

Mémoire présenté devant l'Université de Paris-Dauphine  
pour l'obtention du Certificat d'Actuaire de Paris-Dauphine  
et l'admission à l'Institut des Actuaires

le 08 / 02 / 2021

Par : Céline HOUDAYER

Titre : Tarification en santé sous contrainte budgétaire d'un euro par jour à l'aide de fonctions d'utilité

Confidentialité :  Non  Oui (Durée :  1 an  2 ans)

---

*Les signataires s'engagent à respecter la confidentialité ci-dessus*

*Membres présents du jury de l'Institut  
des Actuaires :*

*Entreprise :*

Nom : Sia Partners

Signature :

Sia Partners SASU  
Capital: 200 000 Euros  
RCS B 423 507 730  
21 rue de Berri - 75008 PARIS

*Membres présents du Jury du Certificat  
d'Actuaire de Paris-Dauphine :*

*Directeur de Mémoire en entreprise :*

Nom : Nicolas SERVAN

Signature :



---

*Autorisation de publication et de mise en ligne sur un site de diffusion de documents  
actuariels (après expiration de l'éventuel délai de confidentialité)*

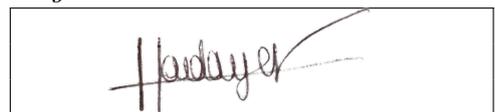
*Secrétariat :*

*Signature du responsable entreprise*



*Bibliothèque :*

*Signature du candidat*





## Résumé

---

En constante évolution, le système de protection sociale français est l'un des plus complets au monde. Le Gouvernement oeuvre à son bon fonctionnement et, dans le but de l'optimiser et d'assurer sa pérennité dans le temps, le réforme grâce à des projets tels que l'Accord National Interprofessionnel (ANI) de 2013 ou plus récemment, la création de la Complémentaire Santé Solidaire (CSS).

Dans la continuité de cette lignée et des aides accordées par l'État, ce mémoire vise à proposer une amélioration de la couverture complémentaire santé des Français aux revenus modestes grâce à la construction d'un contrat collectif à un euro par jour. Une question centrale se pose dans sa création, comment calibrer au mieux les garanties selon cette contrainte budgétaire? Les travaux menés proposent de changer de point de vue et de mettre l'utilité des assurés au cœur de la modélisation.

C'est notamment grâce à des techniques de tarification et des méthodes d'optimisation d'utilité sous contrainte qu'une approche modulable inversée a pu être mise en place afin de répondre à la problématique posée. Partant de la contrainte de prix, elle permet d'obtenir la grille de garanties optimale qui maximise l'espérance d'utilité des assurés. Au travers d'une description actuarielle détaillée, ce mémoire retrace les étapes clés de cette approche qui repose sur des modélisations préliminaires portant sur la dépense engagée, le remboursement obligatoire et les appétences des assurés.

---

*Mots-clés : assurance santé, complémentaire santé, dépense engagée, remboursement obligatoire, remboursement complémentaire, reste à charge, tarification, modèles linéaires généralisés, fonctions d'utilité, optimisation sous contrainte, modèle Cobb-Douglas.*

## Abstract

---

Constantly evolving, the French social protection system is one of the most comprehensive in the world. The Government is working on its proper functioning and, in order to optimize it and ensure its sustainability over time, reforms it through projects such as the National Interprofessional Agreement (ANI) of 2013 or more recently, the creation of the Complementary Solidarity Health Insurance (CSS).

Continuing on from this lineage and the aids granted by the State, this thesis aims to propose an improvement in the complementary health coverage of French people with modest incomes through the construction of a collective contract for one euro per day. A central question arises in its creation, how to best calibrate the guarantees according to this budget constraint? The work carried out proposes to change the prism and put the utility of policyholders at the heart of the modeling.

It is in particular thanks to pricing techniques and constrained utility optimization methods that an inverted modular approach has been put in place in order to respond to the problem posed. Starting from the price constraint, it makes it possible to obtain the optimal guarantee grid which maximizes the expected utility of the insured. Through a detailed actuarial description, this thesis retraces the key stages of this approach, which is based on preliminary models relating to the expenditure incurred, the compulsory reimbursement and the preferences of the insured.

---

*Keywords : health insurance, complementary health insurance, expenditure incurred, compulsory reimbursement, complementary reimbursement, out-of-pocket expense, pricing, generalized linear models, utility functions, constrained optimization, Cobb-Douglas model.*

# Note de Synthèse

Bien que l'impact de la crise inédite de 2020 ne soit pas encore mesurable, elle permet de pouvoir d'ores et déjà attester de la nécessité d'un système de protection sociale efficace. En effet, le modèle français a permis d'amortir une partie des effets économiques et sanitaires de cette crise. Un important maillon de cette protection sociale est le système de santé. Ses deux principaux acteurs que sont la Sécurité sociale, reposant sur le principe de protection universelle, et les assureurs complémentaires, représentés par des organismes privés, oeuvrent pour couvrir au mieux les individus. Cependant, aujourd'hui encore, trois millions de personnes manquent de moyens financiers pour accéder à ce deuxième échelon de protection sociale.

C'est pourquoi ce mémoire porte sur la tarification d'un contrat collectif d'entrée de gamme respectant une contrainte budgétaire d'un euro par jour destiné aux salariés ayant des revenus modestes, ne pouvant s'offrir une complémentaire santé onéreuse mais n'entrant pas dans les conditions d'octroi de la Complémentaire Santé Solidaire (CSS) proposée par la Sécurité sociale. Afin de construire ce contrat de complémentaire santé, une approche audacieuse a été choisie : prendre en compte les appétences en santé des potentiels assurés au travers d'outils mathématiques venant de la théorie de la décision. Contrairement aux méthodes de construction utilisées habituellement, cette alternative permet de renouer avec la demande des clients.

L'ensemble des travaux de ce mémoire a été effectué à la maille sous-poste de soins. Cette maille plus fine que la maille poste de soins mais également plus large que la maille acte de soins permet à la fois de pouvoir mettre en place une modélisation opérationnelle et d'obtenir une analyse assez précise des différents risques en santé.

Au vu de la structure de la protection sociale en santé en France, la dépense engagée ainsi que le remboursement obligatoire ont été modélisés en amont de la création du contrat de complémentaire santé, comme le montre la figure 1 décomposant la dépense engagée d'un individu.

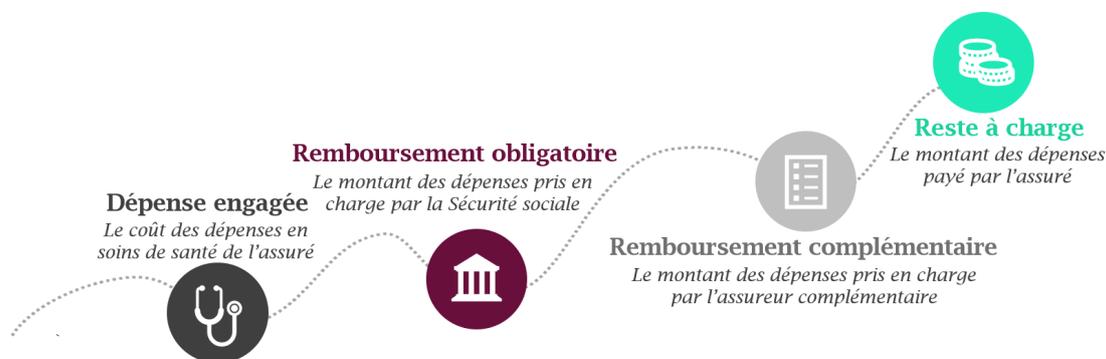


FIGURE 1: Décomposition de la dépense engagée.

## La modélisation de la dépense engagée

Une approche coût-fréquence utilisant des modèles linéaires généralisés a permis de modéliser la dépense engagée des 27 sous-postes de soins. En amont de ces modélisations, les données issues des systèmes de gestion d'un assureur complémentaire, portant sur les exercices 2018 et 2019, ont d'abord été analysées puis retraitées. Ces retraitements, réalisés dans le but d'obtenir des bases de données se prêtant à la réalisation de modèles linéaires, ont fait appel à des données externes venant de l'INSEE et à des outils statistiques tels que les arbres de régression et l'algorithme des *k-means* afin de discrétiser des variables continues, l'âge des assurés et de créer des classes, à l'échelle départementale. De plus, les sinistres extrêmes ont été modélisés à part grâce à une surprime prenant en compte leur coût et leur fréquence pour obtenir un meilleur ajustement des modèles sensibles à ces valeurs. Une fois les données retraitées, les sinistres extrêmes enlevés et les variables non corrélées sélectionnées, les modèles de coût et de fréquence ont pu être réalisés.

Pour chaque modèle, deux lois candidates ont été utilisées et des techniques de sélection de variables telles que la méthode *Forward-Backward* ont été introduites. Le modèle le plus adapté a été retenu à l'aide d'indicateurs, notamment des métriques, et d'outils statistiques comme par exemple des histogrammes. Les coefficients de ces modèles ont permis d'identifier les facteurs explicatifs de la sinistralité associée à chaque sous-poste de soins. De plus, les limites de cette modélisation à la maille sous-poste de soins utilisant des modèles linéaires généralisés ont été analysées grâce à une étude des prédictions et des méthodes permettant de les améliorer, comme le *machine learning* ont été citées. Par exemple, une meilleure segmentation des actes de soins pourrait permettre d'obtenir de meilleurs modèles tarifaires.

