, dont n_{ij} fois le mot i (n_{ij} est donc le coefficient i, j de la matrice de fréquence), la fréquence du mot i est donnée par

$$tf(i, j) = \frac{n_{ij}}{n_j}. \tag{25}$$

Les coefficients i, j de la matrice de fréquence prétraitée par tf-idf sont alors donnés par

$$tf_idf(i, j) = tf(i, j) * idf(i). \tag{26}$$

Cette matrice prétraitée présente généralement deux caractéristiques majeures : elle est de très grande dimension (plusieurs milliers de colonnes, même après l'étape de racinisation), et elle est très creuse. Afin d'être plus facilement exploitable, une méthode de réduction de dimension est appliquée. Généralement, la méthode retenue est l'algorithme SVD, auquel cas l'ensemble du procédé s'appelle Indexation Sémantique Latente (Latent Semantic Indexing, ou LSI), mais d'autres méthodes de réduction de dimension, comme les NMFs, peuvent être choisies [cf (Deerwester et al., 1989) ou (Peter et al., 2009)]. L'approche devient alors similaire avec celle adoptée dans le chapitre 1 de la thèse.

Cette manière de faire offre de très bons résultats et a longtemps été la meilleure disponible. Elle souffre cependant d'un inconvénient majeur : l'utilisation d'une matrice de fréquence écrase l'ordre d'apparition des mots. Afin de capter le contexte, la méthode LSI étudie si les apparitions de deux mots dans un même texte sont corrélées, permettant ainsi de comprendre qu'un texte parlant de musique parlera généralement de notes, de musiciens et d'instruments. Ce n'est pas idéal, car le contexte et une notion locale : avoir le mot "note" dans la même phrase que "instrument" caractérise plus le contexte de ce dernier que si le mot "note" était au début du texte et "instrument" à la fin. Des notions importantes comme les homonymes, voire l'ironie, sont ainsi difficiles à capter par cette méthode.

Réseaux de neurones et text mining

En 2012 a eu lieu une révolution dans l'étude des textes et du langage : Mikolov et al. proposaient une nouvelle architecture de réseaux de neurones, appelé Word2Vec, capable de prendre en compte efficacement le contexte des mots [(Mikolov et al., 2013a) et (Mikolov et al., 2013b)]. Ces travaux et ceux qui ont suivi ont bouleversé l'univers de l'étude informatique du langage, en améliorant drastiquement les performances des algorithmes. Même le test de Turing [9], dont on pensait qu'il ne serait pas atteint avant des années, semble désormais à portée : en 2014, Warwick déclare avoir créé avec son équipe de chercheurs une intelligence artificielle qui a surmonté le test. Même si on a appris depuis que cette annonce a été fortement exagérée [10], les progrès sont immenses et se retrouvent dans la vie de tous les jours, via des interfaces comme Siri ou Alexa.

L'idée de départ de Mikolov et son équipe est simple. Afin d'étudier un mot, l'algorithme va considérer les k mots immédiatement avant et les k mots immédiatement après lui dans le texte. Cette fenêtre va constituer le contexte du mot étudié.

Chacun des mots du contexte est ensuite encodé dans ce qui s'appelle un "one hot vector" afin d'être transmis au réseau de neurones. Un one hot vector est un vecteur de taille n (n étant le nombre de mots différents présents dans le corpus de texte), dont chacune des coordonnées correspond à un mot différent du corpus. Toutes les coordonnées du one hot vector sont arbitrairement fixées à 0, à l'exception de celle associée au mot encodé, qui est fixée à 1 (voir Figure 7).

Tous les one-hot vector (à l'exception de celui lié au mot étudié) sont ensuite présentés en même temps à un réseau de neurones à couches cachées. Ce réseau ne comprend qu'une seule couche cachée possédant k neurones. Le réseau a alors pour tâche d'essayer de retrouver le mot étudié [11]. Cette architecture est appelée CBOW (pour Continuous Bag Of Word).

9. Le test de Turing consiste à faire dialoguer une intelligence artificielle avec des humains. Ceux-ci doivent alors décider si l'entité avec laquelle ils discutent est humaine ou immatérielle. Si la majorité se trompe, alors le test est validé.

10. L'intelligence artificielle était présentée comme un enfant aux évaluateurs, qui n'avaient que 5 minutes

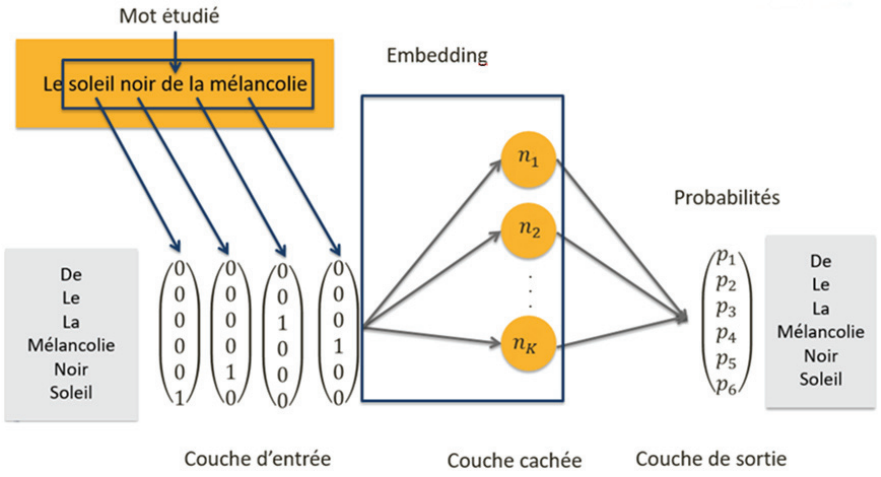

FIGURE 7 – Schéma de l'architecture neuronale Word2Vec

Une fois le réseau entraîné, les poids des neurones de la couche cachée sont récupérés. Ils forment désormais un espace vectoriel des mots : c'est ce qu'on appelle un plongement de mot ("word embedding" en anglais). Cet espace vectoriel capture le sens de chaque mot et permet d'effectuer des opérations sur les mots. Par exemple, au terme de "roi" sont associées des notions de masculin et de règne. Ainsi, dans l'espace nouvellement créé, effectuer les opérations "roi" - "homme" + "femme" donne le résultat "reine". Projeter des textes dans un tel espace permet de les caractériser et réduit fortement la dimensionnalité.

Bien que très efficace, cette méthode souffre de deux limites importantes. D'une part, il n'existe pas d'explication théorique réellement convaincante sur la raison de ces bons résultats. De plus, cette méthode a tendance à être peu stable, ce qui signifie que si l'ordre des textes présentés au réseau de neurones lors de l'apprentissage change, les résultats finaux seront eux aussi différents (Pierrejean and Tanguy, 2018). Une revue de littérature complète sur le sujet a été effectuée par Li et Yang (Li and Yang, 2018).

La méthode Word2Vec a été utilisée dans le chapitre 2, afin de constituer une sorte d'embedding de prestation santé, qui constitue une base d'un espace vectoriel de faible dimension sur laquelle il est possible de projeter les assurés. Elle agit alors comme une méthode de réduction de dimension. Quelques détails techniques sur l'adaptation de cette méthode (dont la méthode de prétraitement des données et la manière dont les assurés sont projetés sur l'embedding) sont présentés en complément, à la fin de ce chapitre.

pour dialoguer avec elle. De plus, seuls 30% des juges se sont trompés sur la nature réelle de leur interlocuteur.

11. Mikolov et al. ont aussi proposé une architecture alternative, appelé "Skip-gram". Dans cette architecture, seul le mot étudié est présenté au réseau de neurones. Sa tâche est alors de retrouver le contexte entier du mot.

IV-D) Résumé des méthodes et de leur utilisation dans cette thèse

Cette partie a pour but de simplifier et de clarifier l'utilisation des méthodes informatiques présentées précédemment.

Trois types d'algorithme sont utilisés dans les chapitres 1 et 2 de la thèse : des techniques de classification non supervisée, de réduction de dimension et d'analyse de texte.

Quatre méthodes de classification non supervisée ont ainsi été présentées. Parmis elle, la méthode des cartes auto organisatrices de Kohonen est la plus utilisée dans cette thèse puisqu'elle représente la technique de classification principale utilisée dans les chapitres 1 et 2. L'utilisation de cette méthode est systématiquement accompagné de l'utilisation d'une Classification Ascendante Hierarchique, afin de réduire le nombre de classe obtenue. Les deux dernières méthodes présentées, les algorithmes des k-means et des fuzzy c-means, ont été utilisé pour challenger la méthode des cartes de Kohonen, mais les résultats n'étant pas probant, ils ne sont pas présentés ici.

Afin d'améliorer les méthodes de classification, des méthodes de réduction de dimension peuvent être utilisées. Quatre d'entre elles ont été présentées. Ainsi, la méthode NMF est utilisée à la fois dans les chapitres 1 et 2. Cette méthode est challengée par la méthode ACP dans le chapitre 1 et par une méthode d'auto encodeur, la méthode mSDA, dans le chapitre 2. Enfin, la méthode SVD a elle aussi été utilisé pour comparer la méthode NMF mais les résultats ne sont pas présentés ici, les performances étant moins bonnes.

Enfin, des méthodes d'analyse de texte sont présentées. En effet, comme développé dans le chapitre 2 et les compléments à ce chapitre et au chapitre 1, il est possible de voir que certains types de données temporelles, comme celles d'assurance santé, sont très proches de données textuelles. Ainsi, les méthodes LSA et de word embedding (et en particulier l'algorithme Word2Vec) sont présentées. Les méthodes LSA ne sont pas utilisés dans cette thèse, mais leur philosophie est proche du processus utilisé dans le chapitre 1. Enfin, l'algorithme Word2Vec est utilisé dans le chapitre 2 afin de proposer une méthode alternative à la méthode présentée dans le chapitre 1.

V) Contributions et principaux résultats

Cette thèse est composée de cinq chapitres, chaque chapitre reprenant un article, et éventuellement d'une annexe en français complétant les résultats. L'un des articles à été accepté à la revue IME, l'un est en cours de révision et les 3 autres sont en attentes de retour de la part des revues.

La classification d'assurés

Les chapitres 1 et 2 traitent tous deux de la difficulté liée au ciblage d'une action de prévention.

Le chapitre 1 [12] propose d'exploiter les bases de prestations afin de classifier les assurés par groupe homogène de risque. À cette fin, il propose une méthode non supervisée en trois temps, inspirée de la méthode LSI :

1. Une matrice de fréquence est constituée
2. Une méthode de réduction de dimension est appliquée
3. Les individus sont ensuite classifiés

Les méthodes d'Analyse en Composante Principale (ACP) et Nonnegative Matrix Factorization (NMF) ont été comparées afin de réaliser l'étape 2, et la méthode NMF a été retenue. De même, les méthodes des cartes de Kohonen et des k-means ont été testées pour l'étape trois, la première s'étant montrée la plus performante.

La méthode a été appliquée à un jeu de données supervisé issu du text mining, montrant des résultats satisfaisants. Ceux-ci sont exposés dans un complément placé en fin de chapitre.

La méthode NMF utilisée lors du chapitre 1 a été challengée dans le chapitre 2 [13] par des méthodes de réduction de dimension issues du text mining.

Ainsi, les méthodes Word2Vec et marginalized Stacked Denoising Autoencoders (mSDA) ont été adaptées afin de pouvoir être appliquées à des données d'assurance. La méthode Word2Vec offre des résultats satisfaisants, mais peu stables. La méthode mSDA est celle qui présente les résultats les plus stables et est la seule à détecter de manière consistante certaines corrélations entre actes, comme les actes de maternités.

Une méthode, proche de la notion de dispersion proposée par Brunet Brunet et al. (2004), permettant de comparer la stabilité des résultats issue d'un algorithme de clustering en calculant une mesure de similarité entre partitions d'un espace est également proposée.

Si les algorithmes employés dans ces chapitres sont très utilisés, ils n'ont jamais été appliqués dans un contexte d'assurance santé. Afin de pouvoir les utiliser dans un tel contexte, il a été nécessaire de voir les données santé comme du texte, ce qui n'est pas la manière traditionnelle de traiter ces données. De manière générale, les bases de données utilisées dans les deux études présentées dans cette thèse sont peu exploitées par les assureurs.Il s'agit des bases de données dites "de prestations".

12. Ce chapitre reprend un article coécrit avec Stéphane Loisel et Jean-Louis Rullière et soumis à la revue European Actuarial Journal.
13. Ce chapitre reprend un article corédigé avec Jean-Pascal Hermet et soumis au journal "Future Generation Computer Systems".

La psychiatrie en assurance

En appliquant les méthodes des chapitres 1 et 2 à différentes bases de données composées d'assurés actifs, il est apparu que la psychiatrie constituait systématiquement une des classes les plus chères pour les actifs. Il n'y avait cependant pas de littérature actuarielle sur le sujet de la psychiatrie en assurance permettant de confirmer ce résultat.

Cette absence de littérature a motivé la constitution du chapitre 3 [14]. Il présente une étude statistique descriptive du risque psychiatrie.

Via l'étude de quatre bases de données, il met en évidence le fait qu'aller chez le psychiatre est corrélé avec une probabilité plus forte d'avoir des dépenses santé (même corrigé des dépenses de psychiatre), et, en cas de dépense, à des montants de dépense plus élevés. Cela est vrai pour la plupart des postes de dépense (kinésithérapie, optique...). Ainsi, en moyenne, un assuré consommant au moins une fois de la psychiatrie demande deux fois plus de remboursements qu'un assuré moyen.

La modélisation de la prévention

Enfin, il est extrêmement difficile d'évaluer une action de prévention en assurance du fait de l'absence de modèle.

Le chapitre 4 [15] propose d'adapter le modèle de Poisson composé afin d'y intégrer un paramètre d'auto-protection. En échange d'une partie des primes (invariante au cours du temps), le rythme de survenance des sinistres diminue. Ce chapitre étudie l'investissement en prévention que l'assureur devrait effectuer afin d'optimiser un certain nombre d'indicateurs, y compris la probabilité de ruine et les bénéfices réalisés. Une analyse de sensibilité du modèle aux différents paramètres est aussi effectuée. Il est notamment montré dans ce chapitre qu'optimiser la probabilité de ruine ne revient pas à optimiser les montants de réserve à un horizon donné, ce qui implique que le gestionnaire doit effectuer un compromis entre rentabilité à court terme et pérennisation de l'entreprise à long terme. Cependant, quel que soit l'indicateur choisi, le montant optimal d'investissement en prévention ne dépend pas des réserves initiales.

Le chapitre 5 [16] propose d'étendre les résultats du chapitre 4 à une entreprise soumise à plusieurs risques, dans un contexte où elle ne prévient que le risque le plus grave. Cela revient aussi à étendre les résultats au cas où la compagnie d'assurance finance un programme de prévention étant à la fois de l'auto-protection et de l'auto-assurance. En utilisant les ordres

14. Ce chapitre reprend un article coécrit avec Jean-Pascal Hermet et soumis au bulletin français d'actuariat.
15. Ce chapitre reprend un article coécrit avec Stéphane Loisel, Jean-Louis Rullière et Julien Trufin et accepté au journal Insurance : Mathematics and Economics.
16. Ce chapitre reprend un article coécrit avec Stéphane Loisel, Jean-Louis Rullière et Julien Trufin et soumis à la revue Journal of mathematical analysis and applications.

HMRL, ce chapitre fournit une condition suffisante pour que la compagnie d'assurance investisse en prévention. En revanche, le montant optimal d'investissement en prévention varie avec les réserves initiales. Les cas où le sinistre prévenu suit une loi à queue lourde ou à queue légère sont distingués, ce qui permet d'obtenir des résultats asymptotiques pour les montants optimaux d'investissement lorsque les réserves initiales deviennent grandes. Les montants optimaux limites obtenus ainsi sont systématiquement plus élevés que les montants optimaux calculés lorsque les réserves sont nulles, ce qui suggère que plus la compagnie d'assurance voit ses réserves augmenter, plus elle devrait investir en prévention.

Bibliographie

Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. (1985). A learning algorithm for boltzmann machines. Cognitive science, 9(1) :147–169.

Adler, R., Feldman, R., and Taqqu, M. (1998). A practical guide to heavy tails : statistical techniques and applications. Springer Science & Business Media.

Andersson, H. and Lundborg, P. (2007). Perception of own death risk. Journal of Risk and Uncertainty, 34(1) :67–84.

Asmussen, S. (2008). Applied probability and queues, volume 51. Springer Science & Business Media.

Asmussen, S. and Albrecher, H. (2010). Ruin probabilities, volume 14. World scientific.

Asmussen, S. and Taksar, M. (1997). Controlled diffusion models for optimal dividend pay-out. Insurance : Mathematics and Economics, 20(1) :1–15.

Assurance Maladie (2018). Améliorer la qualité du système de santé et maîtriser les dépenses–propositions de l'assurance maladie pour 2019. Paris : Caisse Nationale De L'assurance Maladie Des Travailleurs Salariés.

Assurance Maladie (2019). Cartographie des dépenses médicales de santé.

Baicker, K., Cutler, D., and Song, Z. (2010). Workplace wellness programs can generate savings. Health affairs, 29(2) :304–311.

Beaulieu, N., Cutler, D. M., Ho, K., Isham, G., Lindquist, T., Nelson, A., and O'Connor, P. (2006). The business case for diabetes disease management for managed care organizations. In Forum for Health Economics & Policy, volume 9. De Gruyter.

Beeferman, D. and Berger, A. (2000). Agglomerative clustering of a search engine query log. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 407–416.

Bellman, R. E. (2015). Adaptive control processes : a guided tour, volume 2045. Princeton university press.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful ? In International conference on database theory, pages 217–235. Springer.

Boisguérin, B., Bouvet, M., Ferretti, C., Grangier, J., Guibert, G., Lafon, A., Mikou, M., Montaut, A., Mauro, L., Perron-Bailly, E., et al. (2016). Les dépenses de santé en 2015. résultats des comptes de la santé : édition 2016.

Brisson, M., Bénard, É., Drolet, M., Bogaards, J. A., Baussano, I., Vänskä, S., Jit, M., Boily, M.-C., Smith, M. A., Berkhof, J., et al. (2016). Population-level impact, herd immunity, and elimination after human papillomavirus vaccination : a systematic review and meta-analysis of predictions from transmission-dynamic models. The Lancet Public Health, 1(1) :e8–e17.

Briys, E., Schlesinger, H., and vd Schulenburg, J.-M. G. (1991). Reliability of risk management : market insurance, self-insurance and self-protection reconsidered. The Geneva Papers on Risk and Insurance Theory, 16(1) :45–58.

Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the national academy of sciences, 101(12) :4164–4169.

Buchmueller, T. C., Couffinhal, A., Grignon, M., and Perronnin, M. (2004). Access to physician services : does supplemental insurance matter ? evidence from france. Health economics, 13(7) :669–687.

Buss, D. M. (2015). The handbook of evolutionary psychology, volume 2 : Integrations, volume 2. John Wiley & Sons.

Cebul, R. D., Rebitzer, J. B., Taylor, L. J., and Votruba, M. E. (2011). Unhealthy insurance markets : Search frictions and the cost and quality of health insurance. American Economic Review, 101(5) :1842–71.

Centeno, L. (1985). On combining quota-share and excess of loss. ASTIN Bulletin : The Journal of the IAA, 15(1) :49–63.

Centeno, L. and Simões, O. (1991). Combining quota-share and excess of loss treaties on the reinsurance of n independent risks. ASTIN Bulletin : The Journal of the IAA, 21(1) :41–55.

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., and Rong, X. (2015). Data mining for the internet of things : literature review and challenges. International Journal of Distributed Sensor Networks, 11(8) :431047.

Chen, M., Xu, Z., Weinberger, K., and Sha, F. (2012). Marginalized denoising autoencoders for domain adaptation. arXiv preprint arXiv :1206.4683.

Chevreul, K., Prigent, A., Bourmaud, A., Leboyer, M., and Durand-Zaleski, I. (2013). The cost of mental disorders in France. European Neuropsychopharmacology, 23(8) :879–886.

Chiappori, P.-A. and Salanié, B. (2000). Testing for asymmetric information in insurance markets. Journal of political Economy, 108(1) :56–78.

Chikhi, N. F., Rothenburger, B., and Aussenac-Gilles, N. (2007). A comparison of dimensionality reduction techniques for web structure mining. In IEEE/WIC/ACM International Conference on Web Intelligence (WI'07), pages 116–119. IEEE.

Chuang, K.-S., Tzeng, H.-L., Chen, S., Wu, J., and Chen, T.-J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. computerized medical imaging and graphics, 30(1) :9–15.

Courbage, C., Loubergé, H., and Peter, R. (2017). Optimal prevention for multiple risks. Journal of risk and insurance, 84(3) :899–922.

Courbage, C. and Rey, B. (2016). Decision thresholds and changes in risk for preventive treatment. Health economics, 25(1) :111–124.

Courbage, C., Rey, B., and Treich, N. (2013). Prevention and precaution. In Handbook of insurance, pages 185–204. Springer.

Cramér, H. (1930). On the mathematical theory of risk. Centraltryckeriet.

Cramér, H. (1954). On some questions connected with mathematical risk, volume 2. University of California Press.

Cruwys, T., Dingle, G. A., Haslam, C., Haslam, S. A., Jetten, J., and Morton, T. A. (2013). Social group memberships protect against future depression, alleviate depression symptoms and prevent depression relapse. Social science & medicine, 98 :179–186.

De Finetti, B. (1957). Su un'impostazione alternativa della teoria collettiva del rischio. In Transactions of the XVth international congress of Actuaries, volume 2, pages 433–443.

Debreu, D. and Catteau, J. (2003). Approche comparative des populations de déprimés adressés en hospitalisation et des pratiques entre généralistes et psychiatres. Synapse, 199.

Deerwester, S. C., Dumais, S. T., Furnas, G. W., Harshman, R. A., Landauer, T. K., Lochbaum, K. E., and Streeter, L. A. (1989). Computer information retrieval using latent semantic structure. US Patent 4,839,853.

Demailly, L. (2011). Sociologie des troubles mentaux. La Découverte.

Dervaux, B. and Eeckhoudt, L. (2004). Prévention en économie et en médecine. Revue économique, 55(5) :849–856.

Dickson, D. C. and Gray, J. (1984). Exact solutions for ruin probability in the presence of an absorbing upper barrier. Scandinavian Actuarial Journal, 1984(3) :174–186.

DREES (2018). Les dépenses de santé en 2017 - résultats des comptes de la santé - Édition 2018.

Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters.

Ehrlich, I. and Becker, G. S. (1972). Market insurance, self-insurance, and self-protection. Journal of political Economy, 80(4) :623–648.

Etner, J. and Jeleva, M. (2013). Risk perception, prevention and diagnostic tests. Health economics, 22(2) :144–156.

eurostat (2019). Healthy life years and life expectancy at birth, by sex (tps00150).

Fan, G., Filipczak, L., and Chow, E. (2007). Symptom clusters in cancer patients : a review of the literature. Current oncology, 14(5) :173.

Fang, H. and Gavazza, A. (2011). Dynamic inefficiencies in an employment-based health insurance system : Theory and evidence. American Economic Review, 101(7) :3047–77.

Feldman, R. and Sanger, J. (2007). The text mining handbook : advanced approaches in analyzing unstructured data. Cambridge university press.

Gaujoux, R. and Seoighe, C. (2010). A flexible r package for nonnegative matrix factorization. BMC bioinformatics, 11(1) :367.

Gerber, H. U. (1979). An introduction to mathematical risk theory. Number 517/G36i.

Gerber, H. U., Goovaerts, M. J., and Kaas, R. (1987). On the probability and severity of ruin. Astin Bulletin, 17(02) :151–163.

Golub, G. and Kahan, W. (1965). Calculating the singular values and pseudo-inverse of a matrix. Journal of the Society for Industrial and Applied Mathematics, Series B : Numerical Analysis, 2(2) :205–224.

Golub, G. H. and Reinsch, C. (1971). Singular value decomposition and least squares solutions. In Linear Algebra, pages 134–151. Springer.

Grangier, J. (2018). Le vieillissement de la population entraîne une hausse des dépenses de santé liées aux affections de longue durée.

Gruenberg, E. M. (1953). The prevention of mental disease. The ANNALS of the American Academy of Political and Social Science, 286(1) :158–166.

Gubler, T., Larkin, I., and Pierce, L. (2017). Doing well by making well : The impact of corporate wellness programs on employee productivity. Management Science, 64(11) :4967–4987.

Hans-Horst (2011). La prévention est bénéfique à la santé et aux affaires. Les essentiels de la sécurité sociale, Perspectives en politique sociale.

Herring, B. (2010). Suboptimal provision of preventive healthcare due to expected enrollee turnover among private insurers. Health Economics, 19(4) :438–448.

Hesselager, O. (1990). Some results on optimal reinsurance in terms of the adjustment coefficient. Scandinavian Actuarial Journal, 1990(1) :80–95.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6) :417.

Huang, A. (2008). Similarity measures for text document clustering. In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand, pages 49–56.

Huang, R., Xi, L., Li, X., Liu, C. R., Qiu, H., and Lee, J. (2007). Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mechanical systems and signal processing, 21(1) :193–207.

IPSOS and Klesia (2014). Perceptions et représentations des maladies mentales - rapport.

Jamoulle, M. and Roland, M. (1995). Quaternary prevention.

Kahneman, D. and Tversky, A. (1972). Subjective probability : A judgment of representativeness. Cognitive psychology, 3(3) :430–454.

Kalteh, A. M., Hjorth, P., and Berndtsson, R. (2008). Review of the self-organizing map (som) approach in water resources : Analysis, modelling and application. Environmental Modelling & Software, 23(7) :835–845.

Kim, H. and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics, 23(12) :1495–1502.

Knapp, M., McDaid, D., and Parsonage, M. (2011). Mental health promotion and mental illness prevention : The economic case.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological cybernetics, 43(1) :59–69.

Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78(9) :1464–1480.

Konrad, K. A. and Skaperdas, S. (1993). Self-insurance and self-protection : a nonexpected utility analysis. The Geneva Papers on Risk and Insurance Theory, pages 131–146.

Kumar, C. A. (2009). Analysis of unsupervised dimensionality reduction techniques. Comput. Sci. Inf. Syst., 6(2) :217–227.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755) :788.

Lee, K. (1998). Risk aversion and self-insurance-cum-protection. Journal of Risk and Uncertainty, 17(2) :139–151.

Li, Y. and Yang, T. (2018). Word embedding for understanding natural language : a survey. In Guide to Big Data Applications, pages 83–104. Springer.

Liou, C.-Y., Cheng, W.-C., Liou, J.-W., and Liou, D.-R. (2014). Autoencoder for words. Neurocomputing, 139 :84–96.

Lundberg, F. (1903). (1) Approximerad framställning af Sannolikhetsfunktionen :(2) Återförsäkring af Kollekrivrisker. Almquist.

Malmendier, U. M., Della Vigna, S., et al. (2002). Overestimating self-control : Evidence from the health club industry. Technical report.

Mbah, M. L. N., Durham, D. P., Medlock, J., and Galvani, A. P. (2014). Country-and age-specific optimal allocation of dengue vaccines. Journal of theoretical biology, 342 :15–22.

Mihalopoulos, C., Vos, T., Pirkis, J., and Carter, R. (2011). The economic analysis of prevention in mental health programs. Annual review of clinical psychology, 7 :169–201.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. arXiv preprint arXiv :1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pages 3111–3119.

Mingoti, S. A. and Lima, J. O. (2006). Comparing som neural network with fuzzy c-means, k-means and traditional hierarchical clustering algorithms. European journal of operational research, 174(3) :1742–1759.

Ministère des Affaires sociales de la Santé et des Droits des femmes (2014). Etat des lieux du suicide en France.

Morgenstern, O. and Von Neumann, J. (1953). Theory of games and economic behavior. Princeton university press.

Muñoz, R. F., Beardslee, W. R., and Leykin, Y. (2012). Major depression can be prevented. American Psychologist, 67(4) :285.

Murtagh, F. (1995). Interpreting the kohonen self-organizing feature map using contiguity-constrained clustering. Pattern Recognition Letters, 16(4) :399–408.

Newby, L. K., Allen LaPointe, N. M., Chen, A. Y., Kramer, J. M., Hammill, B. G., DeLong, E. R., Muhlbaier, L. H., and Califf, R. M. (2006). Long-term adherence to evidence-based secondary prevention therapies in coronary artery disease. Circulation, 113(2) :203–212.

Nixon, J. and Ulmann, P. (2006). The relationship between health care expenditure and health outcomes. The European Journal of Health Economics, 7(1) :7–18.

Oyelade, O., Oladipupo, O., and Obagbuwa, I. (2010). Application of k means clustering algorithm for prediction of students academic performance. arXiv preprint arXiv :1002.2425.

Paatero, P. and Tapper, U. (1994). Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5(2) :111–126.

Parks, K. M. and Steelman, L. A. (2008). Organizational wellness programs : a meta-analysis. Journal of occupational health psychology, 13(1) :58.

Peter, R., Shivapratap, G., Divya, G., and Soman, K. (2009). Evaluation of svd and nmf methods for latent semantic analysis. International Journal of Recent Trends in Engineering, 1(3) :308.

Pierrejean, B. and Tanguy, L. (2018). Predicting word embeddings variability. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, pages 154–159.

Premalatha, K. and Natarajan, A. (2010). A literature review on document clustering. Information Technology Journal, 9(5) :993–1002.

Quiggin, J. (1991). Comparative statics for rank-dependent expected utility theory. Journal of Risk and Uncertainty, 4(4) :339–350.

Roelandt, J.-L., Caria, A., Defromont, L., Vandeborre, A., and Daumerie, N. (2010). Représentations sociales du «fou», du «malade mental» et du «dépressif» en population générale en france. L'encéphale, 36(3) :7–13.

Rothschild, M. and Stiglitz, J. (1978). Equilibrium in competitive insurance markets : An essay on the economics of imperfect information. In Uncertainty in economics, pages 257–280. Elsevier.

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk : I. a definition. Journal of Economic theory, 2(3) :225–243.

Sandu, A.-L., Artiges, E., Galinowski, A., Gallarda, T., Bellivier, F., Lemaitre, H., Granger, B., Ringuenet, D., Tzavara, E. T., Martinot, J.-L., et al. (2017). Amygdala and regional volumes in treatment-resistant versus nontreatment-resistant depression patients. Depression and anxiety, 34(11) :1065–1071.

Segerdahl, C.-O. (1970). On some distributions in time connected with the collective theory of risk. Scandinavian Actuarial Journal, 1970(3-4) :167–192.

Shaked, M. and Shanthikumar, J. G. (2007). Stochastic orders. Springer Science & Business Media.

Smith, K. A. and Ng, A. (2003). Web page clustering using a self-organizing map of user navigation patterns. Decision Support Systems, 35(2) :245–256.

Sorzano, C. O. S., Vargas, J., and Montano, A. P. (2014). A survey of dimensionality reduction techniques. arXiv preprint arXiv :1403.2877.

Steinhaus, H. (1956). Sur la division des corp materiels en parties. Bull. Acad. Polon. Sci, 1(804) :801.

Stewart, G. W. (1993). On the early history of the singular value decomposition. SIAM review, 35(4) :551–566.

Vagizova, V., Ksenia, L., and Ivasiv, I. B. (2014). Clustering of russian banks : business models of interaction of the banking sector and the real economy.

Van Der Maaten, L., Postma, E., and Van den Herik, J. (2009). Dimensionality reduction : a comparative. J Mach Learn Res, 10(66-71) :13.

Verrall, R. J. and Yakoubov, Y. H. (1999). A fuzzy approach to grouping by policyholder age in general insurance. Journal of Actuarial Practice, 7 :181–204.

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning, pages 1096–1103. ACM.

Wall, M. E., Rechtsteiner, A., and Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In A practical approach to microarray data analysis, pages 91–109. Springer.

Wang, D., Cui, P., and Zhu, W. (2016a). Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1225–1234. ACM.

Wang, Y., Yao, H., and Zhao, S. (2016b). Auto-encoder based dimensionality reduction. Neurocomputing, 184 :232–242.

Welling, M. and Weber, M. (2001). Positive tensor factorization. Pattern Recognition Letters, 22(12) :1255–1261.

Wittchen, H.-U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., et al. (2011). The size and burden of mental disorders and other disorders of the brain in europe 2010. European neuropsychopharmacology, 21(9) :655–679.

Wu, B., Wang, E., Zhu, Z., Chen, W., and Xiao, P. (2018). Manifold nmf with l21 norm for clustering. Neurocomputing, 273 :78–88.

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. Management science, 42(12) :1676–1690.

Xu, R. and Wunsch, D. C. (2005). Survey of clustering algorithms.

Yamin, D., Jones, F. K., DeVincenzo, J. P., Gertler, S., Kobiler, O., Townsend, J. P., and Galvani, A. P. (2016). Vaccination strategies against respiratory syncytial virus. Proceedings of the National Academy of Sciences, 113(46) :13239–13244.

Yang, D., Ma, Z., and Buja, A. (2011). A sparse svd method for high-dimensional data. arXiv preprint arXiv :1112.2433.

Yano, N. and Kotani, M. (2003). Clustering gene expression data using self-organizing maps and k-means clustering. In SICE 2003 Annual Conference (IEEE Cat. No. 03TH8734), volume 3, pages 3211–3215. IEEE.

Zhai, J., Zhang, S., Chen, J., and He, Q. (2018). Autoencoder and its various variants. In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 415–419. IEEE.

Chapitre 1

Health-policyholder clustering using health consumption : A useful tool for targeting prevention plans

Ce chapitre reprend un article coécrit avec Stéphane Loisel et Jean-Louis Rullière et soumis à l'European Actuarial Journal

Abstract

On paper, prevention appears to be a good complement to health insurance. However, its implementation is often costly. To maximize the impact and efficiency of prevention plans, these should target particular groups of policyholders. In this article, we propose a way of clustering policyholders that could be a starting point for the targeting of prevention plans. This two-step method mainly classifies using policyholder health consumption. The dimension is first reduced using a Nonnegative matrix factorization algorithm, producing intermediate health-product clusters. Policyholders are then clustered using Kohonen's map algorithm. This leads to a natural visualization of the results, allowing the simple comparison of results from different databases. The method is applied to two real French health-insurer datasets. The method is shown to be easily understandable, and able to cluster most policyholders efficiently.

Clustering; Health insurance; Kohonen's map; Nonnegative Matrix Factorization; Prevention.

1.1 Introduction

Prevention, as it reduces risk, would appear to be a useful tool for insurers. However, until recently European insurance companies have been wary of prevention, which was generally seen as a marketing product. The first ambitious prevention plan was initiated in 1997 by Discovery in South Africa. It then took until 2016 for a major prevention plan to appear

in Europe: the Vitality program (arising from the merger of Discovery and Generali), first launched in Germany, allows policyholders to win points by adopting healthy lifestyles that can be converted into gifts or discount coupons.

There are a number of reasons why private companies might be reluctant to develop prevention. It is first difficult for insurance companies to achieve high participation rates in prevention plans. And even if they do, they could produce severe adverse selection: individuals who are the most interested in prevention plans are those with specific health risks Beaulieu et al. (2006). Moreover, policyholders can easily switch between health-insurance providers, rendering prevention investment less efficient for insurance companies Herring (2010).

It is also difficult to choose who will benefit from the plan. Data Protection (regarding health data in particular) has always been a touchy subject in Europe. For example, French mutual health-insurance companies cannot ask for any medical information when a policyholder takes out a new contract. Therefore, it is impossible to know precisely the make-up of the policyholders' health, since there is no access to necessary data. Without accessing this information, a supervised method aiming to reconstruct a known partition of the health make-up of policyholders is inconceivable. Consequently, in practice, a useful algorithm to predict health outcomes has to be unsupervised.

In addition, the new European General Data Protection Regulation (GDPR) has limited the possible uses of data. In particular, health data are explicitly considered as very sensitive. Article 22 of the GDPR states that private companies cannot take decisions that significantly affect an individual only based on an automated process. As such, the targeting of prevention plans by insurance companies is complicated.

Nevertheless, there are some exceptions to the GDPR, allowing insurance companies to use data in order to target prevention. First, the GDPR has introduced the notion of individual agreement. If an individual has agreed to a certain use of his personal data for a clearly-defined purpose, the use of these data is allowed. If she does not agree, it is still possible to use these data for regulatory compliance (such as creating financial reserves) or the production of aggregate statistics. The clustering method proposed here purports to meet all of these requirements.

Clustering is frequently used in insurance. For example, credibility theory, which leads to the bonus-malus system, is based on the idea that there exist hidden policyholder clusters (e.g. Bühlmann and Gisler (2006)). Computing class of policyholders with homogeneous risks can also help to define an insurance premium (e.g. Henckaert et al. Henckaerts et al. (2018)). Clustering can be used to identify fraud (Derrig and Ostaszewski Derrig and Ostaszewski (1995)) and classify risks (Yeo et al. Yeo et al. (2001)). It has been used to improve risk premium estimation by Verrall and Yaboukov Verrall and Yakoubov (1999). In

the health-insurance context, Ghoreyshi and Hosseinkhani classify policyholders using the k-means algorithm Ghoreyshi and Hosseinkhani (2015); Peng et al. apply clustering algorithms to detect fraud in a health-insurance dataset Peng et al. (2006); and Kuo et al. cluster Taiwanese respondents in a health-insurance dataset matched to medical information in order to identify the relationships between illnesses Kuo et al. (2007).

However, to the best of our knowledge, the clustering of policyholders based on the relationship between claims has not been addressed. Clustering methods in insurance do not thus fully exploit the available data, as they do not use factors like the individual's treatment program when ill.

A way to capture this kind of relationship consists in pre-processing the data in order to create a frequency matrix, containing one line by policyholder, and one column for each kind of medical act refunded (such as "pharmacy", "general practitioner", etc.). However, this matrix might be of high dimension, notably affecting the quality of classic clustering methods due to the dimension curse (e.g. Aggarwal and Yu (2000)).

The analysis of high-dimensional data has been a prolific research area (e.g. Zhang et al. (1996), Hinneburg and Keim (1999), Indyk and Motwani (1998)). One classic method of dealing with this consists in reducing the dimension before clustering (e.g. Agrawal et al. (1998), or Cheng et al. (1999)). When the dimension is reduced using a singular-value decomposition (SVD), the method is called Latent Semantic Analysis Deerwester et al. (1990). Of course, other dimension-reduction algorithms can be used. For example, Mote et al. Mote et al. (2017) use a Nonnegative Matrix Factorization algorithm (NMF) to reduce the dimension and a self-organizing map to cluster brain-tumor segmentation . To the best of our knowledge, these kinds of methods have not yet been used for the clustering of health policyholders.

The goal of this paper is to present a new clustering method based only on policyholder health consumption, in order to help prevention targeting. The resulting clusters capture particular health profiles. In order to do so, we use a similar process to that in Mote et al., Mote et al. (2017). The dimension is first reduced with the NMF algorithm, preceding the clustering of policyholders using a Kohonen's map. This method is then applied to two real French insurance databases. We demonstrate how to carry out each algorithm step via the detailed analysis of the results obtained for a subpart of one of these databases.

The remainder of the paper is organized as follows: Section 2 describes the different databases used in the presented study. Section 3 sets out the NMF and the Kohonen's map algorithms. Section 4 describes the clustering process and the tests that are carried out, as well as the global results. Examples of a prevention program for psychiatric diseases and falls are provided in Section 5. Lastly, Section 6 discusses the method and proposes some possible extensions.

1.2 Data presentation

This Section aims to present the databases used in this paper and the way they had been pre-processed.

Standard health insurance databases Health insurers usually possess two different databases. The first one is called the policyholder database, and contains all the information the insurer possesses about the policyholder: age, sex, contract details, contact information and so on. This information is typically used to analyze insurance-porfolio profitability.

This database is systematically matched to a second one: the health-consumption database. This latter contains all the information the health insurer needs to reimburse the policyholder when she buys a health product:[1] the nature (sometimes called the medical act), the date and the amount of the expenditure. The elements in this database, and the nature of the consumption in particular, can vary from one insurer to another: some reimburse a product but others do not, and some insurers have more detailed information than others. The health-consumption database can be large: from one million entries for a small mutual health-insurance company to over one billion for national health-insurance systems. This base may also depend on the national health system. For example, in France, medication for long-term diseases (such as cancer) is fully reimbursed by the national public insurer, and does not necessarily appear in these databases.

Presentation of the two databases used in this paper The method has been applied to databases from two different French private health insurance companies: a collective database (**CB**) and an individual database (**IB**). The CB and IB contain policyholders with top-range and mid-range market contracts respectively, observed over one year. Moreover, policyholders with zero health consumption are removed.