Ainsi, ces modèles de fréquence et de coût ont permis de construire la sinistralité associée à chaque sous-poste de soins. Ils ont notamment révélé une forte sinistralité pour quatre sous-postes de soins, qui totalisent à eux seuls 40% de la consommation en santé : ce sont les frais de séjour hospitaliers ayant une queue de distribution lourde, les prothèses dentaires ayant un coût élevé, les consultations chez le spécialiste et les médicaments remboursés à 65% ayant une fréquence élevée. Au niveau des postes de soins, comme le montre la figure 2, ce sont les soins courants caractérisés par une fréquence élevée qui pèsent le plus dans la consommation en santé (34%) et ce poste peut être complété par les médicaments délivrés en pharmacie (11%). De plus, le poste pharmacie représente, en réalité, une part supérieure puisqu'ici les médicaments intégralement remboursés par la Sécurité sociale ne sont pas modélisés. En complément, le dentaire, l'hospitalisation et l'optique, tous trois caractérisés par des coûts élevés, représentent respectivement 23%, 15% et 12% de la consommation en santé.

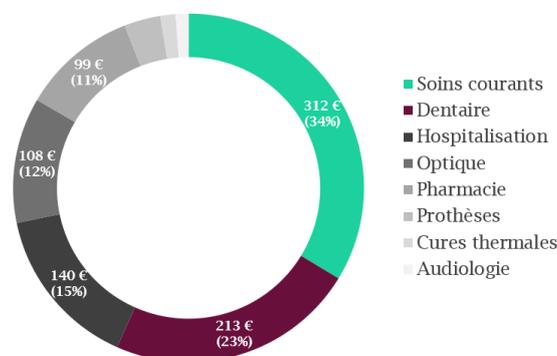


FIGURE 2: Décomposition de la dépense engagée en fonction des différents postes de soins.

## La modélisation du remboursement obligatoire

D'un acte de soins à l'autre, les règles de remboursement de la Sécurité sociale varient de par leur expression et de par leur montant. C'est pourquoi en raison des milliers d'actes de soins référencés et de la modélisation réalisée à la maille sous-poste de soins, un indicateur du taux de remboursement obligatoire à cette même maille a été créé.

Excepté pour les 8 sous-postes de soins caractérisés par un taux de remboursement explicite, un indicateur intuitif est la moyenne des taux de remboursement du sous-poste. Afin de valider cette méthode, la dépense engagée de chaque sous-poste a été segmentée en fonction des taux de remboursement puis le taux de couverture de chaque tranche de dépense engagée a été comparé au taux de couverture moyen. Pour 12 sous-postes de soins, les taux des tranches oscillent légèrement autour de la moyenne sans dévoiler de tendance particulière, comme c'est le cas pour le graphique de gauche de la figure 3, ce qui montre que cet indicateur intuitif est bien adapté. Tandis que pour les 7 autres sous-postes de soins, une tendance linéaire décroissante est apparue, comme pour le graphique de droite. Ce sont majoritairement des sous-postes caractérisés par un faible taux de remboursement de la part de la Sécurité sociale et des coûts hétérogènes pouvant atteindre de fortes valeurs. Ainsi, l'indicateur de ces sous-postes est une droite affine décroissante dépendant de la dépense engagée.

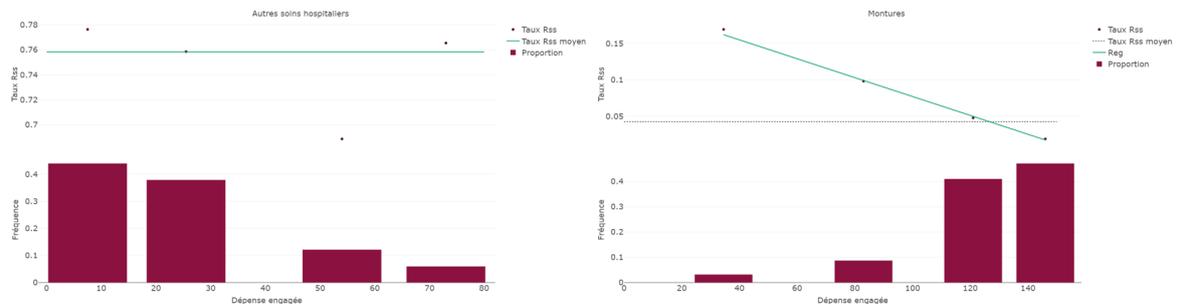


FIGURE 3: Visualisation graphique des taux de couverture des différentes tranches de dépense engagée en fonction de l'indicateur créé pour deux sous-postes de soins.

Il apparaît que les soins les moins bien remboursés sont ceux concernés par la réforme 100% Santé. De plus, les frais de séjour comptent, à nouveau, parmi les sous-postes ayant la plus forte sinistralité bien qu'ils bénéficient d'une prise en charge importante. C'est également le cas des prothèses dentaires et des verres optiques très faiblement remboursés. A une échelle plus large, les soins courants voient leur importance baisser à 27% grâce à une prise en charge importante, a contrario du poste optique qui double et passe à 20% de la consommation en santé.

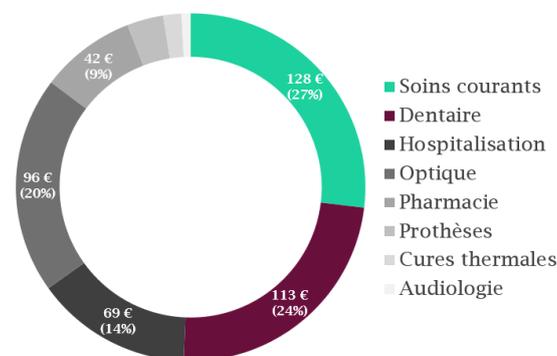


FIGURE 4: Décomposition de la dépense engagée après remboursement obligatoire en fonction des différents postes de soins.

## La modélisation répondant à la problématique posée

La prime commerciale du contrat s'élève à 360€ par an. Puisque ce montant inclus 20% de charges, le budget disponible pour la tarification est donc inférieur et s'élève à 300€. Afin de résoudre la problématique posée dans ce mémoire, le modèle d'optimisation d'utilité sous contrainte budgétaire Cobb-Douglas a été utilisé pour obtenir la répartition optimale de ces 300€ :

$$\max \prod_{i=1}^{27} RC_i^{Poids_i} \text{ s.c. } \sum_{i=1}^{27} RC_i = 300. \quad (1)$$

où  $RC_i$  représente l'espérance de remboursement complémentaire du sous-poste  $i$ . Les poids du modèle ont été obtenus à l'aide d'un questionnaire qui a permis de connaître les importances absolues qu'accordent les sondés pour les postes et sous-postes de soins. Elles ont été retraitées, comme le montre la figure 5, afin d'obtenir les poids utilisés dans le modèle d'optimisation représentant l'utilité. De plus, les analyses descriptives des réponses ont permis de souligner les variations d'appétences entre les sous-postes de soins et ont ainsi confirmé l'importance de ce facteur dans la tarification. Enfin, des intervalles de confiance ont été construits afin de mesurer le biais causé par la fluctuation d'échantillonnage.

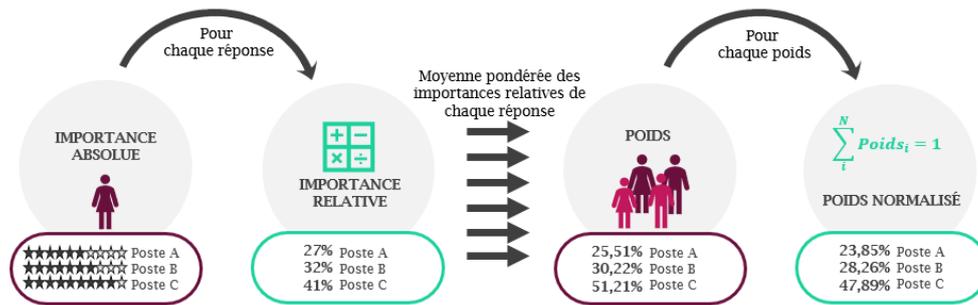


FIGURE 5: Schéma résumant les étapes permettant de calculer les poids des postes de soins.

La répartition obtenue par l'algorithme d'optimisation vérifie les propriétés de la théorie de la décision : tous les sous-postes se voient attribués un budget strictement positif et ceux ayant des poids d'importance plus élevés bénéficient de montants supérieurs.

Des calculs permettant d'égaliser l'espérance de remboursement complémentaire au budget déterminé par l'algorithme ont ensuite été réalisés afin de pouvoir calibrer les garanties du contrat. Ces garanties sont exprimées sous deux formes : le forfait, un montant fixe de remboursement, et le pourcentage de remboursement de la dépense engagée. Suivant le type de garantie, le calcul mis en place diffère pour passer du montant de remboursement complémentaire moyen à la garantie du contrat. Dans le cas d'un forfait, le calcul d'une espérance empirique est effectué pour calculer une espérance non-linéaire. De plus, l'algorithme d'optimisation a été enrichi de contraintes afin de pouvoir obtenir des garanties, c'est pourquoi certaines variables du problème d'optimisation sont saturées.

Ainsi, la garantie permettant d'égaliser le budget donné par l'algorithme d'optimisation dépend de la distribution de la sinistralité ainsi que du remboursement obligatoire. Par exemple, plus un sous-poste aura des coûts élevés après le remboursement obligatoire, plus son remboursement coûtera cher et donc moins la garantie associée sera importante.

## Le contrat optimal

L'application des étapes décrites dans la section précédente a permis de former le contrat optimal. Cependant, la modélisation ne prend pas en compte les effets comportementaux bien connus en assurance santé. Par exemple, la garantie associée aux cures thermales est de 1 905€, ce qui est nettement supérieur au marché et pourrait entraîner un fort risque d'aléa moral. Elle ne permet pas non plus d'obtenir des remboursements rationnels au niveau des médicaments. C'est pourquoi, des contraintes supplémentaires ont été ajoutées au modèle, aisément modulable, grâce à une expertise métier. Elles ont été choisies à l'aide d'une vision globale du marché et d'une analyse de la distribution de la dépense engagée des sous-postes concernés avant et après remboursement obligatoire, permettant ainsi d'obtenir un contrat optimal en cohérence avec les pratiques de place. De plus, une analyse de la sensibilité des résultats par rapport à la calibration des poids a été faite. Elle a permis de montrer que les fluctuations d'échantillonnage du sondage ont un impact quasi-nul sur le contrat créé mais que les variations des poids peuvent entraîner des modifications conséquentes.

Le contrat obtenu a été analysé sous plusieurs angles afin d'obtenir une vision complète de ce dernier. Tout d'abord, la répartition de la prime pure a été analysée. Elle est semblable à celle de la dépense engagée après remboursement obligatoire exceptée quelques variations dues aux poids du modèle d'utilité et aux contraintes imposées : les restes à charge faibles bénéficient d'un remboursement complet tandis que ceux plus élevés bénéficient d'un remboursement partiel. Ensuite, le contrat a été comparé avec ceux des bases de données. Bien que la comparaison soit biaisée du fait de la variation des primes commerciales, elle permet néanmoins de montrer, d'une part, que le contrat est beaucoup moins protecteur au niveau des garanties associées à de fortes dépenses et que, d'autre part, les contrats commercialisés possèdent des garanties similaires exceptées sur les sous-postes de soins servant de leviers concurrentiels tels que les médicaments non remboursés et les cures thermales, ce qui témoigne de l'intensité concurrentielle élevée sur le marché.

Ensuite, les deux dernières analyses ont permis de recentraliser les travaux faits sur la sinistralité en santé : elles ont décrit le reste à charge moyen, totalisant un montant de 173,32€, et résumé la répartition de la dépense engagée des assurés qui souscriraient au contrat créé, comme le montre la figures 6. Ce sont les sous-postes de soins ayant une dépense engagée élevée, caractérisés par un faible remboursement obligatoire et une densité hétérogène, qui composent le reste à charge des assurés. De plus, étant donné que l'algorithme a pris en compte l'utilité des assurés, plusieurs sous-postes ayant une importance faible ont un reste à charge réduit mais strictement positif bien que leur dépense engagée après remboursement obligatoire ne soit pas élevée.

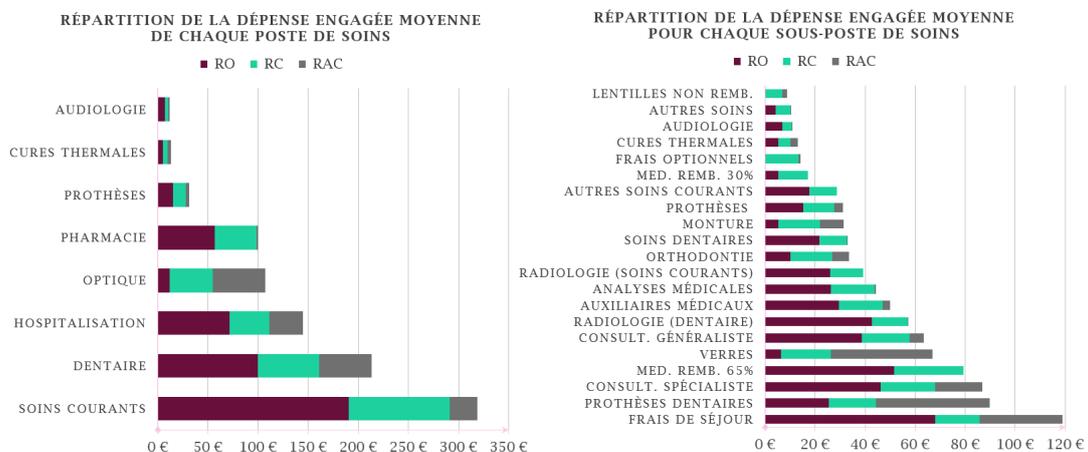


FIGURE 6: Décomposition de la dépense engagée des différents postes et sous-postes de soins.

## **Conclusion**

Les travaux présentés ont permis de mettre en place un cheminement de tarification inversé commençant par la contrainte budgétaire et se terminant par la construction d'une grille de garanties ayant pour critère la maximisation de la satisfaction des assurés. L'approche réalisée à une maille fine est facilement interprétable et aisément modulable. De plus, une expertise métier a été intégrée grâce à l'utilisation de contraintes supplémentaires afin de compléter les travaux réalisés.

Enfin, l'étude peut être poursuivie afin d'obtenir un produit commercialisable. Par exemple, une grille optique détaillée peut être construite et la tarification doit être adaptée à la réforme 100% Santé mais pour cela, des données plus exhaustives sont nécessaires.

En somme, les résultats obtenus dépendent de la modélisation de la dépense engagée et du remboursement de la Sécurité sociale, des poids calculés grâce au sondage et des contraintes choisies pour obtenir un contrat cohérent. Une application concrète de ces travaux à la construction ou à la revue de contrats de complémentaire santé permettrait de pouvoir mettre en pratique les travaux réalisés dans le cadre de ce mémoire et témoigner de la flexibilité de l'approche, mais également dépasser les limites de manque de données ainsi que justifier les contraintes retenues dans la modélisation.

## **Vers une tarification individuelle personnalisée ?**

Les travaux menés dans le cadre de ce mémoire pour tarifier un contrat collectif pourraient être aisément adaptés à la tarification individuelle et mener à la création de contrats sur-mesures en fonction du budget et des besoins des assurés. Néanmoins, cette personnalisation doit rester mesurée car elle engendre des risques comportementaux tels que l'antisélection et l'aléa moral. Ainsi, un équilibre entre mutualisation et individualisation doit être trouvé. Une solution serait, par exemple, de proposer un nombre limité de contrats construit en prenant en compte deux facteurs essentiels, la contrainte tarifaire et les appétences des clients.

# Synthesis note

Although the impact of the unprecedented crisis of 2020 is not yet measurable, it does provide evidence of the need for an effective social protection system. Indeed, the French model cushioned some of the economic and health effects of this crisis. An important link in this social protection is the health system. Its two main players, the Social Security, based on the principle of universal protection, and complementary insurers, represented by private organizations, work to provide the best possible coverage for individuals. However, even today, three million people lack the financial means to access this second level of social protection.

This is why this thesis focuses on the pricing of an entry-level collective contract that respects a budgetary constraint of one euro per day, intended for who cannot afford an expensive supplemental health insurance plan but who do not qualify for the Complementary Solidarity Health Insurance (CSS) offered by the French Social Security system. In order to construct this complementary health insurance contract, a daring approach was chosen: to take into account the health appetites of potential insured through mathematical tools derived from decision theory. Contrary to the construction methods usually used, this alternative makes it possible to reconnect with client demand.

All of the work in this brief was done at the care sub-station level. This mesh, which is finer than the care station mesh but also wider than the care act mesh, makes it possible both to set up an operational model and to obtain a fairly precise analysis of the various health risks.

In view of the structure of social health protection in France, the expenditure incurred and the compulsory reimbursement must be modelled prior to the creation of the supplementary health insurance contract, as shown in figure 7, which breaks down an individual's expenditure.

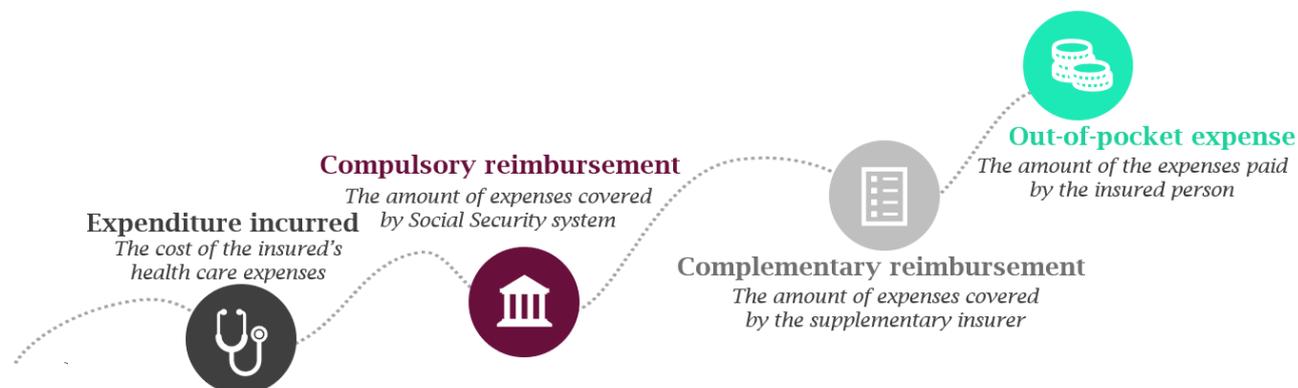


Figure 7: Breakdown of the expenditure incurred.

## Modelling of the expenditure incurred

Expenditures for the 27 care sub-stations were modeled using a cost-frequency approach using generalized linear models. Prior to these models, data from the management systems of a supplementary insurer, covering fiscal years 2018 and 2019, were first analyzed and then restated. These restatements, carried out in order to obtain databases suitable for linear models, used external data from INSEE and statistical tools such as regression trees and the k-means algorithm to discretize continuous variables, the age of the insured and to create classes, at the departmental level. In addition, extreme claims were modeled separately with a premium surcharge taking into account their cost and frequency to obtain a better fit of models sensitive to these values. Once the data was reprocessed, the extreme losses were removed and the uncorrelated variables selected, then, the cost and frequency models could be carried out.