As the two databases come from different companies, there are some small differences. For example, the CB does not distinguish medical specialists with or without prescriptions. The CB also does not separate fee overruns and the legal copayment from the price of health consumption. On the other hand, the IB does not separate biological tests from blood tests. These differences are taken into account when comparing the two databases on Section 1.4.4.

pre-processing step For each database, four different splits are carried out (women over 62, women between 16 and 62, men between 16 and 62 and men over 62[2]), producing a total of eight different subdatabases. Splitting databases this way helps to reduce the heterogeneity

1. The term health product is used for every item of health expenditure that may be refunded by the insurer (such as GP visits, nights at the hospital, medication and glasses).
2. The legal retirement age in France is 62.

between populations, and speeds up the process. Figure 8 contains descriptive figures for the eight populations.

		Individual database		Collective database	
		Headcount	Average age	Headcount	Average age
Men	-62 years	17153	41	15415	43.7
	+62 years	7949	73.9	12691	71.8
Women	-62 years	38386	42.8	15365	43.4
	+62 years	19727	71.8	13269	74.1

Figure 8 – Databases statistics

We set out the results in detail for the eight subdatabases in Section 1.4.4. For the sake of concision, only the subdatabase of the IB containing women over 62 is used to illustrate the results on Sections 1.4.1 and 1.4.2. This subdatabase includes over 500 000 different health reimbursements, and 160 different health products. Most of the entries concern extra charges (such as additional fees for night consultations), apart from extra charges for home-care services that are dropped. We merge identical health products that appeared separately due to spelling mistakes. Last, every optical product (such as glasses and contact lenses) are merged into a single optical product, although this does not change the results. After these changes, a typical health-consumption database contains around 100 different health products.

In order to start the process, each subdatabase has to be transformed into a frequency matrix first. This frequency matrix has one column for each different health products (i.e. around 100 columns) and one row for each policyholder. In this kind of frequency matrix, some columns (such as the column corresponding to the health product "Keratotomy") contain many zeros, whereas some contain almost no zero (such as the one corresponding to "Pharmacy").

This kind of imbalance can influence the results : the dimension reduction method gives a lot of weight to the frequent health consumption and tends to neglect the other one. This can be harmful, since less frequent health consumption might be more discriminating and informative than frequent ones. In order to tackle this subject, two pre-processing methods have been tested on frequency matrices : the tf-idf method (see Robertson (2004) for a general presentation) and applying a logarithm function to the database. On the two tested datasets, the clusters of policyholders obtained from the logarithm treatment are more homogeneous. Thus, the results obtained using the tf-idf pre-processing method are not shown in this paper.

1.3 Algorithms

This section sets out the main algorithms used in this article: the Nonnegative Matrix Factorization and Kohonen's map.

1.3.1 NMF algorithm:

This algorithm is a dimension-reduction method. First introduced by Paatero and Tapper Paatero and Tapper (1994) and Lee and Seung Lee and Seung (1999), it is almost unknown in the insurance sector. The only application of which we are aware is from Nesvijevskaia and Taudau, who announced that they had used this method at the 17th Rencontre MutRé Nesvijevskaia and Taudou (2016).

However, this method is widely used in medicine to analyze the human genome (e.g. Brunet et al. (2004), Lawrence et al. (2013)) and in text mining to extract features (e.g. Lee and Seung (1999), Pauca et al. (2004), Jones and Chung (2016)). The NMF can also be used as a clustering method (e.g. Kuang et al. (2015))

For all $n, m \in \mathbb{N}$, we denote by $\mathcal{NM}(n, m)$ the set of nonnegative matrices with n rows and m columns. Let $n, m, k \in \mathbb{N}$, $V \in \mathcal{NM}(n, m)$. The purpose of the NMF is to find $W \in \mathcal{NM}(n, k)$, $H \in \mathcal{NM}(k, m)$ such that $V \approx WH$.

To do so, a number of algorithms have been proposed (e.g. Lee and Seung (Lee and Seung (1999), Lee and Seung (2001)), Brunet et al. Brunet et al. (2004), Pascual-Montano et al. Pascual-Montano et al. (2006), Badea Badea (2008), Kim and Park Kim and Park (2007) and Pauca et al. Pauca et al. (2006)). Except for the latter, all of these have been tested. The "snmf/l" algorithm, created by Kim and Park, has been used for obtaining the presented results [3].

This method combines the cost function proposed by Pauca et al. Pauca et al. (2006) with the concept of sparseness, first introduced in an NMF algorithm by Hoyer Hoyer (2004).

Kim and Park propose to find $\min_{W,H}(\frac{1}{2}[\|V - WH\|_2^2 + \alpha \|H\|_2^2 + \beta \sum_{i=1}^{n} \|W(i,.)\|_1^2])$, with β being a coefficient controlling for the sparseness of W, α a coefficient reflecting H's smoothness, as suggested by Pauca et al. and $\| \ \|_2$ and $\| \ \|_1$ designing respectively the L_2 and the L_1 norms.

It is important to note that the cost function is not convex. Therefore, the snmf/l algorithm is only able to find local optima and is sensitive to the initial values of W and H. The classic way to deal with this is to randomly initialize the two matrices a number of times and then compare the resulting local minima. Following the work of Utsumi Utsumi (2010), this has been used in this paper, however other methods exist (e.g. Boutsidis and Gallopoulos (2008), Langville et al. (2006)).

3. Six different implementations have been tested: those proposed by Lee and Seung (Lee, Lee and Seung (1999)), Brunet et al. (Brunet, Brunet et al. (2004)), Pascual-Montano et al. (nsNMF, Pascual-Montano et al. (2006)), Badea (Offset, Badea (2008)), and the two of Kim and Park (snmf/l and snmf/r, Kim and Park (2007)). The "snmf/l" algorithm yields one of the best results, while being significantly faster. Implementations from the R package "NMF", developed by Gaujoux and Seoighe Gaujoux and Seoighe (2010), were used in the analysis presented here.

Starting from a random value, Kim and Park propose the following update rules to find a local minimum:

$$H_{n+1} = \min_{H \geq 0} \left\| \begin{pmatrix} W_n \\ \sqrt{\beta} e_{1\times k} \end{pmatrix} H - \begin{pmatrix} V \\ 0_{1\times n} \end{pmatrix} \right\|_2^2,$$

$$W_{n+1} = \min_{W \geq 0} \left\| \begin{pmatrix} H_{n+1}^T \\ \sqrt{\alpha} I_k \end{pmatrix} W - \begin{pmatrix} V \\ 0_{k\times m} \end{pmatrix} \right\|_2^2,$$

with n, m being the dimensions of A, k the final dimension, $e_{1\times k}$ a vector of height k containing only 1, and I_k the identity matrix. This method, from Van Banthem and Keenan Van Benthem and Keenan (2004), is called ANLS (Alternative Nonnegativity constrained Least Squares) and guarantees convergence. The stopping criterion is based on the optimality criterion of Karush Kuhn Tucker.

1.3.2 Kohonen's map algorithm:

Kohonen's map, also called a self-organizing map (SOM), is a clustering method based on neural networks Kohonen (1990). [4]. In this network, every neuron is arranged according to a given topology, usually a two-dimensional grid with square or a hexagonal disposition. This way, each neuron has neighboring neurons.

Say that one wishes to classify N policyholders using a a n-neuron SOM. To start with, each neuron is assigned a random weight m_i [5]. For each learning iteration t, a random policyholder is chosen. It is then possible to determine the neuron that best represents this policyholder, by solving $c = \min_{j \in [\![1,n]\!]} \|x - m_j(t)\|$. The neuron c is called the best machine unit (BMU).

Once the BMU is determined, its weight is adjusted in order to improve policyholder representativeness: $m_c(t+1) = \alpha(t)(x - m_c(t))$. The coefficient $\alpha(t)$ is the learning rate. This falls over time, so that learning is fast at the beginning and meticulous at the end.

In order to create an influence zone, the BMU's neighbors' weights are also changed. To do this, it is necessary to define a neighborhood function $h(c, i, t)$. This function falls with the distance between the neuron i and the BMU c. It also falls over time: at the beginning many neurons are adjusted, while at the end only few are. Last, the weight of neuron i becomes $m_i(t+1) = m_i(t) = \alpha(t)h(c, i, t)(x - m_c(t))$. This process is repeated several times.

4. The R package "Kohonen", developed by Wehrens et al. (Wehrens et al. (2007)), was used in the analysis presented here.

5. a well-known alternative is to choose the starting points via a PCA, however Akinduko et al. show that this is not suitable for non-linear datasets Akinduko et al. (2016)

Kohonen's maps have been used a few times in insurance, for example by Hainaut in order to classify motorcycles insurance policies Hainaut (2019), or by Brockett et al. in order to study claims fraud Brockett et al. (1998).

1.4 The clustering process

The clustering method presented in this paper contains two steps: the dimension is first reduced and then policyholders are clustered. Section 1.4.1 discusses the dimension reduction step.

1.4.1 Dimension reduction using the NMF algorithm

After pre-processing the health database, a matrix with around 100 dimensions is obtained. This is too large for traditional clustering methods, such as k-means, which do not perform very well when there are over 15 dimensions or so (e.g. Beyer et al. (1999)). [6] Thus a dimension reduction is needed.

The sNMF/l algorithm is thus applied to the pre-processed frequency matrix, in order to reduce dimension. After calibrating the algorithm by examining the silhouette, cophenetic and dispersion [7], the final space covers 20 dimensions. This way, two matrices W and H are obtained (see Section 1.3.1), respectively of dimensions closed to 20 000x20 and 20x100. It is important to note that both of them can be interpreted.

The matrix H contains one column for each health product, and 20 lines. There are two ways to make this matrix easier to understand. First, it is possible to normalize each row, by dividing each coefficient by the sum of all the other row coefficients. This normalized matrix can be used to create a heatmap [8] (see Figure 9). On such a heatmap, a dark case reveals a high normalized coefficient. It shows the way each new dimension is represented by each health product.

This matrix shows that each new dimension can be understood as a health-product cluster (HPC). For example, Dimension 7 (row 7 in Figure 9), gathering "Medical apparatus flat fees", "Orthotics" and "Small medical apparatus", is a medical-apparatus HPC. Thus, for the rest of the paper, each new dimension is called an HPC. Each HPC can be easily interpreted.

6. For the databases used here, dimension reduction dramatically improves clustering.

7. These three concepts are three quality indicators commonly measured for a clustering. Cophenetic and dispersion aim to measure the stability, and have been introduce by Brunet et al. Brunet et al. (2004). The silhouette has been presented by Rousseuw in order to test if individuals are well clustered Rousseeuw (1987).

8. The health product "Legal copayment" may be unfamiliar to the reader. In the French health system, many health products are partially reimbursed by the public insurer, "l'Assurance Maladie". The price of health products is fixed by Law (for example, a GP consultation costs 25 Euros). However, the public insurer does not refund all of this amount (only 16.5 Euros for GPs) in order to limit health consumption. We here call the 25-16.5 gap the "legal copayment". Moreover, GPs are allowed to charge higher fees that are not covered by the public insurer. The reimbursement of the legal copayment is usually covered by private insurance.

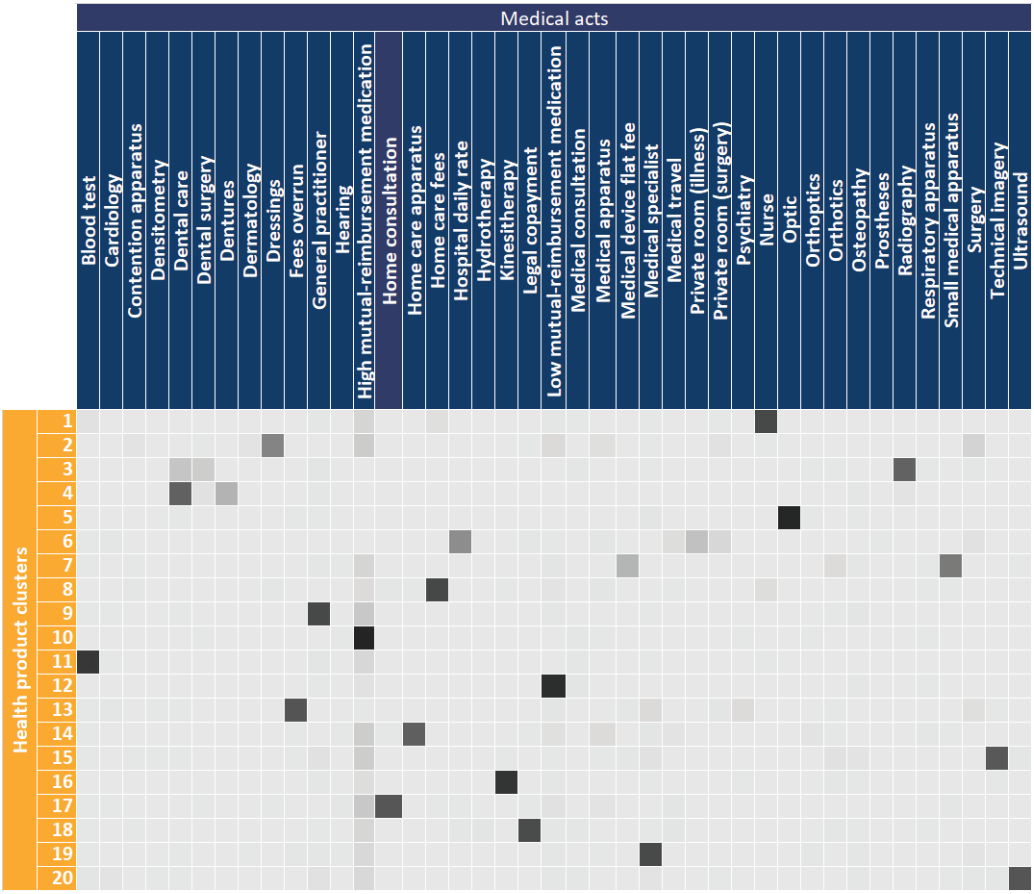

Figure 9 – The horizontal normalization of H. Only the most common health products are shown for clarity reasons.

Furthermore, even though "Technical imagery" looks similar to "Radiology", they do not appear in the same HPC. Whereas a human may have made the mistake of merging these two health products, the algorithm separates between them. This shows that merging applied before using this algorithm can substantially affect the final results, and thus should be carried out with caution.

In Figures 9 and 10 are only represented the most frequent health products for the sake of clarity. Others, such as "Medical travel", are not shown. However, as shown in sub-Section 1.4.4, the most important health products are not necessarily the most common: less-frequent health products (such as "Orthoptics") can be significant.

In order to bring out the role of less-common products, and thus to specify the meaning of each HPC, it is also possible to normalize H by columns (see Figure 10). If each HPC is seen as a cluster, this normalization shows to which cluster each health product belongs. Thus, a dark case means that the health product belongs strongly to the associated HPC.

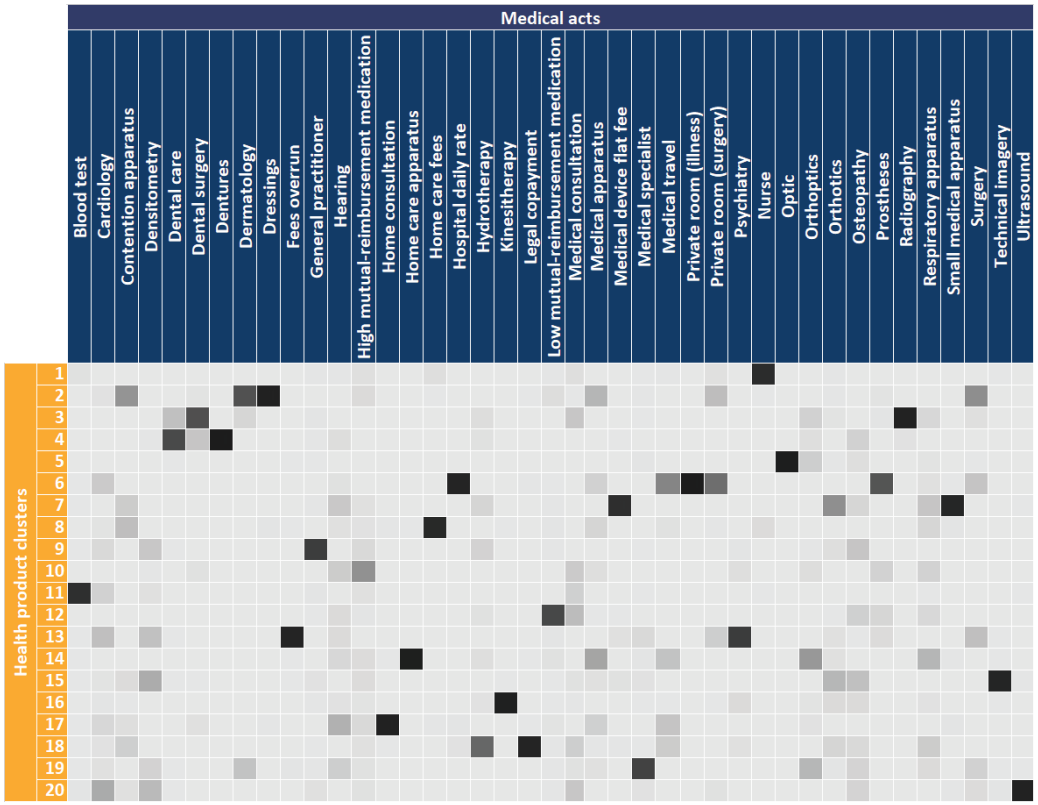

Figure 10 – The vertical normalization of H. This heatmap helps to interpret the meaning of each HPC.

Using this new heatmap allows for a better understanding of each HPC. For example, HPC 3 can now be seen to cover dental surgery, so that "Radiography" means "Dental radiography". HPC 15, containing "Densitometry", "Technical imagery", "Orthotics" and "Osteopathy", can be seen as a fracture HPC. As these results come from data on older respondents, this HPC reflects those who have experienced falls.

Once the matrix H is understood, matrix W is easy to read: this contains one line for each policyholder and 20 columns, one for each new dimension. As the latter are interpreted as HPCs, W shows the HPC consumption for each policyholder. The dimension of this matrix is reasonable, allowing to perform a clustering. For the following, the matrix H is called the HPC matrix and the matrix W the policyholders matrix.

1.4.2 Clustering using Kohonen's map

Once the dimension has been reduced, it is possible to cluster policyholders using the policyholder matrix W.

The classic Kohonen's map method with a linear neighborhood function and a learning rate

decreasing linearly has been used. Gaussian neighborhood functions have also been tested. However, on our datasets, they often produce maps with more empty neurons, although this is not always the case. The map associated to the presented results is given in Figure 11.

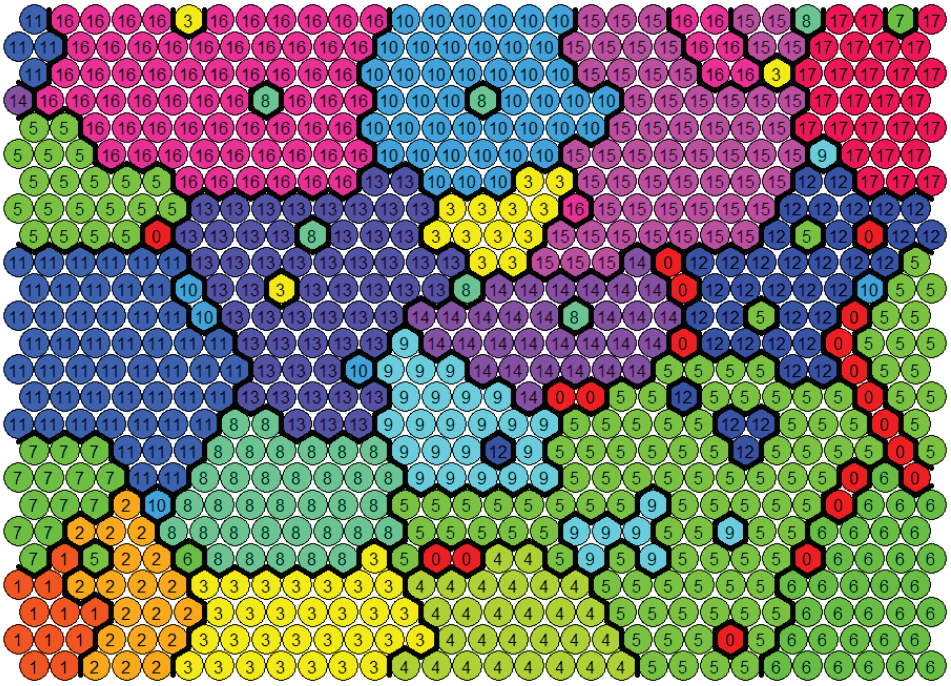

Figure 11 – An example of one of the Kohonen's maps. The associated cluster meanings can be found in Figure 12. Each circle represents a neuron, and each color / number represents a different cluster. The red neurons are empty.

In this map, the neurons are disposed on a hexagonal grid with 20 lines (25 neurons by lines). The hexagonal grid has been chosen following Kohonen's recommendations Kohonen (1990). We have chosen a high number of neurons on purpose. empirically, we might not be able to capture all the characteristics of the dataset if the number of neurons was to low. Moreover, in this case, the amount of clusters is chosen in advance and becomes a parameter. A way to deal with this is to choose a high number of neurons, and then to reduce the amount of clusters by grouping on a same cluster similar neurons (for example by performing a Hierarchical Agglomerative Clustering (HAC)), as first suggested by Murtagh Murtagh (1995). The high number of neurons allows for more flexibility, and the HAC allows to find empirically a suitable number of clusters and to reduce overfitting. After processing the HAC and examining the resulting dendrogram, 17 final clusters are retained.

Before applying the Kohonen's map algorithms, the matrix W is scaled. Usually, the input matrix is scaled by columns. This way, all variables have the same weight. However, empirically and surprisingly, scaling by rows offers much better results on the studied datasets than scaling by columns. Such an operation makes every policyholder comparable : a policyholder

with a lot of health consumption will carry the same weight as the one with only one. This allows the algorithm to focus more on the way a policyholder's health consumption is divided among the different HPCs and less on the overall amount of health consumption made.

It is possible to notice on Figure 12 that some of the neurons are empty (the neurons with a 0). One tries to avoid this in traditional approaches, as it means that some clusters are empty so that there are too many clusters. However, it is acceptable when self-organizing maps are combined with a HAC: the number of neurons is not the number of clusters. Here, empty neurons mainly mark the edge between clusters and low-density areas. To remove empty neurons, we would need to drastically reduce the number of neurons, and cluster quality would suffer.

The visual size of the clusters in Figure 11 is closely correlated with their real size. Cluster 5 is thus the most-populated.

To interpret the clusters, it is possible to calculate the centroid of each class, using the policyholder matrix W. This way, a policyholder of each cluster is computed, representing the whole cluster. This policyholder "type" is described by 20 characteristics, one for each computed HPC. Keeping this in mind, it is easy to interpret each cluster (see Figure 12). Note that a cluster "Fracture" is obtained, which can be read as individuals who are likely to suffer from falls, as the database here covers older respondents.

Policyholders clusters	Health Product Clusters																					
	Blood test	Dental care / denture	Dental surgery	Fees overrun	General Practitioner	Home care device	Home care fees	Home consultation	Hospitalization	Kinesitherapy	Legal copayment	Low refund drugs	Medical device	Medical specialist (with prescription)	Nurse	Optic	Radiography	Surgery	Ultrasound	Well refund drugs	Headcount	
1	0,4	0,1	0,0	0,1	0,3	0,2	1,8	0,6	0,3	0,3	0,3	0,3	0,2	0,2	1,7	0,1	0,1	0,4	0,1	0,4	169	Home care
2	0,5	0,1	0,1	0,2	0,4	0,0	1,1	0,1	0,2	0,2	0,2	0,3	0,2	0,3	1,1	0,1	0,2	0,4	0,2	0,4	295	Nurse home care
3	0,2	0,1	0,1	0,2	0,3	0,0	0,0	0,0	0,1	0,1	0,1	0,4	0,0	0,4	0,1	0,2	0,1	0,7	0,1	0,4	959	Surgery
4	0,3	0,1	0,1	0,2	0,3	0,9	0,0	0,1	0,0	0,1	0,1	0,4	0,1	0,3	0,1	0,2	0,1	0,1	0,1	0,4	577	Home care device
5	0,2	0,1	0,0	0,0	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,5	0,0	0,1	0,1	0,0	0,0	0,0	0,0	0,2	4926	Day-to-day care
6	0,2	0,1	0,0	0,1	0,2	0,0	0,1	0,8	0,1	0,0	0,1	0,5	0,1	0,2	0,1	0,0	0,1	0,1	0,1	0,4	826	Home consultation
7	0,3	0,1	0,1	0,3	0,2	0,2	1,4	0,4	0,3	1,4	0,2	0,3	0,2	0,3	0,3	0,1	0,2	0,2	0,1	0,4	196	Home kinesitherapy
8	0,3	0,1	0,1	0,1	0,4	0,0	0,0	0,0	0,0	0,0	0,6	0,4	0,0	0,2	0,1	0,0	0,1	0,0	0,1	0,3	970	Legal copayment
9	0,3	0,1	0,0	0,3	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,5	0,0	0,3	0,1	0,0	0,0	0,0	0,2	0,3	870	Day-to-day care with fees overrun
10	0,2	0,1	0,0	0,1	0,4	0,0	0,0	0,0	0,0	0,0	0,1	0,4	0,6	0,2	0,1	0,2	0,1	0,1	0,1	0,3	1069	Medical device
11	0,2	0,2	0,1	0,2	0,4	0,0	0,0	0,0	0,0	1,3	0,1	0,4	0,1	0,3	0,1	0,1	0,2	0,0	0,1	0,4	1265	Kinesitherapy
12	0,1	0,0	0,0	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,3	1541	Low consumer
13	0,3	0,1	0,0	0,2	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	0,3	0,1	0,0	0,5	0,0	0,2	0,3	1393	Radiography
14	0,3	0,1	0,0	0,1	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,0	0,3	0,1	0,0	0,0	0,0	0,5	0,3	802	Ultrasound
15	0,2	0,1	0,0	0,2	0,3	0,0	0,0	0,0	0,0	0,0	0,1	0,4	0,0	0,3	0,1	1,1	0,1	0,0	0,1	0,3	1419	Optic
16	0,2	0,5	0,6	0,1	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,4	0,1	0,2	0,1	0,1	0,1	0,0	0,1	0,3	1771	Dental care
17	0,2	0,1	0,0	0,1	0,1	0,0	0,0	0,0	0,8	0,0	0,2	0,3	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,3	679	Hospitalization

Figure 12 – The center of gravity of the 17 final clusters, calculated using policyholder-matrix coefficients. A high coefficient means that the associated cluster is strongly related to the associated HPC.

A Kohonen's map obtained this way also captures some correlations between clusters. For example, clusters 1, 2 and 7 are closed together on Figure 11. It is possible to see on Figure 12 that all of them are "Home-care" clusters.

It is possible to test the consistency of the results by looking at policyholder profiles in each cluster in terms of variables that were not used in the clustering process, such as age and total medical expenditures (see Figure 13). As expected, individuals in the "Occasional consumers" cluster do not cost insurers much, whereas those in the "Home-care" clusters do. It is also possible to check that the average age for comfort-care clusters, such as "Optic" and "Dental care", is higher than the one in the home-care clusters (e.g. Darblade (2015)). As such, the method appears to produce consistent results.

Cluster	Meaning	Age	Private Insurer refund	Public insurer refund	Number of policyholders
15	Optic	69	497	284	1419
9	Everyday care with fees overrun	70	145	301	870
12	Occasional consumer	70	62	52	1541
13	Radiography	70	161	407	1393
14	Ultrasound	70	164	350	802
16	Dental care	70	413	448	1771
5	Everyday-day care	71	97	202	4926
8	Legal copayment	71	220	359	970
11	Kinesitherapy	71	307	603	1265
10	Medical apparatus	73	220	399	1069
2	Nurse home care	74	550	794	295
3	Surgery	74	355	532	959
4	Home care apparatus	75	315	533	577
17	Hospitalization	77	1126	277	679
1	Home care	81	792	1389	169
6	Home consultation	81	260	452	826
7	Home kinesitherapy	82	937	1014	196

Figure 13 – This Figure represents, for each cluster, the average age, and refund associated to the policyholders belonging to the cluster. Clusters are ordered by average age.

This approach can help to choose a risk to reduce using a prevention plan. For example, cluster 17 ("Hospitalization") is expensive for private insurers. It may be of interest to analyze more closely the profile of individuals in this class in order to target prevention plans. Yet, medical advice might be needed to fully exploit the results.

This approach is consistent with the European data regulation GDPR. The GDPR distinguishes between two cases. When the policyholder gives her consent to her data being treated for prevention purposes, individuals in each class can be targeted (individuals in cluster 1 can be proposed a specific prevention plan). Insurance companies do not necessarily have the

consent of all policyholders. However, it is still possible to aggregate data in order to obtain cluster characteristics, and thus obtain a general objective characterization (for example, people in home-care clusters are in general over 80, so that we can target tertiary prevention plans at those over 80, or primary prevention plans towards those aged 70 to 80). In our databases, we only know policyholder age, sex and family situation: with more complete information we could expand the statistical analyses of each cluster.

Please notice that the Kohonen's map algorithm may be sensitive to initialization and input data order, due to the multi-label context [9] (for example, Appendix 2 shows a map obtained in the same way as Figure 11, changing only the original seed). Empirically, it is possible to see that the cluster meanings are very similar between the different maps. However, some policyholders can change cluster from one map to another. One way to tackle this issue would be to construct a number of different Kohonen's maps based on the same NMF result. It is then possible to consider policyholders who appear at least once, or twice, or everytime in a given cluster. We thus obtain for each policyholder the empirical likelihood of belonging to this cluster, performing a kind of fuzzy-clustering.

1.4.3 Other tests

Some additional tests have been carried out to better understand the results produced by this method.

Sensibility to infrequent correlations: First, the model's capacity to identify strong but infrequent correlations is tested, by adding a randomly-chosen number between 3 and 10 to the "Keratotomy", "Hydrotherapy accommodation" and "Orthoptics" consumption of n random policyholders. "Keratotomy" and "Hydrotherapy accommodation" are consumed only infrequently, whereas "Orthoptics" is much more common.The NMF algorithm is then reruned. The goal is to identify the minimum n for which the NMF algorithm detects the new correlation. It turns out that only 60 modified policyholders are necessary (out of 20 000 individuals in the database) in order to detect this correlation.

Challenging this clustering process with a naive approach: Despite all its qualities, the clustering method presented above has to be challenge with a much simpler method. Instead of using all this process in order to create an "Hospitalization" cluster (for example), what if one settles to target every policyholder having at least one hospitalization health product reimbursement, or with hospitalization costs above a certain threshold ?

9. Someone can need glasses and have an operation the same year. Such a people would then belongs to two clusters : the optic cluster and the hospitalization cluster. In this regard, the health sector is linked to a multi-label context.

This approach suffers from at least two limitations. First, it is a fully supervised approach : someone has to fix a threshold. This is not a trivial task and every chosen value could be controversial. Secondly, in contrast to the process describes above, it does not take into account correlation, which can lead to some inappropriate targeting. For example, it appears that psychiatric diseases are often correlated with hospitalization Gauchon and Hermet (2019). However, as discussed in Section 1.5, it is not optimal to target psychiatric diseases with the same prevention plan than other hospitalizations, since they are two very different risks.

In practice, this naive method leads to quite different classes (see Appendix 1 for the detailed results). Most of the time, the NMF-Kohonen method produces clusters with one main health consumption, whereas the clusters obtained with the naive method are more heterogeneous.

Comparison between PCA and NMF: The NMF reduction method is also challenged with the PCA. To do so, the dimension of the IB subdatabases containing only women over 62 has been reduced to 20 dimensions with both methods. Yet, contrary to the pre-processing method presented in Section 1.2, the optical products (such as "Contact lenses" and "Spectacle lenses") have not been merged together. This way, it is possible to see if the dimension reduction method is able to make a consistent optic HPC. As shown in Figure 14, the NMF method performs better at this task [10].

Figure 14 – Comparison between ACP and NMF Optic HPC. A darker case means that the health product is strongly detected has belonging to the Optic HPC.

In addition, a Kohonen's map has been estimated on the data reduced using the PCA. The R^2 obtained for this clustering is only of 0.09. In comparison, the R^2 obtained for the NMF method is usually around 0.25 [11].

10. In fact, almost none of the 20 HPCs made using the PCA offers a satisfying consistency.

11. The R^2 coefficient is given by $R^2 = 1 - \frac{\sum_{i=1}^{k} I_{C_i}}{I_C}$, with I_{C_i} designing the inertia of the cluster C_i. The inertia has been computed using the cosine similarity, following the recommendation of Huang Huang (2008)

The fact that the PCA does not perform well for this task is interesting, since this algorithm has proved to be efficient on a much more high dimensional context Hoyle and Rattray (2003). However, it is well-known that the classical PCA algorithm is very sensitive to the presence of outliers (Xu and Yuille (1995), Xu et al. (2010)), which could explain such a poor quality. As shown by Wu et al., the sNMF/l algorithm used in this paper performs better than the PCA in the presence of outliers and noise Wu et al. (2018).

Test of the method on a database with known clusters. In order to allow a better understanding of the properties of the clustering process, a test on a database with known clusters has been performed. It shows that most of the clusters are pure and consistent, even though the biggest clusters are usually quite heterogeneous.

1.4.4 Presentation of the results obtained on the eight subdatabases

In the previous Sections, the process was explained and illustrated only using the subdatabase of the Individual Database (IB) containing only women over 62. In this Section, results obtained on the eight subdatabases are presented [12]. Since presenting each subdatabases result as precisely as previously would be too long, a more synthetic point of view is adopted.

HPCs discussion The HPCs obtained for the eight subdatabases are summarized in Figure 15. 22 HPCs are made in each of the CB subdatabase, whereas only 20 are found for each of the IB subdatabase. The HPCs obtained for the IB and the IC are essentially the same, with the main differences being due to database construction. The method is thus resistant to a change in the database.

However, some other differences merit discussion. In the CB, the HPCs "Respiratory apparatus" and "Home care apparatus" are more important than in the IB. This can be due to care refusal: these are expensive products that are better-covered by the collective top-range market contract. Also, the "Hospitalization" HPC only concerns older people in the individual base but concerns everybody in the CB. This is due to the "Legal copayment" HPC, which also contains "Hospitalization" health consumption for younger people in the IB. As the "Legal copayment" health product does not exist in the CB, the HPC becomes "Hospitalization" there.

In both databases, the "Osteopathy", "Orthoptics" and "Psychiatry" HPCs do not concern older people. Osteopathy is a modern practice of which the older are less aware, who prefer going to a physiotherapist instead. It is notable that "Orthoptics" [13] is an HPC on its own.

12. See Section 1.2 for a reminding of the subdatabases.

13. According to Wikipedia, "Orthoptics is a profession allied to eye care professions whose primary emphasis is the diagnosis and non-surgical management of strabismus (wandering eyes), amblyopia (lazy eye) and eye movement disorders".

	Individual database			Collective database		
	Men		Women	Men		Women
<62years old		Medical specialist (without prescription) Orthoptics Osteopathy Psychiatry		Surgery	Dental care Osteopathy Psychiatry	Orthoptic
>62 years old		Blood test Dental care / Denture / Surgery Drugs Fees overrun General practitioner Kinesitherapy Legal copayment Medical apparatus Medical specialist Nurse Optic Radiography Surgery Ultrasound		Dental care Respiratory apparatus	Blood / Biology test Dental preventive care / Radiology / Denture Drugs Dressings General practitioner Kinesitherapy Hospitalization Medical specialist Nurse Optic Orthotics / medical apparatus Radiography Technical medical procedures Ultrasound	Home care apparatus
	Respiratory apparatus	Home care fees Home consultation Hospitalization	Home care apparatus		Home general practitioner Home care apparatus Personal medical room	Home nurse Orthopedic kinesitherapy

Figure 15 – Summary of the HPCs

This is a quite narrow health product (in both bases it is consumed by fewer than 2% of policyholders and represents under 0.12% of total expenditure). Orthoptics mainly covers children, which explains why it is an HPC for younger people, with parents paying for their children. Finally, to understand why "Psychiatry" does not appear as an HPC for older people, it is important to underline that dementia in France is usually treated by neurologists or geriatrician specialists rather than psychiatrists. On the other hand, psychiatric diseases such as breakdown often occur before 30 years old, explaining why this risk appears for younger people. An example of a possible Prevention program to perform on the psychiatric cluster is given on Section 1.5.

Some home-care HPCs are specific to older people, due to dependency. However, it is of interest to note that the "Nurse" HPC typically contains some home-care consumption and exists for both younger and older people.

Last, we can also see that the "Respiratory apparatus" HPC concerns only men, as they are more subject to sleep apnea than women. The "Ultrasound" clusters are found for both men and women, as this is not only used for pregnancy but also for heart, blood and musculoskeletal-system radiography.

Clusters discussion Once the HPCs have been discussed, it is possible to process the Kohonen's map algorithm. Figure 16 summarizes all clusters interpretations obtained for each subdatabase.

First, "Everyday care" and "Occasional consumer" are both clusters with no specific consumption and are joint in the CB dataset. Individuals in these clusters are mostly in good health

	Individual database			Collective database		
	Men		Women	Men		Women
<62years old	Blood test	Medical specialist (without prescription) Orthoptics Osteopathy Psychiatry		Blood test Surgery	Dental preventive care / Radiography Hospitalization (with and without personal room) Orthotics / Medical apparatus Osteopathy Psychiatry	Heavy home apparatus Hospitalization Orthoptic
		Dental care Everyday care Home care Kinesitherapy Medical apparatus Legal copayment Optic Radiography Surgery Ultrasound		Dental preventive care / radiography Orthotics / medical apparatus Respiratory apparatus	Biology test Dental care Denture Dressings Everyday care Home care Kinesitherapy Optic Ultrasound	Home care apparatus
>62 years old	Respiratory apparatus	Home care / Kinesitherapy Home consultation Hospitalization Occasional consumer Nurse home care	Everyday care with fees overrun Home care apparatus		Home care device Home general practitioner Hospitalization with personal room Hospitalization without personal room	Orthopedic kinesitherapy Home nurse Hospitalization / Home care

Figure 16 – Interpretation of the final clusters

or refuse treatment.

The HPC cluster-meaning analyses are similar to the one made before (see the comments on "Psychiatry", "Osteopathy", "Orthoptics" and "Respiratory apparatus"). Note that the "Home care" cluster appears in all eight subdatabases. "Orthotics / Medical apparatus" does not form a cluster for women and "Home care apparatus" does not form a cluster for men in the CB. For the younger men in the CB, a "Surgery" cluster in addition to the "Dressings" cluster is found, producing two surgery-like clusters. Surprisingly, in both the IB and CB, younger men also have a "Blood test" cluster, which is not the case in the other databases. Last, "Dentures" appears as a cluster only in the CB, mainly due to differences in contract quality.

The analysis of average age and expenditure by cluster also provides considerable information. For example, the average age in the "Dental care" and "Optic" clusters is lower than the average age for older people in both databases. For the younger, the average age in the "Ultrasound" clusters is lower and contains more policyholders for women than for men, due to motherhood. The "Psychiatry" clusters cover more women than men, which is also a well-known medical fact (see W.H.O. et al. (2017)).

Other findings are more difficult to explain. For young women the "Kinesitherapy" clusters have higher average age than for men, and both are higher than the overall average age. For the IB, the cluster "Medical specialist without prescription" has a lower average age than

"Medical specialist with prescription". Last, the average age in the global "Hospitalization" clusters for the older are above the overall average age, but below the average age in the "Home care" clusters.

From a more global point of view, the "Everyday care" and "Occasional consumer" clusters are very large. On the contrary, there usually exist some very narrow clusters with fewer than 100 policyholders. These usually contain very archetypal consumers and thus can be used to target small prevention plans.

1.5 Prevention of hospitalization risk.

It is now useful to figure out how one could use these results for a prevention plan. This Section focus on the hospitalization risk. Working and retired populations are separated since they represent very different risks.

According to medical experts, hospitalization of policyholders under 62 can be caused by very different causes. However, figure 16 shows that a "Psychiatry" cluster is created for all 4 subdatabases gathering this population. However, psychiatry (and psychiatric hospitalization) has never been identified as a major risk by French private health insurers. Moreover, the average policyholder in a "Psychiatry" cluster needs much more reimbursements than the average policyholders (e.g. for the women under 62 of the CB, they spend 1595 euros on average on health expenditure, as opposed to 724 euros for the average policyholder). Finally, "Psychiatry" clusters usually concern a restricted number of policyholders (around 300 for the same subdatabase as above), limiting the cost of a prevention plan on such a population. Figure 16 also shows that psychiatric clusters are different from hospitalization clusters, underlining the fact that psychiatric and hospitalization are two different risks, even though they seem strongly correlated when one looks at the figures Gauchon and Hermet (2019). Thus, the clustering process reveals an unsuspected risk, which could not be identified by simply applying the naive method presented in Section 1.4.3 on hospitalization.