For each model, two different laws were used and variable selection techniques such as the Forward-Backward method were introduced. The most suitable model was selected using indicators, especially metrics, and statistical tools such as histograms. The coefficients of these models made it possible to identify the explanatory factors for the loss experience associated with each subcategory of care. In addition, the limits of this modeling to the care sub-station grid using generalized linear models were analyzed through a study of predictions and methods to improve them, such as machine learning have been cited. For example, a better segmentation of care acts could lead to improve models and thus pricing.

Thus, these frequency and cost models made it possible to construct the loss rate associated with each care sub-station. In particular, they revealed a high claims rate for four subcategories of care, which together account for 40% of health care consumption: hospital stays with a heavy distribution queue, dental prostheses with a high cost, consultations with specialists and drugs reimbursed at 65% with a high frequency. At the level of health care stations, as shown in the figure 8, it is routine care characterized by high frequency that weighs the most in health consumption (34%) and this item can be completed by drugs dispensed in pharmacies (11%). Moreover, the pharmacy item actually represents a higher share since here the drugs fully reimbursed by the Social Security are not modelled. In addition, dental, hospitalization and optics, all three characterized by high costs, represent respectively 23%, 15% and 12% of health consumption.

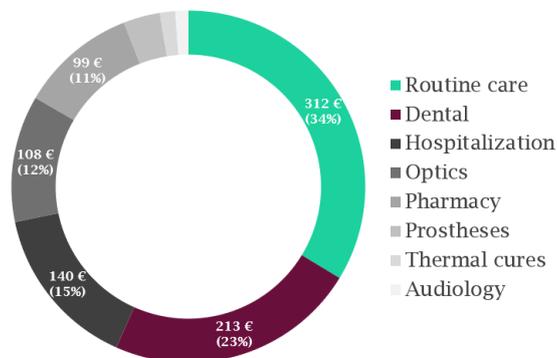


Figure 8: Breakdown of the expenditure incurred according to the different care stations.

### Modelling of compulsory reimbursement

From one act of care to another, the Social Security reimbursement rules vary in terms of their expression and amount. This is why, due to the thousands of care acts referenced and the modeling carried out at the care sub-station level, an indicator of the mandatory reimbursement rate at this same level was created.

Except for the 8 care sub-stations characterized by an explicit reimbursement rate, an intuitive indicator is the average of the reimbursement rates of the sub-station. In order to validate this method, the incurred expenditure of each sub-station was segmented according to the reimbursement rates and then the coverage rate of each slice of incurred expenditure was compared to the average coverage rate. For 12 care sub-stations, the rates of the brackets oscillated slightly around the average without revealing any particular trend, as is the case for the graph on the left of figure 9, which shows that this intuitive indicator is well suited. While for the other 7 care sub-stations, a linear decreasing trend appeared, as for the graph on the right. These are mostly sub-stations characterized by a low rate of reimbursement from Social Security and heterogeneous costs that can reach high values. Thus, the indicator for these sub-stations is a decreasing affine line depending on the expenditure incurred.

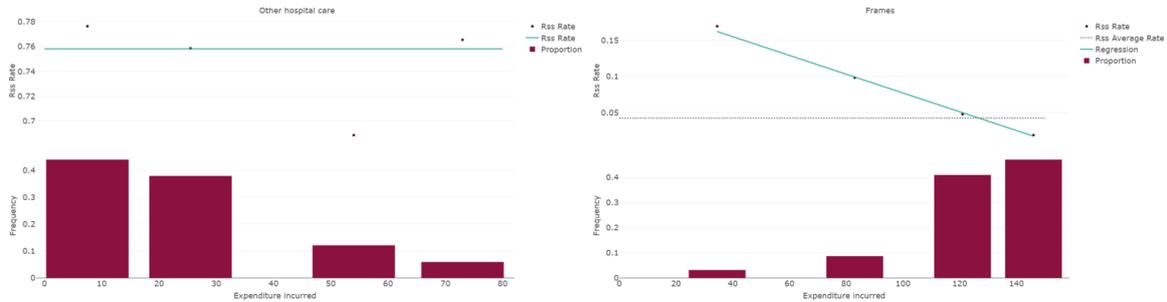


Figure 9: Graphic visualization of the coverage rates of the various expenditure brackets committed according to the indicator created for two care sub-items.

It appears that the least reimbursed treatments are those concerned by the “100% Santé” health-care reform. In addition, hospital stays are once again among the sub-stations with the highest claims rate, even though they are covered to a large extent. This is also the case for dental prostheses and optical glasses, which are poorly reimbursed. On a broader scale, the importance of routine care has dropped to 27% thanks to a high level of coverage, in contrast to the optical post, which doubles to 20% of health consumption.

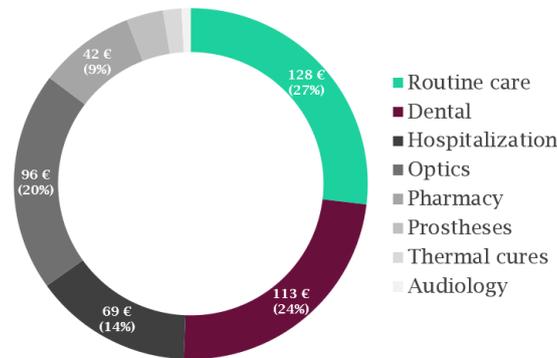


Figure 10: Breakdown of the expenditure incurred after compulsory reimbursement according to the different care stations.

## Modeling responding to the problem posed

The commercial premium of the contract is 360€ per year. Since this amount includes 20% loads, the budget available for pricing is therefore lower and amounts to 300€. In order to solve the problem posed in this thesis, the Cobb-Douglas budget constrained utility optimization model was used to obtain the optimal distribution of these 300€:

$$\max \prod_{i=1}^{27} CR_i^{Weight_i} \text{ s.t. } \sum_{i=1}^{27} CR_i = 300. \quad (2)$$

where  $CR_i$  represents the complementary repayment of the sub-station  $i$ . The weights of the model were obtained by means of a survey. This questionnaire made it possible to know the absolute importance of the care station and sub-station for the respondents. They were reprocessed, as shown in figure 11, in order to obtain the weights used in the optimization model representing the utility. In addition, the descriptive analyzes of the responses made it possible to highlight the variations in appetite between the care sub-stations and thus confirmed the importance of this factor in pricing. Finally, confidence intervals were constructed to measure the bias caused by sampling fluctuation.

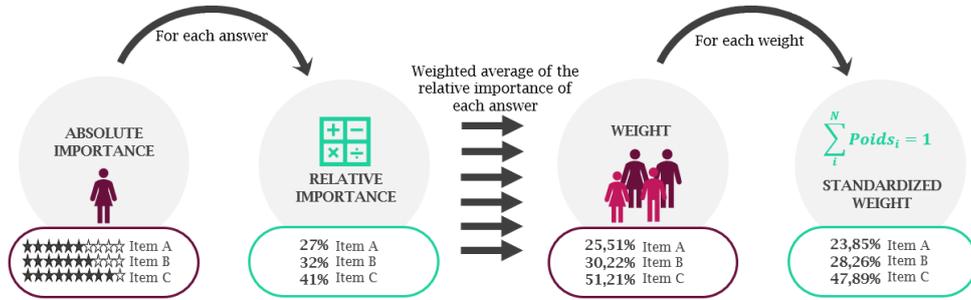


Figure 11: Diagram summarizing the steps to calculate the weights of the care stations.

The distribution obtained by the optimization algorithm checks the properties of decision theory: all the sub-stations are allocated a strictly positive budget and those with higher importance weights receive higher amounts.

Calculations were then made to equalize the expected repayment in addition to the budget determined by the algorithm in order to calibrate the contract guarantees. These guarantees are expressed in two forms: the lump sum, a fixed amount of reimbursement, and the percentage of reimbursement of the expenditure incurred. Depending on the type of guarantee, the calculation differs from the average additional reimbursement amount to the contract guarantee. In the case of a lump sum, an empirical expectation is calculated to calculate a non-linear expectation. Moreover, the optimization algorithm has been enriched with constraints in order to obtain guarantees, which is why some variables of the optimization problem are saturated.

Thus, the guarantee allowing to equalize the given budget strongly depends on the distribution of the claims of the sub-stations as well as on their mandatory reimbursement. For example, the higher the cost of a sub-station after mandatory reimbursement, the more expensive its reimbursement will be and therefore the lower the associated guarantee will be.

### The optimal contract

The application of the steps described in the previous section resulted in the optimal contract. However, the modelling does not take into account the well-known behavioural effects in health insurance. For example, the coverage associated with thermal cures is 1,905€, which is much higher than the market and could lead to a high risk of moral hazard. It also does not allow for rational reimbursements in terms of drugs. This is why additional constraints have been added to the model, which can be easily modulated thanks to our business expertise. They were chosen on the basis of a global vision of the market and an analysis of the distribution of the expenditure incurred by the sub-station concerned before and after compulsory reimbursement, thus making it possible to obtain an optimal contract in line with market practices. In addition, an analysis of the sensitivity of the results in relation to the calibration of the weights was carried out. It showed that fluctuations in the sampling of the survey have almost no impact on the contract created, but that variations in the weights can lead to significant modifications.

The contract obtained was analyzed from several angles in order to obtain a complete picture of the contract. First of all, the distribution of the pure premium was analyzed. It is similar to that of the expense incurred after compulsory reimbursement, except for some variations due to the weight of the utility model and the constraints imposed: the low remaining charges are fully reimbursed, while the higher ones are partially reimbursed. Next, the contract was compared with those of the databases. Although the comparison is biased due to the variation in commercial premiums, it nevertheless shows, on the one hand, that the contract is much less protective in terms of the benefits associated with high expenses and, on the other hand, that the contracts marketed have similar benefits except for the sub-categories of care serving as competitive levers such as non-reimbursed drugs and thermal cures, which reflects the high competitive intensity of the market.

Then, the last two analyses made it possible to recentralize the work done on the health claims experience: they described the average remaining cost, totalling 173.32 €, and summarized the distribution of the expenses incurred by the insured who would subscribe to the created contract, as shown in the figures 12. It is the sub-stations of care with high incurred expenses, characterized by low compulsory reimbursement and heterogeneous density, that make up the remainder to be borne by the insured. Moreover, since the algorithm has taken into account the utility of the insured, several sub-items with low importance have a low but strictly positive remaining charge although their incurred expenditure after compulsory reimbursement is not high.

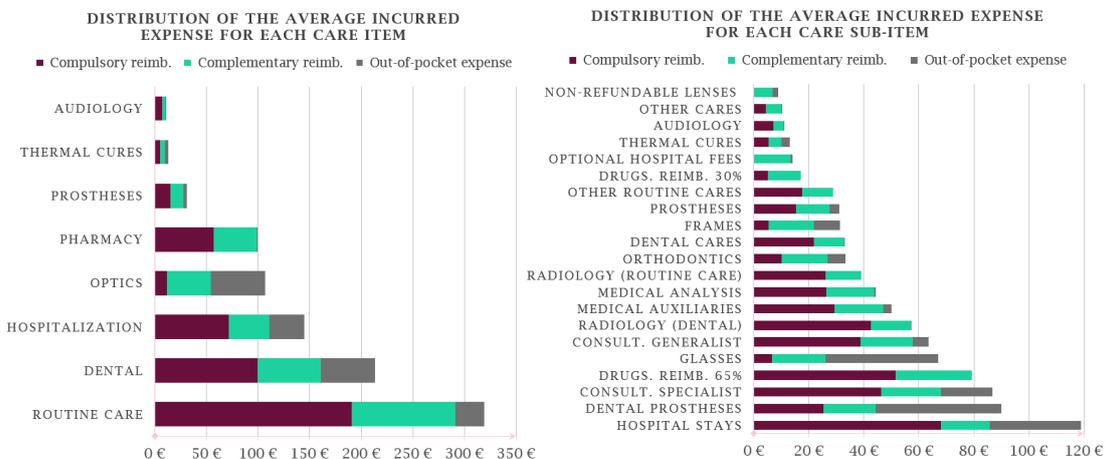


Figure 12: Breakdown of the expenditure incurred for the various care items and sub-items.

## **Conclusion**

The work presents a way to set up a reverse pricing process starting with the budgetary constraint and ending with the construction of a grid of guarantees based on the criteria of maximizing policyholder satisfaction. The fine-mesh approach is easily interpretable and can be conveniently modulated. In addition, a business expertise has been integrated thanks to the use of additional constraints in order to complete the work carried out.

Finally, the study can be continued in order to obtain a marketable product. For example, a detailed optical grid can be built and the pricing must be adapted to the “100% Santé” healthcare reform, but for this, more exhaustive data is necessary.

The results obtained depend on the modelling of the expenditure incurred and the Social Security reimbursement, the weights calculated thanks to the survey and the constraints chosen to obtain a coherent contract. Concrete application of this work to the construction or review of complementary health contracts would put into practice the work carried out in the framework of this thesis and to demonstrate the flexibility of the approach, but also to overcome the limits of the lack of data as well as to justify the constraints retained in the modelling.

## **Towards personalized individual pricing?**

The work carried out in the context of this thesis to price a group contract could be easily adapted to individual pricing and lead to the creation of tailor-made contracts according to the budget and needs of policyholders. Nevertheless, this personalization must remain measured because it generates behavioral risks such as anti-selection and moral hazard. Thus, a balance must be found between mutualization and individualization. One solution would be, for example, to propose a limited number of contracts built by taking into account two essential factors, the budget and the clients' appetites.

# Remerciements

En premier lieu, j'aimerais remercier Nicolas Servan, mon tuteur de mémoire, pour sa bienveillance et l'intérêt qu'il a porté à mes travaux. Il m'a fait confiance et a su se rendre disponible afin de me transmettre ses connaissances en assurance santé. Je voudrais également remercier Andréa Ehui et Baptiste Andrieu qui ont activement participé à l'encadrement de ce mémoire. Leur soutien et leurs conseils avisés m'ont permis d'enrichir mes recherches et ainsi mener à bien ce mémoire.

Je remercie Sia Partners, et plus particulièrement Michaël Donio, pour m'avoir donné l'opportunité de travailler au sein du cabinet. Je tiens également à témoigner ma gratitude envers l'équipe Santé et Prévoyance qui a su m'intégrer chaleureusement.

J'adresse des remerciements envers mes collègues stagiaires, et notamment Martin Dubost, pour leur entraide, leur soutien et leur sympathie.

Je voudrais également remercier l'Université Paris-Dauphine et son corps enseignant pour la qualité de leurs formations.

Enfin, j'aimerais adresser des remerciements sincères à tous mes proches pour m'avoir soutenue tout au long de mon cursus scolaire jusqu'à la réalisation de ce mémoire.



# Table des matières

<b>Résumé</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>Note de Synthèse</b>	<b>5</b>
<b>Synthesis note</b>	<b>11</b>
<b>Remerciements</b>	<b>17</b>
<b>Table des matières</b>	<b>19</b>
<b>Introduction</b>	<b>21</b>
<b>1 Panorama du marché de l'assurance santé en France</b>	<b>23</b>
1.1 Les acteurs du marché . . . . .	23
1.2 Le fonctionnement de l'assurance santé . . . . .	32
1.3 Les récentes réformes . . . . .	37
1.4 Le cadre et les enjeux de ce mémoire . . . . .	40
<b>2 Dépense engagée et remboursement de la Sécurité sociale</b>	<b>43</b>
2.1 Pré-requis théoriques sur les outils statistiques classiques . . . . .	44
2.2 Modélisation de la dépense engagée . . . . .	53
2.3 Modélisation du remboursement obligatoire . . . . .	74
<b>3 Création du contrat optimal sous contrainte budgétaire</b>	<b>81</b>
3.1 Pré-requis : la théorie de l'utilité et le modèle Cobb-Douglas . . . . .	82
3.2 Description des étapes de modélisation . . . . .	88
3.3 Construction du contrat optimal et analyse des résultats . . . . .	98
3.4 Ouverture : vers une tarification individuelle personnalisée ? . . . . .	112
<b>Conclusion</b>	<b>117</b>
<b>Bibliographie</b>	<b>119</b>

<b>Annexes</b>	<b>I</b>
A.1 Annexe 1 : Variables présentes dans les bases de données . . . . .	I
A.2 Annexe 2 : Classes d'âge créées . . . . .	IV
A.3 Annexe 3 : Classes départementales créées . . . . .	V
A.4 Annexe 4 : Résultats des modèles linéaires généralisés . . . . .	VIII
A.5 Annexe 5 : Indicateur du remboursement de la Sécurité sociale . . . . .	XVII
A.6 Annexe 6 : Le sondage . . . . .	XX

# Introduction

Plus de 95% de la population française possède, en 2019, un contrat de complémentaire santé selon l'étude publiée en avril 2019 par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees). De ce fait, plus de trois millions de Français sont uniquement couverts par l'Assurance Maladie obligatoire. Ce sont principalement des travailleurs précaires, chômeurs ou encore inactifs renonçant à une assurance santé complémentaire majoritairement pour des raisons budgétaires.

L'objectif principal de ce mémoire est de calibrer les garanties d'un contrat d'assurance complémentaire santé collectif sous une contrainte de prix, à savoir un euro par jour. Afin de tarifier ce contrat d'entrée de gamme, des fonctions d'utilité sont utilisées dans le cadre d'un algorithme d'optimisation prenant en compte les appétences des potentiels assurés mesurées à l'aide d'un sondage.

Tout d'abord, afin de contextualiser le sujet et comprendre les enjeux liés à ce mémoire, un panorama de la structure et du fonctionnement du marché de l'assurance santé français actuel est présenté.

Dans un second temps, en amont de la création du contrat de complémentaire santé optimal, deux modélisations faisant appel aux données d'un organisme d'assurance sont nécessaires : la première concerne la dépense engagée des différents sous-postes de soins et la seconde porte sur le remboursement de la Sécurité sociale de ces derniers.

Enfin, le contrat de complémentaire santé collectif est construit en utilisant le modèle Cobb-Douglas permettant d'introduire la notion d'utilité au travers de poids calibrés par une enquête. Ce modèle d'optimisation a pour but de maximiser l'espérance d'utilité des assurés sous la contrainte budgétaire imposée : le contrat optimal est celui qui satisfait le plus les assurés tout en respectant la contrainte d'un euro par jour. Le calcul d'optimisation dépend également du sous-poste de soins concerné de par le type de garantie associé et de par sa sinistralité.

Une ouverture concernant l'application des techniques utilisées tout au long de ce mémoire sur la personnalisation des contrats de complémentaire santé individuels est également présentée dans une ultime partie.



# Chapitre 1

## Panorama du marché de l'assurance santé en France

La protection sociale est un ensemble de mécanismes de prévoyance ayant pour but d'aider les individus et les ménages à faire face à des risques sociaux pouvant provoquer une hausse de dépenses ou bien une baisse de ressources. Par exemple, une hospitalisation peut entraîner des frais de santé conséquents et une maladie professionnelle peut diminuer le revenu d'un individu.

Ce mémoire va se focaliser sur les risques liés à la santé. Une première partie de ce chapitre présente l'organisation du marché à travers ses différents acteurs. Ensuite, puisqu'il s'agit d'un marché spécifique, ses termes techniques ainsi que ses mécanismes propres seront détaillés. Enfin, les réformes rythmant les modifications du secteur de l'assurance santé seront présentées afin d'exposer les objectifs des travaux menés.