However, it is remarkable that once the psychiatric risk has been identified, it is possible to use the simpler method presented in Section 1.4.3 in order to target every policyholders with at least one psychiatric consultation. People selected using this simpler method present a similar consumption profile as the one in the "Psychiatry" cluster, even though the number of policyholders targeted is bigger. In the case of psychiatric diseases, the method is thus mostly useful to detect the risk or if one has a tightened budget.

Identifying psychiatric risks as important for private insurers has never been done in France Gauchon and Hermet (2019). This is due to the fact that in France, people suffering from a chronic disease benefit from a special status allowing for a better refunding from the French public insurer. The fact that chronic diseases are mainly impacting the social security is

thus a popular belief, and private insurers tend to neglect this kind of risk. However, due to co-morbidity with other diseases and fees overrun, it appears that this belief is false, at least for psychiatric diseases. The data driven method presented in this paper has thus revealed something practitioners did not know.

After discussing these facts with psychiatrists, it appears that people in a "Psychiatry" cluster mainly suffer from severe depression or anxiety disorders, and that they would heavily benefit from a prevention plan. There may be a few schizophrenic people among them. However, since schizophrenic people are usually unemployed, they usually do not benefit from a private health insurance.

It also appears that a population going to a psychiatric is often benefiting from a sick leave, and specially the one in the psychiatric cluster, since it concerns people mainly described by a strongly psychiatric oriented profile. Thus, it is acceptable to suppose that a lot of people on this cluster are actually on labour disruption. From all of these, an appropriate prevention plan emerges: initiating a return to work. It consists on setting meetings with the employer and with an advisor before going back to work, and on assisting them into setting up a personalized return-to-work schedule, in order to limit anxiety and the risk of relapse. Moreover, it would also be helpful to train coworkers in appropriately welcoming back there colleague and reacting in case of relapse. Reducing stigmatization is a key point in the fight against psychiatric diseases.

This kind of example illustrates three points. Firstly, since the whole clustering process is unsupervised, the method allows for a new point of view on data and can reveal some unsuspected class of risks. Secondly, some clusters are sufficiently small to allow a targeted prevention plan. Finally, this clustering process might be a useful tool, but is not enough to target a prevention plan. In order to design a good prevention program it is necessary to join the medical and the actuarial community together.

Concerning retired people, it is possible to see that a cluster "Radiography" is created. According to experts, these clusters gather policyholders who might have suffered a fall. Indeed, when a senior fall, the medical procedure is to systematically look for a fracture in hospital. This screening phase is very similar to the medical acts consumed by policyholders in this cluster. Thus, some hospitalization can be successfully prevented by fighting against falls, and this cluster seems to be a good target for this kind of prevention program.

Prevention, and in particular sport has shown to be efficient against falls Dargent-Molina and Cassou (2008). Examples of statistically efficient programs are Tai Chi Chuan or other physical activity sessions supervised by a physiotherapist.

Some organizations, such as "Siel bleu", are specialized in this field and proposed adapted programs for all kinds of population. A possible prevention program would then be to propose

to policyholders gathered in the "Fracture" cluster to subscribe to a plan managed by this kind of organization. It would mostly consist of practicing adapted physical activity, in order to improve the confidence, the balance and the faculty to stand up of participants after a fall.

1.6 Conclusion

We have presented a method for clustering policyholders based on their health consumption. This first reduces the dimension problem by carrying out a Nonnegative Matrix Factorization. This stage improves the results and helps interpret the clustering, by identifying meaningful health-product clusters.

In the second stage, policyholders are clustered using Kohonen's maps. This algorithm offers a readable visualization of the results. It allows the simple comparison of clusterings carried out on different databases. Moreover, these can be interpreted as different risk clusters. The method has been subjected to a number of tests, revealing its reliability and the quality of the results. Except for the "Everyday care" clusters (composed of occasional consumers or very particular policyholders), most clusters are pure and can be used in practice. These clusters are established using common insurance data, as possessed by every health insurer. By constituting clusters, we aggregate the data and so respect the legislation. The method applied here can thus be used to target prevention plans aimed at policyholders, and our tests have shown that this process is accurate when we do not have a clear idea of the prevention plans to be instigated. To illustrate this point, an example of a potential prevention program on psychiatric disease, resulting of a talk with a psychiatrist, is proposed. Indeed, thanks to the presented clustering process, psychiatry has been identified has an important risk for private health insurer, which was unknown from practitioner.

There are a number of ways in which the method can be improved. First, this is a mono-label clustering, and multi-label clustering would usually be more appropriate for health-risk profiles. This multi-label clustering can be carried out via fuzzy clustering (such as fuzzy c-means) instead of Kohonen's map.

Moreover, the NMF method has considerable advantages, but one main disadvantage: due to the initialization requirements, it can be quite slow. In order to accelerate dimension reduction, other methods may be more appropriate (e.g. word-embedding methods).

The method applied here does not take into account the temporality of health consumption. Knowing that a policyholder consumes a great deal over a short period or regularly throughout the year could be useful.

The results from this method are very dense, and are sometimes difficult to interpret. Medical advice for other clusters than psychiatric ones would be useful in understanding the potential scope of this method.

Lastly, this method is not specific to health-policyholder clustering, and could for example also be applied to customer clustering.

1.7 Appendix

Appendix 1: Comparing the clusters obtained from the proposed method to those from a very basic approach

Health consumption type	Class	1 - Home care	2 - Nurse home care	3 - Surgery	4 - Home care apparatus	5 - Everyday care	6 - Home consultation	7 - Home kinesitherapy	8 - Legal copayment	9 - Everyday care - fees overrun	10 - Medical apparatus	11 - Kinesitherapy	12 - Occasional consumer	13 - Radiography	14 - Ultrasound	15 - Optic	16 - Dental care	17 - Hospitalization
	Drugs	409	375	281	300	187	288	319	253	236	252	262	65	253	213	194	199	163
	Blood test	74	105	43	47	27	46	52	49	44	42	44	17	52	50	36	38	36
	Medical apparatus	155	120	96	184	47	66	207	55	75	154	95	34	64	57	52	84	105
	Other	76	31	24	22	7	15	38	56	24	17	26	8	14	14	17	14	23
	Nurse / medical auxiliaries	423	60	4	13	2	7	25	2	3	2	3	1	2	2	2	2	1
	Surgery	200	235	223	50	11	27	265	22	28	26	34	13	38	28	32	25	167
	Dental care	21	88	62	89	67	42	80	86	25	57	107	36	67	81	62	369	41
	Home care	319	37	6	11	1	93	112	3	2	3	4	1	2	2	2	2	5
	General practitioner	69	124	86	81	46	45	77	84	91	82	101	28	97	83	71	77	33
	Hospitalization	347	220	122	102	29	126	495	70	50	51	61	31	68	58	55	44	983
	Kinesitherapy	77	48	12	22	3	11	391	8	5	6	257	3	7	5	5	7	8
	Optic	113	146	162	181	20	42	109	46	21	170	142	32	32	35	985	138	38
	Denture	19	108	74	129	88	60	127	108	29	70	141	28	86	119	70	492	46
	Radiography	44	70	44	34	5	24	59	34	9	43	57	6	102	83	33	53	13
	Medical specialist	39	58	56	47	11	20	61	30	70	30	46	13	45	38	39	33	18
	Total	2386	1824	1296	1311	553	913	2417	905	712	1006	1380	315	931	868	1655	1580	1679
	Size	169	295	959	577	4926	826	196	970	870	1069	1265	1541	1393	802	1419	1771	679

Figure 17 – Detailed statistics for all classes.

Appendix 2: An example of another map obtained from the same data

Health consumption type	At least one consumption of : Drugs	Blood test	Medical apparatus	Surgery	Dental care	Home care	General practitioner	Hospitalization	Kinesitherapy	Optic	Radiography	Medical specialist	Global mean
Drugs	227	274	302	311	241	332	239	270	290	241	266	265	216
Blood test	40	78	53	60	46	65	45	53	52	45	54	53	39
Medical apparatus	77	92	266	124	88	123	79	118	112	92	98	94	74
Other	18	23	27	29	22	28	20	24	31	23	23	25	18
Nurse / medical auxiliaries	8	10	16	24	6	38	7	11	13	6	8	9	7
Surgery	47	66	81	272	51	111	54	97	85	67	71	85	45
Dental care	92	104	103	106	284	87	103	101	116	113	168	107	92
Home care	11	14	20	23	8	65	9	14	23	8	11	11	11
General practitioner	70	91	90	101	86	84	89	91	105	84	98	97	68
Hospitalization	94	115	167	204	86	196	85	228	144	98	110	110	96
Kinesitherapy	28	34	45	43	33	57	31	37	245	33	40	39	27
Optic	137	153	165	183	169	123	153	192	159	843	169	204	136
Denture	120	139	135	134	367	118	134	131	158	140	219	139	118
Radiography	35	48	50	55	49	45	41	52	63	43	84	53	34
Medical specialist	32	42	45	63	40	40	38	48	52	48	48	72	31
Total	1036	1281	1564	1730	1575	1513	1127	1466	1647	1884	1467	1362	1013
Size	18770	9761	5516	3277	6363	3245	15099	8339	2213	3193	8072	8441	19727

Figure 18 – Consumption if consuming at least some of a particular health product, and overall consumption.

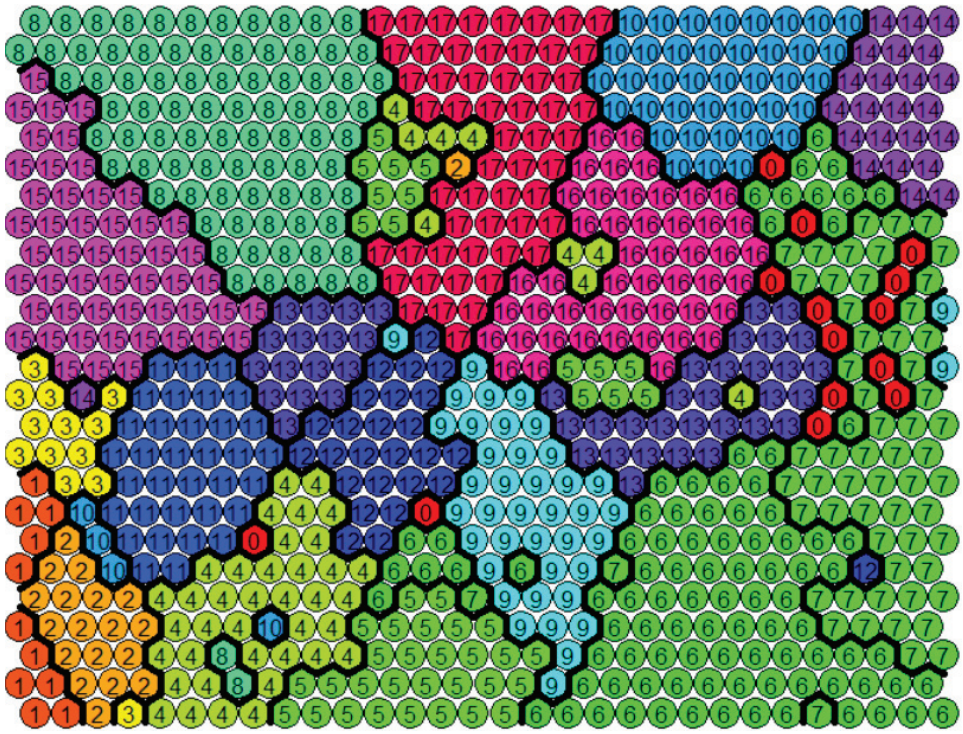

Figure 19 – Self-organizing map obtained from the same data as in Figure 11, using a different seed.

Bibliographie

Aggarwal, C. C. and Yu, P. S. (2000). Finding generalized projected clusters in high dimensional spaces, volume 29. ACM.

Agrawal, R., Gehrke, J., Gunopulos, D., and Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications, volume 27. ACM.

Akinduko, A. A., Mirkes, E. M., and Gorban, A. N. (2016). Som : Stochastic initialization versus principal components. Information Sciences, 364 :213–221.

Badea, L. (2008). Extracting gene expression profiles common to colon and pancreatic adenocarcinoma using simultaneous nonnegative matrix factorization. In Biocomputing 2008, pages 267–278. World Scientific.

Beaulieu, N., Cutler, D. M., Ho, K., Isham, G., Lindquist, T., Nelson, A., and O'Connor, P. (2006). The business case for diabetes disease management for managed care organizations. In Forum for Health Economics & Policy, volume 9. De Gruyter.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful ? In International conference on database theory, pages 217–235. Springer.

Boutsidis, C. and Gallopoulos, E. (2008). Svd based initialization : A head start for nonnegative matrix factorization. Pattern Recognition, 41(4) :1350–1362.

Brockett, P. L., Xia, X., and Derrig, R. A. (1998). Using kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. Journal of Risk and Insurance, pages 245–274.

Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the national academy of sciences, 101(12) :4164–4169.

Bühlmann, H. and Gisler, A. (2006). A course in credibility theory and its applications. Springer Science & Business Media.

Cheng, C.-H., Fu, A. W., and Zhang, Y. (1999). Entropy-based subspace clustering for mining numerical data. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 84–93. ACM.

Darblade, M. (2015). Analyse de profils de consommation et tarification des futures garanties sur-complémentaire santé. Master's thesis, ISFA.

Dargent-Molina, P. and Cassou, B. (2008). Prévention des chutes et des fractures chez les femmes âgées. Gérontologie et société, 31(2) :65–78.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6) :391–407.

Derrig, R. A. and Ostaszewski, K. M. (1995). Fuzzy techniques of pattern recognition in risk and claim classification. Journal of Risk and Insurance, 62(3) :447–482.

Gauchon, R. and Hermet, J.-P. (2019). La psychiatrie : un risque important en assurance santé ?

Gaujoux, R. and Seoighe, C. (2010). A flexible r package for nonnegative matrix factorization. BMC bioinformatics, 11(1) :367.

Ghoreyshi, S. and Hosseinkhani, J. (2015). Developing a clustering model based on k-means algorithm in order to creating different policies for policyholders in insurance industry. International Journal of Advanced Computer Science and Information Technology (IJACSIT), 4(2) :46–53.

Hainaut, D. (2019). A self-organizing predictive map for non-life insurance. European Actuarial Journal, 9(1) :173–207.

Henckaerts, R., Antonio, K., Clijsters, M., and Verbelen, R. (2018). A data driven binning strategy for the construction of insurance tariff classes. Scandinavian Actuarial Journal, 2018(8) :681–705.

Herring, B. (2010). Suboptimal provision of preventive healthcare due to expected enrollee turnover among private insurers. Health Economics, 19(4) :438–448.

Hinneburg, A. and Keim, D. A. (1999). Optimal grid-clustering : Towards breaking the curse of dimensionality in high-dimensional clustering. pages 506–517. 25 th International Conference on Very Large Databases.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. Journal of machine learning research, 5(Nov) :1457–1469.

Hoyle, D. and Rattray, M. (2003). Pca learning for sparse high-dimensional data. EPL (Europhysics Letters), 62(1) :117.

Huang, A. (2008). Similarity measures for text document clustering. In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand, pages 49–56.

Indyk, P. and Motwani, R. (1998). Approximate nearest neighbors : towards removing the curse of dimensionality. In Proceedings of the thirtieth annual ACM symposium on Theory of computing, pages 604–613. ACM.

Jones, B. W. and Chung, W. (2016). Topic modeling of small sequential documents : Proposed experiments for detecting terror attacks. In Intelligence and Security Informatics (ISI), 2016 IEEE Conference on, pages 310–312. IEEE.

Kim, H. and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics, 23(12) :1495–1502.

Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78(9) :1464–1480.

Kuang, D., Choo, J., and Park, H. (2015). Nonnegative matrix factorization for interactive topic modeling and document clustering. In Partitional Clustering Algorithms, pages 215–243. Springer.

Kuo, R., Lin, S., and Shih, C. (2007). Mining association rules through integration of clustering analysis and ant colony system for health insurance database in taiwan. Expert Systems with Applications, 33(3) :794–808.

Langville, A. N., Meyer, C. D., Albright, R., Cox, J., and Duling, D. (2006). Initializations for the nonnegative matrix factorization. In Proceedings of the twelfth ACM SIGKDD international conference on knowledge discovery and data mining, pages 23–26. Citeseer.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature, 499(7457) :214.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755) :788.

Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In Advances in neural information processing systems, pages 556–562.

Mote, S. R., Baid, U. R., and Talbar, S. N. (2017). Non-negative matrix factorization and self-organizing map for brain tumor segmentation. In Wireless Communications, Signal

Processing and Networking (WiSPNET), 2017 International Conference on, pages 1133–1137. IEEE.

Murtagh, F. (1995). Interpreting the kohonen self-organizing feature map using contiguity-constrained clustering. Pattern Recognition Letters, 16(4) :399–408.

Nesvijevskaia, A. and Taudou, B. (2016). La data science au service de la prévention santé et prévoyance : nouveaux paradigmes - 17eme rencontre mutré, 14-15 november - nantes. Technical report, Malakoff Mederic.

Paatero, P. and Tapper, U. (1994). Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5(2) :111–126.

Pascual-Montano, A., Carazo, J. M., Kochi, K., Lehmann, D., and Pascual-Marqui, R. D. (2006). Nonsmooth nonnegative matrix factorization (nsnmf). IEEE transactions on pattern analysis and machine intelligence, 28(3) :403–415.

Pauca, V. P., Piper, J., and Plemmons, R. J. (2006). Nonnegative matrix factorization for spectral data analysis. Linear algebra and its applications, 416(1) :29–47.

Pauca, V. P., Shahnaz, F., Berry, M. W., and Plemmons, R. J. (2004). Text mining using non-negative matrix factorizations. In Proceedings of the 2004 SIAM International Conference on Data Mining, pages 452–456. SIAM.

Peng, Y., Kou, G., Sabatka, A., Chen, Z., Khazanchi, D., and Shi, Y. (2006). Application of clustering methods to health insurance fraud detection. In Service Systems and Service Management, 2006 International Conference on, volume 1, pages 116–120. IEEE.

Robertson, S. (2004). Understanding inverse document frequency : on theoretical arguments for idf. Journal of documentation, 60(5) :503–520.

Rousseeuw, P. J. (1987). Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20 :53–65.

Utsumi, A. (2010). Evaluating the performance of nonnegative matrix factorization for constructing semantic spaces : Comparison to latent semantic analysis. In 2010 IEEE International Conference on Systems, Man and Cybernetics, pages 2893–2900. IEEE.

Van Benthem, M. H. and Keenan, M. R. (2004). Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. Journal of Chemometrics : A Journal of the Chemometrics Society, 18(10) :441–450.

Verrall, R. J. and Yakoubov, Y. H. (1999). A fuzzy approach to grouping by policyholder age in general insurance. Journal of Actuarial Practice, 7 :181–204.

Wehrens, R., Buydens, L. M., et al. (2007). Self-and super-organizing maps in r : the kohonen package. Journal of Statistical Software, 21(5) :1–19.

W.H.O. et al. (2017). Depression and other common mental disorders : global health estimates.

Wu, B., Wang, E., Zhu, Z., Chen, W., and Xiao, P. (2018). Manifold nmf with l21 norm for clustering. Neurocomputing, 273 :78–88.

Xu, H., Caramanis, C., and Sanghavi, S. (2010). Robust pca via outlier pursuit. In Advances in Neural Information Processing Systems, pages 2496–2504.

Xu, L. and Yuille, A. L. (1995). Robust principal component analysis by self-organizing rules based on statistical physics approach. IEEE Transactions on Neural Networks, 6(1) :131–143.

Yeo, A. C., Smith, K. A., Willis, R. J., and Brooks, M. (2001). Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry. Intelligent Systems in Accounting, Finance and Management, 10(1) :39–50.

Zhang, T., Ramakrishnan, R., and Livny, M. (1996). Birch : an efficient data clustering method for very large databases. In ACM Sigmod Record, volume 25, pages 103–114. ACM.

1.8 Complément : Utilisation d'un jeu de données labélisé de text mining pour tester la qualité de classification

Les jeux de données utilisés jusqu'à présent dans ce chapitre étaient entièrement non supervisés, c'est-à-dire qu'ils n'incluaient pas la classe théorique dans lequel chaque assuré devrait être rangé. Si cette caractéristique n'a pas empêché de tester en partie la qualité de la classification proposée par l'algorithme (via des métriques telles que l'inertie intra-groupe), elle empêche de calculer objectivement un score d'erreur. Afin de pallier cette limite, il est souhaitable de tester cette méthode sur un jeu de données supervisé (ou les classes à retrouver sont précisées dans la base de données).

Cependant, des principes légaux, comme le secret médical ou le RGPD, empêche les assureurs santé privés de constituer une telle base de données. Afin d'effectuer cette étude, il a donc été nécessaire de se tourner vers un autre type de jeu de données. Par chance, il se trouve que les jeux de données provenant du text mining présentent des similarités avec les bases de données utilisées jusqu'à présent [14].

Il existe de nombreux jeux de données supervisés en text mining. Généralement, ils contiennent deux variables : le texte en lui-même, et une information sur le texte (exemple : type de document, thème du texte, provenance du texte...). C'est le cas du jeu de données **"20NewsGroup"**, utilisé dans le cadre de cette étude.

Ce jeu de données contient 18 821 messages postés sur internet, provenant de 20 forums différents (voir le détail Figure 20). Il est généralement séparé en une base d'apprentissage de 11 293 documents, et une base de test de 7 528 documents. Afin d'accélérer les traitements, seule la base d'apprentissage a été utilisée ici.

Les textes sont rarement utilisés tels quels : les mots les moins informatifs sont généralement supprimés, les fautes d'orthographes corrigées, etc... Afin d'éviter cette phase fastidieuse qui n'est pas le sujet de l'étude présentée ici, le jeu de donnée "no short", prétraité par Ana Cardoso Cachopo et mis à disposition sur son site web, a été utilisé (Cardoso-Cachopo, 2007). Il contient les données nettoyées, dont les mots d'une ou deux lettres ont été retirés.

Ce jeu de données a été utilisé dans des études provenant de toutes les branches du text-mining, que ce soit le word embedding (e.g. (Hinton and Salakhutdinov, 2009), (Wang et al., 2016)), la classification non supervisée (e.g. (Ding and He, 2004), (Huang, 2008)) ou la classification supervisée (e.g. (Settles et al., 2008), (Rennie et al., 2003)). Il est donc possible de comparer les résultats obtenus avec ceux présents dans la littérature. Dans le cadre de cette étude, les résultats ont été comparés à ceux obtenus par Huang (Huang, 2008). Comme l'objectif n'est pas de battre l'état de l'art en termes de clustering, mais simplement de vérifier si les résultats obtenus sont cohérents, un seul jeu de paramètre à été testé.

14. Voir Chapitre 2

Sujet du forum	Nombre de documents
Athéisme	480
Graphiques digitaux	584
Windows (général)	572
Ordinateurs IBM	590
Ordinateurs mac	578
Fenêtre Windows	593
Vente générale	585
Voitures	594
Motos	598
Baseball	597
Hockey	600
Cryptographie	595
Électronique	591
Médecine	594
Espace	593
Christianisme	598
Politique/arme à feu	545
Politique/Moyen-Orient	564
Politique/divers	465
Religion/divers	86

FIGURE 20 – Titres des forums et nombre de messages associés

La base de données a été d'abord transformée en matrice de fréquence (avec une colonne par mots et une ligne par texte). Conformément à la pratique habituelle en text mining, cette matrice à ensuite été transformée via la méthode tf-idf. L'algorithme NMF a ensuite été utilisé afin de réduire la dimension. La taille de la matrice de fréquence étant nettement supérieure à celle des matrices de fréquences obtenues à partir de base d'assurés santé (environ 6000 mots uniques VS environ 150 prestations), la dimension a été réduite à 60 (au lieu de 20 habituellement).

Une fois la dimension réduite, l'algorithme des cartes de Kohonen a été appliqué. Le regroupement des neurones via l'utilisation d'une classification ascendante hiérarchique permet d'obtenir le dendrogramme suivant :

À la vue de ce dendrogramme, il n'est pas clair qu'il existe 20 différentes classes sous-jacentes. En revanche, choisir un nombre de 19 classes aurait été plus cohérent. Cependant, afin d'essayer de reconstruire toutes les classes théoriques, il a été choisi de conserver 20 classes. Finalement, la carte de Kohonen obtenue est celle présentée Figure 22.

Les interprétations des clusters ont été faites à partir de la matrice de confusion présentée Figure 23.

Cette matrice de confusion est plutôt satisfaisante, puisqu'elle présente une pureté de 60%, et une entropie [15] de 40%. À titre de comparaison, Huang obtenait une pureté moyenne de 50%

15. Les définitions de Huang (Huang, 2008) ont été utilisées pour faire les calculs, afin de permettre la comparaison entre les résultats

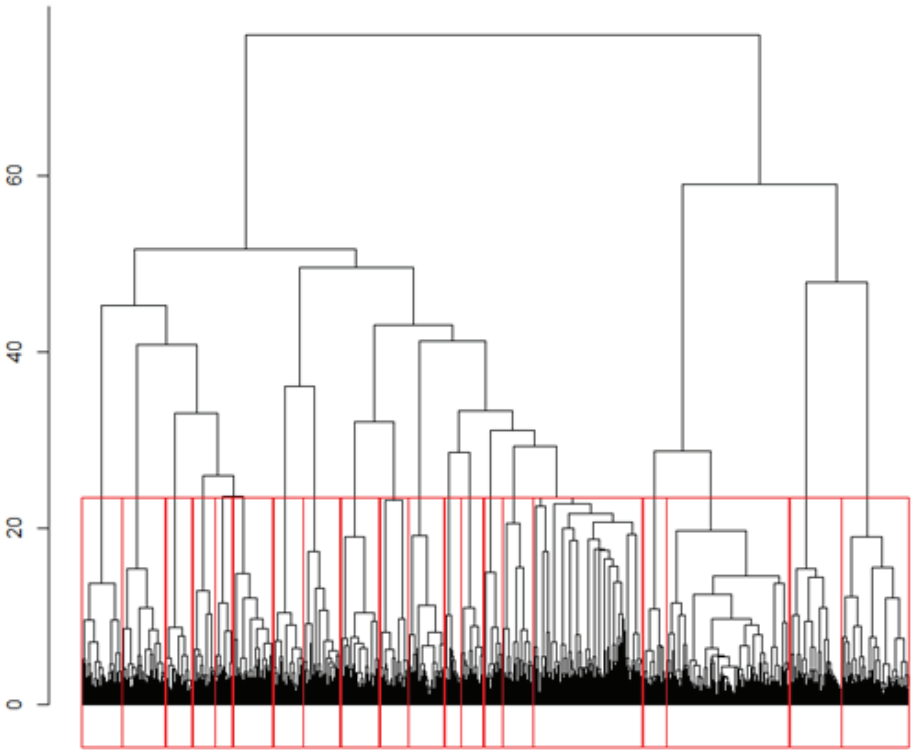

FIGURE 21 – Dendrogramme de regroupement des neurones - 20 Newsgroups. Le trait rouge coupe le dendrogramme afin de créer 20 classes différentes

.

et une entropie de 49%, ce qui est significativement moins bon (Huang, 2008). Du point de vue des mesures retenues, la méthode de classification non supervisée semble donc efficace.

L'utilisation d'un jeu de données supervisé permet aussi d'analyser les forces et les faiblesses du process. L'étude de la matrice de confusion permet de constater que l'algorithme à du mal à distinguer les forums aux thèmes très proches. Par exemple, la classe 6 mélange des textes provenant des forums "Fenêtre windows", "Windows", et "Graphique digitaux" (qui sont tous liés à des logiciels), quand la classe 9 mélange "Ordinateur Mac" et "Ordinateur IBM". On remarque aussi que le forum "Religion/divers", qui contient moins de documents que les autres forums, est mal distingué. Il est cependant impossible de dire si cela vient d'une confusion avec les forums "Athéisme" et "Christiannisme", aux thèmes assez proches, ou de la faible quantité de documents provenant de ce forum. En revanche, les forums plus spécifiques, comme "Cryptographie" ou "Motos", sont très bien reconstitués. À noter aussi que les forums "Hockey" et "Baseball", pourtant contenant tous deux un champ lexical commun, sont eux aussi bien séparés par l'algorithme. Enfin, on peut observer que la méthode à tendance à constituer quelques très grosses classes hétérogènes (classe 2 et 6), mais que, à l'exception des classes 10 et 18, les classes plus petites sont généralement assez homogènes.

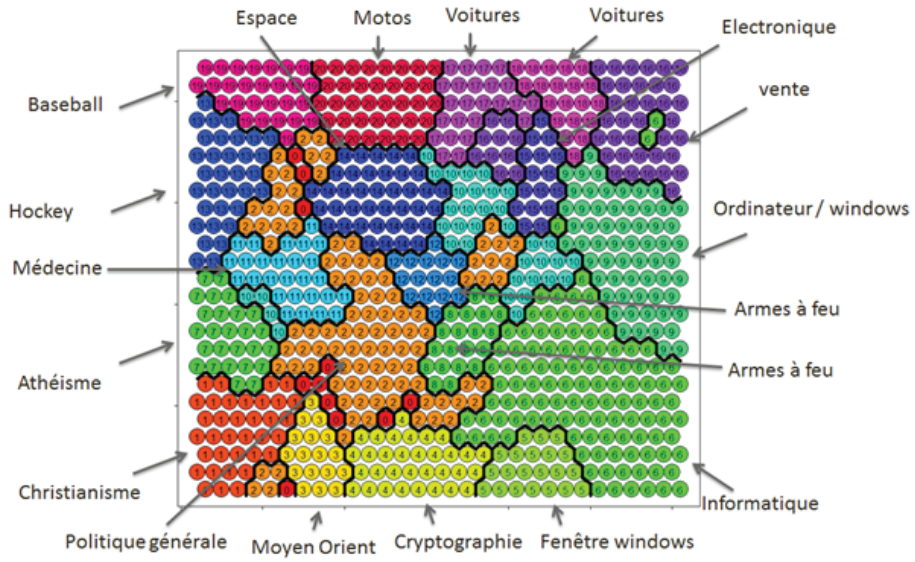

FIGURE 22 – Carte de Kohonen obtenue pour le jeu de données 20newsgroup. La classe 10 n'a pas pu être interprétée.

	Christianisme	Politique / divers	Politique / moyen orient	Cryptographie	Fenêtre Windows	Windows (général)	Athéisme	Politique / arme à feu	Ordinateurs IBM	Graphique digitaux	Médecine	Religion / divers	Hockey	Espace	Electronique	Vente générale	Voitures	Ordinateurs mac	Baseball	Motos
1	455	5	7	0	2	0	95	2	1	1	1	38	0	4	2	0	2	0	0	0
2	57	337	251	42	11	4	50	121	4	17	150	5	39	31	25	17	17	12	47	18
3	3	1	279	0	0	0	2	1	0	0	0	2	0	1	0	0	1	1	0	0
4	0	2	0	504	1	2	0	3	2	7	3	0	1	1	9	1	0	0	0	0
5	0	0	0	1	326	18	0	0	1	13	2	0	0	0	0	0	0	1	0	0
6	16	5	6	16	221	396	3	8	134	385	41	4	9	28	121	57	31	125	36	7
7	40	12	4	3	3	2	295	3	2	9	33	26	0	8	6	0	2	2	3	1
8	0	37	2	7	0	0	2	205	0	0	0	0	0	0	2	0	13	1	0	2
9	0	0	0	2	12	118	0	0	370	64	0	0	0	1	72	10	2	288	0	1
10	9	27	6	6	7	11	10	7	9	49	35	7	8	48	73	9	21	11	13	6
11	2	5	0	1	0	2	3	0	2	2	293	1	2	3	2	1	2	0	1	0
12	3	8	1	2	0	0	2	177	1	0	2	0	0	2	0	0	2	0	1	0
13	0	0	0	0	1	0	0	0	0	0	0	0	527	2	4	0	0	0	34	0
14	0	6	1	1	3	0	1	0	1	8	4	0	0	454	11	0	0	2	0	1
15	0	0	0	0	2	2	0	0	9	9	6	0	0	1	144	1	11	7	0	1
16	3	3	0	1	1	5	0	3	44	14	5	0	4	5	70	477	65	105	5	23
17	0	1	2	0	0	5	0	0	2	0	0	0	0	1	21	4	308	1	0	7
18	10	11	4	8	2	7	16	14	8	6	17	2	2	2	26	5	90	20	10	12
19	0	3	1	0	1	0	0	0	0	0	1	0	7	1	2	0	0	0	447	1
20	0	2	0	1	0	0	0	1	0	0	1	1	1	0	1	2	27	2	0	517

FIGURE 23 – Matrice de confusion obtenue pour le jeu de données 20newsgroups.

En conclusion, le procédé NMF/Kohonen semble être un procédé de Clustering pertinent, même s'il présente un certain nombre de limites : les plus grosses classes sont souvent hétérogènes et les classes similaires sont souvent mal distinguées. Du point de vue de l'application

au clustering d'assurés santé, cela peut impliquer que les classes de sous consommateurs / consommateurs de soin courants sont très hétérogènes, et qu'il sera impossible grâce à cette méthodologie de retrouver la pathologie des individus. En revanche, la plupart des classes semblent fiables et peuvent donc être exploitées en pratique, par exemple, dans le cadre d'une action de prévention.

Bibliographie

Cardoso-Cachopo, A. (2007). Improving methods for single-label text categorization. PdD Thesis, Instituto Superior Tecnico, Universidade Tecnica de Lisboa.

Ding, C. and He, X. (2004). K-means clustering via principal component analysis. In Proceedings of the twenty-first international conference on Machine learning, page 29. ACM.

Hinton, G. E. and Salakhutdinov, R. R. (2009). Replicated softmax : an undirected topic model. In Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., editors, Advances in Neural Information Processing Systems 22, pages 1607–1614. Curran Associates, Inc.

Huang, A. (2008). Similarity measures for text document clustering. In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand, pages 49–56.

Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In Proceedings of the 20th international conference on machine learning (ICML-03), pages 616–623.

Settles, B., Craven, M., and Ray, S. (2008). Multiple-instance active learning. In Advances in neural information processing systems, pages 1289–1296.

Wang, D., Cui, P., and Zhu, W. (2016). Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1225–1234. ACM.

Chapitre 2

Individuals tell a fascinating story : using unsupervised text mining methods to cluster policyholders based on their medical history

Ce chapitre reprend un article coécrit avec Jean-Pascal Hermet et soumis au journal "Future Generation Computer Systems.

Abstract

Background and objective : Classifying people according to their health profile is crucial in order to propose appropriate treatment. However, the medical diagnosis is sometimes not available. This is for example the case in health insurance, making the proposal of custom prevention plans difficult. When this is the case, an unsupervised clustering method is needed. This article aims to compare three different methods by adapting some text mining methods to the field of health insurance. Also, a new clustering stability measure is proposed in order to compare the stability of the tested processes.

Methods : Nonnegative Matrix Factorization, the word2vec method, and marginalized Stacked Denoising Autoencoders are used and compared in order to create a high-quality input for a clustering method. A self-organizing map is then used to obtain the final clustering. A real health insurance database is used in order to test the methods.

Results : the marginalized Stacked Denoising Autoencoder outperforms the other methods both in stability and result quality with our data.

Conclusions : The use of text mining methods offers several possibilities to understand the context of any medical act. On a medical database, the process could reveal unexpected correlation between treatment, and thus, pathology. Moreover, this kind of method could exploit the refund dates contained in the data, but the tested method using temporality,

word2vec, still needs to be improved since the results, even if satisfying, are not as better as the one offered by other methods.

Keywords : Unsupervised learning; Clustering; Health insurance; Word embedding; Nonnegative Matrix Factorization; Self Organizing Map.

2.1 Introduction

Healthcare circuits are a keystone of the medical system in most countries. They represent the succession of treatments every patient goes through. Thus, they tell the story of the patient's medical history : previous and current diseases, the way they have been treated, the effects on the patient's life... Obviously, the current patient's health status depends on past cares and on their efficiency. Studying healthcare circuits could also improve the understanding of the reasons for a treatment failure, for example.

It could also indicate how a patient should be treated in the future. For instance, the success of a tertiary prevention action could depend on the different steps the patient has gone through. Thus, the healthcare circuit seems to be a useful input for an algorithm seeking to target a custom prevention plan based on specific patient needs.

Targeting such a prevention plan is a major challenge for health insurers. By improving policyholders' health, they improve their reputation and reduce their health costs efficiently. However, they are not aware of the health status of their policyholders. They are only aware of a crude vision of the healthcare circuit, due to every health expense's being refunded. Thus, health insurers offer a prime example of how healthcare circuit data could be used for medical purposes, since they barely possess other data due to regulations (such as the GDPR).

A way to use this data is to create clusters of people. Classifying is a common practice in the medical field since formulating a diagnosis is already a classification task (e.g. (Ibrahim et al., 2015), (Azar et al., 2013), (Silva et al., 2016)). It consists in finding homogeneous groups underlying a given population, in order to later compare two populations or target as a specific cluster for appropriate care.

A clustering method based on the healthcare circuit in the insurance industry has already been proposed in (Gauchon et al., 2019). It is based on a two-steps method : the dimension is first reduced using Non-negative Matrix Factorization (NMF), then individuals are clustered using a self-organizing Kohonen's map. However, this method suffers some limitation: for instance, it cannot create a pregnancy cluster.

The goal of this article is to challenge the NMF method with two other unsupervised methods inspired by text mining algorithms, based respectively on Word2Vec word embedding (W2V) and autoencoders (mSDA) methods, as proposed by Baldassini and Serrano (Baldassini and

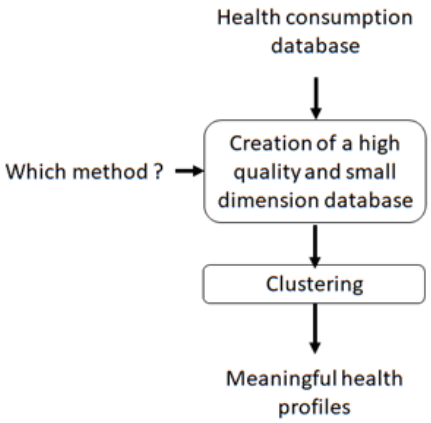

Figure 24 – The main stages of the process

Serrano, 2018). Both improve cluster quality. Like the NMF method, mSDA offers meaningful medical acts clusters. Moreover, W2V can take into account the timeline of the healthcare circuit to reduce dimensions. Methods are compared using cluster consistency and stability. In order to quantify the stability, a new similarity measure between two clusters is proposed.

2.2 Materials and methods

2.2.1 Dimension reduction methods

Dimension reduction makes it possible to identify patterns in a dataset, and thus improves clustering quality by allowing to deal with the dimension curse. These methods have been particularly successful at studying languages: dimension reduction methods are often able to understand general word contexts. It could also improve spectacularly the quality of a clustering method (Gauchon et al., 2019). In this section, we present three dimension reduction methods: NMF, W2V and mSDA. Since these methods are mainly used in the native language processing field, we assume in this section that we work with a text corpus of m different texts counting n different words. The reader could refer to Simpson and Demner-Fushman (Simpson and Demner-Fushman, 2012) for a survey of text-mining applications in biomedicine.

Before explaining these methods, we need to introduce a common concept in text mining: frequency matrices. Given a text corpus, a frequency matrix A counts every occurrence of each word for each text. These matrices are usually very large non-negative sparse matrices. Dimension reduction methods aims to find a matrix A' of dimension $m * k$, with $k < n$, minimizing the loss of information from the matrix A.

Non-negative Matrix Factorization

Lee and Seung (Lee and Seung, 1999) proposed a method for factorizing the matrix A into a produce of two matrices W and H with m (resp. k) rows and k (resp. n) columns. Since it is not always possible to find W and H such as $WH = A$, NMF usually amounts to finding W and H minimizing $d(WH, A)$, with $d()$ a function measuring differences between two matrices.

The matrix H is the basis of the reduced dimensional space. It can be seen as a fuzzy word clustering: it represents the frequency of apparition of word pairs in the same text. Since this basis is not orthogonal, it makes it possible for a word to belong to two different clusters, and thus takes into account the different meanings of a word. Matrix W is the projection of matrix A in the reduced dimensional space. W shows if a text is well represented by each word cluster created this way.

NMF has been used both in text mining (Ding et al., 2008), (Kuang et al., 2015) and in medicine (Brunet et al., 2004), (Chen et al., 2006).Since it has performed better than other algorithms in the field of insurance (Gauchon et al., 2019), the snmf/l algorithm [1] as proposed by Kim and Park has been used (Kim and Park, 2007).