### 1.1 Les acteurs du marché

Plusieurs acteurs composent le marché de la protection sociale. Ils ont pour rôle de protéger socialement les citoyens français. Cette protection sociale peut se décomposer en plusieurs couches. La première couche de base est la Sécurité sociale avec notamment la branche Maladie. À celle-ci viennent s'ajouter les régimes complémentaires représentés par des organismes privés qui délivrent une protection sociale à adhésion obligatoire ou bien à adhésion facultative au travers de contrats collectifs ou individuels.

#### 1.1.1 La Sécurité sociale

##### Présentation de la Sécurité sociale

Créée en 1945, la Sécurité sociale est un service public de l'Etat français visant à assurer l'ensemble des personnes travaillant et résidant en France de manière stable et régulière. Au coeur de la protection sociale en France, son rôle est de couvrir les pertes de revenus ainsi que les dépenses des Français dues à des risques sociaux. Elle fournit les prestations prévues par le Code de la Sécurité sociale et est subdivisée en différents régimes et différentes branches. Tout d'abord, la Sécurité sociale est divisée en différents régimes obligatoires de base suivant l'activité professionnelle des assurés :

➡ Le **régime général** et le **régime Alsace-Moselle** :

Le régime général est le plus important, puisqu'il couvre 88% de la population française. Il concerne les salariés du secteur privé ainsi que les travailleurs indépendants depuis le 1er janvier 2018, ces derniers étant auparavant rattachés au Régime Social des Indépendants (RSI).

➡ Le **régime agricole** (MSA) :

La mutualité sociale agricole (MSA) couvre l'ensemble de la population agricole et des ayants droit (exploitants, salariés, employeurs de main d'oeuvre, etc.). Elle compte plus de trois millions d'assurés, soit plus de 5% de la population française. Il s'agit du deuxième régime de protection sociale en France.

➡ Les **régimes spéciaux** :

Il existe une trentaine de régimes spéciaux regroupant au total 4,7 millions de cotisants en 2018, soit 7% de la population française. Ils rassemblent notamment les fonctionnaires, les salariés de la SNCF, les salariés de EDF-GDF, les parlementaires, les clercs et les employés de notaires, etc. Cette liste n'est pas exhaustive. Ces régimes possèdent leurs propres spécificités, la plupart d'entre eux offrent une protection sociale plus avantageuse que le régime général. En effet, certains corps sociaux avaient déjà mis en place leur propre système de protection sociale au moment de la création de la Sécurité sociale et décidèrent de garder leurs acquis sociaux.

Le régime général est lui-même composé de 5 branches :

➡ La **branche Famille**, gérée par la Caisse Nationale des Allocations Familiales (CNAF) :

Cette branche couvre l'accompagnement des familles, les aides au logement ainsi que les prestations de solidarité. Pilotée par la CNAF, elle accompagne socialement les assurés via des aides financières ou bien des services.

➡ La **branche Accidents du travail – maladies professionnelles**, gérée par la Caisse Nationale de l'Assurance Maladie (CNAM) :

Elle a pour fonction d'indemniser les salariés victimes d'accidents du travail ou de maladies professionnelles. Les cotisations, versées par les entreprises, financent les indemnisations intégralement prises en charge par l'Assurance Maladie en cas d'accidents du travail et de maladies professionnelles.

➡ La **branche Retraite**, gérée par la Caisse Nationale de l'Assurance Vieillesse (CNAV) :

Elle gère la retraite du régime général de la Sécurité sociale. Il s'agit de la retraite de base des salariés de l'industrie, du commerce et des services. Elle assure le suivi de carrière, la préparation et le versement des retraites de base du régime général.

➡ La **branche Maladie**, gérée par la Caisse Nationale de l'Assurance Maladie (CNAM) :

La branche Maladie assure la prise en charge des dépenses de santé des assurés et garantit l'accès aux soins. Elle couvre notamment les risques de maladie, de maternité, d'invalidité et de décès en plus du risque santé.

➡ La **branche Recouvrement**, gérée par les URSSAF :

Elle a un rôle de gestionnaire de trésorerie de la Sécurité sociale : la branche Recouvrement collecte les cotisations et contributions sociales auprès des entreprises, salariés et particuliers et les redistribue ensuite aux autres branches pour financer les prestations diverses.

Cette décomposition actuelle du fonctionnement de la Sécurité sociale en cinq branches n'est pas immuable. En collaboration avec le gouvernement, les partenaires sociaux réfléchissent à la création d'une sixième branche servant à couvrir les risques liés à la dépendance et à la perte d'autonomie. Cette nouvelle branche permettrait d'anticiper une hausse des dépenses liée à la prise en charge des personnes âgées dépendantes. En effet, la Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques (Drees) a publié une étude en octobre 2017 estimant que la dépense publique en faveur des personnes dépendantes pourrait atteindre plus de 2% du PIB français en 2060, DREES (2017). C'est pourquoi, le gouvernement mène actuellement des projets ayant pour objectif de créer une branche à part entière pour ce risque.

Chaque branche est gérée par une caisse nationale ayant pour fonction d'assurer le pilotage des caisses locales. Ces dernières ont un rôle de proximité avec les assurés. Au total, la France comptabilise plus d'une centaine de caisses contribuant à former un réseau étendu sur l'ensemble de son territoire.

## **L'Assurance Maladie**

### Le rôle :

La branche Maladie au travers de l'Assurance Maladie a pour but de garantir l'accès universel aux droits de protection sociale en matière de santé et de permettre l'accès aux soins. Pour cela, elle accompagne les assurés tout au long de leurs démarches de santé.

Elle a également pour but d'améliorer l'efficacité du système de santé français, à l'image de la promulgation de programmes de prévention et d'éducation à la santé. Favoriser l'accès à la santé des personnes démunies, à l'aide de la Complémentaire Santé Solidaire (CSS), fait également partie de ses prérogatives.

Enfin, en plus de gérer les dépenses de santé, la branche Maladie a un rôle de gestionnaire du risque : son but est de maîtriser l'évolution des dépenses tout en améliorant l'état de santé de la population. A cette fin, chaque année un rapport d'activité retraçant les résultats de l'année encourue et détaillant les objectifs de l'année à venir est publié par l'Assurance Maladie sur le site [ameli.fr](http://ameli.fr).

### L'organisation :

La branche Maladie est gérée par la Caisse Nationale de l'Assurance Maladie (CNAM) qui pilote des caisses à l'échelle départementale et à l'échelle régionale. Elle pilote notamment le réseau des Caisses Primaires d'Assurance Maladie (CPAM). Les 102 CPAM, présentes au niveau départemental, ont pour rôle d'assurer la mise en oeuvre des actions menées par l'Assurance Maladie en accompagnant les assurés dans leurs démarches quotidiennes.

### La PUMa :

Entrée en vigueur en 2016, la Protection Universelle Maladie (PUMa) a été créée afin de remplacer la Couverture Médicale Universelle (CMU). Elle assure aux personnes exerçant une activité professionnelle en France ou résidant en France de façon stable et régulière, la prise en charge des frais de santé en cas de maladie ou de maternité, à titre personnel et de manière continue tout au long de leur vie.

### L'activité :

Les contributions sociales permettent de financer les prestations de l'Assurance Maladie. Ces prestations peuvent être en nature, correspondant par exemple à la mise à disposition d'appareils médicaux ou bien à un service d'aide à domicile, ou en espèce, s'agissant alors de remboursements de frais de santé indemnisables, de revenus de remplacement en cas d'arrêt maladie, congés maternité ou paternité ou bien d'une pension d'invalidité.

### **La Complémentaire Santé Solidaire (CSS)**

Entrée en vigueur le 1er novembre 2019, elle remplace et regroupe la Couverture Maladie Universelle Complémentaire (CMU-C) et l'Aide au paiement d'une Complémentaire Santé (ACS). Il s'agit d'une couverture maladie complémentaire proposée par l'Assurance Maladie au travers de contrat gérés par des organismes complémentaires partenaires visant à faciliter l'accès aux soins des personnes à faibles revenus. Elle est accordée pour un an et doit être renouvelée chaque année, à l'exception des bénéficiaires du RSA (Revenu de Solidarité Active) et de l'ASPA (Allocation de Solidarité aux Personnes Âgées) pour qui le renouvellement est automatique. De plus, elle dispense les bénéficiaires de l'obligation, imposée à l'ANI\*, d'adhérer au contrat collectif obligatoire de leur entreprise.

Les individus bénéficiant de la Complémentaire Santé Solidaire se voient exemptés de l'avancement de leurs frais de santé. De plus, leur ticket modérateur est remboursé par la Complémentaire Santé Solidaire et ils ne sont pas soumis à la participation forfaitaire de 1€<sup>†</sup>.

Afin de bénéficier de cette aide, il faut être couvert par la PUMa et répondre à des critères de ressources : les revenus doivent être inférieurs à un plafond fixé dépendant de la composition du foyer familial. Selon les ressources, une participation financière peut être demandée. Pour les revenus les plus faibles, aucune participation financière n'est exigée. Par exemple, depuis le 1er avril 2020, une personne seule résidant en métropole et ayant un plafond annuel de référence inférieur à 9 032€ est exemptée de la participation financière. Entre 9 032€ et 12 193€, une participation financière est exigée. Cette dernière croît en fonction de l'âge de chaque personne du foyer familial. Elle est comprise entre 96€ et 360€ par an, soit au maximum 30€ par mois, soit moins de 1€ par jour.