Word2Vec

W2V is a neural network proposed by Mikolov et al. (Mikolov et al., 2013a). It aims to create a word vector space of k dimensions, called word embedding, capturing each component of each word and allowing operations on words [2]. This vector space can be seen as a reduction in k dimensions of the initial space. W2V has been used in many fields, such as medical literature study (Minarro-Giménez et al., 2014), (De Vine et al., 2014) or text clustering (Lilleberg et al., 2015).

We implemented the Continuous Bag of Word (CBOW) architecture for this article. It is a neural network with one hidden layer of k neurons. Contrary to mSDA and NMF algorithms, CBOW does not use frequency matrices and focuses on the context of each word: given a studied word, CBOW takes the neighbors of this word as an input.

In order to counter the imbalance between the rare and frequent words, Mikolov et al. proposed to improve the model by subsampling based on the overall word frequency (Mikolov et al., 2013b).

The neural network is trained to guess the studied word by only knowing its context. Usually, it performs badly at this task. However, the weights of the hidden layer of neurons form the desired word embedding.

1. The R package "NMF", developed by Gaujoux and Seoighe has been used to obtain results of this article (Gaujoux and Seoighe, 2010)
2. As an example, Mikolov et al. wrote that in that kind of space: King – Man + Women = Queen.

Marginalized stacked denoising autoencoders

Autoencoders are usually a kind of neural network trying to compress then reconstruct an input. To do so, the compression must be as efficient as possible. Thus, autoencoders make it possible to capture relationships between data features and to create a reduced dimensional space.

The method called marginalized stacked denoising autoencoders (mSDA) has been proposed by Chen et al. in order to speed up traditional algorithms (Chen et al., 2012). Denoising autoencoders work the same way as classical autoencoders, except that the input is deliberately corrupted. For example, it is possible to set each feature to 0 with probability p. Chen et al. pointed out that for this kind of noise, the law of large numbers makes it possible to explicitly determine the result of an infinite amount of corrupted input. Thus, mSDA is a deterministic method and only depends on two parameters: the probability of noise p and the number of stacked autoencoders.

2.2.2 Clustering with self organizing maps

Self organizing map

Self Organizing Maps[3] (SOM) are a clustering neural network architecture popularized by Kohonen (Kohonen, 1990). Its clustering quality is similar to k-means algorithms one, with the added benefit of a natural result visualization. However, results are very sensitive to the input dimension, which justifies the need for a previous dimension reduction step (Mingoti and Lima, 2006).

In this neural network, each neuron is arranged following a given topology (for example, a two-dimensional hexagonal grid). It is thus possible to define a neighborhood for each neuron, using a so-called neighborhood function h.

Suppose that one wants to cluster M individuals into a m neurons SOM. Each neuron i is initialized with initial random weights w_i^0. The first individual k is then presented to each neuron, in order to determine the neuron with the weights closest to the individual features, called the Best Machine Unit (BMU). Then the BMU weights, but also neighboring neurons' weights are adjusted, in order to create an area which attracts similar individuals. The process is then iterated.

Once the map is trained, we reduce the cluster number by proceeding to a hierarchical clustering (Ambroise et al., 2000).

3. The R package "Kohonen", developed by Wehrens et al. has been used to obtain the results of this article (Wehrens et al., 2007)

Projection

The NMF method results in two matrices, one of them being the data projected into a reduced dimensional space. However, W2V only provides an embedding, which can be seen as the basis W of the reduced dimensional space. Thus, it is still needed to project the inputs on this space.

In order to do that, we first normalize the basis H. We then project all individuals by computing $A^{'} = AH$. This is the classical way to project individuals using the Euclidean distance. Finally, the matrix A' is normalized row by row. This way, the whole process amount to projecting individuals using the cosine similarity [4].

Moreover, mSDA does not directly give an embedding. Instead, it results in a matrix Z with m rows and $m + 1$ columns (the last column being called the bias). To compute the embedding, we have to drop the bias and to take the absolute value of this matrix. This way, we can use a NMF [5] in order to reduce this matrix into a matrix with k rows and m columns, which play the role of H.

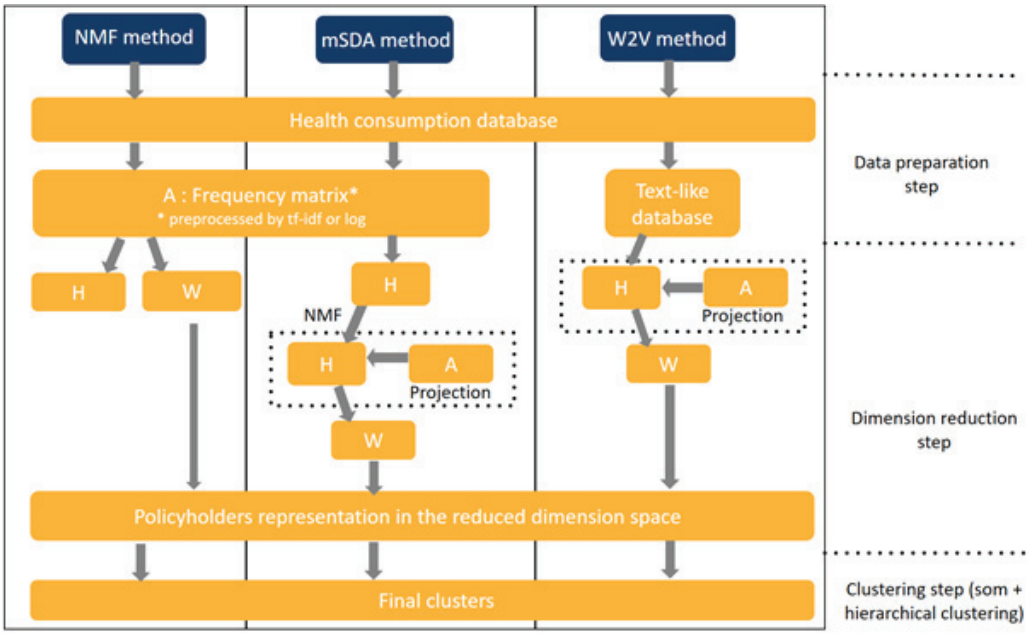

Figure 25 – simplified display of tested methods.

4. Voir 2.5.3 pour une démonstration

5. PCA and SVD have also been tested, resulting in worst results. Also, it is possible to duplicate the rows of the matrix Z a few times with a very small noise, which makes it possible for NMF to find a better correlation between medical acts and result in a better embedding.

2.2.3 From health insurance data to text

A health insurer usually possesses two databases. The first one groups global information about all policyholders, such as age or sex. Due to European data regulation, this database often does not contain a lot of features. For the purpose of focusing only on healthcare circuit, we use this database only to analyze results, and not for clustering.

This database is joined to a second one: the health consumption database. This database contains all the information needed by the health insurer to refund policyholders. It contains the date, the cost and the nature of all policyholder health expenditures (called medical act).

It is then possible to transform these data into a database similar to a text mining dataset. Each policyholder becomes a text, where words are the reimbursed medical acts, and where text order is given by the chronological order of each medical act. For example, a policyholder using drugs on the 01/02 and hospitalization on the 05/05 becomes the text "drugs hospitalization".

As in text mining, some medical acts are much more common than others (for example, pharmaceutical spending), and a given medical act could carry numerous meanings (e.g. ultrasound can be used both for maternity and cardiology).

This dataset does not take into account whether two successive medical acts are separated by one day or by one month. To address this point, it is possible to add a "word" capturing the temporal difference. Hence, the previous example can be represented by the text "drugs 3_ 6_ months hospitalization".

As in text mining, it is still possible to construct a frequency matrix, and then to use NMF or mSDA methods. Since some words are so common that they carry almost no information, text mining frequency matrices are usually transformed using tf-idf method. Since we meet the same problems with drugs, and, to a lesser extent, generalist and specialist practitioners, we also implement this pretreatment, which improves results effectively. An example of application of tf-idf for text classification in biomedicine can be found in Srivastava et al (Srivastava et al., 2019). Also, as shown by Huang, cosine similarity is more adapted to text comparison than the Euclidean distance (Huang, 2008), thus, henceforth all presented metrics are based on the cosine similarity.

Database : the database used in these studies comes from a private French health insurer. It is composed of 28,540 women aged between 16 and 62 years. Most of them are working women. There are more than 1,300,000 medical acts observed over one year, distributed into a hundred different medical acts. Women without at least one medical consumption have been removed.

2.3 Results

2.3.1 Dimension reduction

The first step of the proposed method consists in creating a reduced dimensional space by obtaining a health consumption embedding. In order to make comparison easier, the chosen dimension of the final space is 20, regardless of the dimension reduction method. By normalizing the embedding by columns, it is possible to see each embedding vector as a Medical Act Clusters (MAC) (see Figure 26).

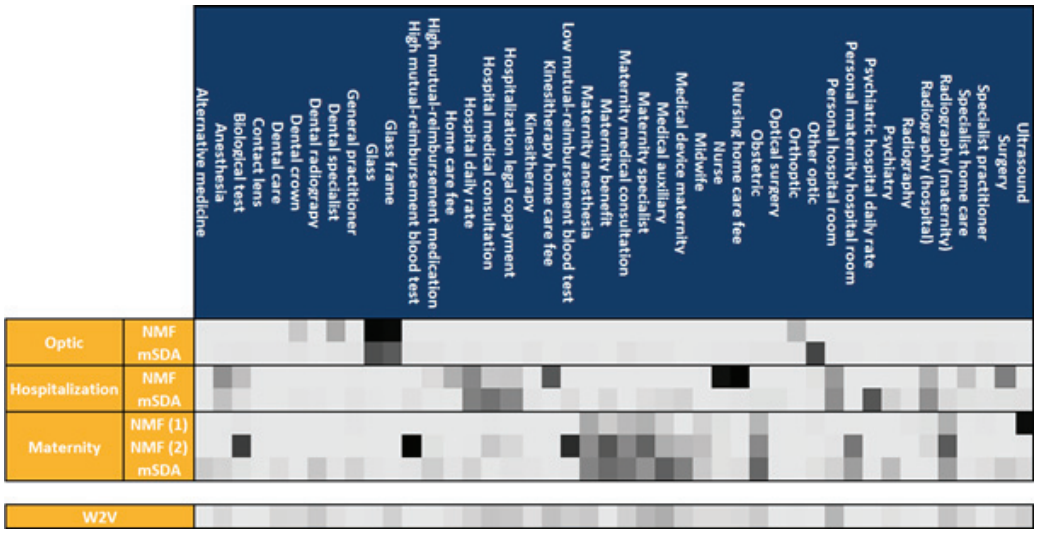

Figure 26 – Dimension reduction leads to Medical Act Clusters. The last line is a vector of the embedding offered by W2V, illustrating that this embedding is not meaningful. NMF(1) and NMF(2) are two medical act classes produced by the NMF method and probably linked to maternity.

First, the embedding obtained with the W2V dimension reduction is not meaningful. It is due to the faculty of W2V to capture nonlinear relationships between medical acts.

On the contrary, MACs obtained with NMF and mSDA are easily understandable. However, MACs obtained with NMF seem to be less accurate than the ones obtained with mSDA, especially for uncommon medical acts, like “Other optic” or “optical surgery” (consumed less than 200 times).

Both algorithms do not focus on the same effects. For example, NMF clusters hospitalization and physical therapy together, whereas mSDA combines psychiatric and classical hospitalization. As for maternity MACs, NMF is not able to dissociate maternity medical acts from biological analysis or ultrasound. Instead, mSDA is able to dissociate maternity from both ultrasounds and biological analysis. However, it is interesting to notice that mSDA puts psychiatry into maternity MAC, this likely being due to postpartum depression.

Finally, notice that for both algorithms, a same medical act can belong to multiple MACs.

2.3.2 Clustering

The dimension reduction makes it possible to obtain a small number of high quality features before clustering. The SOM method has been chosen to take advantage of its natural visualization (an example is given in Appendix A). To make the comparison between each algorithm easier, 20 clusters have been done for each classification. For the sake of concision, we call the whole process by the name of "dimension reduction method".

To understand the meaning of a given cluster, it is possible to compute the rate of people consuming these acts for every medical act (Example given in Figure 4).

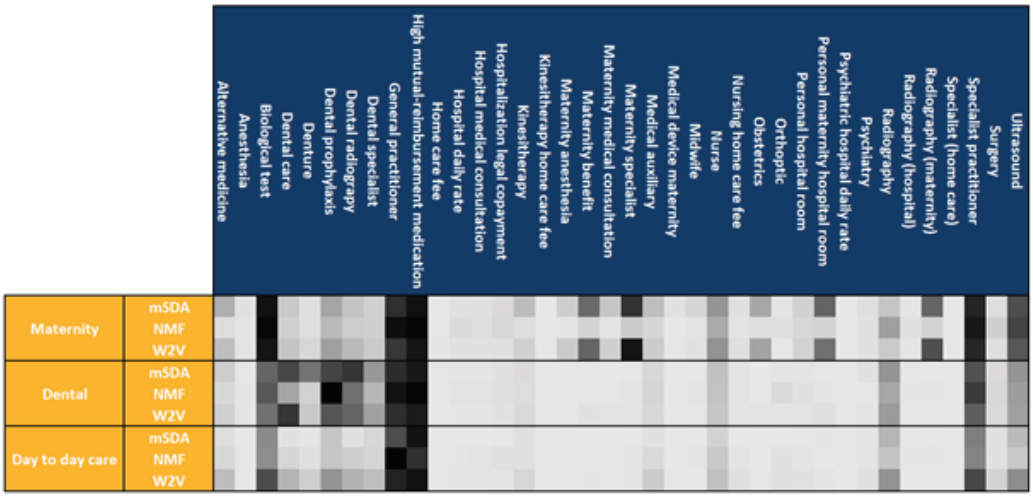

Figure 27 – Comparison of final clusters obtained. A dark square means that most of the policyholders in the associated cluster are consuming the associated medical act.

We can notice that W2V gives understandable final results, even if the embedding is not meaningful. However, the clusters created this way usually seem to be less pure. NMF systematically fails to create a maternity cluster, whereas mSDA rarely creates a dentures cluster. All three methods create several day-to-day care clusters. They also create a huge low-consuming policyholder cluster of around 8 000 people, containing healthy women.

There are two main qualities for a clustering method: clustering quality and stability. Clustering quality are compared using the cluster inertia. Classical similarity measure, such as Gower's distance, cannot be used to compare two partitions. To compare stability, we introduce a new similarity, inspired by the Jaccard coefficient as suggested by Hennig (Hennig, 2007). To the best of the authors knowledge, this similarity has not been seen in this form before. It can be used for a totally unsupervised context, with a nondeterministic algorithm such as a Kohonen's map.

Let N be a set of features. Let C_1 and C_2 two partitions in k subsets (clusters) of N. Intuitively, for each couple of features in a subset in C_1, the clustering method is stable if they also belong to the same subset in C_2.

Formally : Let n_i^1 (resp n_i^2) be the cardinal of subset number i in the partition C_1 (resp C_2). The number of possible couple in a subset is given by $\binom{n_i^1}{2}$. Let $n_{i,j}^{1,2}$ be the number of features in the subset number i of C_1 and in the subset j of C_2.

We compute $sim(C_1, C_2) = \frac{\frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{n_{i,j}^{1,2}}{2}}{\sum_{i=1}^{k}\binom{n_i^1}{2}} + \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{n_{i,j}^{2,1}}{2}}{\sum_{i=1}^{k}\binom{n_i^2}{2}}}{2}$. It is easy to verify that sim is a partition similarity.

This way, we can test if two clusters are similar, and thus test if the method is stable (see Figure 28 for an example).

Since the NMF and mSDA dimension reduction algorithms are almost deterministics, we have constructed 10 different SOMs for both method. We have also constructed 5 different embeddings using W2V, and each of them were used to construct 5 different SOMs. Two stability measures have been computed for W2V : one comparing all 25 SOMs, and one comparing two SOMs only if they are obtained from the same embedding, in order to test the variability caused by the SOM algorithm alone.

Both mSDA and W2V offer better cluster quality than NMF (in terms of inertia). Moreover, mSDA is the most stable process. It is interesting to notice that a big part of W2V's variability comes from the dimension reduction, which has also been observed by Baldassini and Serrano (Baldassini and Serrano, 2018).

2.3.3 Numerical analysis

Analyzing each cluster is necessary for a better understanding of their underlying population. Here, we take the example of clusters mainly described by a high consumption of biological tests. Purity score shows the number of people in the clusters consuming biological tests. mSDA produces a single biological test cluster whereas W2V and NMF produce two biological test clusters.

The mSDA cluster is quite centered on biological test consumption. These profiles could be linked to either chemotherapy control, chronic fatigue or urinary infection. Maternity profiles are contained in another cluster. This is not the case for NMF, which fails to create a maternity cluster. Thus, maternity profiles can be found in the NMF (B) cluster. NMF (A) and W2V

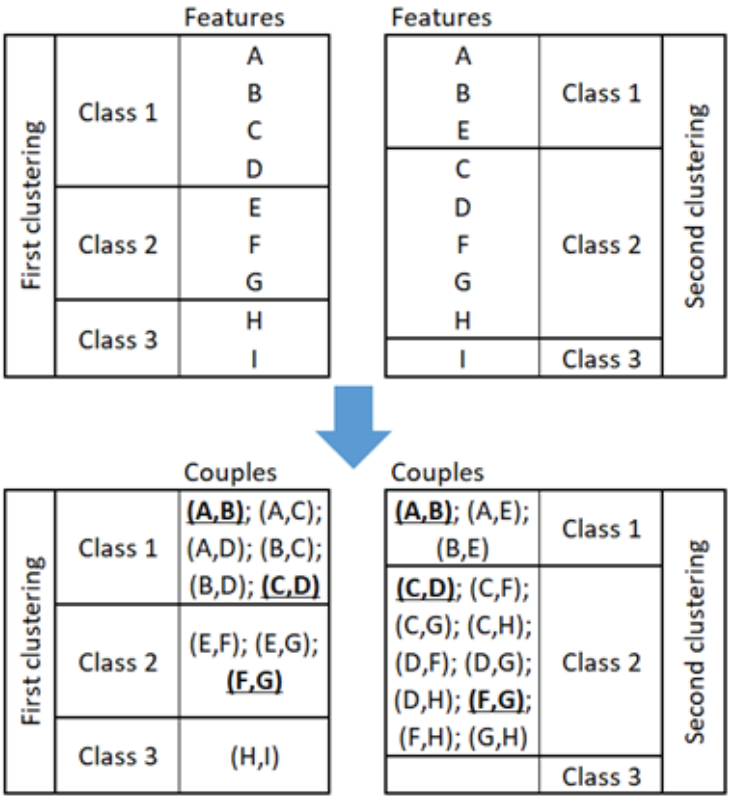

Figure 28 – Two clusterings in 3 classes are made on a set of 9 features $(A, B, \dots, I)$. It appears that A and B; C and D; F and G are on the same class in the first AND in the second clustering. Thus, $\sum_{i=1}^{k}\sum_{j=1}^{k}\binom{n_{i,j}^{1,2}}{2} = \sum_{i=1}^{k}\sum_{j=1}^{k}\binom{n_{i,j}^{2,1}}{2} = 3$. Moreover, $\sum_{i=1}^{k}\binom{n_i^1}{2} = 6+3+1 = 10$ and $\sum_{i=1}^{k}\binom{n_i^2}{2} = 3+10 = 13$. Finally, the similarity beetween the two clusters is $\frac{\frac{3}{10}+\frac{3}{13}}{2}$.

	NMF	mSDA	W2V	W2V (for a given embedding)
Stability mean	47%	**56%**	46%	**58%**
R^2 mean	35%	**43%**	45%	
R^2 maximum	36%	45%	**48%**	

Figure 29 – clustering quality and stability

(A) are similar to the mSDA cluster, while penalizing optic and denture consumption less. Finally, W2V (B) is probably a cluster of people subject to disease monitoring.

	People having at least one biological test	W2V (A)	W2V (B)	mSDA	NMF (A)	NMF (B)	All basis average
Biological tests	**105**	**154**	**113**	**163**	**112**	**157**	58
Medical device	51	27	65	22	47	31	42
Others	21	19	24	10	25	16	16
Medical auxiliary	12	5	8	4	7	4	8
Surgery	43	15	26	9	47	36	33
Dental	60	57	76	23	71	31	59
General practitioner	103	107	146	73	**211**	117	78
Hospitalisation	54	14	23	10	71	36	40
Kinesitherapy	73	11	22	6	89	6	58
Maternity	43	15	2	6	19	**126**	29
Optic	216	294	286	110	263	32	203
Drugs	174	147	**413**	119	242	173	132
Dentures	129	211	228	11	165	45	126
Psychiatry	28	2	8	2	23	4	23
Radiography	84	76	81	67	102	113	60
Specialist practioner	100	119	110	78	122	135	78
Sum	1296	1272	1631	712	1617	1063	1042
Headcount	15686	971	858	1505	814	1362	28540
Average age	41.5	38.8	47.7	38.5	40.2	39.2	40.8
Purity	100%	90%	86%	100%	80%	97%	

Figure 30 – "biological tests" clusters, numerical details. Both NMF and W2V produce two "biological test" clusters.

Conclusion

Three clustering processes based on the use of the healthcare circuit have been presented and compared. All three processes are in twofold: first the dimension is reduced, and then a clustering is performed using a self-organizing map. The dimension reduction step makes it possible for the self-learning process to understand the correlation between medical acts, and creates high quality features for the clustering step. The method based on the W2V principle offer the best results. However, results offered by the one based on mSDA are close, but this method is much more meaningful. All methods offer a useful clustering and a more precise understanding of the makeup of a population's health.

Moreover, a new stability measure has been proposed, in order to compare all three methods. It results that the mSDA method is the most stable methods for our data.

The method has been tested on a health insurance database. Moreover, no other medical information has been added in the clustering process. However, it would be of interest to use them in a more general medical context, such as for studying and clustering low back pain diseases. Adding features to the clustering algorithm could improve results easily and effectively.

Conflict of interest statement : Jean-Pascal Hermet and Romain Gauchon work for ADDACTIS France, an actuarial consulting firm.

Acknowledgements : The authors would like to thank ADDACTIS France for providing the database. This paper was realized within the framework of the Chair Prevent'Horizon, supported by the risk foundation Louis Bachelier and in partnership with Claude Bernard Lyon 1 University, ADDACTIS France, AG2R La Mondiale, G2S, Covea, Groupama Gan Vie, Groupe Pasteur Mutualité, Harmonie Mutuelle, Humanis Prévoyance and La Mutuelle Générale.

2.4 Appendix

2.4.1 Appendix A : Example of a Kohonen's map obtained following the whole process

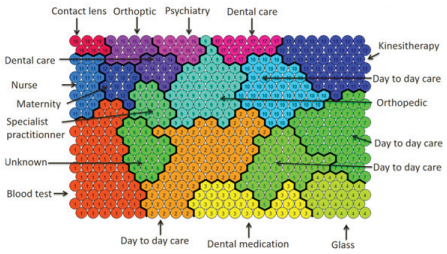

Figure 31 – Example of a Kohonen's map. Data has been preprocessed using the W2V method.

Bibliographie

Ambroise, C., Sèze, G., Badran, F., and Thiria, S. (2000). Hierarchical clustering of self-organizing maps for cloud classification. Neurocomputing, 30(1-4) :47–52.

Azar, A. T., El-Said, S. A., and Hassanien, A. E. (2013). Fuzzy and hard clustering analysis for thyroid disease. Computer methods and programs in biomedicine, 111(1) :1–16.

Baldassini, L. and Serrano, J. A. R. (2018). client2vec : towards systematic baselines for banking applications. arXiv preprint arXiv :1802.04198.

Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the national academy of sciences, 101(12) :4164–4169.

Chen, M., Xu, Z., Weinberger, K., and Sha, F. (2012). Marginalized denoising autoencoders for domain adaptation. arXiv preprint arXiv :1206.4683.

Chen, Z., Cichocki, A., and Rutkowski, T. M. (2006). Constrained non-negative matrix factorization method for eeg analysis in early detection of alzheimer disease. In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, volume 5, pages V–V. IEEE.

De Vine, L., Zuccon, G., Koopman, B., Sitbon, L., and Bruza, P. (2014). Medical semantic similarity with a neural language model. In Proceedings of the 23rd ACM international conference on conference on information and knowledge management, pages 1819–1822. ACM.

Ding, C., Li, T., and Peng, W. (2008). On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. Computational Statistics & Data Analysis, 52(8) :3913–3927.

Gauchon, R., Loisel, S., and Rullière, J.-L. (2019). Health-policyholder clustering using health consumption.

Gaujoux, R. and Seoighe, C. (2010). A flexible r package for nonnegative matrix factorization. BMC bioinformatics, 11(1) :367.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis, 52(1) :258–271.

Huang, A. (2008). Similarity measures for text document clustering. In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand, pages 49–56.

Ibrahim, S., Chowriappa, P., Dua, S., Acharya, U. R., Noronha, K., Bhandary, S., and Mugasa, H. (2015). Classification of diabetes maculopathy images using data-adaptive neuro-fuzzy inference classifier. Medical & biological engineering & computing, 53(12) :1345–1360.

Kim, H. and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics, 23(12) :1495–1502.

Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78(9) :1464–1480.

Kuang, D., Choo, J., and Park, H. (2015). Nonnegative matrix factorization for interactive topic modeling and document clustering. In Partitional Clustering Algorithms, pages 215–243. Springer.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755) :788.

Lilleberg, J., Zhu, Y., and Zhang, Y. (2015). Support vector machines and word2vec for text classification with semantic features. In 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), pages 136–140. IEEE.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. arXiv preprint arXiv :1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pages 3111–3119.

Minarro-Giménez, J. A., Marin-Alonso, O., and Samwald, M. (2014). Exploring the application of deep learning techniques on medical text corpora. Studies in health technology and informatics, 205 :584–588.

Mingoti, S. A. and Lima, J. O. (2006). Comparing som neural network with fuzzy c-means, k-means and traditional hierarchical clustering algorithms. European journal of operational research, 174(3) :1742–1759.

Silva, L. F., Santos, A. A. S., Bravo, R. S., Silva, A. C., Muchaluat-Saade, D. C., and Conci, A. (2016). Hybrid analysis for indicating patients with breast cancer using temperature time series. Computer methods and programs in biomedicine, 130 :142–153.

Simpson, M. S. and Demner-Fushman, D. (2012). Biomedical text mining : a survey of recent progress. In Mining text data, pages 465–517. Springer.

Srivastava, S. K., Singh, S. K., and Suri, J. S. (2019). Effect of incremental feature enrichment on healthcare text classification system : A machine learning paradigm. Computer methods and programs in biomedicine, 172 :35–51.

Wehrens, R., Buydens, L. M., et al. (2007). Self-and super-organizing maps in r : the kohonen package. Journal of Statistical Software, 21(5) :1–19.

2.5 Complément : Comment transposer l'algorithme Word2Vec à des données d'assurance santé

L'algorithme Word2Vec n'ayant pas été créé afin de traiter des données issues de l'assurance santé, il a été nécessaire de l'adapter au contexte assurantiel. Ce complément a pour objectif de détailler certains choix effectués dans le cadre de cette thèse, ainsi que de proposer quelques pistes afin de potentiellement en améliorer les performances.

2.5.1 Transformer des données de prestations santé en texte

Comme souligné plus haut dans le chapitre 2, il existe de nombreuses similarités entre des données de texte et des données d'assurance santé. Notamment, les prestations santé sont assimilables à des mots. Afin d'appliquer l'algorithme Word2Vec à des prestations santé, il est nécessaire d'utiliser ces similarités afin de transformer la base de données en un texte exploitable pour l'algorithme. Cependant, il existe aussi quelques différences entre un texte et une base de prestations santé, qui doivent être prises en compte afin de transformer correctement les données santé en texte et d'appliquer l'algorithme :

— Une base de données d'assurance santé contient jusqu'à 200 types de prestations différents. C'est beaucoup moins que le nombre de mots différents que compte généralement un corpus de texte.

  Or, à partir d'un texte, l'algorithme Word2Vec a pour but de créer un espace de dimension réduite appelé embedding. Généralement, cet espace est de dimension 200 environ. Cette dimension n'est donc évidemment pas adaptée à des données d'assurance santé. Afin de calibrer, il a donc été nécessaire de créer des embeddings de différentes dimensions afin de trouver celles qui offraient les meilleurs résultats. Empiriquement, une dimension aux alentours de 20 permettait d'obtenir les classifications présentant les meilleurs R^2.

— Il est impossible de savoir dans quel ordre placer les prestations qui ont eu lieu un même jour. En effet, les dates de prestation ne sont précises qu'à la journée près. En termes de données de texte, cela revient à avoir des mots qui se superposent.

  Dans les tests réalisés, dans de tels cas, il a été décidé d'ordonner les prestations aléatoirement. L'algorithme Word2Vec considérant tous les mots placés dans une fenêtre autour d'un mot cible, ce choix n'a une influence que lorsque la fenêtre choisie est trop petite pour contenir toutes les prestations réalisées un même jour. Une fenêtre de 7 ayant été retenue après différents tests empiriques, ce cas est resté marginal.

— Dans un texte, un espace entre deux mots a un sens bien défini. Or, pour des données de prestations santé, le délai entre deux prestations santé joue le rôle de l'espace. Cependant, un délai d'un jour entre deux prestations santé ne contient pas la même information qu'un délai d'un mois. Il est donc nécessaire de tenir compte de ces délais.

En pratique, deux méthodes ont été proposées afin de tenir compte de ce phénomène (voir l'exemple 32). Soit une prestation fictive est insérée tous les jours, à la suite des prestations du jour (y compris lorsqu'il n'y a pas de prestation). Ces prestations fictives, notées "N" (pour nuit), servent de marqueurs de séparation des jours entre eux (traitement "N"). Soit une unique prestation fictive est insérée entre deux prestations, lorsqu'elles sont séparées d'un ou plusieurs jours. Idéalement, cette prestation fictive dépend de la durée séparant les deux prestations (traitement "XN"). Ces deux manières de constituer un texte ont été testées. Empiriquement, les résultats obtenus avec le traitement "XN" semblaient légèrement meilleurs.

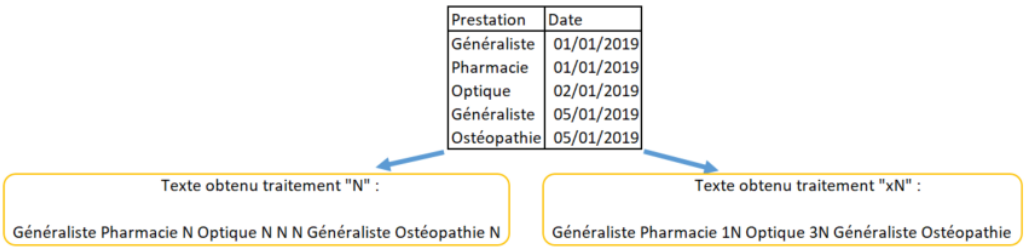

FIGURE 32 – Deux manières de transformer des prestations en texte

2.5.2 Subsampling

Tous les mots ne sont pas utilisés à la même fréquence dans le langage. Certains, comme les articles, sont utilisés très fréquemment mais apportent très peu d'information. De même, dans le domaine de l'assurance santé, certaines consommations (comme la pharmacie) sont très fréquentes mais regroupent des soins si divers que l'information qu'elles contiennent est très diluée. Cependant, comme pointé par Mikolov et al. dans l'article (Mikolov et al., 2013), considérer les mots fréquents de la même manière que les mots rares peut potentiellement affecter la qualité des résultats. Afin de résoudre ce problème, Mikolov et al. proposent un prétraitement des données d'entrée, appelé subsampling.

Ce prétraitement consiste à supprimer aléatoirement des mots. Pour un mot w_i donné, qui a une fréquence d'apparition $f(w_i)$, Mikolov et al. propose de le supprimer selon une probabilité $P(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}}$. Le paramètre t est un paramètre fixé par l'utilisateur permettant de contrôler le nombre de mots supprimés : plus ce coefficient est élevé, moins la probabilité de supprimer un mot est forte. Mikolov et al. le fixent empiriquement à 10^{-5}. Cependant, cette valeur empirique n'est pas forcément justifiée pour des données issues de l'assurance santé : le nombre de consommations différentes étant nettement inférieur au nombre de mots dans un corpus, les ordres de grandeur des fréquences ainsi calculées sont différents et la valeur idéale du paramètre t peut changer. Un travail de test sur différentes valeurs de t a donc été réalisé.

Il est apparu que sur les jeux de données utilisés, une valeur de 10^{-6} offrait des résultats convenables.

2.5.3 Projeter les assurés sur l'espace de dimension réduite

Contrairement à la méthode NMF, l'algorithme Word2Vec ne permet pas d'obtenir directement la représentation des assurés dans l'espace de dimension réduite : seule la base de cet espace est disponible [6].

Afin de résoudre ce problème, il a été décidé de projeter les assurés, représentés sous forme de matrice de fréquence, dans l'espace de dimension réduite. Autrement dit, si on note E l'espace de dimension non réduite, et $\tilde{E}$ l'espace de dimension réduite, pour tout assuré $x_i \in E$, on recherche $\tilde{x_i} \in \tilde{E}$ tel que $\tilde{x_i} = \underset{y \in \tilde{E}}{\text{Argmin}}\, d(y, x_i)$, avec d une distance choisie.

Or, comme souligné par Huang (Huang, 2008), la distance choisie peut avoir une grande influence sur les résultats. En particulier, la distance euclidienne n'est souvent pas adaptée à des données de type textes. Deux mesures sont généralement utilisées : la divergence de Kullback Leibler et la similarité cosinus.

Selon la distance choisie, il n'existe pas forcément de formule explicite afin de calculer les projections de chacun des points dans l'espace de dimension réduite. Il est alors généralement nécessaire d'approximer les projections via des méthodes numériques, ce qui peut être coûteux en temps. Cependant, ce n'est pas nécessaire pour la similarité cosinus. En effet, projeter les données avec la similarité cosinus revient à projeter selon la norme euclidienne, puis à normer les projections (afin qu'elles soient de norme un).

En effet, notons

$$\begin{array}{rccl} P_{euc} & : & E & \to & \tilde{E} \\ & & X & \mapsto & \underset{Y}{\text{Argmin}}\, d_{euc}(Y, x_i) \end{array}$$

la projection selon la distance euclidienne, et

$$\begin{array}{rccl} P_{cos} & : & E & \to & \tilde{E} \\ & & X & \mapsto & \underset{Y \in \tilde{E} \ tq \ ||Y||=1}{\text{Argmin}} \left(\frac{|X \cdot Y|}{||X||\ ||Y||}\right) \end{array}$$

la projection selon la similarité cosinus [7]. Soit $X \in E$.

En utilisant la relation de Chasles et la distributivité du produit scalaire, il est facile de voir que pour tout $Y \in \tilde{E}$, $\frac{X \cdot Y}{||X||\ ||Y||} = \frac{(Y+X-Y)\cdot Y}{||X||\ ||Y||} = \frac{||Y||+(X-Y)\cdot Y}{||X||\ ||Y||}$. Ainsi, trouver $\underset{Y \in \tilde{E} \ tq \ ||Y||=1}{\text{Argmin}} \left(\frac{|X \cdot Y|}{||X||\ ||Y||}\right)$

6. Une variante de l'algorithme Word2Vec, appellée Doc2Vec, a été proposée par Le et Mikolov afin de régler ce problème (Le and Mikolov, 2014). Cependant, cette variante n'a pas été testée.

7. Imposer la condition $||Y|| = 1$ est nécessaire afin d'assurer l'unicité de la projection. En effet, minimiser la similarité cosinus revient à minimiser le cosinus d'un angle. Or, la notion d'angle entre deux vecteurs ne dépend pas de la norme de chacun de ces vecteurs. Une infinité de vecteurs peuvent donc minimiser la similarité cosinus.

revient à trouver $\underset{Y \in \tilde{E} \ tq \ ||Y||=1}{\text{Argmin}} |(X-Y) \cdot Y|$. Or, le théorème de Pythagore assure que $p_{euc}(X)$ est la projection orthogonale de X sur $\tilde{E}$, donc $|(X - p_{euc}(X)) \cdot p_{euc}(X)| = 0$. Finalement, $\underset{Y \in \tilde{E} \ tq \ ||Y||=1}{\text{Argmin}} |(X-Y) \cdot Y| = \frac{p_{euc}(X)}{||p_{euc}(X)||}$, donc projeter selon la norme euclidienne et normer la projection revient bien à projeter selon la similarité cosinus. C'est pourquoi il a été choisi de projeter selon la norme euclidienne, puis de normer la matrice des individus projetés par ligne et non par colonne (ce qui va à l'encontre de ce qu'il se fait d'habitude) avant de classifier les individus entre eux.

La manière dont les assurés ont été projetés pour la méthode mSDA est la même que pour la méthode W2V.

Bibliographie

Huang, A. (2008). Similarity measures for text document clustering. In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand, pages 49–56.

Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In International conference on machine learning, pages 1188–1196.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pages 3111–3119.

Chapitre 3

La psychiatrie : un risque important en assurance santé ?

Ce chapitre reprend un article coécrit avec Jean-Pascal Hermet et soumis au bulletin français d'actuariat

Résumé

La psychiatrie a été identifiée par la Sécurité Sociale comme un enjeu majeur. Pourtant, celle-ci n'est pas identifiée comme un risque important par les organismes d'assurance santé complémentaire, si bien qu'aucune étude n'a été publiée en France sur sa prise en charge en dehors de l'Assurance Maladie. Cet article présente le résultat de travaux réalisés à partir de quatre bases de données regroupant les prestations versées par des complémentaires santé. L'objectif de cette analyse consiste à mieux comprendre ce risque et évaluer son coût pour les complémentaires santé. Cette étude permet de conclure qu'un assuré identifié comme étant suivi par un psychiatre présente un coût médical à la charge de l'assureur deux fois supérieur à celui d'un assuré moyen, notamment à cause de fortes dépenses en hospitalisation non psychiatrique. Les personnes ayant entre 15 et 30 ans et les séniors sont les plus concernés par ce risque. Une analyse à partir des jeux de données gouvernementaux Open DAMIR et Open MEDIC est également proposée.

Abstract

The French Social Security has highlighted psychiatric diseases as a major risk. However, French private health insurance companies do not consider psychiatric care as an important topic. In this article is presented a study (based on four real health insurance databases) of the cost of a policyholder benefiting from a psychiatric follow-up care. It aims to provide a better understanding of this risk in an insurance context. It is shown that such a policyholder costs twice as much as an average policyholder, notably due to important expenses in hospitalization. Policyholders aged from 15 to 30 years, as well as the ones over 70 years old are the most concerned by this risk. An analysis using the Open DAMIR and Open MEDIC datasets is also provided.

Keywords : Assurance santé; psychiatrie

3.1 Introduction

20,3 milliards d'euros. Cela représente le coût en 2017 des maladies psychiatriques et des psychotropes pour l'Assurance Maladie Assurance Maladie (2019), soit près de 5 milliards de plus que celui des cancers, et 6 de plus que celui des soins courants ou celui des pathologies cardio-vasculaires. Il atteint à 109 milliards d'euros si l'on inclut les coûts indirects de ces pathologies liés par exemple à la perte de qualité de vie et d'autonomie (voir Chevreul et al. (2013)).

L'Assurance Maladie évalue à 7,2 millions le nombre de Français touchés par une maladie psychiatrique ou consommant des psychotropes, ce qui représente environ 12,5% de la population qu'elle couvre. Cette estimation ne prend pas en compte le non-recours aux soins : le chiffre réel dépasserait plus de 20% (c.f. Wittchen et al. (2011)).

Ces chiffres significatifs ont conduit l'Assurance Maladie à produire d'importantes études statistiques afin de mieux appréhender ce risque. Par exemple, un individu touché par une pathologie psychiatrique a coûté en moyenne 6 586 euros lors de l'année 2016 à la Sécurité Sociale [1], 61% de cette somme provenant de frais d'hospitalisation, 15% de prestations en espèces [2], et 7% de la pharmacie (c.f. Assurance Maladie (2018a)). Ces études statistiques très complètes ne s'intéressent cependant pas aux coûts des maladies psychiatriques pour les complémentaires santé.

Étonnamment, il n'existe pas d'étude publiée sur le coût du risque psychiatrique pour les assurances complémentaires santé. Ce risque n'a jamais été réellement évalué par des actuaires en France et reste très méconnu. Or, les **assurés qui ont recours aux actes de psychiatrie (AP)** coûtent en moyenne très cher aux complémentaires santé. Le but de cet article est de montrer que le risque psychiatrique est un risque important pour les organismes complémentaires santé. À partir de l'étude statistique de quatre **Bases De Données (BDD)** constituant un benchmark, on met en évidence le fait que **les assurés bénéficiant de remboursements en psychiatrie sont quasiment ceux qui coûtent le plus cher à un organisme complémentaire santé**, dépassant notamment ceux qui ont des remboursements en optique ou en kinésithérapie. Une des raisons de l'importance de ce risque est **sa corrélation avec la quasi-totalité des postes de dépenses de l'assureur**.