### **Les sources de financement et dépenses du régime général de la Sécurité Sociale**

Le régime général de la Sécurité sociale utilise plusieurs sources de financement afin de faire face aux dépenses. Toutefois, malgré des ressources élevées, le régime est en situation de déficit, et ce depuis plusieurs années. Bien qu'il diminue d'année en année, le déficit subsiste. Les chiffres exposés dans cette partie concernent l'année 2018, ils sont issus de l'édition 2019 des chiffres clés de la Sécurité sociale en 2018, DIRECTION DE LA SÉCURITÉ SOCIALE (2019) et du rapport de la situation financière de la Sécurité sociale en 2018, COUR DES COMPTES (2019).

#### Les sources de financement :

Les ressources du régime général de la Sécurité sociale s'élèvent à 394,6 milliards d'euros en 2018 et se répartissent en six catégories :

---

\*L'Accord National Interprofessionnel (ANI) est présenté page 39.

<sup>†</sup>Le ticket modérateur ainsi que la participation forfaitaire de 1€ sont présentés page 35.

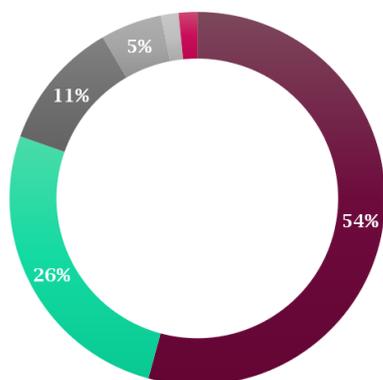
- ➡ cotisations sociales (54,2% , soit 213,9 milliards d’euros) ;
- ➡ Contribution Sociale Généralisée (CSG) (26,2%, soit 103,4 milliards d’euros) ;
- ➡ impôts et taxes et autres contributions sociales (11,2%, soit 44,2 milliards d’euros) ;
- ➡ transferts (5,3%, soit 44,2 milliards d’euros) ;
- ➡ contributions prises en charge par l’Etat (1,5%, soit 5,9 milliards d’euros) ;
- ➡ autres produits (1,6%, soit 6,3 milliards d’euros).

Les cotisations sociales sont des versements obligatoires effectués par les salariés, leurs employeurs mais aussi les non-salariés. Il s’agit, par exemple, des cotisations d’Assurance Maladie-maternité-invalidité-décès ou d’assurance vieillesse.

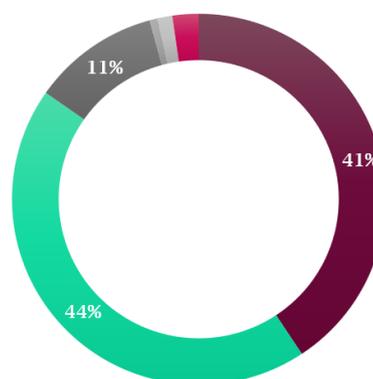
La contribution sociale généralisée (CSG) est un impôt créé en 1990 pour financer la Sécurité sociale. Elle concerne toute personne résidant en France et est prélevée sur les revenus d’activités, de patrimoine, de placement et de remplacement bruts. Les contributions publiques de l’Etat financent des dépenses de solidarité (fonds de solidarité vieillesse, RSA,...), une partie des exonérations de cotisations employeurs pour les bas salaires et subventionnent des régimes de façon permanente ou en cas de difficultés.

On retrouve une répartition semblable pour les recettes de la branche Maladie du régime général, avec des cotisations sociales à hauteur de 44,7%, une CSG couvrant 41,3% des recettes et 9,7% des recettes provenant des impôts et des taxes. Une description complète de la répartition des recettes du régime général et de la branche Maladie est exposée sur la figure 1.1.

**Structure des recettes du régime général en 2018**



**Structure des recettes de la branche maladie en 2018**



■ Cotisations  
 ■ CSG  
 ■ Contributions sociales diverses, impôts et taxes, hors CSG  
 ■ Transferts  
 ■ Cotisations prises en charge par l'état  
 ■ Autres produits

FIGURE 1.1: Répartition des recettes du régime général et de la branche Maladie en 2018.

#### Les dépenses :

Les dépenses du régime général de la Sécurité sociale, servant à la fois pour couvrir les prestations mais aussi les dépenses de fonctionnement, s’élèvent à 395,7 milliards d’euros en 2018, soit environ 17% du PIB français.

Parmi ces dépenses, plus de la moitié concerne la branche Maladie de la Sécurité sociale (51,9%, soit 205,4 milliards d'euros). Un tiers correspond à la branche Retraite (32,8%, soit 129,8 milliards d'euros). Enfin, les branches Famille et ATMP représentent respectivement 12,3%, soit 48,7 milliards d'euros et 3%, soit 11,9 milliards d'euros.

Les prestations versées par la Cnam en 2018 s'élèvent à 198,3 milliards d'euros, soit une hausse de 12% par rapport aux dépenses de 2017. Concernant la répartition de ces dépenses, deux postes prédominent : il s'agit des soins de villes et des établissements de santé qui représentent respectivement 45,7% et 41,2% des dépenses de l'Assurance Maladie. En effet, ces postes sont en majeure partie pris en charge par la Sécurité sociale. De plus, les soins de ville sont caractérisés par une fréquence élevée et les actes en établissement de santé ont un coût significatif. Un troisième poste, le médico-social, représente 10,5% des dépenses. Il s'agit des dépenses de structures accompagnant les personnes handicapées, dépendantes ou en situation d'exclusion sociale.

La figure 1.2 ci-dessus présente l'ensemble de la répartition des dépenses du régime général et de la branche Maladie.

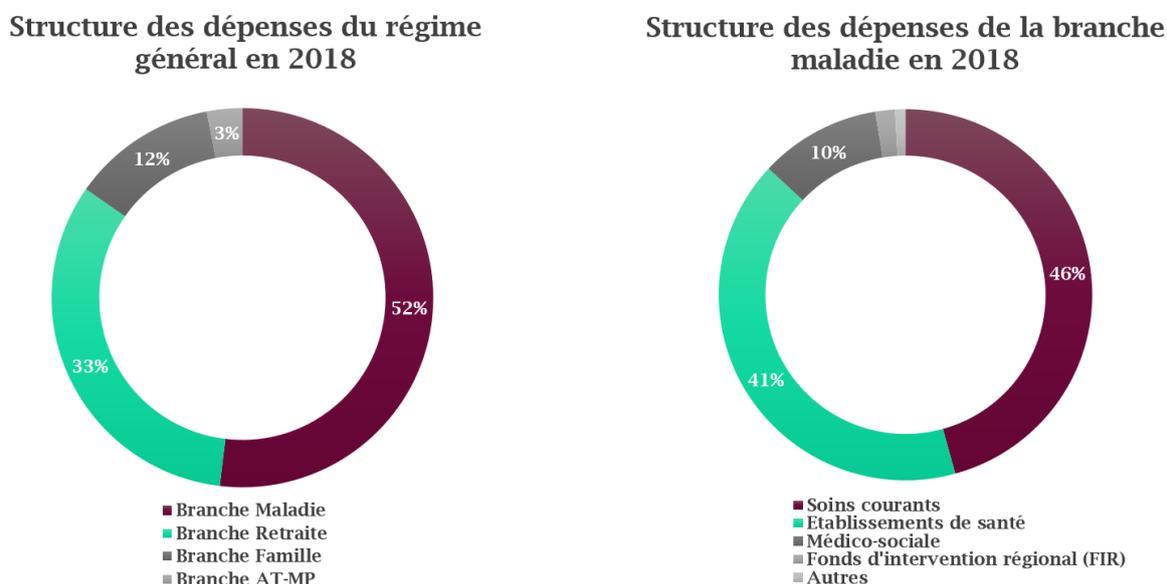


FIGURE 1.2: Répartition des dépenses du régime général et de la branche Maladie en 2018.

### La résorption du déficit :

Le déficit de la Sécurité sociale est un sujet majeur en France depuis une vingtaine d'années. En effet, les dépenses totales de santé représentent une part non-négligeable du PIB français (11,2% en 2018) et la Sécurité sociale est en situation de déficit depuis 2001. Cependant, elle est aujourd'hui à son plus bas niveau. En 2018, le déficit du régime général de la Sécurité sociale s'élève à 1,1 milliards d'euros contre 5,1 milliards en 2017 et plus de 20 milliards en 2010. Il s'agit d'un résultat très encourageant étant donné qu'il est inférieur au déficit initialement voté dans la loi de financement de la Sécurité sociale (LFSS) pour 2018. Cette situation s'explique par un surplus de recettes malgré une progression des dépenses : les recettes de la Sécurité sociale ont progressé plus rapidement que les dépenses (respectivement 3,1% et 2,4%), principalement sous l'effet de la progression de la masse salariale. De plus, il s'agit de la plus forte réduction du déficit, avec 4 milliards d'euros en moins par rapport à l'année précédente.

Seule la branche Maladie demeure déficitaire, toutefois son déficit s'est réduit de manière substantielle et atteint 0,7 milliard d'euros en 2018, contre 5 milliards d'euros en 2017. La branche Maladie a considérablement réduit son déficit ces dernières années et ce grâce aux différentes réformes mises en place en ce sens.

Les chiffres exposés précédemment détaillent la situation financière de la Sécurité sociale en 2018. La crise sanitaire de la Covid-19, débutée au premier trimestre 2020, a entraîné une forte dégradation des comptes sociaux. Fin avril 2020, le ministre des comptes publics, Gérard Darmanin, a déclaré que le déficit de 2020 devrait atteindre plus de 41 milliards d'euros, soit un déficit jusque-là encore jamais atteint. Le 2 juin 2020, il a annoncé une dégradation de plus 10 milliards d'euros par rapport au déficit annoncé en avril 2020, soit un déficit estimé à 52,2 milliards d'euros. Les prévisions du gouvernement étant à ce stade encore dépendantes de l'évolution de la pandémie, le projet de loi de financement de la Sécurité sociale n'a pas encore été publié par le gouvernement.