Du fait de l'utilisation de bases de données telles que possédées en pratique par les assureurs santé, les résultats peuvent parfois sembler incomplets. L'objectif de cet article n'est pas de trancher définitivement sur le sujet du coût de la psychiatrie pour les compagnies d'assurance. En revanche, elle vise à montrer que la psychiatrie est un sujet qui devrait être considéré sérieusement par les acteurs de l'assurance santé, en raison des coûts importants qui peuvent

1. A titre de comparaison, le coût moyen par individu était de 1 126 euros en 2016
2. Les prestations en espèces sont les montants versés par la Sécurité Sociale afin de compenser une perte de revenu, due par exemple à un arrêt de travail

lui être reliés. En effet, ces acteurs évitent souvent de s'emparer du sujet, évoquant le mythe que les coûts relatifs aux Affections Longues Durées (dont fait partie la plupart des pathologies psychiatriques) sont majoritairement pris en charge par la Sécurité Sociale et donc restent négligeables pour les complémentaires santé. Au vu de l'étude qui suit, il apparait que ce n'est pas le cas en matière de psychiatrie, et que le sujet mérite d'être approfondi. Cet article a donc pour vocation d'éveiller la curiosité des acteurs et d'être suivi par d'autres études.

Cet article est composé de cinq parties. Une première partie résume le contexte de la psychiatrie en France. Une deuxième partie présente les bases de données et plusieurs statistiques descriptives sur la psychiatrie en assurance. Une troisième partie expose différents modèles linéaires permettant de mettre en avant la corrélation de la psychiatrie aux divers postes de dépenses. Une quatrième partie exploite les jeux de données Open DAMIR et Open MEDIC afin de compléter les données disponibles, notamment en étudiant la part des ALDs dans la dépense et le montant réels de dépense en matière de psychotropes. Enfin, une dernière partie conclut et propose des pistes de recherches ultérieures.

3.2 Présentation rapide de la psychiatrie en France

Dans cette section sont présentés un certain nombre d'éléments de contexte autour du milieu de la psychiatrie et des pathologies associées. La psychiatrie peut être vue comme l'étude des pathologies mentales. Elle recouvre généralement quatre grandes classes de maladies[3] pouvant entraîner la consultation d'un psychiatre :

- Les troubles graves du comportement (troubles anxieux, addictions graves, anorexie...)
- Les troubles de l'humeur (troubles dépressifs, bipolaires)
- Les psychoses (schizophrénies, délires, troubles maniaques...)
- Les démences (Alzheimer...)

Parmi elles, les troubles dépressifs et les troubles anxieux sont les pathologies les plus courantes, avec une prévalence globale supérieure à 10%, et sont plus fréquents chez les femmes que chez les hommes (d'après Vaiva et al. (2008) et Wittchen et al. (2011)).

Une des spécificités des maladies psychiatriques est leur durée. Certaines, comme les schizophrénies, ne sont *a priori* jamais guéries. D'autres, comme les dépressions, peuvent être soignées, mais les taux de rechutes sont très importants Cruwys et al. (2013). C'est pourquoi, à l'exception notable des dépressions non chroniques, ces pathologies sont classées en Affections Longues Durées (ALD)[4], expliquant qu'une grande partie des coûts soit supportée par la Sécurité Sociale.

Ce sont généralement les médecins généralistes qui détectent une pathologie psychiatrique. Si la maladie n'est pas jugée trop grave, le médecin généraliste peut décider de la suivre

3. Certains handicaps comme les retards mentaux peuvent aussi être suivis par des psychiatres
4. Les démences sont rassemblées dans l'ALD 15, les autres maladies sont groupées dans l'ALD 23

lui-même, éventuellement en coordination avec un psychologue. Sinon, il redirige le patient vers un psychiatre, qui peut lui aussi choisir de se faire seconder par un psychologue. Les psychiatres et les généralistes, à la différence des psychologues, sont habilités à prescrire des médicaments (comme des psychotropes).

Les pathologies les plus sévères sont prises en charge dans le cadre d'une hospitalisation, mais la majorité sont traitées de façon ambulatoire, avec ou sans médication. Les établissements prenant en charge les patients sont généralement des hôpitaux psychiatriques ou des cliniques privées, mais il existe aussi des divisions spécialisées qui assurent le suivi, après prise en charge aux urgences, des personnes ayant fait une tentative de suicide (près de 200 000 tentatives de suicide médicalisées ont été enregistrées en France en 2014, voir Ministère des Affaires sociales de la Santé et des Droits des femmes (2014)).

Même s'il existe de nombreux acteurs capables de suivre des pathologies psychiatriques, beaucoup d'entre elles ne sont jamais soignées. Il existe en effet un très fort décalage entre le nombre de maladies psychiatriques soignées et le nombre de maladies estimées. Or, le baromètre du renoncement aux soins dans le Gard Warin and Chauveaud (2014) montre que ce dernier est très faible vis-à-vis des pathologies psychiatriques. On peut donc conclure qu'il s'agit principalement de non-recours aux soins [5], une partie des personnes concernées refusant de se reconnaître comme malades. Les spécialistes du secteur soulignent souvent la volonté de ne pas commencer une thérapie par peur "d'être pris pour un fou" (voir par exemple Demailly (2011), qui estime par ailleurs entre 40 et 66% la part des dépressions non soignées), la folie étant souvent confondue à tort avec les maladies mentales. Ces dernières sont d'ailleurs spontanément assimilées à la folie par de nombreux Français (c.f. IPSOS and Klesia (2014) ou Roelandt et al. (2010)).

Aux soins directs s'ajoute la prise en charge de pathologies annexes, de nombreuses études pointant le fait que les maladies psychiatriques sont corrélées à de nombreuses autres maladies. Ainsi, la Sécurité Sociale montre une proportion beaucoup plus forte de cancers et de maladies cardio-vasculaires chez les personnes atteintes d'une maladie psychiatrique (e.g. Assurance Maladie (2018a)). D'autres études montrent que les troubles psychiatriques sont souvent la cause de divers problèmes dentaires (e.g. Denis and Coquaz (2013), Kisely (2016)). Plusieurs raisons peuvent expliquer ces corrélations. D'une part, certaines maladies peuvent aussi favoriser l'apparition de dépressions, soit du fait de la chronicité, soit du fait des effets secondaires de certains médicaments (par exemple les béta-bloquants dans le cas de maladies cardio-vasculaires). D'autre part, les maladies psychiatriques peuvent diminuer l'efficacité du système immunitaire, favorisant ainsi l'apparition de diverses pathologies. De plus, les

5. Le renoncement au soin fait référence à un besoin de soin identifié (mais pas forcément avéré), et non satisfait. Il se distingue du non recours au soin, qui fait référence à un besoin de soin avéré (mais pas forcément identifié) et non satisfait. Par exemple, un individu dépressif qui n'a pas conscience qu'il a besoin d'un suivi médical ferait preuve de non recours au soin mais pas de renoncement au soin.

psychotropes, utilisés dans le cadre du traitement des troubles psychiatriques, peuvent eux aussi avoir des effets secondaires nécessitant des soins spécifiques.

La prise en charge financière de ces soins est majoritairement supportée par la Sécurité Sociale. Ainsi, le régime général, comme le régime local, ne distingue pas en termes de remboursement la psychiatrie des autres spécialités médicales si elle s'inscrit dans le cadre du parcours de soins. Il est aussi possible de consulter un psychiatre sans prescription préalable, mais dans ce cas la séance est moins bien prise en charge par l'Assurance Maladie pour les plus de 25 ans. De plus, les hospitalisations psychiatriques sont légèrement mieux remboursées que les hospitalisations standards, puisque le forfait journalier d'un établissement psychiatrique s'élève à 15 euros (contre 20 euros en général). Enfin, sauf dans le cadre de certaines expérimentations récentes, comme en Haute-Garonne depuis mars 2018, et cas particuliers comme l'intervention dans un hôpital, les psychologues ne sont pas remboursés par la Sécurité Sociale.

Les organismes complémentaires interviennent ensuite en complément de l'Assurance Maladie afin de compléter les remboursements. Cependant, ils se contentent généralement de rembourser des soins mal couverts par la Sécurité Sociale, comme les consultations de psychologues, et s'investissent peu sur le terrain des maladies psychiques. Seules deux initiatives majeures financées par des compagnies privées et portant sur la psychiatrie ont pu être idéentifiées : l'institut de prévoyance Klesia a ainsi collaboré avec l'institut FondaMental afin de réaliser un sondage sur la perception des maladies mentales en France IPSOS and Klesia (2014). Klesia a aussi mis en place un partenariat avec l'association IAF Réseau afin de mieux accompagner les aidants de personnes atteintes de maladies psychiatriques. De son côté, la Mutuelle Générale de l'Education Nationale a elle aussi réalisé un sondage avec Opinion Way (en distinguant ses adhérents de la population française globale) afin de mieux comprendre comment la psychiatrie était perçue par ses assurés Opinion Way and MGEN (2014).

Si une partie des coûts dus à la prise en charge de la psychiatrie est supportée par les assureurs santé, deux autres grands secteurs peuvent aussi être impactés par les maladies mentales : la prévoyance et la vie. Pour le premier, les arrêts de travail liés à ces pathologies souvent longues peuvent être très importants. Certaines dépressions ("burn out") peuvent exceptionnellement être classées comme maladies professionnelles (600 cas en 2016, c.f. Assurance Maladie (2018b)), mais la majorité reste considérée comme des incapacités classiques. En vie, les maladies psychiatriques sont corrélées avec une baisse de l'éspérance de vie et avec un très fort risque de suicide. Celui-ci fait cependant généralement partie des exclusions au cours de la première année de vie d'un contrat.

Cependant, les études statistiques présentées dans les parties suivantes concernent uniquement les dépenses santé. Cet article propose une vision centrée sur les complémentaires santé et vient donc compléter les études effectuées par la Sécurité Sociale.

3.3 Résultats statistiques

3.3.1 Présentation des bases de données

L'étude a été réalisée à partir de quatre bases de données (BDD) regroupant les prestations versées par des complémentaires santé, totalisant plus de 500 000 individus uniques. Chacune de ces bases à ses spécificités, détaillées dans cette Partie. Utiliser quatre bases de données permet ainsi de tester la sensibilité des résultats aux spécificités des portefeuilles sous-jacents, en vérifiant qu'ils ne sont pas dus par exemple à un biais lié à la population qui le compose. A noter qu'afin d'alléger la lecture, et puisque les résultats observés sont similaires entre toutes les BDD, une attention particulière sera portée aux résultats issus de la base A. Les principales statistiques descriptives de chacune des bases sont disponibles dans le tableau 33 (les effectifs de la BDD A par âge sont également disponibles dans l'Annexe 3.6).

Chacune des bases contient, pour chaque acte santé, 4 dépenses différentes : le coût total en euros de l'acte (dépense totale, ou DT), le montant remboursé par la Sécurité Sociale (remboursement obligatoire, ou RO), le remboursement de l'assureur (remboursement complémentaire, ou RC), et les restes à charge pour l'assuré (RAC). Sauf mention contraire, les coûts présentés dans la suite de cet article sont les RC des assurés, afin que cette étude soit complémentaire avec celles de la Sécurité Sociale qui portent sur la DT et le RO (de plus, la quasi-totalité des résultats obtenus sur les RC peuvent être généralisés aux DT observées).

	Base A	Base B	Base C	Base D
Nombre de consommateurs	113 000	201 600	188 200	219 900
- Dont femmes	48%	53%	69%	?
- Dont +62 ans	13%	20%	26%	24%
Nombre d'AP	3 900	8 200	4 900	7 100
- Dont femmes	56%	59%	77%	?
- Dont +62 ans	12%	16%	18%	20%
DT moyenne par assuré	1 174	1 355	886	1 331
RC moyen par assuré	602	630	216	426
DT moyenne par AP	2 680	3 050	1 773	2 880
RC moyen par AP	1421	1280	404	916
Age moyen	37,6	40,3	42	43,5
Age moyen AP	42,3	41,9	43,4	47,3

Figure 33 – Statistiques descriptives des bases

— La base A est une base santé collective. Elle contient l'âge et le sexe des assurés ainsi que la totalité des prestations versées par l'assureur pendant un an.
— La base B est une base santé collective. Elle contient l'âge, le sexe et la région d'habitation des assurés, ainsi que la totalité des prestations versées pendant deux

ans par l'assureur, à l'exception des prestations pharmacie.

— La base C est une base santé individuelle. Il s'agit d'une surcomplémentaire, complétant les remboursements d'un premier contrat comportant des garanties très faibles qui concernent essentiellement l'optique et le dentaire. Elle contient l'âge et le sexe des assurés, ainsi que la totalité des prestations versées pendant deux ans par l'assureur. L'exposition n'est pas renseignée, impliquant la présence d'assurés dans la base qui ne sont pas observés sur la période complète.

— La base D est une base collective. Elle contient l'âge (mais pas le sexe) des assurés, ainsi que la totalité des prestations versées pendant un an par l'assureur.

Lorsque la BDD couvre un laps de temps de deux ans, les individus présents pendant la période complète et qui n'ont pas changé de contrat ont été considérés deux fois. Ainsi, un assuré de la base B qui serait présent en 2017 et 2018 serait considéré comme deux assurés distincts, avec un an de différence d'âge. Les assurés n'ayant demandé aucun remboursement santé dans l'année ont été systématiquement retirés. Chaque BDD possède plusieurs contrats santé ayant plusieurs niveaux de garantie, qui n'étaient pas connus pour cette étude.

Les remboursements d'actes de soin ont été ventilés en différents grands postes de dépense afin d'être analysés, à savoir : psychiatrie (y compris hospitalisation psychiatrique et psychologue), soins courants (analyse, généraliste, spécialiste, pharmacie), optique et dentaire, hospitalisation (y compris chirurgie et maternité, mais **hors hospitalisation psychiatrique**), et autres (dont la kinésithérapie). Selon les BDD, il est possible que des psychiatres ou des hospitalisations psychiatriques soient respectivement renseignés en tant que spécialiste ou hospitalisation classique. Ceci entraîne une sous-estimation du nombre d'assurés demandant des remboursements pour des actes de psychiatrie, et peut potentiellement créer un biais dans les résultats. Un certain nombre d'hospitalisations ont été reclassées en hospitalisation psychiatrique en utilisant les montants de forfaits journaliers, mais, au vu des informations disponibles, aucun autre retraitement n'a pu être effectué.

Pour la suite, les **assurés ayant au moins une fois un acte de psychiatrie remboursé (AP)** seront souvent comparés aux **assurés qui n'ont pas de soins psychiatriques remboursés par la complémentaire santé (NAP)**.

Les statistiques descriptives montrent que les populations étudiées sont majoritairement actives, et que les BDD A et B sont bien équilibrées en fonction du sexe. Il est important de noter qu'aucune de ces bases n'est *a priori* représentative de la population française, mais le fait que les principaux résultats se retrouvent d'une BDD à l'autre laisse penser qu'ils pourraient être étendus à d'autres BDD.

Les AP représentent entre 2,5% [6] et 4 % des assurés présents dans les bases, avec une sur-

6. Moins d'AP sont identifiables dans la BDD C, à cause de la nature de surcomplémentaire des contrats.

représentation des femmes[7]. Ces chiffres sont du même ordre de grandeur que le nombre de personnes touchées par une maladie psychiatrique en France d'après la Sécurité Sociale (3,6 %, voir
Ministère des Affaires sociales de la Santé et des Droits des femmes (2014)), mais nettement inférieurs au nombre de personnes bénéficiant d'un traitement psychotrope chronique (8.9%) et au nombre d'individus concernés par une maladie psychiatrique tel qu'estimé par les médecins (au moins 20%, Wittchen et al. (2011)). Plusieurs effets peuvent causer ces forts décalages. Tout d'abord, un phénomène de non-recours aux soins peut expliquer une grande partie des décalages entre les évaluations de la Sécurité Sociale et celles des médecins. Ensuite, quatre raisons peuvent expliquer la différence entre les chiffres observés par les sociétés d'assurance et ceux de la Sécurité Sociale : d'une part, l'assureur n'a pas forcément connaissance de tous les soins, notamment ceux pris en charge au titre d'une ALD ou ceux ne pouvant être reliés à un acte psychiatrique (certaines hospitalisations psychiatriques peuvent être renseignées comme hospitalisation classique dans les BDD étudiées, et certaines dépressions sont soignées uniquement par des médecins généralistes). D'autre part, les complémentaires santé s'adressent essentiellement à des actifs, quand les personnes avec des pathologies psychiatriques sont plus souvent au chômage (c.f. Butterworth et al. (2012)). Enfin, les études de la Sécurité Sociale observent les individus sur cinq ans, alors que cette étude n'observe les assurés que sur un an.

Les DT par AP sont elles aussi nettement inférieures à celles observées par la Sécurité Sociale (6 600 euros), ce qui peut s'expliquer par le fait que les remboursements effectués dans le cadre d'une ALD peuvent être inconnus de l'assureur. En revanche, en se basant sur les données collectées par un assureur santé, il est notable qu'**un AP coûte plus de deux fois plus qu'un assuré moyen**, que ce soit en termes de DT observées par l'assureur, ou en termes de RC. Enfin, l'âge moyen des AP est systématiquement plus élevé que celui des assurés moyens. La Sécurité Sociale constate une forte augmentation de la prévalence des maladies psychiatriques pour les âges les plus élevés, ce qui est cohérent avec ce résultat.

3.3.2 Statistiques descriptives détaillant le risque psychiatrie

Cette section vise à analyser le poids financier de la psychiatrie pour une société d'assurance, via un certain nombre de statistiques descriptives. Sur les BDD étudiées, seule la base A possédait à la fois l'exposition, l'âge et le sexe des assurés ainsi que la totalité des prestations des assureurs. Cependant, les résultats varient très peu d'une BDD à l'autre. Aussi, afin de faciliter la lecture, seuls les résultats obtenus sur la base A seront exposés ici.

Le tableau 33 met en évidence le coût élevé des AP en termes de RC moyen, et ce pour chaque assureur. Cette différence pourrait être entièrement due aux coûts des soins psychiatriques.

7. Cette surreprésentation est observée sur toutes les BDD, en cohérence avec la littérature médicale, bien que cette surreprésentation soit ici peu marquée.

Or, la Figure 34 montre que même si l'on corrige les dépenses totales des AP en retirant les majorations, les soins de spécialistes et les soins psychiatriques, ceux-ci coûtent toujours plus cher que les NAP [8].

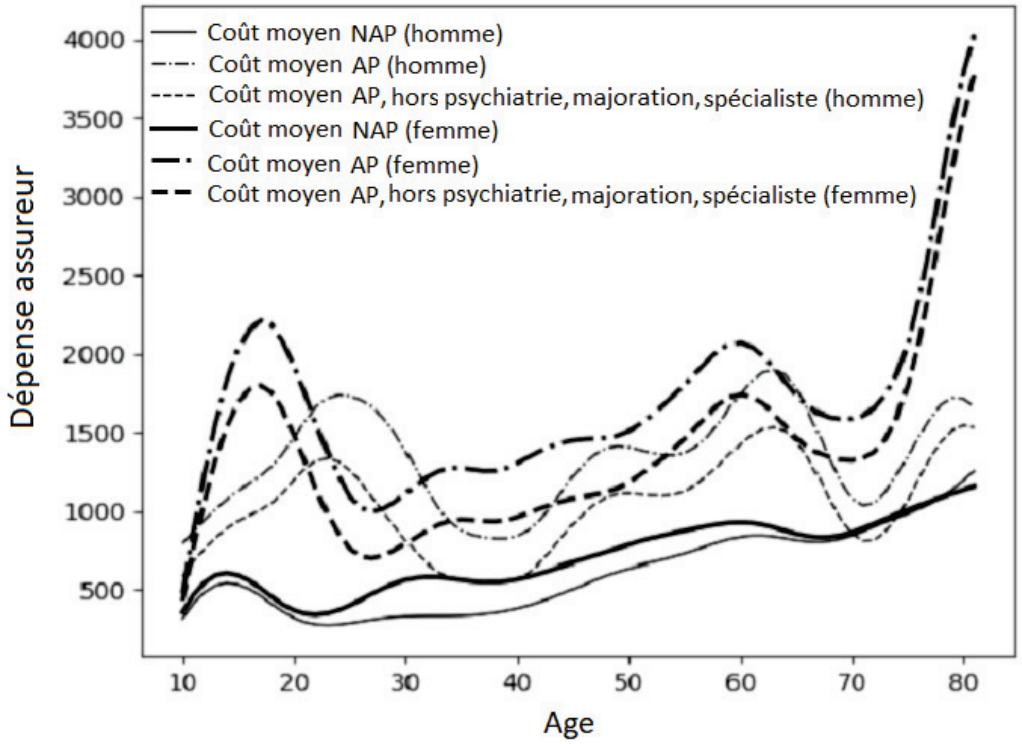

Figure 34 – Courbe des RC annuels moyens pour les hommes (traits fins) et les femmes (traits épais) en fonction de l'âge, pour les NAP, les AP et les AP corrigés des dépenses en psychiatrie, en spécialistes et majorations. Les courbes ont été lissées avec la méthode de Whittaker-Henderson.

Outre le fait que les AP, même corrigés de certains postes de dépenses majeurs, coûtent toujours plus cher que les NAP, il est possible de faire plusieurs constats à partir de ce graphe. Tout d'abord, il apparaît que le risque psychiatrique concerne l'ensemble des âges, bien qu'il soit plus important pour les adolescents / jeunes adultes et pour les séniors.

Ensuite, une hausse des coûts apparaît pour les femmes AP de 15-20 ans, et pour les hommes AP de 20-30 ans. Cette bosse chez les femmes semble très bien s'expliquer par une augmentation importante des tentatives de suicide à cet âge, et donc des coûts d'hospitalisation associés (voir Institut de veille sanitaire (2014)). En revanche, cela ne peut pas suffire à expliquer la hausse sur la courbe des hommes. Celle-ci coïncide avec la crise du quart de vie [9], ce

8. Les effectifs par âge peuvent être trouvés en Annexe
9. La crise du quart de vie est un phénomène psychologique apparaissant entre 20 et 30 ans, lié notamment à l'entrée dans la vie active (c.f. Robbins and Wilner (2001))

qui pourrait être une explication. Une autre explication possible pourrait être le fait que les maladies psychiatriques sont diagnostiquées plus tard chez les hommes que chez les femmes.

Il est aussi possible d'observer une baisse des dépenses peu après 60 ans, tant chez les femmes que chez les hommes. Les départs en retraite sont une des causes possibles. D'une part, le départ à la retraite est souvent accompagné d'une hausse du bien-être psychique et pourrait donc être lié à une baisse à court terme des dépressions (c.f. Kim and Moen (2002)). D'autre part, le passage à la retraite peut entraîner une diminution des revenus, contraignant parfois les assurés à changer de contrat d'assurance, ce qui peut provoquer une modification de la démographie du portefeuille et pourrait expliquer ce phénomène. Cependant, la baisse observée s'estompe très vite. Cette baisse des dépenses coïncide aussi avec la baisse des hospitalisations psychiatriques observée lors d'une étude réalisée par la région Pays de la Loire (p.221, ORS Pays de la Loire (2017)). Ses auteurs évoquent la possibilité qu'une évolution du parcours de soins ou qu'une évolution naturelle des maladies psychiatriques puissent causer cette diminution.

Les courbes observées pour la BDD B et les hommes de la BDD C sont similaires. La courbe obtenue pour les femmes de la BDD C ne montre pas de bosse pour les jeunes femmes, mais présente bien la baisse des RC pour les femmes de 60 ans. Enfin, la courbe obtenue pour la BDD D est beaucoup plus lisse, ce qui est certainement dû au fait que cette BDD ne contienne pas d'information sur le sexe des adhérents.

La Figure 35 permet d'observer que les écarts constatés sur la Figure 34 sont répartis sur tous les postes de soins, avec un effet majeur de l'hospitalisation [10].

Une partie de l'écart des dépenses d'hospitalisation peut être expliquée par la prise en charge des tentatives de suicide. Une autre partie peut découler de comorbidités entre les maladies psychiatriques et d'autres pathologies, comme les pathologies cardiaques.

Ces résultats rejoignent ceux de la Sécurité Sociale, pour laquelle la majorité des coûts des AP est issue des hospitalisations, principalement psychiatriques. Pourtant, dans toutes les BDD étudiées, les séjours en établissement psychiatrique sont peu présents (même après reclassement de certaines hospitalisations standards en hospitalisations psychiatriques en utilisant les montants de forfaits journaliers). Cela peut provenir d'une mauvaise qualité générale des données fournies par les intermédiaires médicaux ou des regroupements effectués par les sociétés d'assurance.

Si la psychiatrie est un risque important, il semble pertinent de le comparer aux autres risques majeurs auxquels sont exposées les sociétés d'assurance santé.

10. Seule la BDD B présente des RC en psychiatrie au même niveau que les RC d'hospitalisation. Les trois autres BDD présentent une nette prédominance des dépenses en hospitalisation. Cela peut être dû au fait que la BDD B identifie mieux les hospitalisations psychiatriques.

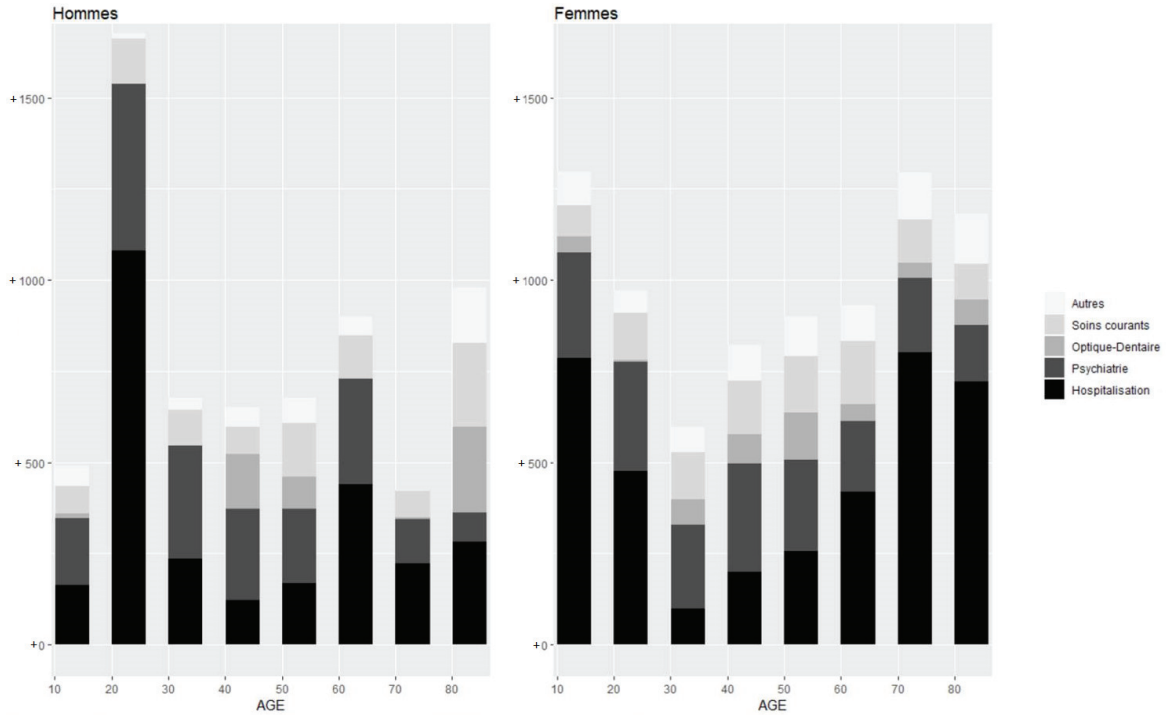

Figure 35 – Écart de RC entre AP et NAP, par groupes d'actes et par âge

La Figure 36 montre, pour divers actes santé (optique, kinésithérapie, ...), **la différence de dépense moyenne par individu** entre, d'une part, les personnes actives qui demandent au moins une fois un remboursement pour ces actes et, d'autre part, ceux qui ne le font pas, en termes de RAC, RO et RC [11].

Les AP constitue la deuxième population la plus coûteuse, juste derrière les utilisateurs de prothèses dentaires [12], mais devant les assurés demandant des remboursements en hospitalisation, en optique et en kinésithérapie, qui sont pourtant des risques beaucoup mieux identifiés en pratique. On peut aussi remarquer que la population qui a recours à des soins en psychiatrie coûte individuellement beaucoup plus cher que celle qui a recours à des soins de spécialistes [13], ce qui souligne la spécificité de la psychiatrie par rapport aux autres maladies.

En revanche, les AP font face à relativement peu de RAC. Ce chiffre est cependant biaisé, car une partie du RAC n'est pas connue des compagnies d'assurance et n'est pas présente dans ces chiffres. Notamment, les consultations des psychologues peuvent être exclues des contrats, et lorsqu'elles sont comprises dans les garanties, seules quelques séances sont remboursées. De

11. Les résultats obtenus pour la psychiatrie sur les actifs et les retraités sont similaires.
12. Quelle que soit la BDD étudiée, la psychiatrie présente la deuxième plus grande différence de RC. En revanche, en fonction des BDD, la première place est soit occupée par les prothèses dentaires, soit par l'hospitalisation. Les écarts de niveau de garantie entre les BDD peuvent expliquer cette différence.
13. Les psychiatres n'ont pas été considérés comme des spécialistes pour la réalisation de ces statistiques.

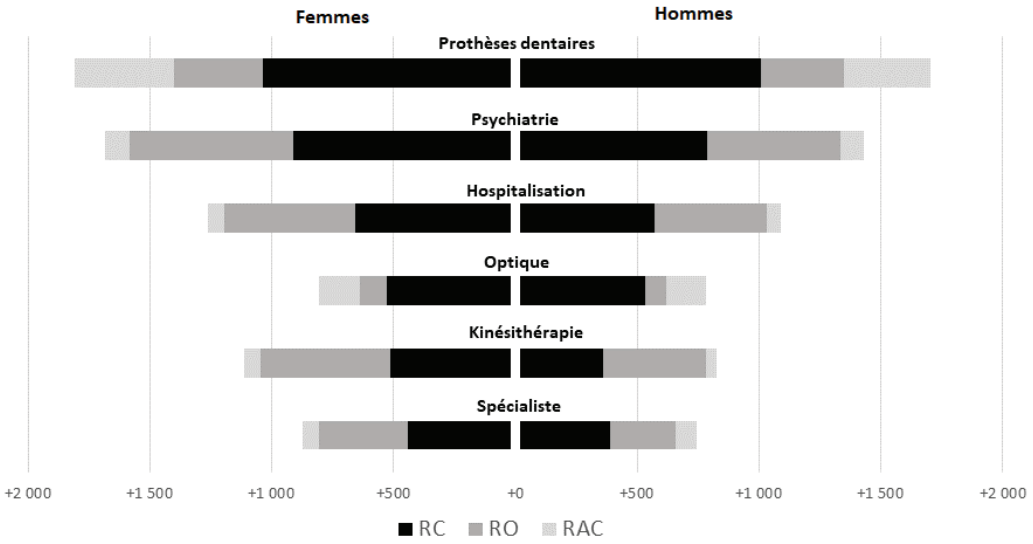

Figure 36 – **Différences de dépenses** entre les personnes actives bénéficiant d'au moins un remboursement pour certains actes santé particuliers et celles n'en bénéficiant pas. Les femmes sont sur la partie gauche du graphe et les hommes sur la partie droite.

même, la prise en charge de certaines garanties, comme les chambres particulières, est limitée par un quota de jours, qui est facilement atteint en psychiatrie car les séjours en hôpital sont dans ce cas plus longs.

Les résultats présentés dans cette partie ne tiennent pas compte des différences de caractéristiques observées entre les populations AP et NAP (la population AP est plus âgée par exemple). La Section 3.4 propose des modèles GLM permettant de contrôler ces différences et de mieux estimer l'effet marginal de la psychiatrie.

3.4 Modèles linéaires et probit

Afin de souligner la corrélation entre le remboursement d'actes de psychiatrie et celui de la quasi-totalité des autres grands postes de dépenses, une série de modèles économétriques a été réalisée. **La population AP n'étant pas représentative de la population générale, ces modèles permettent d'isoler les effets du sexe et de l'âge de celui de la psychiatrie.**

Ces modèles s'intéressent aux corrélations entre la psychiatrie et différents groupes d'actes. Quatre grands groupes d'actes ont été constitués : l'hospitalisation (y compris chirurgie et maternité, mais hors hospitalisation psychiatrique), l'optique et le dentaire, les soins courants (pharmacie, généraliste, spécialiste et analyses), et les autres dépenses (y compris kinésithérapie). Douze groupes d'actes plus fins ont aussi été étudiés afin d'augmenter le niveau de détail des résultats (les modèles les concernant sont exposés dans l'annexe 3.6).

Pour chaque groupe d'actes, deux modèles ont été estimés. Un premier modèle linéaire s'intéresse au RC individuel dû uniquement à ce groupe d'actes. Afin d'estimer les modèles, les assurés n'ayant reçu aucun remboursement pour ce groupe d'acte ont été retirés : il s'agit donc d'un modèle estimant le RC sachant que celui-ci est non nul. Un modèle portant sur le RC total tous groupes d'actes confondus a aussi été estimé. Les résultats sont présentés dans la partie 3.4.1.

Afin de compléter cette approche s'intéressant au montant de RC, la partie 3.4.2 présente des modèles probit s'intéressant à la probabilité qu'un individu ait un RC non nul pour chaque groupe d'acte.

Pour chaque modèle, l'âge a été segmenté en tranches de 20 ans, afin de prendre en compte le fait que l'évolution des dépenses en fonction de l'âge n'est pas linéaire. La variable "PSY" vaut 1 si l'assuré est un AP, 0 sinon.

3.4.1 Modèles sur les montants de RC

Dans cette section sont présentés les modèles portant sur les RC par individu pour chaque groupe d'actes. Les résultats sont présentés dans le tableau 37.

	RC Total	RC Hospitalisation	RC Optique dentaire	RC Soins courants	RC Autres
Référence	778 ***	291 ***	544 ***	190 ***	155 ***
AGE (+ 60 ans)	160 ***	156 ***	-94 ***	74 ***	64 ***
AGE (10-20 ans)	-281 ***	-93 **	-101 ***	-92 ***	-68 ***
AGE (20-40 ans)	-281 ***	32	-225 ***	-45 ***	-44 ***
SEXE	-142 ***	-29	-14 .	-56 ***	-28 ***
PSY	**855 *****	**492 *****	**94 *****	**142 *****	**86 *****
AGE (+60 ans)* SEXE	98 ***	32	20	33 ***	8
AGE (10-20 ans)* SEXE	77 ***	-29	-6	38 ***	31 ***
AGE (20-40 ans) * SEXE	-23	-37	-20	-6 *	13 **
AGE (>60 ans) * PSY	164 ***	297 ***	-78 *	-1	15
AGE (10-20 ans) * PSY	142 **	998 ***	-70	-41 ***	1
AGE (20-40 ans)* PSY	28	284 ***	-81 *	-8	-30 **
SEXE * PSY	-177 ***	-165 **	8	-29 ***	-52 ***

Note : . p < 0,1; * p < 0,05; ** p < 0,01; *** p < 0,001

Référence : Femme NAP de 40 à 60 ans

Figure 37 – Etude des RC des différents groupes d'actes

La psychiatrie est bien corrélée avec d'autres risques puisque **la variable PSY est significative pour chacun des 5 modèles** : la différence de RC observée entre les AP et les NAP n'est pas seulement due au décalage d'âge entre ces deux populations. Un AP coûte ainsi 855 euros de plus qu'un NAP à l'assureur, soit une augmentation de près de 110%. Cette hausse est très

marquée pour l'hospitalisation : un AP qui consomme de l'hospitalisation coûtera 492 euros de plus[14] qu'un NAP qui consomme de l'hospitalisation.

Comme expliqué précédemment, afin de détailler les résultats, des modèles ont aussi été réalisés en effectuant un regroupement plus fin des actes santé (voir annexe 3.6). Ces modèles laissent apparaître que **la psychiatrie est corrélée significativement (à 95%)** à la quasi-totalité de ces groupes de dépenses, à l'exception de la chirurgie (la psychiatrie n'est significative que pour deux BDD sur quatre) et des soins dentaires (la psychiatrie n'est jamais significative). Le fait que la variable "PSY" soit significative dans les modèles expliquant les RC en radiologie, en analyse ou en kinésithérapie peut être dû aux comorbidités des maladies psychiatriques avec d'autres pathologies. Comme espéré, la variable "PSY" est aussi corrélée avec les dépenses en prothèses dentaires, ce qui est conforme aux résultats de la littérature médicale sur le sujet (e.g. Denis and Coquaz (2013), Kisely (2016)), qui observe que les AP vont moins souvent chez le dentiste et ont, en conséquence, une moins bonne hygiène dentaire.

Les résultats sur la significativité de la psychiatrie sont valables pour toutes les BDD, y compris pour la BDD B qui permet de contrôler la région d'habitation, ce qui souligne la robustesse des résultats et leur faible sensibilité au niveau de garanties des contrats santé. De plus, les résultats restent vrais si les modèles portent sur les DT à la place des RC. De même, les significativités des âges se généralisent à l'ensemble des BDD, à l'exception de la significativité de la tranche 10-20 ans en hospitalisation, qui n'est observée que pour les BDD A et D. Ce n'est pas le cas de la significativité du sexe, qui varie selon les BDD.

3.4.2 Modèle sur la probabilité d'avoir eu au moins un remboursement

Afin de compléter la partie précédente, quatre modèles probit[15] portant sur la probabilité de bénéficier d'un remboursement ont été réalisés. Le modèle sur la fréquence de consommation globale n'a pas été effectué, puisque la totalité des individus considérés dans cette étude a au moins bénéficié d'un remboursement dans la période étudiée.

La psychiatrie est bien corrélée positivement et significativement à la probabilité d'avoir bénéficié d'au moins un remboursement pour chacun des grands types d'actes. Contrairement aux modèles étudiant les montants remboursés, les variables de type AGE * PSY sont peu significatives.

De même que précédemment, 12 autres modèles plus fins ont aussi été réalisés (voir en annexe 3.6). La variable PSY est significative dans la totalité des modèles réalisés (y compris pour les soins dentaires), et ce pour la totalité des BDD. Cela va à l'encontre de la littérature pour les soins dentaires hors prothèses (e.g. Valtat (2005), McCreadie et al. (2004)). En moyenne,

14. En dépense complémentaire d'hospitalisation uniquement.
15. Les modèles logit ont aussi été testé, sans différence notable dans les résultats, et ne sont donc pas présentés.

	RC Hospitalisation	RC Optique dentaire	RC Soins courants	RC Autres
Référence	0,22 ***	0,74 ***	0,95 ***	0,83 ***
AGE (+ 60 ans)	0,14 ***	-0,02 **	0,03 ***	0,05 ***
AGE (10-20 ans)	-0,06 ***	-0,05 ***	0,00	-0,09 ***
AGE (20-40 ans)	0,04 ***	-0,14 ***	0,00 .	-0,06 ***
SEXE	0,00	-0,06 ***	0,00	-0,11 ***
PSY	0,16 ***	0,06 ***	0,05 ***	0,13 ***
AGE (+60 ans)* SEXE	0,03 ***	0,02 **	0,00	0,04 ***
AGE (10-20 ans)* SEXE	0,04 ***	0,01	-0,01 **	0,06 ***
AGE (20-40 ans) * SEXE	-0,09 ***	0,01 .	-0,03 ***	-0,02 **
AGE (>60 ans) * PSY	-0,02	0,03	-0,04	-0,09 **
AGE (10-20 ans) * PSY	0,00	-0,03	-0,12 ***	0,00
AGE (20-40 ans)* PSY	-0,01	0,01	-0,09 ***	0,02
SEXE * PSY	0,01	-0,05 **	-0,01	0,00

Note : . p < 0,1; * p < 0,05; ** p < 0,01; *** p < 0,001

Référence : Femme NAP de 40 à 60 ans

Figure 38 – Etude des probabilités d'avoir bénéficié d'un remboursement - Effets marginaux des variables

les personnes souffrant de maladies psychiatriques vont moins souvent chez le dentiste (ce qui explique qu'elles ont besoin de plus de prothèses dentaires). Ce résultat contre-intuitif pourrait être dû au biais de constitution des BDD assurantielles : les troubles psychiques les plus forts, comme la schizophrénie, entraînent souvent une invalidité. Les personnes qui en souffrent sont donc peu présentes dans les BDD étudiées. Ce n'est en revanche pas le cas des dépressions. Ce résultat pourrait donc être dû à une différence entre les AP étudiés et la population générale souffrant de ces pathologies.

Ainsi, au fait que les remboursements liés à chaque groupe d'actes soient plus élevés pour les AP, s'ajoute le fait que la probabilité qu'un AP ait besoin de ces remboursements est elle aussi plus élevée.

3.5 Apports des bases de données libres

Comme précisé plus haut, les données possédées par des assureurs sont incomplètes. L'absence des informations concernant les ALDs empêche par exemple de mener un certain nombre d'analyses. Afin de compléter les résultats ci-dessus obtenus sur environ 500 000 individus uniques, deux bases de données mises à disposition par le gouvernement français ont été exploitées : la base Open MEDIC et la base Open DAMIR.

3.5.1 La base Open MEDIC

La base de données Open MEDIC regroupe l'ensemble des médicaments vendus en pharmacie de ville entre les années 2014 et 2018. Les médicaments fournis dans un cadre hospitalier sont donc exclus. En revanche, les médicaments prescrits par des praticiens en hôpital, mais achetés en dehors de l'hôpital sont présents. De plus, la base ne distingue pas les praticiens en hôpital psychiatrique des praticiens en hôpital standard.

Cette base contient des informations précises sur le médicament (nom du médicament, type de médicament, dosage), le prescripteur du médicament, ainsi que le montant remboursé, la base de remboursement, le nombre de boîtes délivrées et le nombre de consommateurs. Ces informations permettent donc d'identifier les dépenses propres aux psychotropes. Quelques informations sur l'assuré qui achète la boîte sont aussi disponibles, à savoir l'âge (par intervalle de 10 ans), le sexe et la région d'habitation. Il n'est en revanche pas possible avec cette base de données de déterminer le nombre de boîtes achetées au titre d'une ALD. De plus, l'information sur un éventuel remboursement supplémentaire provenant d'un organisme de complémentaire santé n'est pas disponible. Nous parlerons donc de Reste A Charge Avant Complémentaire (RACAC).

La base de données propre à l'année 2018 a été utilisée pour cette étude. Cela représente au total environ 22,9 milliards d'euros de dépense dont environ 19,3 milliards ont été remboursés par la Sécurité Sociale (le RACAC total est donc d'environ 3,6 milliards d'euros, soit 15,8% des dépenses).

Résultats

Au total, 92% des médicaments prescrits par des psychiatres sont des psychoanaleptiques ou psycholeptiques[16]. La majorité des autres médicaments prescrits par les psychiatres sont principalement pour lutter contre les effets secondaires, comme des antihistaminiques ou des anti-acides. Les prescriptions des psychiatres représentent environ 170 millions d'euros, dont un RACAC représentant 17,2% des dépenses. Il est de plus intéressant de noter que les pédiatres prescrivent peu de boîtes de psychotropes (0,03% du nombre de boîtes total), mais que les boîtes qu'ils prescrivent coûtent très cher (un coût moyen par boîte de 17,4 euros et un RACAC par boîte de 5,2 euros, contre 3,8 et 0,9 euros pour les généralistes, et 2,3 et 0,7 euros pour les psychiatres).

Les psychotropes représentent en 2018 une dépense totale de 1 milliard d'euros, dont 18,2% de RACAC avant complémentaire. 49,7% de cette dépense totale est due à des généralistes, 34,4% a des praticiens hospitaliers, et 13,3% ont des psychiatres. Cela confirme l'importance de la place des généralistes dans le soin des maladies psychiatriques. En termes de volume, ce sont près de 69,3% des boîtes de psychotropes qui sont prescrites par des généralistes, contre

16. Les psychoanaleptiques sont des excitants psychiques et les psycholeptiques sont des sédatifs psychiques

17,4% par des hôpitaux et 11,2% par des psychiatres. En revanche, les psychotropes prescrits par des généralistes sont moins bien remboursés, puisqu'ils présentent un taux de RACAC de près de 24,5%, contre 16,8% pour ceux prescrits par un spécialiste et 0,10% pour ceux prescrits par des practiciencs hospitaliers.

En conclusion, l'étude de la base de données Open MEDIC permet de confirmer que les généralistes jouent un rôle important dans le traitement des maladies psychiatriques, ce qui n'est pas mesurable dans les bases en provenance d'assurance. L'étude permet aussi de confirmer que les restes à charge avant complémentaire dus aux médicaments psychoanaleptiques ou psycholeptiques représentaient en 2018 près de 200 millions d'euros, soit une charge non négligeable pour les complémentaires santé. La base de données ne permet en revanche pas de mesurer le coût des comorbidités liées à la psychiatrie, puisque les psychiatres prescrivent essentiellement des psychotropes, redirigeant certainement les patients vers d'autres spécialistes pour le traitement des comorbidités.

3.5.2 La base Open DAMIR

Le jeu de données Open DAMIR est une extraction du Système National Inter Régimes d'Assurance Maladie (SNIIRAM). Il regroupe l'ensemble des remboursements agrégés effectués par l'Assurance Maladie, de 2009 à nos jours, ventilés selon plusieurs axes d'analyses :

- Analyse des prestations : Contient notamment des variables portant sur la nature de la prestation remboursée, le motif d'exonération du ticket modérateur le cas échéant, le montant de la dépense, le montant remboursé et la quantité de remboursement.
- Analyse des bénéficiaires : Contient notamment des variables sur le sexe, la tranche d'âge, la région d'habitation des bénéficiaires d'une prestation donnée.
- Analyse de l'exécutant de la prestation : Contient notamment la spécialité, le statut juridique et la région d'exercice de l'exécutant.
- Analyse du prescripteur de la prestation : Contient notamment la spécialité, le statut juridique et la région d'exercice du prescripteur.

Les dépenses hospitalières liées à cette base de données sont très incomplètes [17], puisque Open DAMIR ne regroupe que les données propres aux cliniques privées, et les prestations faisant l'objet d'une facturation directe à l'Assurance Maladie (comme les prestations au titre de la CMU). Les montants obtenus, et notamment les RACAC, ne tiennent donc presque pas compte des dépenses d'hospitalisation, pourtant principal poste de dépense en matière de psychiatrie. Les lignes étant agrégées, il n'est pas possible de remonter à un individu ni d'obtenir des informations sur les comorbidités. Il est possible de savoir si les prestations sont liées à une ALD, sans savoir la nature de l'ALD associée. La base de données complète étant volumineuse, il est possible de la télécharger mois par mois sur data.gouv.fr. Par exemple,

17. Une analyse complète sur l'hospitalisation aurait nécessité un accès aux bases de données du PMSI

la base de données du mois de décembre 2018 compte environ 31 350 000 lignes, pour 10,6 milliards d'euros de remboursement.

Pour mener l'étude présentée ici, les 12 bases propres à l'année 2018 ont été utilisées. Les lignes vérifiant au moins l'une des propriétés suivantes ont été conservées pour l'analyse :

— Le prescripteur est un psychiatre
— Le réalisateur est un psychiatre
— L'établissement exécutant est un hôpital psychiatrique
— L'établissement prescripteur est un hôpital psychiatrique
— La nature de la prestation est en lien avec la psychiatrie/psychologie [18].

De plus, l'ensemble des lignes avec des montants de remboursements négatifs ont été supprimées. L'ensemble des lignes propres à des indemnités journalières a aussi été supprimé, afin de se concentrer uniquement sur les dépenses de santé (les informations disponibles sur les indemnités journalières ne sont pas assez complètes pour permettre de mener une analyse sur la prévoyance). Enfin, il n'y a pas d'information sur les complémentaires santé, le reste à charge avant complémentaire santé (RACAC) sera donc utilisé.

Ces prétraitements permettent de retrouver environ 188 millions d'actes, soit 2,2 milliards de dépenses pour l'année 2018, dont 1,7 milliard de remboursé par la Sécurité Sociale. Ces chiffres sont loin de ceux évoqués dans les analyses de la Sécurité Sociale, mais cela peut s'expliquer par l'absence de la majorité des données d'hospitalisation. À cet effet majeur s'ajoutent l'impossibilité de retrouver les comorbidités et la non-prise en compte des indemnités journalières.

Résultats

Environ 55,5% des dépenses peuvent être rattachées à des femmes, ce qui est cohérent avec les résultats observés en Figure 33. Sur les 2,2 milliards de dépenses dans la base reconstituée, environ 1,3 milliard (soit 57%) proviennent de prescriptions de psychiatres. Les dépenses associées sont principalement liées à de la pharmacie (390 millions d'euros, dont 60 millions de RACAC) et les consultations de médecins (690 millions d'euros, dont 250 millions de RACAC). Parmi les autres prestations, il est remarquable de voir que 300 000 séances d'orthophonie ont été prescrites par des psychiatres, pour une dépense totale de 10 millions d'euros. C'est par exemple le même ordre de grandeur que les actes de biologie prescrivent par des psychiatres (14 millions d'euros).

Il est aussi possible de faire des études sur les ALDs. Sur les 188 millions d'actes, 58% le sont au titre d'ALD (109 millions). Cela représente 47% de la dépense totale (un milliard

18. Les prestations suivantes ont été conservées : "Avis ponctuel de consultant psychiatre", "Consultation de psychiatre cotée C2,5", "Avis ponctuel de consultant psychiatre (visite)", "Différentiel psy réglementaire", "Forfait psychiatre de sécurité : hospitalisation avec hébergement", "Forfait psychiatrie séance", "Entretien évaluation psychologue", "Accompagnment psychologique de soutien", "Psychotérapie structurée".

d'euros environ), mais seulement 20% du RACAC total (110 millions d'euros). Les actes pris en charge au titre d'une ALD sont donc très bien remboursés, mais il est faux de dire que les ALDs ne coutent rien aux assureurs santé. Les prestations versées aux bénéficiaires d'une ALD au titre de cette ALD représentent la majorité de cette somme (environ 960 millions d'euros, pour 66 millions de RACAC).

Il est enfin possible d'analyser le poids de la couverture maladie universelle (CMUC) dans ces dépenses, ce qui n'est pas possible avec des données issues d'une complémentaire santé. Pour cette analyse, 6 millions d'actes, soit 173 millions d'euros de dépense, ont été supprimés, car il n'était pas spécifié s'ils étaient effectués au titre de la CMUC ou non. Au total, 11,2% des actes sont versés au titre de la CMUC. Cela représente une dépense totale de 280 millions d'euros (soit 13,5% des dépenses). Le RACAC associé est de 120 millions d'euros, soit un taux de RACAC de 44,1%. Les personnes bénéficiant de la CMUC sont donc moins bien remboursées vis-à-vis des dépenses psychiatriques considérées.

3.6 Conclusion

L'étude présentée dans cet article s'intéresse au coût de la psychiatrie pour une complémentaire santé en France. Ces travaux ont porté sur quatre bases de données observées sur un ou deux ans. Ils permettent de conclure que les assurés, qui ont recours au moins une fois à des soins psychiatriques, coûtent en moyenne deux fois plus cher qu'un assuré moyen, et ce malgré le potentiel statut d'ALD. Ils constituent ainsi l'une des catégories d'assurés demandant le plus de remboursement, devant ceux qui ont recours à la kinésithérapie ou à l'optique. Ce coût est essentiellement dû aux frais d'hospitalisation et aux soins psychiatriques, mais la comorbidité des pathologies psychiatriques avec les diverses pathologies somatiques entraîne aussi une hausse des coûts globaux. De plus, la psychiatrie est corrélée à la fois aux montants de dépenses et à la fréquence de consommation des assurés, pour la quasi-totalité des grands actes santé. Les adolescents / jeunes adultes et les séniors sont les assurés qui ont le plus recours à la psychiatrie. Des analyses complémentaires sur des bases de la Sécurité Sociale Open DAMIR et Open MEDIC confirment un important montant de reste à charge concernant les soins psychiatriques. Si le statut d'ALD semble bien réduire les dépenses à la charge des complémentaires santé, celles-ci restent non négligeables. Cela va à l'encontre des croyances d'une partie de la communauté actuarielle en santé qui considère que les ALD ne concernent pas ou peu les assurances.

L'une des limites de l'étude concerne la qualité des données. Il est probable que certains actes santé soient mal renseignés, et que certains assurés souffrant de maladies psychiatriques ne soient pas identifiés comme tels, entraînant une sous-évaluation du nombre de personnes concernées. De plus, les garanties exactes des contrats ne sont pas connues et ne sont donc pas contrôlées. Cependant, la stabilité des résultats obtenus pour les quatre bases de données (qui

proviennent toutes d'assureurs différents proposant des garanties différentes) laisse à penser que le contrôle des garanties n'aurait pas changé la significativité de la psychiatrie dans les modèles.

Le périmètre de l'étude ne concernait que les dépenses de santé. Une possible extension de ces travaux serait d'élargir ce périmètre en couplant des bases de données issues de l'assurance santé avec des bases de données issues de l'assurance prévoyance ou de l'assurance vie, afin d'obtenir le coût total de ces assurés pour les sociétés d'assurance. En effet, les personnes souffrant de maladies psychiatriques sont associées à une surmortalité importante ainsi qu'à des arrêts de travail plus longs (c.f. Assurance Maladie (2018a)).

Le faible nombre d'assurés identifié comme ayant recours à la psychiatrie (entre 2,5 et 4% des portefeuilles) et leur facilité d'identification ouvrent potentiellement la possibilité de mener des actions de prévention ciblées et efficaces sur ce terrain. Le déploiement d'aides personnalisées via des applications spécialisées comme iSMART ou Woebot®, de campagnes d'informations des assurés sur ces risques mal connus ou encore un accompagnement financier via un meilleur remboursement des séances de psychothérapie pourraient à la fois apporter une aide aux personnes concernées et assurer une maîtrise des coûts associés. Il pourrait aussi être envisagé des actions de prévention coordonnées entre l'employeur et la complémentaire santé. Par exemple, il pourrait être pertinent de faire venir un psychiatre sur un lieu de travail afin, dans un premier temps, d'animer une formation sensibilisant les employés sur les maladies psychiatriques, puis dans un second temps, de proposer des diagnostics "bien être" permettant de dépister les maladies psychiatriques.

Cependant, les données utilisées et le faible historique disponible pour l'étude restreignent celle-ci à l'étude des corrélations entre la psychiatrie et une augmentation des dépenses sur la majorité des autres postes de soins. Des études complémentaires sont nécessaires afin d'étudier les causalités et l'état de santé précis des assurés bénificiant d'un remboursement de psychiatrie. Cela permettrait ainsi d'estimer précisément l'impact potentiel de la mise en oeuvre d'une action de prévention.

Remerciements : Cet article a été fait dans le cadre de la Chaire Prevent'Horizon, soutenue par la fondation du risque Louis Bachelier et en partenariat avec l'université Claude Bernard Lyon 1, ADDACTIS France, AG2R la mondiale, G2S, Covéa, Groupama Gan vie, Groupe Pasteur Mutualité, Harmonie mutuelle, Humanis prévoyance et La Mutuelle Générale. Les auteurs voudraient aussi remercier les différents relecteurs, ainsi qu'Alexandra Dima, Luiza Prado et Sofia El Oussoul pour leur très précieux point de vue médical.

Annexe 1 : Effectifs de la base A détaillés par âge

	Moins de 20 ans	20-30 ans	30-40 ans	40-50 ans	50-60 ans	60-70 ans	70-80 ans	Plus de 80 ans
Effectif	31500	10000	15100	17200	20700	11400	4700	2400
Dont femmes	15400	5200	7200	8200	9300	5000	2200	1300
Dont AP	700	300	600	800	900	400	200	100
Dont femmes AP	300	200	400	500	500	200	100	50

Figure 39 – Effectifs détaillés par âge - BDD A.

Annexe 2 : Modèles détaillés par grands postes de dépense

	RC Analyse	RC Autres	RC Chirurgie	RC Soins dentaires	RC Généraliste	RC Pharmacie	RC Spécialiste	RC Hospitalisation	RC Radiologie	RC Kinésithérapie	RC Prothèses dentaires	RC Optique
Référence	44 ***	99 ***	181 ***	60 ***	47 ***	86 ***	64 ***	277 ***	54 ***	157 ***	886 ***	493 ***
AGE (+ 60 ans)	-1	52 ***	24 *	-13 **	2 **	55 ***	8 ***	202 ***	7 ***	5	-127 ***	-109 ***
AGE (10-20 ans)	-6 ***	-18 ***	-42 *	298 ***	-18 ***	-41 ***	-25 ***	-100 *	-26 ***	-46 ***	-413 **	-176 ***
AGE (20-40 ans)	13 ***	-22 ***	10	-9 *	-11 ***	-33 ***	-9 ***	30	-2 *	-37 ***	-214 ***	-130 ***
SEXE	-6 ***	-2	-35 ***	-8 *	-12 ***	-17 ***	-20 ***	-9	-13 ***	-23 ***	21	8
PSY	**17 *****	**43 *****	**-21**	**4**	**27 *****	**66 *****	**25 *****	**663 *****	**19 *****	**24 ** **	**166 * **	**51 *****
AGE (+60 ans)* SEXE	5 ***	12 *	28 .	9	6 ***	16 ***	7 ***	13	4 **	0	-34	13
AGE (10-20 ans)* SEXE	2	7	0	5	8 ***	13 ***	17 ***	-50	11 ***	13 .	512 *	-17 .
AGE (20-40 ans) * SEXE	-11 ***	-4	-14	0	3 ***	6 *	2	-50	-4 **	23 ***	-114 .	-3
AGE (>60 ans) * PSY	0	-3	26	0	0	-2	0	359 ***	-1	15	-241 *	-32
AGE (10-20 ans) * PSY	-8 *	37 *	46	25	-9 ***	-16 *	-5	1006 ***	-15 ***	5	-533	-34
AGE (20-40 ans)* PSY	1	-17	21	-1	-1	-11 .	-2	295 **	-5 .	-13	-37	-34
SEXE * PSY	-7 ***	-30 **	-16	-2	-7 ***	-2	-4 .	-136 .	-4 .	-9	41	-32 .

Note : . p < 0,1; * p < 0,05; ** p < 0,01; *** p < 0,001

Référence : Femme NAP de 40 à 60 ans

Figure 40 – Montants de RC des différents groupes d'actes

	Analyse	Autres	Chirurgie	Soins dentaires	Généraliste	Pharmacie	Spécialiste	Hospitalisation	Radiologie	Kinésithérapie	Prothèses dentaires	Optique
Référence	0,56 ***	0,41 ***	0,15 ***	0,50	0,73 ***	0,87 ***	0,74 ***	0,14 ***	0,58 ***	0,22 ***	0,16 ***	0,47 ***
AGE (+ 60 ans)	0,13 ***	-0,01	0,11 ***	0,03 ***	0,08 ***	0,08 ***	0,04 ***	0,09 ***	0,04 ***	0,03 ***	0,04 ***	-0,09 ***
AGE (10-20 ans)	-0,27 ***	-0,07 ***	-0,07 ***	0,04 ***	0,02 ***	-0,04 ***	-0,12 ***	-0,02 ***	-0,17 ***	-0,11 ***	-0,15 ***	-0,13 ***
AGE (20-40 ans)	0,01 *	-0,05 ***	-0,06 ***	-0,12 ***	-0,04 ***	-0,01	-0,07 ***	0,08 ***	-0,17 ***	-0,07 ***	-0,10 ***	-0,11 ***
SEXE	-0,08 ***	-0,14 ***	0,00	-0,02 ***	0,02 ***	-0,03 ***	-0,17 ***	-0,01 *	-0,19 ***	-0,05 ***	0,00	-0,09 ***
PSY	0,13 ***	**0,39 *****	**0,09 *****	**0,10 *****	**0,15 *****	**0,10 *****	**0,13 *****	**0,13 *****	**0,13 *****	**0,13 *****	**0,02 ***	**0,05 ****
AGE (+60 ans)* SEXE	0,10 ***	0,07 ***	0,02 **	0,01	-0,04 ***	0,00	0,07 ***	0,02 ***	0,09 ***	-0,01	0,00	0,04 ***
AGE (10-20 ans)* SEXE	-0,07 ***	0,07 ***	0,03 ***	0,01	-0,04 ***	-0,01 .	0,07 ***	0,04 ***	0,15 ***	0,04 ***	-0,02	-0,02 .
AGE (20-40 ans) * SEXE	-0,19 ***	-0,06 ***	-0,01 *	0,01	-0,02 ***	-0,07 ***	-0,05 ***	-0,07 ***	0,06 ***	0,05 ***	0,01	0,00
AGE (>60 ans) * PSY	-0,07 *	-0,11 ***	-0,01	-0,03	-0,07 *	-0,03	-0,01	-0,01	0,00	0,01	0,00	0,06 *
AGE (10-20 ans) * PSY	0,07 **	-0,02	-0,02	-0,08 **	-0,08 **	-0,11 ***	-0,01	0,02	-0,06 *	-0,02	0,00	0,02
AGE (20-40 ans)* PSY	0,02	0,04 .	0,03 .	0,03	0,03	-0,03	0,05 **	-0,02 *	0,01	-0,01	-0,02	0,00
SEXE * PSY	-0,01	0,08 ***	0,01	-0,05 **	-0,04 *	-0,02	-0,05 **	0,00	-0,02	-0,04 **	0,00	-0,02

Note : . p < 0,1; * p < 0,05; ** p < 0,01; *** p < 0,001

Référence : Femme NAP de 40 à 60 ans

Figure 41 – Probabilités de remboursement pour les différents groupes d'actes

Bibliographie

Assurance Maladie (2018a). Améliorer la qualité du système de santé et maîtriser les dépenses–propositions de l'assurance maladie pour 2019. Paris : Caisse Nationale De L'assurance Maladie Des Travailleurs Salariés.

Assurance Maladie (2018b). Les affections psychiques liées au travail : éclairage sur la prise en charge actuelle par l'assurance maladie - risques professionnels.

Assurance Maladie (2019). Cartographie des dépenses médicales de santé.

Butterworth, P., Leach, L. S., Pirkis, J., and Kelaher, M. (2012). Poor mental health influences risk and duration of unemployment : a prospective study. Social psychiatry and psychiatric epidemiology, 47(6) :1013–1021.

Chevreul, K., Prigent, A., Bourmaud, A., Leboyer, M., and Durand-Zaleski, I. (2013). The cost of mental disorders in France. European Neuropsychopharmacology, 23(8) :879–886.

Cruwys, T., Dingle, G. A., Haslam, C., Haslam, S. A., Jetten, J., and Morton, T. A. (2013). Social group memberships protect against future depression, alleviate depression symptoms and prevent depression relapse. Social science & medicine, 98 :179–186.

Demailly, L. (2011). Sociologie des troubles mentaux. La Découverte.

Denis, F. and Coquaz, C. (2013). Santé buccale et psychiatrie. eisai ch-la charteuse.

Institut de veille sanitaire (2014). Hospitalisations et recours aux urgences pour tentative de suicide en France métropolitaine à partir du pmsi-mco 2004-2011 et d'Oscour® 2007-2011.

IPSOS and Klesia (2014). Perceptions et représentations des maladies mentales - rapport.

Kim, J. E. and Moen, P. (2002). Retirement transitions, gender, and psychological well-being : A life-course, ecological model. The Journals of Gerontology Series B : Psychological Sciences and Social Sciences, 57(3) :P212–P222.

Kisely, S. (2016). No mental health without oral health. The Canadian Journal of Psychiatry, 61(5) :277–282.

McCreadie, R., Stevens, H., Henderson, J., Hall, D., McCaul, R., Filik, R., Young, G., Sutch, G., Kanagaratnam, G., Perrington, S., et al. (2004). The dental health of people with schizophrenia. Acta Psychiatrica Scandinavica, 110(4) :306–310.

Ministère des Affaires sociales de la Santé et des Droits des femmes (2014). Etat des lieux du suicide en France.

Opinion Way and MGEN (2014). Enquête santé mentale.

ORS Pays de la Loire (2017). La santé des habitants des pays de la loire.

Robbins, A. and Wilner, A. (2001). Quarterlife crisis : The unique challenges of life in your twenties. Penguin.

Roelandt, J.-L., Caria, A., Defromont, L., Vandeborre, A., and Daumerie, N. (2010). Représentations sociales du «fou», du «malade mental» et du «dépressif» en population générale en france. L'encéphale, 36(3) :7–13.

Vaiva, G., Jehel, L., Cottencin, O., Ducrocq, F., Duchet, C., Omnes, C., Genest, P., Rouillon, F., and Roelandt, J.-L. (2008). Prévalence des troubles psychotraumatiques en France métropolitaine. L'Encéphale, 34(6) :577–583.

Valtat, M. (2005). État de santé bucco-dentaire des patients hospitalisés en psychiatrie. PhD thesis.

Warin, P. and Chauveaud, C. (2014). Le baromètre du renoncement aux soins dans le gard.

Wittchen, H.-U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., et al. (2011). The size and burden of mental disorders and other disorders of the brain in europe 2010. European neuropsychopharmacology, 21(9) :655–679.

3.7 Complément : coefficient obtenu pour la variable psychiatrie pour chacun des modèles

Ce complément a pour objectif de compléter la présentation des résultats effectués dans le Chapitre 3. Il présente deux tableaux : un pour les modèles linéaires et un pour les modèles probit. Dans chacun de ces tableaux, il est possible de trouvé la valeur de l'effet marginal de la psychiatrie, pour chacun des modèles détaillés (comme dans l'Annexe 2), et pour chacune des 4 bases de données, ainsi que le niveau de significativité de ce coefficient. Afin de permettre la comparaison, la valeur de l'intercept associé à chaque modèle est également présente entre parenthèse.

Coefficient de la psychiatrie pour les modèles de prédiction des montants de dépenses

	Base A		Base B		Base C		Base D	
	Coefficient	(Intercept)	Coefficient	(Intercept)	Coefficient	(Intercept)	Coefficient	(Intercept)
Général	855 ***	(778)	1289 ***	(1558)	183 ***	(200)	800 ***	(700)
Analyse	17 ***	(44)	20 ***	(27)	6 ***	(22)	15 ***	(40)
Chirurgie	-21	(181)	15 ***	(63)	2 ***	(9)	-23	(162)
Dentaire	4	(60)	9	(46)	0	(13)	4	(56)
Généraliste	27 ***	(47)	40 ***	(52)	7 ***	(16)	26 ***	(40)
Pharmacie	66 ***	(86)	NA	NA	30 ***	(39)	67 ***	(77)
Spécialiste	25 ***	(64)	21 ***	(63)	2 ***	(8)	26 ***	(55)
Hospitalisation	663 ***	(277)	369 ***	(446)	216 ***	(246)	610 ***	(272)
Radiologie	19 ***	(54)	23 ***	(52)	4 ***	(14)	18 ***	(48)
Kinésithérapie	24 **	(157)	72 ***	(148)	10 ***	(56)	25 ***	(146)
Prothèses dentaires	166 *	(886)	216 ***	(1088)	40 **	(222)	182 **	(898)
Optique	51 ***	(493)	51 ***	(639)	0	(69)	38 ***	(497)

Note : . $p < 0{,}1$; * $p < 0{,}05$; ** $p < 0{,}01$; *** $p < 0{,}001$

Référence : Femme NAP de 40 à 60 ans

FIGURE 42 – Coefficients obtenus dans chacun des modèles linéaires pour la variable psychiatrie, ainsi que la valeur obtenu pour l'intercept du modèle

Effet marginal de la psychiatrie pour les modèles de prédiction des probabilités de dépenses

	Base A		Base B		Base C		Base D	
	Coefficient	(Intercept)	Coefficient	(Intercept)	Coefficient	(Intercept)	Coefficient	(Intercept)
Analyse	0,13 ***	(0,56)	0,08 ***	(0,79)	0,08 ***	(0,78)	0,15 ***	(0,55)
Chirurgie	0,09 ***	(0,15)	0,08 ***	(0,82)	0,08 ***	(0,76)	0,03 ***	(0,14)
Dentaire	0,10 ***	(0,5)	0,04 ***	(0,78)	0,06 ***	(0,71)	0,08 ***	(0,41)
Généraliste	0,15 ***	(0,73)	0,05 ***	(0,9)	0,04 ***	(0,94)	0,07 ***	(0,86)
Pharmacie	0,10 ***	(0,87)			0,01 ***	(0,98)	0,07 ***	(0,9)
Spécialiste	0,13 ***	(0,74)	0,07 ***	(0,85)	0,07 ***	(0,84)	0,13 ***	(0,54)
Hospitalisation	0,13 ***	(0,14)	0,15 ***	(0,24)	0,14 ***	(0,23)	0,14 ***	(0,25)
Radiologie	0,13 ***	(0,58)	0,11 ***	(0,7)	0,07 ***	(0,83)	0,14 ***	(0,60)
Kinésithérapie	0,13 ***	(0,22)	0,13 ***	(0,25)	0,13 ***	(0,25)	0,10 ***	(0,18)
Prothèses dentaires	0,02 *	(0,16)	0,03 ***	(0,33)	0,02 **	(0,23)	0,03 ***	(0,14)
Optique	0,05 **	(0,47)	0,06 ***	(0,69)	0,06 ***	(0,64)	0,02 *	(0,29)

Note : . $p < 0,1$; * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Référence : Femme NAP de 40 à 60 ans

FIGURE 43 – Effets marginaux obtenus dans chacun des modèles probits pour la variable psychiatrie, ainsi que la valeur obtenu pour l'intercept du modèle

Chapitre 4

Optimal prevention strategies in the classical risk model

Ce chapitre reprend un article coécrit avec Stéphane Loisel, Jean-Louis Rullière et Julien Trufin et accepté à Insurance : Mathematics and Economics.

Abstract

In this paper, we propose and study a first risk model in which the insurer may invest into a prevention plan which decreases claim intensity. We determine the optimal prevention investment for different risk indicators. In particular, we show that the prevention amount minimizing the ruin probability maximizes the adjustment coefficient in the classical ruin model with prevention, as well as the expected dividends until ruin in the model with dividends. We also show that the optimal prevention strategy is different if one aims at maximizing the average surplus at a fixed time horizon. A sensitivity analysis is carried out. We also prove that our results can be extended to the case where prevention starts to work only after a minimum prevention level threshold.

Keywords: Ruin theory; Prevention; Optimal prevention strategy; Insurance.

4.1 Introduction

An insurance company has several levers to manage risk and to optimize its risk/return profile, including reinsurance, investment and dividends. Optimal reinsurance strategies have been studied by Hesselager (Hesselager, 1990), Højgaard et al. (Højgaard and Taksar, 1998) among others. Optimal investment strategies have been studied independently from optimal reinsurance, or simultaneously like in Schmidli (Schmidli, 2002). To better reflect the strategy of the insurer, a dividend barrier has been introduced by De Finetti (De Finetti, 1957). Optimal dividend strategy has become a classical topic in ruin theory, see Avanzi (Avanzi, 2009) for a survey. Another way to manage risk is to update the pricing thanks to credibility techniques (see for example Afonso et al. (Afonso et al., 2010)). In this paper, we propose a

new risk model that takes into account the possibility for the insurer to carry out prevention plans, and we determine optimal prevention strategies in ruin theory.

Prevention has become a key, trendy topic in insurance. In 1997, Discovery was the first insurance company to implement an important prevention plan. Since then, a growing number of insurers have launched their own programs (e.g. John Hancock or OscarHealth in the USA, Generali Vitality or AXA prevention in Europe). However, according to the OECD, prevention spendings only represent 3% of the overall health expenditures in the world. Moreover, this financial effort is mainly made by public authorities (e.g. 94% of preventive spending is due to publics organism in the USA, 90% in Germany, 77% in the United Kingdom and 64% in France). These low figures suggest that nowadays, prevention is essentially a marketing tool for insurers.

In 1972, Ehrlich and Becker (Ehrlich and Becker, 1972) proposed to separate two types of prevention: self-protection and self-insurance. If one tried to adapt this work to our insurance company, self-insurance would correspond to reinsurance. Self-protection would consist in reducing claims frequency thanks to (costly) prevention plans. Optimal self-protection strategies have not yet been studied in risk theory and determining them is an interesting problem that differs from optimal reinsurance, as we shall see in the sequel. In the sequel, the word‘"prevention" refers to self-protection only.

In the present paper, we use the notation of Bowers et al. (Bowers et al., 1984). We consider an insurance company with initial surplus $U(0) = u$. This company receives premiums at a rate c per unit of time and invests a fixed amount p in prevention per unit of time. Prevention strategies could be in theory dynamic and depend on the history of the claim process. In practice prevention is a long term strategy, therefore we assume in the present paper that it is defined from the beginning and that it is not a control variable. The aggregate claim amount up to time t is given by a compound Poisson process $S(t) = \sum_{i=1}^{N(t)} X_i$, where N is a Poisson process of arrival intensity $\lambda(p)$ and the $(X_i)_{i\in\mathbb{N}^*}$ are independent and identically distributed (i.i.d.) random variables, independent from N, and with cumulative distribution function F_X, such that $\mathbb{E}(X) = \mu < \infty$.

We assume that $\lambda(.)$ is a decreasing, strictly convex, positive, and $\mathcal{C}^2$ function defined on $[0, c]$. Let us further comment these assumptions.

Choosing $\lambda(.)$ positive means that one cannot prevent all risk. This hypothesis is explained by the fact that if $\lambda(.)$ could be equal to zero, it would allow some arbitrage opportunity.

Having $\lambda(.)$ decreasing means that prevention always reduces the claim arrival intensity. This is a classical economic hypothesis, done by Ehrlich and Becker (Ehrlich and Becker, 1972) or Dionne and Eeckhoudt (Dionne and Eeckhoudt, 1985) for example. However, a prevention investment not ambitious enough could produce non significant effects. For example, smoking

29 cigarettes a day instead of 30 cigarettes leads to almost the same risk. Thus, this hypothesis is relaxed later on in this paper, by allowing $\lambda(.)$ to be first constant on some interval and then decreasing.

Assuming $\lambda(.)$ strictly convex means that the more one spends in prevention, the smaller the additional reduction of claims frequency is. This is also a common economic hypothesis. This assumption has been made by Ehrlich and Becker (Ehrlich and Becker, 1972) and by Courbage (Courbage, 2001) among others.

It is also implicitly assumed that $\lambda(.)$ does not change over time. This is quite a strong hypothesis since in reality, there is no reason for prevention to be efficient from the first moment you start implementing it. In this paper, we do not treat this subject. Optimal control of prevention planning strategies in presence of varying prevention efficiency is left for further research.

The surplus process is thus given by

$$U(t,p) = u + (c-p)t - \sum_{i=1}^{N(t)} X_i. \tag{4.1}$$

We denote by $\varphi(u,p) = \mathbb{P}(\forall t \in \mathbb{R}^+, U(t) \geq 0)$ the non-ruin probability in infinite time ($\psi(u,p) = 1 - \varphi(u,p)$ denoting the ruin probability). In order to lighten formulas, we denote by $\varphi'(u,p) = \frac{\partial \varphi(u,p)}{\partial p}$ when there is no ambiguity. We have $\varphi(u,p) > 0$ if and only if the safety loading condition is verified, i.e.

$$1 - \frac{\lambda(p)\mu}{c-p} > 0. \tag{4.2}$$

We suppose that Equation (4.2) holds true. Thus, we only consider $p \in D = [0, p_{lim}(1-\varepsilon)]$, with p_{lim} verifying $1 - \frac{\lambda(p_{lim})\mu}{c-p_{lim}} = 0$, and $\varepsilon \in [0,1[$. The function $\lambda(.)$ being decreasing and continuous on $[0,c]$, the intermediate value theorem guarantees the existence of p_{lim}.

For this model, it is easy to extend several classical results of the classical compound Poisson model. For $u = 0$, the non-ruin probability is given by

$$\varphi(0,p) = 1 - \frac{\lambda(p)\mu}{c-p}, \tag{4.3}$$

and for all u, the non-ruin probability is given by the Pollaczeck-Khinchin formula

$$\varphi(u,p) = \varphi(0,p) \sum_{n=0}^{\infty} (1-\varphi(0,p))^n F_{e,X}^{*n}(u), \tag{4.4}$$

where $F_{e,X}$ designates the equilibrium cumulative distribution function given by

$$F_{e,X}(u) = \frac{\int_0^u 1 - F_X(x)dx}{\mu}. \tag{4.5}$$

As mentioned earlier, the optimal prevention problem is different from the optimal reinsurance one. Consider for example proportional reinsurance in the classical ruin model without prevention. If the insurer cedes a proportion $\alpha \in [0,1]$ of the claims with a reinsurance loading rate $\theta > 0$, then its classical surplus process

$$U(t) = u + ct - \sum_{i=1}^{N^\gamma(t)} X_i \tag{4.6}$$

where N^γ is a Poisson process with constant intensity γ becomes

$$\tilde{U}(t,\alpha) = u + (c - \alpha(1+\theta)\gamma\mu)t - (1-\alpha)\sum_{i=1}^{N^\gamma(t)} X_i. \tag{4.7}$$

The ruin probability for the surplus process $\tilde{U}(t,\alpha)$ is the same as the one of the process

$$V(t,\alpha) = \frac{u}{1-\alpha} + \frac{(c - \alpha(1+\theta)\gamma\mu)}{1-\alpha}t - \sum_{i=1}^{N^\gamma(t)} X_i. \tag{4.8}$$

If one now considers an insurer using prevention but no reinsurance, then, thanks to a time change, the ruin probability is similar to the one of the process

$$W(t,p) = u + (c-p)\frac{\gamma}{\lambda(p)}t - \sum_{i=1}^{N^\gamma(t)} X_i. \tag{4.9}$$

One can see that the initial surplus level remains homogeneous to the claim amounts X_i's in the prevention case (4.9), while it is divided by the retention rate $1-\alpha$ in the proportional reinsurance case (4.8). Another way to see easily that the two problems are different is provided in Section 2.2.

The main contributions of this paper are to propose a first risk model with prevention and to determine the optimal prevention investment for different risk indicators. In particular, we show that the prevention amount minimizing the ruin probability maximizes the adjustment coefficient in the classical ruin model with prevention, as well as the expected dividends until ruin in the model with dividends. We also show that the optimal prevention strategy is different if one aims at maximizing the average surplus at a fixed time horizon. We carry out some sensitivity analysis of the optimal prevention investment level with respect to the premium income rate on the one hand, and with respect to the average claim size in the exponential case on the other hand. We also prove that our results can be extended to the case where prevention starts to work only after a minimum prevention level threshold.

The paper is organized as follows. In Section 2, two optimal prevention amounts are studied. More specifically, in Section 2.1, we discuss a prevention amount which maximizes the non ruin probability, the adjustment coefficient and the expected dividends until ruin, while in Section 2.2, we study a prevention amount maximizing the average surplus level on a given time horizon. In Section 3, we relax the assumption on λ and tackle the case where it is first constant and then decreasing. Some research perspectives are given in the conclusion.

4.2 Optimizing prevention strategies

An insurance company interested in prevention can pursue several purposes. In this section are presented two optimal prevention strategies. The first one consists in maximizing the non-ruin probability. Performing a time change in the model as in (4.9) enables to highlight that maximizing the non-ruin probability amounts to optimizing the premium rate. We show that the prevention amount maximizing the non-ruin probability also maximizes the adjustment coefficient and the expected dividend amount earned before ruin. That is why all these optimal prevention amounts are discussed in the same section (i.e. in Section 2.1). The second one consists in maximizing the average surplus on a given time horizon and is discussed in Section 2.2.

4.2.1 Non-ruin probability

A first goal for the insurer can be reducing its risk. In this section, we aim to determine the prevention amount that maximizes the non-ruin probability of the insurance company. We call it the “optimal prevention amount" in the following.

Let

$$p^* : \left| \begin{array}{ccc} \mathbb{R}^+ & \longmapsto & \mathbb{R}^+ \\ u & \longmapsto & min(\underset{p\in D}{argmax}(\varphi(u,p))) \end{array} \right.$$

be the function associating to an initial surplus u the prevention amount maximizing the non-ruin probability $\varphi(u,p)$. Hence, $p^*(u)$ represents the optimal prevention amount when the initial surplus is equal to u.

The next proposition shows that the optimal prevention amount does not depend on the value of the initial surplus, such that $p^*(.)$ is actually constant. We denote by p_0^* the optimal prevention amount $p^*(u)$ in the remaining of the paper. Furthermore, the next result provides an equation which enables to determine p_0^*.

Proposition 1.

The optimal prevention amount $p^*(u)$ does not depend on u.

Moreover, p_0^* is positive if and only if

$$-\lambda'(0) - \frac{\lambda(0)}{c} > 0. \tag{4.10}$$

In this case, we have

$$p_0^* = c + \frac{\lambda(p_0^*)}{\lambda'(p_0^*)}. \tag{4.11}$$

Proof.

The proof is divided into two parts. In a first time, we find p_0^* maximizing the function $\varphi(0,.)$. In a second time, we show that p_0^* also maximizes the function $\varphi(u,.)$ for all $u > 0$.

From (4.3), we directly get

$$\varphi'(0,p) = -\frac{\lambda'(p)\mu}{c-p} - \frac{\lambda(p)\mu}{(c-p)^2} \tag{4.12}$$

and

$$\varphi''(0,p) = \frac{2}{c-p}\varphi'(0,p) - \frac{\lambda''(p)\mu}{c-p}. \tag{4.13}$$

Hence, by Equation (4.13), thanks to the strict convexity of $\lambda(.)$, we notice that $\varphi'(0,p) \leq 0$ implies $\varphi''(0,p) < 0$ for all $p \in \mathbb{R}^+$.

Let us now prove that if $\varphi'(0,0) \leq 0$, then we have $\varphi'(0,p) < 0$ and $\varphi''(0,p) < 0$ for all $p > 0$. We can restrict the prove to the case $\varphi'(0,0) < 0$ since $\varphi'(0,0) = 0$ implies $\varphi''(0,p) < 0$ for small $p > 0$ which in turn implies that $\varphi'(0,.)$ is a decreasing function in a neighborhood of 0.

In that goal, let us define the interval $I \subset \mathbb{R}^+$ such that

— $0 \in I$;

— $\varphi''(0,p) \leq 0$ for all $p \in I$;

— If $J = [a,b] \subset \mathbb{R}^+$ such that $0 \in J$ and $\varphi''(0,p) \leq 0$ for all $p \in J$, then $J \subset I$.

Then, if $I = \mathbb{R}^+$, we have proved the desired result. Otherwise, it means that there would exist $a > 0$ such that $I = [0,a]$. But in that case, since $\varphi''(0,.)$ is continuous, the intermediate value theorem tells us that we would have $\varphi''(0,a) = 0$. However, by definition, $\varphi''(0,.)$ is negative on the interval I and so $\varphi'(0,.)$ is decreasing on I. Now, since $\varphi'(0,0) < 0$, we necessarily have $\varphi'(0,a) < 0$ and hence $\varphi''(a) < 0$ as shown previously, which contradicts the result $\varphi''(0,a) = 0$ that we would have if $I = [0,a]$. Finally, we necessarily have $I = \mathbb{R}^+$.

In conclusion, if $\varphi'(0,0) \leq 0$ holds, $\varphi(0,.)$ is a decreasing function meaning that we should not spend money on prevention. On the contrary, if $\varphi'(0,0) > 0$, which is equivalent to the condition $-\lambda'(0) - \frac{\lambda(0)}{c} > 0$, then $\varphi(0,.)$ increases in a neighborhood of 0, which means that prevention reduces the risk.

In such a situation, the optimal prevention amount is solution of $\varphi'(0,p_0^*) = 0$. From (4.12), we directly get

$$p_0^* = c + \frac{\lambda(p_0^*)}{\lambda'(p_0^*)},$$

which is a maximum since $\varphi''(0,p_0^*) < 0$.

We now prove that the optimal prevention amount does not depend on u. Let $u \in \mathbb{R}^+$ and $p \in D$, and let us denote by N^* the Poisson process of arrival intensity $\lambda(p_0^*)$. We now consider the two surplus processes

$$U(t,p) = u + (c-p)t - \sum_{i=1}^{N(t)} X_i, \tag{4.14}$$

with N a Poisson process of parameter $\lambda(p)$ and

$$U(t, p_0^*) = u + (c - p_0^*)t - \sum_{i=1}^{N^*(t)} X_i. \tag{4.15}$$

It is well-known that changing the time scale (here multiplying the time by $\lambda(p_0^*)/\lambda(p)$) does not influence the infinite-time ruin probability, meaning that

$$\psi(u, p) = \mathbb{P}(\exists t \text{ such as } U(t, p) < 0) = \mathbb{P}(\exists t \text{ such as } U_2(t, p) < 0), \tag{4.16}$$

where the surplus process $U_2(t, p)$ is defined as

$$U_2(t, p) = u + (c - p)\frac{\lambda(p_0^*)}{\lambda(p)}t - \sum_{i=1}^{N^*(t)} X_i. \tag{4.17}$$

So, the only difference between the surplus processes $U_2(t, p)$ and $U(t, p_0^*)$ is the drift. Thus, it comes $\varphi(u, p) < \varphi(u, p_0^*)$ if and only if $(c-p)\frac{\lambda(p_0^*)}{\lambda(p)} < (c-p_0^*)$. Now, it suffices to notice that

$$(c - p)\frac{\lambda(p_0^*)}{\lambda(p)} - (c - p_0^*) = \mu\lambda(p_0^*)\left[\frac{c - p}{\lambda(p)\mu} - \frac{c - p_0^*}{\lambda(p_0^*)\mu}\right] < 0 \tag{4.18}$$

since p_0^* minimizes $\psi(0, p) = \left(\frac{c-p}{\lambda(p)\mu}\right)^{-1}$, which completes the proof.

■

Notice that if there is no p_0^* in D such that $\varphi'(0, p_0^*) = 0$, then the optimal prevention amount is $p_{lim}(1 - \varepsilon)$ as $\varphi(0, .)$ increases with p.

In Figure 44, we illustrate Proposition 1. We assume $c = 10$, $\lambda(p) = \frac{0,5}{2} + \frac{0,5e^{-p}}{2}$ and we consider claim amounts that are exponentially distributed with mean $\mu = 10$. We see that the optimal prevention amount is always the same whatever the value taken by the initial surplus.

From Proposition 1, we observe that maximizing $\varphi(u, .)$ amounts to find p minimizing the loss ratio $\frac{\lambda(p)\mu}{c-p}$, which is the expected claim amount $\lambda(p)\mu$ over one year divided by the annual net premium $c - p$. In other words, it amounts to find p that maximizes the implicit safety loading.

Moreover, notice that equations (4.10) and (4.11) imply

$$\lambda'(p) \leq \frac{-\lambda(p)}{c - p} \quad \text{for all } p \in [0, p_0^*)], \tag{4.19}$$

which yields

$$\frac{\lambda(p)}{\lambda(0)} \leq \frac{c - p}{c} \quad \text{for all } p \in [0, p_0^*]. \tag{4.20}$$

In words, investing in prevention makes sense as long as the claim frequency decreases more than the premium rate.

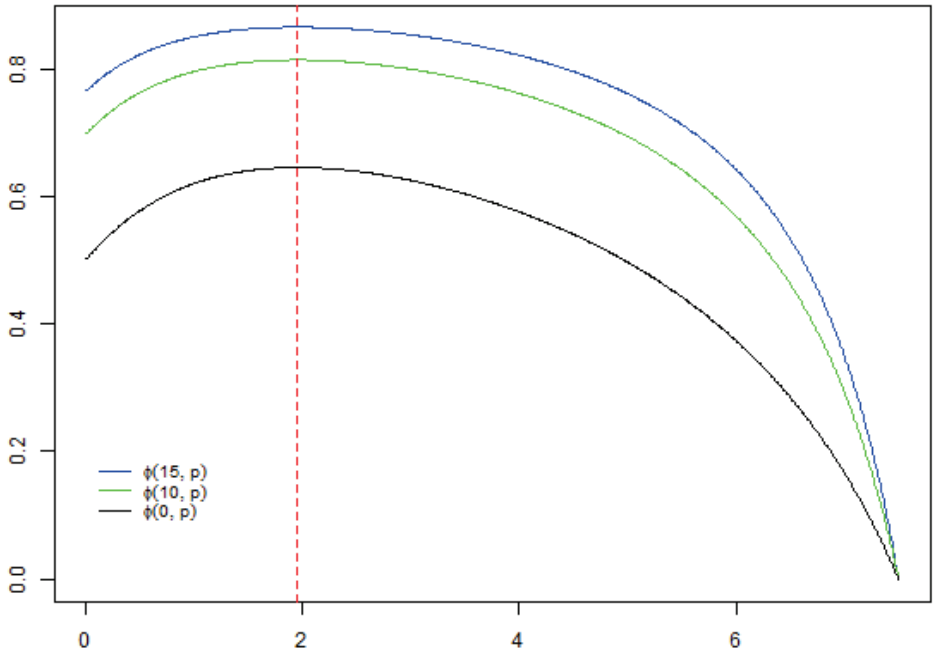

Figure 44 – Non-ruin probabilities $\varphi(0,p)$, $\varphi(10,p)$ and $\varphi(15,p)$ for $c=10$, $\lambda(p)=\frac{0{,}5}{2}+\frac{0{,}5e^{-p}}{2}$ and claim amounts that are exponentially distributed with mean $\mu=10$

It is also interesting to remark that p_0^* can be found with a geometric construction. Indeed, Equation (4.11) shows that the tangent line of the function $\lambda(.)$ at point p_0^* passes through the point $(c,0)$, as illustrated in Figure 1.

Adjustment coefficient

In this section, we assume that the moment generating function $M_X(.)$ for the claim amount is well defined. Namely, we assume that

$$M_X(s)<\infty \quad \text{for all } s\in\mathbb{R} \tag{4.21}$$

or that there exists $s^*>0$ such that

$$M_X(s)<\infty \quad \text{for all } s<s^* \quad \text{and} \quad M_X(s)=\infty \quad \text{for all } s\geq s^*. \tag{4.22}$$

For all $p\in D$, we denote by $\kappa(p)$ the adjustment coefficient defined as the positive solution of the equation

$$\lambda(p)+(c-p)s=\lambda(p)M_X(s). \tag{4.23}$$

The adjustment coefficient is thus well defined regardless of the prevention amount.

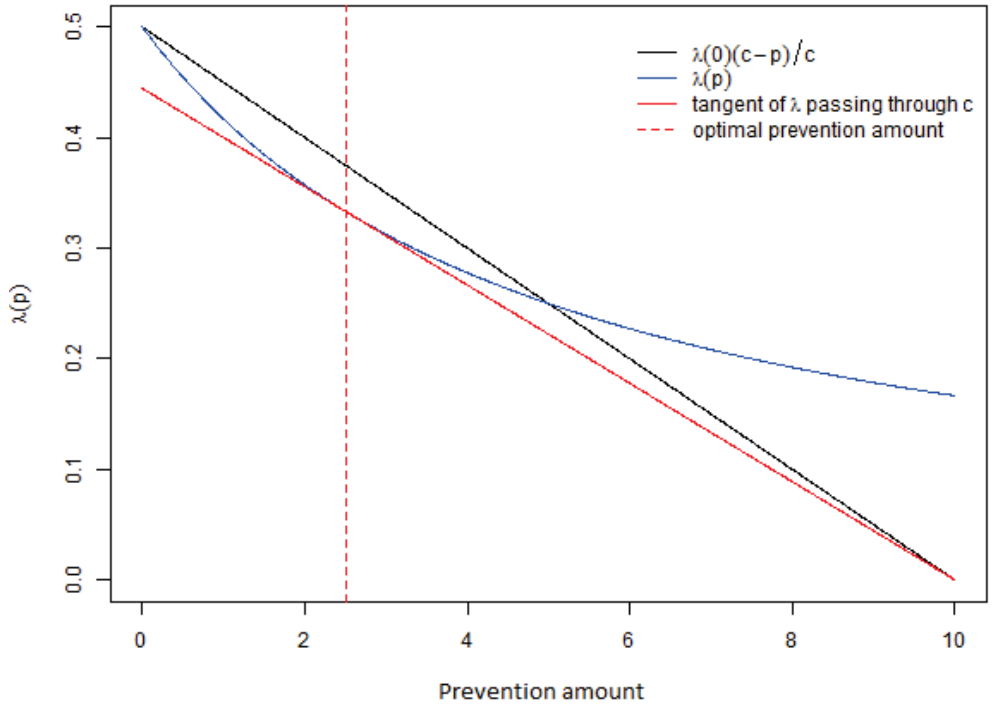

Figure 45 – Geometric construction to find the optimal prevention amount (with $c = 10$)

It is well known that if the adjustment coefficient exists, inequality $1-\varphi(u,p) \leq e^{-\kappa(p)u}$ holds true, which means that the ruin probability decreases exponentially with the initial surplus, the adjustment coefficient controlling the slope of the decrease. Therefore, the adjustment coefficient can be seen as an indicator of the initial surplus efficiency in reducing the ruin probability.

A prevention amount can be defined as optimal if it maximizes the adjustment coefficient of the insurance company. The next result shows that this optimal prevention amount coincides with the one that maximes the non-ruin probability. It is a direct consequence of Proposition 1 which highlights the fact that maximizing the non-ruin probability with respect to p amounts to find p maximizing the implicit safety loading.

Corollary 2.

If Equation (4.10) holds, then p_0^* maximizes the function $\kappa(.)$.

Proof. Let $p \in D$. Equation (4.23) can be rewritten as

$$1 + \frac{(c-p)s}{\lambda(p)} = M_X(s). \tag{4.24}$$

Geometrically, Equation (4.24) tells us that the adjustment coefficient is the point where the line passing through (0,1) with slope $\frac{(c-p)}{\lambda(p)}$ crosses the moment generating function $M_X(.)$, as depicted in Figure 46. So, one can see that the adjustment coefficient increases with the

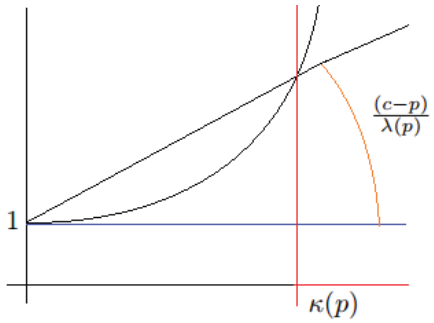

Figure 46 – Geometric construction of $\kappa(p)$

drift of the surplus process $U(t,p)$. Let us now change the time scale to obtain the surplus process described in (4.17). Of course, the adjustment coefficient is not impacted by such a transformation. Now, if we compare the drifts of surplus processes (4.17) and (4.15), one sees by Equation (4.18) that p_0^* leads to the largest drift for (4.17) (which corresponds in that case to the drift of (4.15)) and hence maximizes $\kappa(.)$. ■

Hence, the prevention amount that maximizes the adjustment coefficient is equivalent to the prevention amount that minimizes the ruin probability.

Dividends

In 1957, De Finetti introduced a dividend parameter in the classical Cramer-Lundberg model (De Finetti, 1957). In this paper, we consider the case of a constant dividend barrier as explored by Segerdahl (Segerdahl, 1970) and Dickson and Gray (Dickson and Gray, 1984), among others.

Denote the constant dividend barrier level by $K > u$. When the surplus process hits K, it remains constant until a new claim happens and dividends are being transferred to shareholders. We denote by L_t the total amount of dividends distributed up to time t. In such a setting, ruin occurs almost surely.

In Segerdahl (Segerdahl, 1970), we can find an expression for the probability of being ruined before reaching dividend barrier K, which is

$$\xi(u,K,p) = \frac{\varphi(K,p) - \varphi(u,p)}{\varphi(K,p)}. \tag{4.25}$$

From that expression, we can easily obtain an expression for the expected amount of dividends earned before being ruined, namely

$$\mathbb{E}(L_t)(p) = \frac{(1-\xi(u,K,p))(c-p)}{\lambda(p)\theta}, \tag{4.26}$$

where $\theta = \int_0^K \xi(K-x,K,p)dF_X(x)$. We are now in position to prove the following result, where we use the same change of time than in the proof of Proposition (1).

Corollary 3. *The function $\mathbb{E}(L_t)(.)$ reaches its maximum in p_0^*.*

Proof. Let $u < K \in \mathbb{R}^+$ and $p \in D$. Using the same change of time than in (4.17) and comparing again the deterministic parts of the surplus processes (4.17) and (4.15), one easily sees that p_0^* minimizes the function $\xi(u,K,.)$.

Now, combining (4.25) and (4.26), we get

$$\mathbb{E}(L_t)(p) = \frac{\varphi(u,p)}{\int_0^K 1-\varphi(K-x,p)dF_X(x)}\frac{c-p}{\lambda(p)}. \tag{4.27}$$

We have seen in the proof of Proposition 1 that p_0^* maximizes $\frac{c-p}{\lambda(p)}$. Furthermore, we also know by Proposition 1 that p_0^* maximizes $\varphi(u,p)$ and therefore minimizes $\int_0^K 1-\varphi(K-x,p)dF_X(x)$ as well, which ends the proof. ■

Thus, p_0^* introduced in Proposition 1 is also the prevention amount that maximizes the expected amount of dividends.

Impact of c on p_0^*

We denote the optimal prevention amount by $p_0^*(c)$ to make explicit the dependence with the premium rate c and we investigate the impact of c on $p_0^*(c)$. We get the following result.

Proposition 4. *If condition (4.10) is fulfilled, then $p_0^*(.)$ increases with c.*

Proof. Recalling Equation (4.11),

$$p_0^*(c) = c + \frac{\lambda(p_0^*(c))}{\lambda'(p_0^*(c))} \tag{4.28}$$

yields

$$p_0^{*\prime}(c) = \frac{\lambda'(p_0^*(c))^2}{\lambda''(p_0^*(c))\lambda(p_0^*(c))} > 0. \tag{4.29}$$

Equation (4.29) guarantees that $p_0^*(.)$ increases with c. ■

So, if the insurer gets a higher premium amount, it is always optimal to invest a part of this premium increase for prevention. In the extreme case where the premium rate c goes to infinity, the next result shows that $p_0^*(c)$ tends to infinity as well.

Proposition 5.

We have $\lim_{c\to\infty} p_0^*(c) = +\infty$. *Furthermore,*

(i) If $\lim_{c\to\infty} \lambda'(c)c \neq 0$, *then* $\lim_{c\to\infty} \frac{p_0^*(c)}{c} \neq 0$;

(ii) If $\lim_{c\to\infty} \lambda'(c)c = 0$ *and* $\lim_{c\to\infty} \lambda(c) > 0$, *then* $\lim_{c\to\infty} \frac{p_0^*(c)}{c} = 0$;

(iii) If $\lim_{c\to\infty} \lambda(c) = 0$, *then* $\lim_{c\to\infty} \frac{p_0^*(c)}{c} = 0$ *if and only if* $\lim_{c\to\infty} \frac{\lambda(c)-\lambda'(c)c}{\lambda(c)} = 1$.

Proof. Let us first prove that $\lim_{c\to\infty} p_0^*(c) = +\infty$. Equation (4.11) can be rewritten as

$$\frac{p_0^*(c)}{c} = \frac{\lambda(p_0^*(c))}{\lambda'(p_0^*(c))c} + 1, \tag{4.30}$$

where $c \in \mathbb{R}^+$ is such that the safety loading condition is fulfilled. Because $0 \leq \frac{p_0^*(c)}{c} < 1$, we know that $\frac{\lambda(p_0^*(c))}{\lambda'(p_0^*(c))c}$ is bounded.

Let us assume that $\lim_{c\to\infty} p_0^*(c) = \alpha \in \mathbb{R}^+$. We have $\lambda'(\alpha) < 0$ since $\lambda(.)$ is a decreasing function. It comes

$$\lim_{c\to\infty} \frac{p_0^*(c)}{c} = 1 + \lim_{c\to\infty} \frac{\lambda(p_0^*(c))}{\lambda'(p_0^*(c))c} = 1 \tag{4.31}$$

as $\lim_{c\to\infty} \frac{\lambda(p_0^*(c))}{\lambda'(p_0^*(c))c} = \lim_{c\to\infty} \frac{\lambda(\alpha)}{\lambda'(\alpha)c} = 0$. Now, since $\lim_{c\to\infty} p_0^*(c) = \alpha$, we also get that $\lim_{c\to\infty} \frac{p_0^*(c)}{c} = 0$, which contradicts our previous result. Hence, as $p_0^*(.)$ is an increasing function, we necessarily have $\lim_{c\to\infty} p_0^*(c) = +\infty$.

Let us now turn to items (i)-(iii). Since $p_0^*(.)$ is strictly increasing, the inverse function $p_0^{*-1}(.)$ exists. Hence, Equation (4.30) evaluated in $p_0^{*-1}(c)$ gives

$$p_0^{*-1}(c) = c - \frac{\lambda(c)}{\lambda'(c)}. \tag{4.32}$$

Obviously, we have $\lim_{c\to\infty} p_0^{*-1}(c) = +\infty$, so that $\lim_{c\to\infty} \frac{\lambda(p_0^*(c))}{\lambda'(p_0^*(c))c}$ exists and is equal to l if and only if

$$\lim_{c\to\infty} \frac{\lambda(c)}{\lambda'(c)p_0^{*-1}(c)} = \lim_{c\to\infty} \frac{\lambda(c)}{\lambda'(c)c - \lambda(c)} = l. \tag{4.33}$$

Moreover, by assumption on $\lambda(.)$, we know that $\lim_{c\to\infty} \lambda(c) = l_1$.

In a first time, let us assume that $l_1 > 0$.

Since $0 \leq \frac{p_0^*(c)}{c} < 1$, Equation (4.30) directly yields

$$\lim_{c\to\infty} \lambda'(p_0^*(c))c \neq 0 \tag{4.34}$$

and

$$\frac{\lambda'(p_0^*(c))c}{\lambda(p_0^*(c))} < -1. \tag{4.35}$$

Since $\lambda(.)$ is decreasing, Equations (4.29) and (4.35) lead to

$$(\lambda'(p_0^*(c))c)' = \lambda'(p_0^*(c))\left[1 + \frac{\lambda'(p_0^*(c))c}{\lambda(p_0^*(c))}\right] > 0. \tag{4.36}$$

Thus, $\lambda'(p_0^*(c))c$ is negative and increases with c. The monotone convergence theorem and (4.34) prove that there exists $l_2 < 0$ such that $\lim_{c\to\infty} \lambda'(p_0^*(c))c = l_2$, which in turn implies that $\lim_{c\to\infty} \frac{p_0^*(c)}{c} = 1 + \frac{l_1}{l_2}$. Hence, from Equation (4.33), one sees that we have $l_2 = -l_1$ if and only if $\lim_{c\to\infty} \lambda'(c)c = 0$, which proves item (ii). Also this proves item (i) in the case where $l_1 > 0$.

In a second time, let us assume that $l_1 = 0$ in order to end the proof for item (i) and to prove item (iii).

We learn from Equation (4.33) that if $\lim_{c\to\infty} \lambda'(c)c \neq 0$, then we have $l = 0$ and $\lim_{c\to\infty} \frac{p_0^*(c)}{c} = 1$, which ends the proof of item (i). Finally, turning to item (iii), we know that if $\lim_{c\to\infty} \frac{\lambda(c)-\lambda'(c)c}{\lambda(c)} = 1$, then there exists $c^* \in \mathbb{R}$ such that for all $c > c^*$, one has $\lambda(c) - \lambda'(c)c > 0$. Hence, $\lim_{c\to\infty} \frac{\lambda(c)-\lambda'(c)c}{\lambda(c)} = 1$ is equivalent to $\lim_{c\to\infty} \frac{\lambda(c)}{\lambda'(c)c-\lambda(c)} = -1$, so that from Equations (4.31) and (4.33), one sees that $\lim_{c\to\infty} \frac{\lambda(c)-\lambda'(c)c}{\lambda(c)} = 1$ is equivalent to $\lim_{c\to\infty} \frac{p_0^*(c)}{c} = 0$.

■

Notice that the case $\lim_{c\to\infty} \lambda(c) = 0$ is of practical relevance if one thinks about vaccination campaigns for instance. Indeed, in this setting, investing enough money in prevention to vaccinate the entire population will enable one to eradicate the associated disease, as it was the case for the smallpox virus.

Remark 6. When $\lim_{c\to+\infty} \{(\lambda(c) - \lim_{c\to+\infty} \lambda(c))c\}$ exists[1] (which is generally true in our setting), condition $\lim_{c\to\infty} \lambda'(c)c \neq 0$ cannot be fulfilled. Indeed, let us show that the existence of $\lim_{c\to+\infty} \{(\lambda(c) - \lim_{c\to+\infty} \lambda(c))c\}$ necessarily implies $\lim_{c\to+\infty} \lambda'(c)c = 0$. Notice that since $\lambda'(c) \leq 0$, we get $\lim_{c\to+\infty} \lambda'(c)c \leq 0$ when it exists.

In a first time, we consider the case where $\lim_{c\to+\infty} \lambda(c) = 0$. We have

$$\lambda'(c)c = (\lambda(c)c)' - \lambda(c). \tag{4.37}$$

One distinguish two situations. Firstly, we suppose that $\lim_{c\to\infty} \lambda(c)c = +\infty$. Hence, we cannot find a constant $\tilde{c}$ such that $\lambda(c)c$ is decreasing on $[\tilde{c}, +\infty[$. However, if $\lim_{c\to+\infty} \lambda'(c)c < 0$, then Equation (4.37) leads to $\lim_{c\to+\infty} (\lambda(c)c)' < 0$, which contradicts the previous sentence. Thus, we necessarily have $\lim_{c\to+\infty} \lambda'(c)c = 0$. Secondly, we assume that $\lim_{c\to\infty} \lambda(c)c = \gamma \geq 0$. Since for all $c > 0$, we have $\lambda(c)c = \int_0^c (\lambda(x)x)'dx$ and $\lim_{c\to\infty} (\lambda(c)c)' = 0$, one sees that Equation (4.37) yields $\lim_{c\to+\infty} \lambda'(c)c = 0$.

1. including here the case where $\lim_{c\to+\infty} \{(\lambda(c) - \lim_{c\to+\infty} \lambda(c))c\} = +\infty$

In a second time, we consider the case where $\lim_{c\to+\infty} \lambda(c) = \beta > 0$. Let

$$\lambda_2(c) = \lambda(c) - \beta. \tag{4.38}$$

One sees that $\lambda_2(.)$ is decreasing, convex, $\mathcal{C}_2$ and such that $\lim_{c\to+\infty} \lambda_2(c) = 0$. The assumptions on $\lambda(.)$ implies that $\lim_{c\to+\infty} \lambda_2(c)c$ exists. Applying the same arguments to $\lambda_2(.)$ than in the first case leads to $\lim_{c\to+\infty} \lambda_2'(c)c = 0$, which in turn implies that $\lim_{c\to+\infty} \lambda'(c)c = 0$ by deriving Equation (4.38).

Remark 7. Equation (4.11) tells us that p_0^* does not depend on μ while the non-ruin probability does. If the claim amounts are exponentially distributed, so that their distribution is entirely characterized by its mean, then we have $\frac{\partial \frac{\varphi(u,p_0^*)}{\varphi(u,0)} - 1}{\partial\ \mu} > 0$, which means that prevention is still important when you deal with claim amounts that are higher in average. Indeed, in the exponential case, the non-ruin probability writes

$$\varphi(u,p) = 1 - \frac{\lambda(p)\mu}{c-p} e^{-(\frac{1}{\mu} - \frac{\lambda(p)}{c-p})u} \tag{4.39}$$

for all $u > 0$ and $p \in [0, p_{lim}]$, so that

$$\frac{\partial \varphi(u,p)}{\partial \mu} = -\left(1 + \frac{u}{\mu}\right) \frac{\lambda(p)}{c-p} e^{-\left(\frac{1}{\mu} - \frac{\lambda(p)}{c-p}\right)u}. \tag{4.40}$$

Now, as $\frac{\partial \frac{\varphi(u,p_0^*)}{\varphi(u,0)} - 1}{\partial \mu} = \frac{\frac{\partial \varphi(u,p_0^*)}{\partial \mu}\varphi(u,0) - \frac{\partial \varphi(u,0)}{\partial \mu}\varphi(u,p_0^*)}{\varphi^2(u,0)}$, Equation (4.40) leads to

$$\frac{\partial \frac{\varphi(u,p_0^*)}{\varphi(u,0)} - 1}{\partial \mu} = \frac{\left(1 + \frac{u}{\mu}\right)}{\varphi^2(u,0)} \left(\frac{\lambda(0)}{c} e^{-\left(\frac{1}{\mu} - \frac{\lambda(0)}{c}\right)u} - \frac{\lambda(p_0^*)}{c-p_0^*} e^{-\left(\frac{1}{\mu} - \frac{\lambda(p_0^*)}{c-p_0^*}\right)u} \right). \tag{4.41}$$

Therefore, since $\frac{\lambda(p_0^*)}{c-p_0^*} < \frac{\lambda(0)}{c}$ by definition of p_0^*, it comes $\frac{\partial \frac{\varphi(u,p_0^*)}{\varphi(u,0)} - 1}{\partial \mu} > 0$, as previously announced.

4.2.2 Surplus expectancy

An insurance company may also want to maximize its expected surplus on a given time horizon t, which amounts to find the prevention amount maximizing the expectation of $U(t)$, denoted $\tilde{p}$ in the following.

Obviously, we have

$$\mathbb{E}(U(t))(p) = u + [c - p - \lambda(p)\mu]t. \tag{4.42}$$

Hence, by deriving (4.42) with respect to p, we easily see that the optimal prevention amount $\tilde{p}$ is positive if and only if $-\lambda'(0) > \frac{1}{\mu}$. In such a case, $\tilde{p}$ solves

$$-\lambda'(\widetilde{p}) = \frac{1}{\mu}. \tag{4.43}$$

This time, the optimal prevention amount $\tilde{p}$ is different from p_0^*. One sees that $\tilde{p}$ does not depend on the premium rate c but well on the expected claim amount μ. As a consequence, an insurance company has to make a choice between maximizing its expected wealth on a given time horizon and decreasing its risk on the long run.

Note that this particular optimisation result provides an additional way to demonstrate that, as stated in the introduction, the present problem differs from the optimal reinsurance problem: in the proportional reinsurance context where the surplus process is given by

$$V(t,\alpha) = u + (c - \alpha(1+\theta)\gamma\mu)t - \sum_{i=1}^{N^\gamma(t)} X_i, \tag{4.44}$$

as in Equation (4.8), maximizing the expected surplus at the end of the first period leads to maximizing

$$u + c - \gamma\mu(1 + \alpha\theta),$$

which is achieved for $\alpha = 0$ as long as $\theta > 0$.

In the sequel, we focus our study on the optimal prevention amount p_0^*.

4.3 $\lambda(.)$ constant then decreasing

Assuming $\lambda(p)$ decreasing with p can be too restrictive in practice. Indeed, as mentioned in the introduction, we can have situations where we must at least invest a certain amount P in prevention to start to observe an impact on $\lambda(.)$. That is why, in this section, we consider the case where $\lambda(.)$ is first constant for prevention amounts smaller than a certain threshold P. Beyond P, $\lambda(.)$ is still assumed to be decreasing and strictly convex.

Formally, let $0 < P < c$ and let

$$\lambda : \left| \begin{array}{ccl} [0,c] & \longmapsto & \mathbb{R}^+ \\ p & \longmapsto & \lambda(0) \; if \; p \leq P \\ & & \lambda(p) < \lambda(0) \; otherwise. \end{array} \right.$$

Moreover, for all $p \geq P$, we suppose $\lambda \in \mathcal{C}^2$, $\lambda'(p) < 0$ and $\lambda''(p) > 0$.

We look for the prevention amount p_0^* maximizing $\varphi(0,p) = 1 - \frac{\lambda(p)\mu}{c-p}$. The reasoning is conducted in two steps. Firstly, we look for the optimal prevention amount $\overline{p^*}$ on the interval $[P, p_{lim}[$, where prevention is useful. Secondly, we compare the ruin probability obtained by investing this amount in prevention with the one obtained without prevention.

Proposition 8. *If condition* $\frac{c-\overline{p^*}}{c}\lambda(0) > \lambda(\overline{p^*})$ *is verified, then we have* $p_0^* = \overline{p^*}$. *Otherwise, we have* $p_0^* = 0$.

Proof. In a first time, we consider $p \in [P, p_{lim}]$. In this case, $\lambda(.)$ verifies all the assumptions stated in Section 2 and condition (4.10) becomes

$$-\lambda'(P) - \frac{\lambda(P)}{c-P} > 0. \tag{4.45}$$

Thus, when condition (4.45) is fulfilled, Proposition 1 tells us that $\overline{p^*}$ does not depend on u and is given by $\overline{p^*} = c + \frac{\lambda(\overline{p^*})}{\lambda'(\overline{p^*})}$). Otherwise, if (4.45) is not verified, it comes
$overline p^* = P$ and hence $p_0^* = 0$.

In a second time, when condition (4.45) is fulfilled, we have to compare $\varphi(0,0)$ and $\varphi(0,\overline{p^*})$. It comes

$$\begin{aligned} \varphi(0,0) - \varphi(0,\overline{p^*}) < 0 \quad &\Leftrightarrow \quad -\frac{\lambda(0)\mu}{c} + \frac{\lambda(\overline{p^*})\mu}{c-\overline{p^*}} < 0 \\ &\Leftrightarrow \quad \frac{c-\overline{p^*}}{c}\lambda(0) > \lambda(\overline{p^*}). \end{aligned}$$

Hence, if condition $\frac{c-\overline{p^*}}{c}\lambda(0) > \lambda(\overline{p^*})$ holds true, then we have $p_0^* = \overline{p^*}$. Otherwise, we have $p_0^* = 0$. ■

Let us notice that condition $\frac{c-\overline{p^*}}{c}\lambda(0) > \lambda(\overline{p^*})$ is closed to (4.10). It can be rewritten as $\frac{\lambda(\overline{p^*})}{\lambda(0)} < \frac{c-\overline{p^*}}{c}$, so that the relative decrease of the premium rate (from c to $c - \overline{p^*}$) must be smaller than the relative decrease of the claim frequency (from $\lambda(0)$ to $\lambda(\overline{p^*})$). Moreover, the proofs of Proposition 1 and Corollaries 2 and 3 can easily be adapted to this section.

4.4 Conclusion

In this paper, a first risk model with prevention has been introduced. The optimal prevention strategies have been identified for several optimisation problems. In further works, it would be interesting to consider different impacts of prevention for different categories of claims. The uncertainty on prevention efficiency, its evolution over time, as well as the impact of impulses of prevention strategies (represented by an instantaneous, one-shot drop in the surplus instead of a steady prevention effort which decreases the premium income rate) could be taken into account. It would also be interesting to consider optimal prevention as a stochastic control problem, and to jointly use optimal prevention and reinsurance. This is left for further research.

Acknowledgements : This paper was realized within the framework of the Chair Prevent'Horizon, supported by the risk foundation Louis Bachelier and in partnership with Claude Bernard Lyon 1 University, Actuaris, AG2R La Mondiale, G2S, Covea, Groupama Gan Vie, Groupe Pasteur Mutualité, Harmonie Mutuelle, Humanis Prévoyance and La Mutuelle Générale.

Bibliographie

Afonso, L. B., dos Reis, A. D. E., and Waters, H. R. (2010). Numerical evaluation of continuous time ruin probabilities for a portfolio with credibility updated premiums. ASTIN Bulletin : The Journal of the IAA, 40(1) :399–414.

Avanzi, B. (2009). Strategies for dividend distribution : a review. North American Actuarial Journal, 13(2) :217–251.

Bowers, N. L., Gerber, H., Hickman, J., Jones, D., and Nesbitt, C. (1984). Actuarial mathematics.

Courbage, C. (2001). Self-insurance, self-protection and market insurance within the dual theory of choice. The GENEVA Papers on Risk and Insurance-Theory, 26(1) :43–56.

De Finetti, B. (1957). Su un'impostazione alternativa della teoria collettiva del rischio. In Transactions of the XVth international congress of Actuaries, volume 2, pages 433–443.

Dickson, D. C. and Gray, J. (1984). Approximations to ruin probability in the presence of an upper absorbing barrier. Scandinavian Actuarial Journal, 1984(2) :105–115.

Dionne, G. and Eeckhoudt, L. (1985). Self-insurance, self-protection and increased risk aversion. Economics Letters, 17(1-2) :39–42.

Ehrlich, I. and Becker, G. S. (1972). Market insurance, self-insurance, and self-protection. Journal of political Economy, 80(4) :623–648.

Hesselager, O. (1990). Some results on optimal reinsurance in terms of the adjustment coefficient. Scandinavian Actuarial Journal, 1990(1) :80–95.

Højgaard, B. and Taksar, M. (1998). Optimal proportional reinsurance policies for diffusion models with transaction costs. Insurance : Mathematics and Economics, 22(1) :41–51.

Schmidli, H. (2002). On minimizing the ruin probability by investment and reinsurance. Annals of Applied Probability, pages 890–907.

Segerdahl, C.-O. (1970). On some distributions in time connected with the collective theory of risk. Scandinavian Actuarial Journal, 1970(3-4) :167–192.

Chapitre 5

Optimal prevention of large risks with two types of claims

Ce chapitre reprend un article coécrit avec Stéphane Loisel, Jean-Louis Rullière et Julien Trufin et soumis au "Journal of Mathematical Analysis and Applications"

Abstract

In this paper, we propose and study a risk model with two types of claims in which the insurer may invest into a prevention plan which decreases the large claims intensity without impacting the small claims. In this setting, we prove that prevention is advantageous when claim severities for small and large claims are ordered in the sense of the Harmonic-Mean-Residual-Lifetime (HMRL) order. In addition, we show that the optimal prevention amount is the lowest when there is no initial surplus. Finally, we characterize the asymptotic optimal prevention strategy when the initial surplus tends to infinity in the two main cases where both claim types are light-tailed and where one of them is light-tailed and the other one is heavy-tailed.

keyword : Ruin theory; Prevention; Optimal prevention strategy; Insurance.

5.1 Introduction

Prevention has become very popular in insurance. It is more and more used by insurers for marketing purposes, in relation with their corporate social responsibility. However, it is not yet regarded as a powerful risk management tool by insurance managers, in particular because they feel that prevention in insurance has more impact on extreme risks than on more frequent claims. As some of the extreme health risks (like long term affections) are taken care of by social security in several countries, and as the policyholder might switch insurer in the meantime, insurers fear that prevention might benefit to the state or to their competitors. In the near future, it might become possible for insurers of some countries to claim reimbursement of their prevention expenses by the state if they prove that their efforts

have an impact on the claims paid by social security. In property and casualty insurance, prevention is often efficient to manage extreme claims arising from consequences of dramatic fire or flooding episodes. Prevention has however less impact on less severe events. It is therefore interesting to study optimal prevention strategies in presence of two types of claims: large claims for which prevention has an impact and small claims for which the effect of prevention is limited or does not exist.

The impact of prevention (sometimes combined with self-insurance, that can be assimilated to reinsurance in our framework) has been investigated in economics by Ehrlich and Becker (Ehrlich and Becker, 1972), Dionne and Eeckhoudt (Dionne and Eeckhoudt, 1985) and Courbage (Courbage, 2001), among others. In ruin theory, a first risk model with prevention has been recently proposed by Gauchon et al. (Gauchon et al., 2019), with a unique set of claims for which prevention has some effect. The authors consider an insurance company with initial surplus $U(0) = u$. This company receives premiums at a rate c per unit of time and invests a fixed amount p in prevention per unit of time. The aggregate claim amount up to time t is given by a compound Poisson process $S(t) = \sum_{i=1}^{N(t)} X_i$, where N is a Poisson process of arrival intensity $\lambda(p)$ and the $(X_i)_{i\in\mathbb{N}^*}$ are i.i.d. random variables, independent from N, and with cumulative distribution function F_X, such that $\mathbb{E}(X) = \mu < \infty$. (Gauchon et al., 2019) assume that $\lambda(.)$ is a decreasing, strictly convex, positive, and $\mathcal{C}^2$ function defined on $[0, c]$. They determine the optimal prevention investment for different risk indicators, and in particular for the ruin probability. This optimal prevention strategy does not depend on the initial surplus level.

A common point of the approaches of Ehrlich and Becker (Ehrlich and Becker, 1972), Gauchon et al. (Gauchon et al., 2019) and of all other papers (except Courbage and Rey who consider a second risk with lotteries on a two periods model (Courbage and Rey, 2012)) addressing prevention (to the best of our knowledge) is that they consider that prevention has the same effect on all claims. In the present paper, we introduce two types of claims, and assume that prevention only works for the large claims. We show that in this case, the optimal prevention investment does depend on the initial surplus level. One could think that it would be optimal to do less prevention if one has a high initial surplus level. In the present paper, we show that it is the opposite: when the initial surplus is very large, under very reasonable conditions, optimising the ruin probability leads to implement more prevention than without an initial surplus. In the conclusion, we propose an intuitive explanation for this surprising result.

Our main contribution is threefold. First, we propose and analyse a first risk model with two types of claims, where prevention has some impact on the large claims only. Second, we show that prevention is advantageous when claim severities of small claims and large claims are ordered in the sense of the so-called Harmonic Mean Residual Lifetime (HMRL) order, and that the optimal prevention effort is higher with a positive initial surplus than when the initial surplus is zero. Third, we characterize the asymptotic optimal prevention strategy when the

initial surplus tends to infinity in the two main cases where both claim types are light-tailed and where one of them is light-tailed and the other one is heavy-tailed. In addition, on the occasion of a natural example of the second case, we provide a necessary and sufficient condition to order a translated Exponential random variable and a Pareto one in the HMRL order.

The paper is organized as follows. In Section 2, we describe our model with two types of claims. In Section 3, we recall the definition of the HMRL order and we notably provide a necessary and sufficient condition to order a translated Exponential random variable and a Pareto one in the HMRL order. In Section 4, we show that, under some HMRL ordering condition, the optimal prevention level is positive and higher than when the initial surplus is zero. In Section 5, we study the asymptotic optimal prevention strategy when both claim types are light tailed. Section 6 is devoted to the case where the second type of claims is heavy tailed.

5.2 The model

Let us consider an insurer facing two types of claims, namely small claims and large claims. In such a case, the number of claims $N(t)$ up to time t can be written as $N_1(t)+N_2(t)$, where N_1 (resp. N_2) is a Poisson process of arrival intensity $\lambda_1(p)$ (resp. $\lambda_2(p)$) that represents the number of small claims (resp. large claims), such that $\lambda(p) = \lambda_1(p) + \lambda_2(p)$. The Poisson processes N_1 and N_2 are assumed to be independent. Hence, the aggregate claim amount $S(t) = \sum_{k=1}^{N(t)} X_k$ up to time t can be decomposed as

$$S(t) = \sum_{k=1}^{N(t)} X_k = \sum_{i=1}^{N_1(t)} X_i^{(1)} + \sum_{j=1}^{N_2(t)} X_j^{(2)},$$

where the $X_i^{(1)}$s (resp. $X_j^{(2)}$s) are the claim severities for small claims (resp. large claims) and are assumed to be independent and identically distributed as $X^{(1)}$ (resp. $X^{(2)}$) with mean μ_1 (resp. μ_2) such that $\mu_1 \leq \mu_2$. Furthermore, the claim severities $X_i^{(1)}$ and $X_j^{(2)}$ are supposed to be independent and independent of the number of claims $N_1(t)$ and $N_2(t)$. Such a model amounts to consider the claim severities X_k with cumulative distribution function

$$F_X(x) = \frac{\lambda_1(p)F_{X^{(1)}}(x) + \lambda_2(p)F_{X^{(2)}}(x)}{\lambda(p)}, \quad x \geq 0, \tag{5.1}$$

where $F_{X^{(1)}}(.)$ and $F_{X^{(2)}}(.)$ are the cumulative distribution functions of $X^{(1)}$ and $X^{(2)}$, respectively.

Considering that the prevention amount p only influences $\lambda(p)$ and not $F_X(.)$, as supposed in Gauchon et al. (Gauchon et al., 2019), means that $\lambda_1(p)$ and $\lambda_2(p)$ must be impacted in the same way by p in order to keep constant both ratios $\frac{\lambda_1(p)}{\lambda(p)}$ and $\frac{\lambda_2(p)}{\lambda(p)}$ in (5.1). In other words, a decrease of $x\%$ for $\lambda(.)$ because of prevention also means a decrease of $x\%$ for both $\lambda_1(.)$ and $\lambda_2(.)$.

In practice, we can have some situations where prevention only impacts the number of large claims. Let us think about DNA test leading to preventive mastectomy, only preventing breast cancer and not smaller risks, like influenza. That is why in this paper the function $\lambda_1(.)$ is considered to be constant, meaning that prevention only impacts the large claims. So, in the following, we consider an insurer with a surplus process given by

$$\begin{aligned} U(t,p) &= u + (c-p)t - \sum_{k=1}^{N(t)} X_k \\ &= u + (c-p)t - \sum_{i=1}^{N_1(t)} X_i^{(1)} - \sum_{j=1}^{N_2(t)} X_j^{(2)} \end{aligned} \tag{5.2}$$

where $\lambda_1(p) = \lambda_1 > 0$ and $\lambda_2(p)$ is a decreasing, strictly convex, positive, and $\mathcal{C}^2$ function defined on $[0, c]$. These constraints on $\lambda_2(.)$ are similar to the ones imposed on $\lambda(.)$ in (Gauchon et al., 2019) and we refer the reader to this latter paper for a discussion about these assumptions. In order to avoid ruin with certainty, we define $p_{lim} \in [0, c]$ as the solution of

$$\frac{\lambda_1\mu_1 + \lambda_2(p_{lim})\mu_2}{c - p_{lim}} = 1 \tag{5.3}$$

and we require $p < p_{lim}$. Since $\lambda_2(.)$ is continuous and decreasing, the intermediate value theorem ensures the existence of p_{lim} when $\frac{\lambda_1\mu_1+\lambda_2(0)\mu_2}{c} < 1$, which is assumed in the rest of the paper.

In this model, we notice that the prevention does not only act on the number of claims but also on the distribution F_X of the claim severities X_k. In the following, we denote by $X(p)$ the random variable X to make explicit the link with the prevention amount p.

In this paper, we are interested in the prevention amount $p^*(u)$ that minimizes the ruin probability

$$\psi(u,p) = \mathbb{P}\left(\exists t > 0 \text{ such that } U(t,p) < 0\right).$$

It is well-known that $\psi(u,p)$ coincides with the tail function of a compound geometric distribution, namely

$$\psi(u,p) = \mathbb{P}\left(\sum_{j=1}^{M} D_j > u\right), \tag{5.4}$$

where M follows the geometric distribution with success probability $\varphi(0,p) = 1 - \psi(0,p)$ and where the random variables $D_1, D_2, \ldots$ are called ladder heights and are independent and identically distributed as the random variable D, also denoted $D(p)$ in the following to make explicit the dependence with respect to p. In our compound Poisson model, the cumulative distribution function of $D(p)$ is given by the integrated tail distribution

$$F_{D(p)}(u) = \int_0^u \frac{1 - F_X(x)}{\mathbb{E}(X)} dx, \quad u > 0. \tag{5.5}$$

From (5.4), we get the well-known Pollaczeck Khinchin formula

$$\varphi(u,p) = \varphi(0,p)\sum_{n=0}^{\infty}(1-\varphi(0,p))^n F_{D(p)}^{*n}(u), \tag{5.6}$$

where $F_{D(p)}^{*n}$ is the n-fold convolution of $F_{D(p)}$. Let us remark that this formula guarantees the existence of the derivative $\frac{\partial\varphi(u,p)}{\partial p}$ when $\frac{\partial F_{D(p)}(u)}{\partial p}$ exists.

Denoting by $M_X(.)$ the moment generating function of X, we can define for our model the well-known adjustment coefficient $\kappa(p)$, which verifies

$$\lambda(p) + (c-p)\kappa(p) = \lambda(p)M_X(\kappa(p)). \tag{5.7}$$

It always exists when

$$M_X(s) < \infty \quad \text{for all } s \in \mathbb{R} \tag{5.8}$$

or when there exists $s^* > 0$ such that

$$M_X(s) < \infty \quad \text{for all } s < s^* \quad \text{and} \quad M_X(s) = \infty \quad \text{for all } s \geq s^*. \tag{5.9}$$

The reader could refer to the book of Asmussen and Albrecher (Asmussen and Albrecher, 2010) for more details.

Notice that models with multiple risks have been introduced by Cramér in 1955 (Cramér, 1955) and have been used for example by Dickson and Gray (Dickson and Gray, 1984), Bowers et al. (Bowers et al., 1984) or Gerber et al. (Gerber et al., 1987).

5.3 Preliminaries

Before studying the optimal prevention amount $p^*(u)$, we recall in this section the notion of Harmonic Mean Residual Life (HMRL) order (more details can be found in Shaked and Shanthikumar (Shaked and Shanthikumar, 2007)). Also, we prove two results that will be useful in the subsequent analysis

Definition 9. Given two non-negative random variables $X^{(1)}$ and $X^{(2)}$ with respective cumulative distribution functions $F_{X^{(1)}}$ and $F_{X^{(2)}}$, recall that $X^{(1)}$ is said to be smaller than $X^{(2)}$ in the harmonic mean residual life order (denoted as $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$) when

$$\frac{\int_t^\infty \overline{F}_{X^{(1)}}(u)\,du}{\mathbb{E}(X^{(1)})} \leq \frac{\int_t^\infty \overline{F}_{X^{(2)}}(u)\,du}{\mathbb{E}(X^{(2)})} \quad \text{for all } t \geq 0, \tag{5.10}$$

where $\overline{F}_{X^{(1)}} = 1 - F_{X^{(1)}}$ and $\overline{F}_{X^{(2)}} = 1 - F_{X^{(2)}}$. The inequality in (5.10) can be equivalently written as

$$\frac{\mathbb{E}((X^{(1)}-t)_+)}{\mathbb{E}(X^{(1)})} \leq \frac{\mathbb{E}((X^{(2)}-t)_+)}{\mathbb{E}(X^{(2)})} \quad \text{for all } t \geq 0. \tag{5.11}$$

Below we give some examples taken from Heilmann and Schröter (Heilmann and Schröter, 1991) where $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ holds true:

— If $X^{(1)}$ is Uniformly distributed over the interval $[a, b]$ and $X^{(2)}$ is Uniformly distributed over the interval (a', b'), then $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ if and only if $a + b \leq a' + b'$ and $b \leq b'$.

— If $X^{(1)}$ is Exponentially distributed with mean $1/a$ and $X^{(2)}$ is Exponentially distributed with mean $1/a'$ then $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ if and only if $a \geq a'$.

— If $X^{(1)}$ is Pareto distributed with parameters a and b, that is, $X^{(1)}$ has distribution function $1 - (\frac{a}{x})^b$, $x > a$, and if $X^{(2)}$ is Pareto distributed with parameters a' and b' with $\min(b, b') > 1$, then $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ if and only if $\frac{b-1}{b'-1} \geq \max(\frac{a}{a'}, 1)$

Next to these examples, we can also add the next result that will be of interest later on.

Proposition 10. *If $X^{(1)}$ follows a translated Exponential distribution over the interval $[a, \infty]$ with mean $a + \frac{1}{\lambda}$ and $X^{(2)}$ is Pareto distributed with parameters a and b, then $X^{(1)} \preceq_{hmrl} X^{(2)}$ if and only if $b \leq \lambda a + 1$.*

Proof. Since

$$F_{X^{(1)}}(x) = 1 - e^{-\lambda(x-a)}, \quad x \geq a \tag{5.12}$$

and

$$F_{X^{(2)}}(x) = 1 - \left(\frac{a}{x}\right)^b \quad x \geq a, \tag{5.13}$$

inequality (5.10) tells us that $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ if and only if

$$\frac{e^{\lambda(a-t)}}{\lambda a + 1} \leq \left(\frac{a}{t}\right)^{b-1} \frac{1}{b}, \quad \text{for all } t \geq a. \tag{5.14}$$

In particular, inequality (5.14) must hold for $t = a$. A necessary condition for $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ is thus given by

$$\frac{1}{\lambda a + 1} \leq \frac{1}{b}. \tag{5.15}$$

Let us now show that (5.15) is actually a sufficient condition for $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$. Inequality 5.14 can be rewritten as

$$log\left(\frac{b}{\lambda a + 1}\right) + \lambda(a - t) - (b - 1)\left[log(a) - log(t)\right] \leq 0 \quad \text{for all } t \geq a. \tag{5.16}$$

Considering $t > a$ and introducing the function $g(t) = log(\frac{b}{\lambda a+1}) + \lambda(a-t) - (b-1)[log(a) - log(t)]$, we easily see that $g'(t) < 0$ if and only if $\frac{b-1}{\lambda} < t$, which is always true since (5.15) yields $\frac{b-1}{\lambda} \leq a < t$. So, it comes $g'(t) < 0$ for all $t > a$. Now, as (5.15) implies that $g(a) < 0$, we finally get $g(t) < 0$ for all $t \geq a$. ■

Notice that if $b > 1$, $\mathbb{E}(X^{(1)}) \leq \mathbb{E}(X^{(2)})$ if and only if $1 + \lambda a \leq b\frac{a\lambda}{b-1}$. Indeed, multiplying both sides of the last inequality by $b - 1$ shows that $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ is equivalent to $\mathbb{E}(X^{(1)}) \leq \mathbb{E}(X^{(2)})$.

The next proposition will be useful to prove Proposition 13. Recall that one says that $X^{(1)}$ is smaller than $X^{(2)}$ in the usual stochastic order, denoted $X^{(1)} \preceq_{\text{st}} X^{(2)}$, when $F_{X^{(1)}}(t) \geq F_{X^{(2)}}(t)$ for all t.

Proposition 11. *Let $p_1 \leq p_2$. We have $X^{(1)} \preceq_{hmrl} X^{(2)}$ if and only if $D(p_2) \preceq_{st} D(p_1)$.*

Proof. We know from Theorem 2.B.2. in Shaked and Shanthikumar (Shaked and Shanthikumar, 2007) that

$$X(p_2) \preceq_{\text{hmrl}} X(p_1) \Leftrightarrow D(p_2) \preceq_{\text{st}} D(p_1). \tag{5.17}$$

So, it suffices to show that $X^{(1)} \preceq_{\text{hmrl}} X^{(2)} \Leftrightarrow X(p_2) \preceq_{\text{hmrl}} X(p_1)$. Let us denote $l(p) = \frac{\lambda_1\mu_1}{\lambda_1\mu_1+\lambda_2(p)\mu_2}$ for $p \in (0,1)$. We notice that $l(p_1) \leq l(p_2)$ since $\lambda_2(p_1) \geq \lambda_2(p_2)$. Considering the "$\Rightarrow$" part, as (5.10) holds, we clearly have

$$\begin{aligned}
\frac{\int_t^\infty \overline{F}_{X(p_2)}(u)\,du}{\mathbb{E}(X(p_2))} &= \frac{\lambda_1 \int_t^\infty \overline{F}_{X^{(1)}}(u)\,du + \lambda_2(p_2)\int_t^\infty \overline{F}_{X^{(2)}}(u)\,du}{\lambda_1\mu_1 + \lambda_2(p_2)\mu_2} \\
&= l(p_2)\frac{\int_t^\infty \overline{F}_{X^{(1)}}(u)\,du}{\mu_1} + (1-l(p_2))\,\frac{\int_t^\infty \overline{F}_{X^{(2)}}(u)\,du}{\mu_2} \\
&\leq l(p_1)\frac{\int_t^\infty \overline{F}_{X^{(1)}}(u)\,du}{\mu_1} + (1-l(p_1))\,\frac{\int_t^\infty \overline{F}_{X^{(2)}}(u)\,du}{\mu_2} \\
&= \frac{\int_t^\infty \overline{F}_{X(p_1)}(u)\,du}{\mathbb{E}(X(p_1))}.
\end{aligned}$$

Turning to the "$\Leftarrow$" part, it suffices to notice that $\frac{\int_t^\infty \overline{F}_{X(p_2)}(u)\,du}{\mathbb{E}(X(p_2))} \leq \frac{\int_t^\infty \overline{F}_{X(p_1)}(u)\,du}{\mathbb{E}(X(p_1))}$ yields

$$(l(p_2)-l(p_1))\left(\frac{\int_t^\infty \overline{F}_{X^{(2)}}(u)\,du}{\mu_2} - \frac{\int_t^\infty \overline{F}_{X^{(1)}}(u)\,du}{\mu_1}\right) \geq 0, \tag{5.18}$$

which ends the proof. ■

5.4 Optimal prevention amount $p^*(u)$

Let us first consider an insurer with no initial surplus. We then have the following result.

Proposition 12. *The optimal prevention amount $p^*(0)$ is positive if and only if*

$$-\lambda_2'(0) > \frac{\lambda_1\mu_1 + \lambda_2(0)\mu_2}{\mu_2 c}. \tag{5.19}$$

In this case, $p^*(0)$ satisfies

$$-\lambda_2'(p^*(0)) = \frac{\lambda_1\mu_1 + \lambda_2(p^*(0))\mu_2}{\mu_2(c-p^*(0))}. \tag{5.20}$$

Proof. Starting from

$$\varphi(0,p) = 1 - \frac{\mu_1\lambda_1 + \mu_2\lambda_2(p)}{c-p}, \tag{5.21}$$

we get

$$\frac{\partial\varphi(0,p)}{\partial p} = -\frac{\lambda_1\mu_1}{(c-p)^2} - \frac{\lambda_2(p)\mu_2}{(c-p)^2} - \frac{\lambda_2'(p)\mu_2}{c-p} \tag{5.22}$$

and

$$\frac{\partial^2\varphi(u,p)}{\partial p^2} = \frac{2}{c-p}\frac{\partial\varphi(0,p)}{\partial p} - \frac{\lambda_2''(p)\mu_2}{c-p}. \tag{5.23}$$

Now, similarly to the proof of Proposition 1 in (Gauchon et al., 2019), it suffices to see that Equations (5.22) and (5.23) enable us to prove that if $\frac{\partial\varphi(0,0)}{\partial p} \leq 0$, then $\frac{\partial\varphi(0,p)}{\partial p} < 0$ and $\frac{\partial^2\varphi(0,p)}{\partial p^2} < 0$ for all $p > 0$. ■

Hence, if Condition (5.19) is fulfilled, an insurer with no initial surplus can decrease its risk by investing a part of its premiums in prevention.

Notice that Condition (5.19) is slightly easier to fulfill than Condition (10) in (Gauchon et al., 2019), suggesting that in some cases, it is better to concentrate the effort for preventing large claims rather than trying to invest money for preventing both small and large claims.

Next to Condition (5.19), the following result gives an additional sufficient condition ensuring that an insurer with a positive initial surplus can also decrease its risk through prevention.

Proposition 13. *If Condition (5.19) is verified and if* $X^{(1)} \preceq_{hmrl} X^{(2)}$*, then* $p^*(u) > 0$ *for all* $u \geq 0$.

Proof. By definition, $\varphi(u, p_{lim}) = 0$ and $p < p_{lim}$ ensures that $\varphi(u,0) > 0$. As Condition (5.19) holds true, we know from the boundedness theorem that $\varphi(u,.)$ reaches a maximum for one $p \in [0, p_{lim}]$.

Now, it suffices to prove that $\frac{\partial\varphi(0,p)}{\partial p} > 0$ implies $\frac{\partial\varphi(u,p)}{\partial p} > 0$ and the proof will be completed. From the Pollaczeck Khinchin formula (5.6), we get

$$\begin{aligned}\frac{\partial\varphi(u,p)}{\partial p} &= \frac{\partial\varphi(0,p)}{\partial p}\sum_{n=0}^{\infty}(1-\varphi(0,p))^n F_{D(p)}^{*n}(u)\left(1 - \frac{\varphi(0,p)}{1-\varphi(0,p)}n\right) \\ &\quad +\varphi(0,p)\sum_{n=0}^{\infty}(1-\varphi(0,p))^n\frac{\partial F_{D(p)}^{*n}(u)}{\partial p}.\end{aligned} \tag{5.24}$$

In a first time, let us show that

$$\sum_{n=0}^{\infty}(1-\varphi(0,p))^n F_{D(p)}^{*n}(u)\left(1 - \frac{\varphi(0,p)}{1-\varphi(0,p)}n\right) \geq 0. \tag{5.25}$$

We have

$$\begin{aligned}
\sum_{n=0}^{\infty}(1-\varphi(0,p))^n\left(1-\frac{\varphi(0,p)}{1-\varphi(0,p)}n\right) &= \sum_{n=0}^{\infty}(1-\varphi(0,p))^n\left(1+n-\frac{n}{1-\varphi(0,p)}\right)\\
&= \sum_{n=0}^{\infty}(1-\varphi(0,p))^n(1+n)-\sum_{n=0}^{\infty}(1-\varphi(0,p))^{n-1}n\\
&= 1+\sum_{n=1}^{\infty}(1-\varphi(0,p))^n(1+n)-\sum_{n=1}^{\infty}(1-\varphi(0,p))^{n-1}n\\
&= 1-1+\sum_{n=1}^{\infty}(1-\varphi(0,p))^{n-1}n-\sum_{n=1}^{\infty}(1-\varphi(0,p))^{n-1}n\\
&= 0\\
&= 1+\sum_{n=1}^{\infty}(1-\varphi(0,p))^n\left(1-\frac{\varphi(0,p)}{1-\varphi(0,p)}n\right).
\end{aligned}$$

Moreover, we notice that $F_{D(p)}^{*n+1}(u) \leq F_{D(p)}^{*n}(u) \leq 1$ for all $n \in \mathbb{N}^+$. Hence,

$$\sum_{n=0}^{\infty}(1-\varphi(0,p))^n F_{D(p)}^{*n}(u)\left(1-\frac{\varphi(0,p)}{1-\varphi(0,p)}n\right) = 1+\sum_{n=1}^{\infty}(1-\varphi(0,p))^n F_{D(p)}^{*n}(u)\left(1-\frac{\varphi(0,p)}{1-\varphi(0,p)}n\right) \geq 0, \tag{5.26}$$

so that

$$\sum_{n=0}^{\infty}(1-\varphi(0,p))^n F_{D(p)}^{*n}(u)\left(1-\frac{\varphi(0,p)}{1-\varphi(0,p)}n\right) \geq 0. \tag{5.27}$$

In a second time, it is easy to see that

$$\varphi(0,p)\sum_{n=0}^{\infty}(1-\varphi(0,p))^n\frac{\partial F_{D(p)}^{*n}(u)}{\partial p} \geq 0. \tag{5.28}$$

Indeed, as the usual stochastic order is closed under convolution, Proposition 11 directly leads to $\frac{\partial F_{D(p)}^{*n}(u)}{\partial p} \geq 0$, which completes the proof. ■

Thus, in case (5.19) and $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ hold true, we know that $p^*(u) > 0$ for all u. We learn from the next proposition that the minimum of $p^*(u)$ is reached in $u = 0$. Notice that contrary to (Gauchon et al., 2019), one observes that $p^*(u)$ is not constant anymore in the present context.

Proposition 14. *If (5.19) holds true and if we have $X^{(1)} \preceq_{hmrl} X^{(2)}$, then $p^*(0) < p^*(u)$ for all $u > 0$.*

Proof. From Equation (5.24) and inequalities (5.27) and (5.28), one sees that $\frac{\partial\varphi(0,p^*(u))}{\partial p} < 0$ is a necessary condition for $\frac{\partial\varphi(u,p^*(u))}{\partial p} = 0$. Now, Equations (5.22) and (5.23) imply that $\varphi(0,.)$ is increasing on $[0,p^*(0)]$ and decreasing on $[p^*(0),p_{lim}]$. Thus, $\frac{\partial\varphi(0,p^*(u))}{\partial p} < 0$ is only possible when $p^*(u) > p^*(0)$. ■

In particular, as we will see later on, the asymptotic values derived for $p^*(.)$ in Propositions 15 and 16 are indeed larger than $p^*(0)$, which enables to think that $p^*(.)$ is increasing with u. However, we are not in position to prove this assertion.

5.5 $p^*(u)$ when $\kappa(p)$ exists for all $p \in [0, c]$

In this Section, we consider claim severities $X^{(1)}$ and $X^{(2)}$ such that the adjustment coefficient $\kappa(p)$ exists for all $p \in [0, c]$. This requirement is for example met when both random variables $X^{(1)}$ and $X^{(2)}$ are Exponentially distributed, which is a particular case of models considered in Gerber et al. (Gerber et al., 1987) and Dufresne and Gerber(Dufresne and Gerber, 1988).

When the initial surplus u goes to infinity, we show in the next proposition that the optimal prevention amount $p^*(u)$ converges to the one that maximizes the adjustment coefficient $\kappa(p)$, denoted p^*_κ.

Proposition 15. *The prevention amount p^*_κ maximizing the adjustment coefficient $\kappa(p)$ is solution of the equation*

$$-\lambda_2'(p)[(c-p)+\lambda_1\mathbb{E}(e^{\kappa(p)X(p)}(X^{(1)}-X^{(2)}))] = \lambda_1+\lambda_2(p). \tag{5.29}$$

Moreover, we have

$$\lim_{u\to\infty} p^*(u) = p^*_\kappa. \tag{5.30}$$

Proof. The adjustment coefficient $\kappa(p)$ verifies

$$\lambda_1+\lambda_2(p)+(c-p)\kappa(p) = (\lambda_1+\lambda_2(p))M_{X(p)}(\kappa(p)). \tag{5.31}$$

Deriving (5.31) with respect to p yields

$$\begin{aligned}\lambda_2'(p)\left(1-M_X(\kappa(p))\right)-\kappa(p) \quad &= \quad (\lambda_1+\lambda_2(p))\left(\frac{\lambda_1\kappa(p)}{\lambda_1+\lambda_2(p)}\right)' M'_{X^{(1)}}\left(\frac{\lambda_1\kappa(p)}{\lambda_1+\lambda_2(p)}\right) M_{X^{(2)}}\left(\frac{\lambda_2(p)\kappa(p)}{\lambda_1+\lambda_2(p)}\right)\\ &\quad +(\lambda_1+\lambda_2(p))\left(\frac{\lambda_2(p)\kappa(p)}{\lambda_1+\lambda_2(p)}\right)' M_{X^{(1)}}\left(\frac{\lambda_1\kappa(p)}{\lambda_1+\lambda_2(p)}\right) M'_{X^{(2)}}\left(\frac{\lambda_2(p)\kappa(p)}{\lambda_1+\lambda_2(p)}\right).\end{aligned} \tag{5.32}$$

Now, Equation (5.31) can be rewritten as

$$(M_X(\kappa(p))-1) = \frac{(c-p)\kappa(p)}{\lambda_1+\lambda_2(p)}. \tag{5.33}$$

So, combining (5.32) and (5.33) and taking $p = p^*_\kappa$, we get

$$-\lambda_2'(p^*_\kappa)[(c-p^*_\kappa)+\lambda_1\mathbb{E}(e^{\kappa(p^*_\kappa)X(p^*_\kappa)}(X^{(1)}-X^{(2)}))] = \lambda_1+\lambda_2(p^*_\kappa) \tag{5.34}$$

since $\kappa'(p^*_\kappa) = 0$. Notice that Equation (5.34) admits a unique solution when Condition (5.19) holds true.

Let us now prove that $\lim_{u\to\infty} p^*(u) = p^*_\kappa$. It is well-known that $\varphi(u,p)$ can be written as

$$\varphi(u,p) = 1-e^{-\kappa(p)u}\mathbb{E}_{\kappa(p)}(e^{-\kappa(p)\xi(u)}), \tag{5.35}$$

where $\xi(u)$ is the random variable representing the overshoot in case of ruin and $\mathbb{E}_{\kappa(p)}$ is the expected value computed under a change of probability measure using the exponentials families (for more details, we refer the reader to Section 4.4 and Equations (5.4) to (5.6) in Section 4.5 of Albrecher and Asmussen (Asmussen and Albrecher, 2010)). Thus, for all $u > 0$, we have

$$\begin{aligned}\frac{\partial\varphi(u,p^*(u))}{\partial p} &= -\kappa'(p^*(u))ue^{-\kappa(p^*(u))u}\mathbb{E}_{\kappa(p)}(e^{-\kappa(p^*(u))\xi(u)}) \\ &\quad +e^{-\kappa(p^*(u))u}\frac{\partial\mathbb{E}_{\kappa(p)}(e^{-\kappa(p^*(u))\xi(u)})}{\partial p} \\ &= 0, \end{aligned} \tag{5.36}$$

from which we deduce

$$\frac{\frac{\partial\mathbb{E}_{\kappa(p)}\left(e^{-\kappa(p^*(u))\xi(u)}\right)}{\partial p}}{u\mathbb{E}_{\kappa(p)}(e^{-\kappa(p^*(u))\xi(u)})} = \kappa'(p^*(u)). \tag{5.37}$$

Moreover, the Cramér-Lundberg approximation shows that

$$\lim_{u\to\infty}\mathbb{E}_{\kappa(p)}\left(e^{-\kappa(p)\xi(u)}\right) = C(p) = (c-p)\frac{\varphi(0,p)}{(\lambda_1+\lambda_2(p))M_X'(\kappa(p)) - c + p} > 0. \tag{5.38}$$

Now, since p is bounded, $C(p)$ is bounded as well. Furthermore, we get

$$\lim_{u\to\infty}\frac{\partial\mathbb{E}_{\kappa(p)}(e^{-\kappa(p_\kappa^*)\xi(u)})}{\partial p} = C'(p_\kappa^*) \in \mathbb{R}. \tag{5.39}$$

Then, combining (5.38) and (5.39) leads to

$$\lim_{u\to\infty}\frac{\frac{\partial\mathbb{E}_{\kappa(p)}(e^{-\kappa(p^*(u))\xi(u)})}{\partial p}}{u\mathbb{E}_{\kappa(p)}(e^{-\kappa(p^*(u))\xi(u)})} = \lim_{u\to\infty}\frac{C'(p^*(u))}{uC(p^*(u))} = 0. \tag{5.40}$$

As Equation (5.37) holds true for all $u > 0$, (5.40) shows that

$$\lim_{u\to\infty}\kappa'(p^*(u)) = 0.$$

Because p_κ^* is the unique positive solution of $\kappa'(p) = 0$, we then get $\lim_{u\to\infty} p^*(u) = p_\kappa^*$. ■

5.6 $p^*(u)$ when $\kappa(p)$ does not exist

Let us now consider the case where $X^{(2)}$ follows a sub-exponential distribution and $X^{(1)}$ a light-tailed distribution. This is for example the case when $X^{(1)}$ follows a translated Exponential distribution over the interval $[a,\infty]$ with mean $a + \frac{1}{\lambda}$ and $X^{(2)}$ a Pareto distribution of parameters a and b. Then, the stochastic inequality $X^{(1)} \preceq_{\text{hmrl}} X^{(2)}$ holds true when $b \leq \lambda a + 1$, as proved in Proposition 10.

In such a situation, the moment generating function $M_X(.)$ does not exist anymore. However, we can still determine the optimal prevention amount when u goes to infinity, as shown in the next proposition.

Proposition 16. *If Condition (5.19) holds true, if* $X^{(1)} \preceq_{hmrl} X^{(2)}$ *and if* $D(p)$ *follows a sub-exponential distribution for all* p*, then* $\lim_{u\to\infty} p^*(u) = p_\infty > 0$*, where* p_∞ *is solution of the equation*

$$-\lambda_2'(p)\left[\frac{\mu_2}{\varphi(0,p)} + \frac{\lambda_1\mu_1}{\lambda_2(p)}\right] = \frac{\lambda_2(p)\mu_2 + \lambda_1\mu_1}{\varphi(0,p)(c-p)}. \tag{5.41}$$

Proof. Proposition 13 guarantees the existence of an optimal prevention amount for all u. As seen in (5.24), we have

$$\begin{aligned}\frac{\partial\psi(u,p)}{\partial p} &= \frac{\partial\varphi(0,p)}{\partial p}\sum_{n=0}^{\infty}(1-\varphi(0,p))^n\overline{F}^{*n}_{D(p)}(u))\left(1-\frac{\varphi(0,p)}{1-\varphi(0,p)}n\right)\\ &\quad+\varphi(0,p)\sum_{n=0}^{\infty}(1-\varphi(0,p))^n\frac{\partial\overline{F}^{*n}_{D(p)}(u)}{\partial p}.\end{aligned} \tag{5.42}$$

Moreover,

$$F_{D(p)}(u) = \frac{1}{\lambda_1\mu_1+\lambda_2(p)\mu_2}\int_0^u \lambda_1\overline{F}_{X^{(1)}}(t) + \lambda_2(p)\overline{F}_{X^{(2)}}(t)dt, \tag{5.43}$$

which yields

$$\frac{\partial F_{D(p)}(u)}{\partial p} = \frac{\lambda_2'(p)\lambda_1\mu_1}{\lambda_2(p)\left(\lambda_1\mu_1+\lambda_2(p)\mu_2\right)}\left(F_{D(p)}(u) - F_{D_{X^{(1)}}}(u)\right) \tag{5.44}$$

with $F_{D_{X^{(1)}}}(u) = \frac{1}{\mu_1}\int_0^u \overline{F}_{X^{(1)}}(t)dt$.

Also, we have

$$\frac{\partial\frac{F^{*n}_{D(p)}(u)}{F_{D(p)}(u)}}{\partial p} = \frac{\frac{\partial F^{*n}_{D(p)}(u)}{\partial p}F_{D(p)}(u) - \frac{\partial F_{D(p)}(u)}{\partial p}F^{*n}_{D(p)}(u)}{F^2_{D(p)}(u)}, \tag{5.45}$$

which leads to

$$\frac{\frac{\partial F^{*n}_{D(p)}(u)}{\partial p}}{F_{D(p)}(u)} = \frac{\frac{\partial\frac{F^{*n}_{D(p)}(u)}{F_{D(p)}(u)}}{\partial p}F^2_{D(p)}(u) + \frac{\partial F_{D(p)}(u)}{\partial p}F^{*n}_{D(p)}(u)}{F^2_{D(p)}(u)}. \tag{5.46}$$

Hence, combining (5.42), (5.44) and (5.46) and letting u goes to infinity, we finally obtain

$$\begin{aligned}\lim_{u\to\infty}\frac{\frac{\partial\psi(u,p)}{\partial p}}{\overline{F}_{D(p)}(u)} &= \frac{\partial\varphi(0,p)}{\partial p}\sum_{n=0}^{\infty}(1-\varphi(0,p))^n n - \frac{\partial\varphi(0,p)}{\partial p}\sum_{n=0}^{\infty}(1-\varphi(0,p))^{n-1}n^2\varphi(0,p)\\ &\quad+\varphi(0,p)\frac{\lambda_2'(p)\lambda_1\mu_1}{\lambda_2(p)(\lambda_1\mu_1+\lambda_2(p)\mu_2)}\sum_{n=1}^{\infty}(1-\varphi(0,p))^n n\\ &= -\frac{\frac{\partial\varphi(0,p)}{\partial p}}{\varphi(0,p)^2} + \frac{\lambda_2'(p)\lambda_1\mu_1}{\lambda_2(p)(\lambda_1\mu_1+\lambda_2(p)\mu_2)}\frac{(1-\varphi(0,p))}{\varphi(0,p)}.\end{aligned} \tag{5.47}$$

Now, taking $\frac{\partial\psi(u,p)}{\partial p} = 0$ gives the announced result. ■

5.7 Conclusion

In this paper, we have proposed a risk model with two types of risks where prevention reduces the claim intensity of the most severe risk. In this context, we have studied the ruin probability as a function of the prevention effort. A sufficient condition for prevention to be efficient has been provided, and both cases of light and heavy-tailed claims have been considered.

Moreover, we have shown that an insurer should invest more in prevention when its initial reserves are large enough. It is likely due to the fact that, when the reserves are huge, the small claims are less important since ruin is most likely to occur due to an extreme claim.

Finally, let us notice that it would be interesting to consider a model where the impact of prevention is not constant over time and where the uncertainty on prevention efficiency is taken into account. This is left for further research.

Acknowledgements : This paper was realized within the framework of the Chair Prevent'Horizon, supported by the risk foundation Louis Bachelier and in partnership with Claude Bernard Lyon 1 University, Actuaris, AG2R La Mondiale, G2S, Covea, Groupama Gan Vie, Groupe Pasteur Mutualité, Harmonie Mutuelle, Humanis Prévoyance and La Mutuelle Générale.

Declaration of interest : Romain Gauchon is employed by Actuaris. The paper has been supported by the Chair Prevent'Horizon.

Bibliographie

Asmussen, S. and Albrecher, H. (2010). Ruin probabilities, volume 14. World scientific.

Bowers, N. L., Gerber, H., Hickman, J., Jones, D., and Nesbitt, C. (1984). Actuarial mathematics.

Courbage, C. (2001). Self-insurance, self-protection and market insurance within the dual theory of choice. The GENEVA Papers on Risk and Insurance-Theory, 26(1) :43–56.

Courbage, C. and Rey, B. (2012). Optimal prevention and other risks in a two-period model. Mathematical Social Sciences, 63(3) :213–217.

Cramér, H. (1955). Collective risk theory : A survey of the theory from the point of view of the theory of stochastic processes. Nordiska bokhandeln.

Dickson, D. C. and Gray, J. (1984). Exact solutions for ruin probability in the presence of an absorbing upper barrier. Scandinavian Actuarial Journal, 1984(3) :174–186.

Dionne, G. and Eeckhoudt, L. (1985). Self-insurance, self-protection and increased risk aversion. Economics Letters, 17(1-2) :39–42.

Dufresne, F. and Gerber, H. U. (1988). The probability and severity of ruin for combinations of exponential claim amount distributions and their translations. Insurance : Mathematics and Economics, 7(2) :75–80.

Ehrlich, I. and Becker, G. S. (1972). Market insurance, self-insurance, and self-protection. Journal of political Economy, 80(4) :623–648.

Gauchon, R., Loisel, S., Rulliere, J.-L., and Trufin, J. (2019). Optimal prevention strategies in the classical risk model.

Gerber, H. U., Goovaerts, M. J., and Kaas, R. (1987). On the probability and severity of ruin. Astin Bulletin, 17(02) :151–163.

Heilmann, W.-R. and Schröter, K. J. (1991). Orderings of risks and their actuarial applications. Lecture Notes-Monograph Series, pages 157–173.

Shaked, M. and Shanthikumar, J. G. (2007). Stochastic orders. Springer Science & Business Media.

Conclusion et perspectives

Tout au long de cette thèse, divers aspect de la prévention du point de vue d'une société d'assurance ont été examinés.

Les deux premiers chapitres se sont intéressés à la classification d'assurés dans le domaine de l'assurance santé, afin par exemple de faciliter le ciblage d'une action de prévention. Si les méthodes de classification proposées permettent d'obtenir des classes d'une qualité satisfaisante, elles souffrent toutes d'une limitation majeure : elles sont monolabélisées. Or, le domaine de la santé appelle plutôt à une classification multilabélisée : il est plausible qu'un individu appartienne à la fois à la classe de consommateur d'optique et celle de bénéficiaire de prothèses dentaires. Cette limitation peut réduire la qualité du ciblage d'une action de prévention. Afin de compléter ces travaux, une approche floue ou multilabélisée pourrait donc être pertinente.

De plus, si le chapitre 2, à l'aide de l'algorithme Word2Vec, propose une méthode permettant de prendre en compte la dimension temporelle du problème, celle-ci souffre de problèmes de stabilité. D'autres approches, notamment basées sur la théorie des graphes, pourraient être envisagées afin d'utiliser au mieux les dates des prestations et potentiellement de retrouver les parcours de soin.

Le troisième chapitre s'intéresse au poids de la psychiatrie pour une compagnie d'assurance. En se basant sur des données d'assurance uniquement, il permet de conclure qu'un assuré qui va chez le psychiatre coûte en moyenne deux fois et demie plus cher qu'un assuré moyen.

Ce chapitre souffre de deux limitations. La première est la nature des données possédées par des assureurs. Du fait de leur rôle d'assureur payeur et de l'observation des données sur de faibles périodes temporelles, il n'est possible de détecter que des corrélations entre coûts et il est impossible de détecter certains soins psychiatriques effectués par les médecins généralistes. La seconde est la nécessité de suivre ce chapitre par des actions de prévention concrètes menées conjointement par des organismes assureurs et des psychiatres, couplées avec une analyse coût-bénéfice. Ceci permettrait de vérifier expérimentalement que la psychiatrie est bien un terrain sur lequel les sociétés d'assurance sont légitimes à intervenir.

Les chapitres 4 et 5 proposent une extension du modèle de Poisson composé afin d'étudier l'impact d'un paramètre de prévention. De nombreux travaux complémentaires peuvent être envisagés. L'un d'entre eux, par exemple, serait de démontrer que le montant de prévention optimale permettant de réduire la probabilité de ruine croît avec l'augmentation des réserves initiales.

De plus, le modèle proposé dans le chapitre 4 peut être étendu de nombreuses manières différentes, dont une est d'ailleurs explorée dans le chapitre 5. Une extension possible serait d'aller plus loin encore que le chapitre 2 et de proposer une réelle modélisation ou l'action de prévention est à la fois une action d'auto-assurance et d'auto-protection. D'autres extensions possibles peuvent consister à supposer que la prévention à un impact aléatoire et non plus déterministe, qu'elle a un effet variable au cours du temps, ou à imaginer des actions de prévention "one-shot", dont le coût serait payé en t=0 plutôt que de manière continue au cours du temps.

Enfin, le sujet de la prévention en assurance est extrêmement diversifié et comporte de très nombreuses problématiques. De ce fait, les différents chapitres de cette thèse n'ont pu traiter de toutes les problématiques associées. De nombreuses autres études doivent encore être menées afin de lever certains verrous pratiques. Par exemple, il serait intéressant de mettre en regard le financement d'une action de prévention et l'adhésion ou la participation des participants à celle-ci. Une action financée par la sécurité sociale ou par une ONG médicale aurait-t-elle plus de succès qu'une action financée par une société d'assurance ? Une action financée à 100 % par un organisme assureur serait-elle suffisament efficace ou faudrait-il demander systématiquement une participation aux assurés ?

Cette thèse s'inscrit finalement dans un cycle d'étude beaucoup plus large qui, il faut l'espérer, pourrait à terme convaincre les organismes assureurs d'assumer pleinement un rôle social dont bénéficierait l'ensemble de la société.