Ainsi, grâce au principe de couverture universelle, la Sécurité sociale est le premier échelon de la protection sociale française. On observe des assurés principalement concentrés sur le régime général malgré la survivance de régimes spéciaux. L'Assurance Maladie est en charge de la protection sociale liée au risque santé. Elle demeure déficitaire malgré la réduction de l'écart entre les recettes et les dépenses de celle-ci au fil des années et voit son solde financier se dégrader considérablement des suites de l'épidémie Covid-19. Un second échelon de la protection sociale liée au risque santé concerne les organismes complémentaires.

### 1.1.2 Les organismes complémentaires

La prise en charge de l'intégralité des frais de santé par la Sécurité sociale dépendant de l'équilibre général de cette dernière à travers la réglementation en vigueur et des actes concernés, le rôle des couvertures de santé complémentaires est, à l'image de leur nom, d'apporter un appoint aux remboursements versés par l'Etat. Les organismes d'assurance complémentaires peuvent également prendre en charge des dépenses ne bénéficiant d'aucun remboursement de la part de l'Assurance Maladie obligatoire tels que des actes de confort comme les chambres particulières en séjours hospitaliers, certains médicaments ou encore des lentilles de contact. Il s'agit en général, de soins onéreux non pris en charge par l'Assurance Maladie pour des raisons financières ou bien de soins visant à améliorer le confort du patient, bien que ces remboursements soient marginaux étant donné que les complémentaires santé proposent en majorité des contrats responsables\*.

#### Différents organismes

Les organismes complémentaires peuvent être répertoriés en quatre grands types d'organismes : les mutuelles, les mutuelles d'assurance, les compagnies d'assurance et les institutions de prévoyance. Les deux premières représentent la moitié du marché de l'assurance santé complémentaire français (52%) tandis que les deux dernières représentent respectivement 30% et 18% du marché.

Le tableau 1.1 expose leurs similitudes ainsi que leurs disparités à différents niveaux. Ces entités juridiques ont des structures et des objectifs variés et proposent leurs produits à des publics différents. Les mutuelles vont majoritairement vendre leurs contrats à des particuliers, il s'agit de "B-to-C" (Business-to-Customer) tandis que les institutions de prévoyance vont essentiellement se concentrer sur le secteur de la prévoyance collective des salariés, il s'agit alors de "B-to-B" (Business-to-Business).

---

\*La définition d'un contrat responsable est présentée page 38.

Enfin, ces spécificités commerciales impliquent des contrats spécifiques dépendants du public visé.

	LES COMPAGNIES D'ASSURANCE	LES MUTUELLES D'ASSURANCE	LES MUTUELLES	LES INSTITUTIONS DE PRÉVOYANCE
CODE	Le code des assurances	Le code des assurances	Le code de la mutualité	Le code de la Sécurité Sociale
FORME JURIDIQUE	Société Anonyme	Société Anonyme Mutuelle	Mutuelle	Institution de Prévoyance
OBJECTIFS	But lucratif Redistribution des bénéfices aux actionnaires	But non lucratif Couverture des risques des sociétaires		But non lucratif
FINANCEMENT	Appel public à l'épargne Augmentation du capital social	Absence de capital social		
TARIFS	Selon le profil des adhérents		Tarif mutualisé	

TABLE 1.1: Présentation des différents organismes complémentaires.

## Contrats collectifs et contrats individuels

Il existe deux types de complémentaires santé : la complémentaire individuelle souscrite individuellement auprès d'un acteur d'assurance et la complémentaire d'entreprise souscrite par une entreprise pour ses salariés.

Les **contrats collectifs** à adhésion facultative ou obligatoire, co-financés par l'employeur et l'employé, permettent de proposer des tarifs collectifs. Puisqu'ils sont collectifs, les garanties sont les mêmes pour la catégorie de salarié concernée et donc il n'y a pas de différence tarifaire suivant l'âge, la rémunération ou bien l'état de santé. Ils sont mis en place par référendum, accord collectif ou décision unilatérale soumis aux différentes conventions collectives nationales.

Suite à l'Accord National Interprofessionnel (ANI) du 11 janvier 2013 relatif à la compétitivité des entreprises, tout employeur du secteur privé, que ce soit une entreprise ou une association, a l'obligation de proposer une couverture complémentaire santé collective obligatoire à ses salariés. De plus, l'employeur a l'obligation de cotiser à hauteur minimale de 50%. Certains salariés peuvent être dispensés de l'obligation de souscrire une complémentaire santé. Il s'agit des bénéficiaires de la Complémentaire Santé Solidaire, des bénéficiaires d'une assurance santé individuelle ou encore des personnes assurées par une complémentaire santé gratuite d'une autre entreprise.

Les **contrats individuels**, quant à eux, sont souscrits par des particuliers et proposent des tarifs adaptés aux profils des assurés. Ils concernent principalement les étudiants, les retraités, les chômeurs, les travailleurs indépendants mais aussi les personnes qui souhaitent avoir une couverture plus importante en souscrivant une surcomplémentaire santé.

Les contrats à adhésion facultative permettent d'améliorer la couverture santé des assurés. Ils présentent un risque de sélection adverse : les personnes ayant des frais de santé importants veulent une couverture santé élevée tandis que les personnes ayant peu de frais de santé ne seront pas intéressées par ces offres. Ce type de contrat génère donc des effets d'antisélection, les assureurs doivent adapter leur tarification afin de se prémunir contre ce risque.

## Différentes grilles de garanties

Les organismes complémentaires proposent plusieurs grilles de garanties. Ces grilles peuvent être classées de la façon suivante :

- ➔ **Grilles d'entrée de gamme** : ces contrats satisfont les conditions d'un contrat responsable et sont caractérisés par des garanties restreintes, par exemple les dépassements d'honoraires ne sont pas pris en charge ;
- ➔ **Grilles de milieu de gamme** : ces contrats offrent une plus grande couverture que les contrats d'entrée de gamme, ils prennent en charge une partie des dépassements d'honoraires et proposent un meilleur remboursement notamment sur les actes moins bien remboursés par la Sécurité sociale ;
- ➔ **Grilles haut de gamme** : ces contrats offrent une plus grande couverture que les contrats de milieu de gamme et remboursent même des actes non pris en charge par la Sécurité sociale, comme les actes de médecine douce.

Il existe donc une large gamme de contrats complémentaires santé sur le marché français. Ils sont délivrés par divers organismes et de diverses manières.

Ainsi, en France, plusieurs couches de protection sociale en santé se dessinent. La première, celle de base, est la couverture offerte par la Sécurité sociale au travers de la branche Maladie. Elle concerne tous les résidents français. À celle-ci vient s'ajouter la couverture complémentaire collective obligatoire imposée par l'ANI pour les salariés. Une dernière couche est celle des couvertures complémentaires facultatives. Il peut s'agir alors de contrats collectifs ou individuels. Le schéma 3.9 représentant ces couches est présenté ci-dessus.

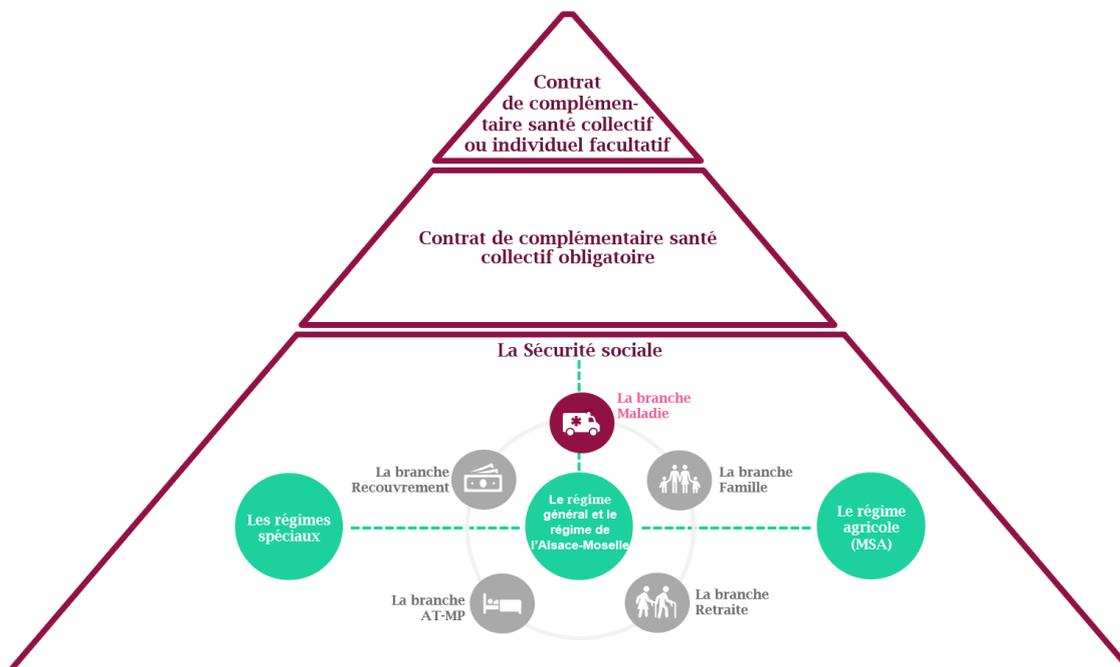


FIGURE 1.3: Présentation de la structure de la protection sociale en France.

La protection sociale des individus est segmentée entre ces divers acteurs. Il s'agit d'un emboîtement de mécanismes dépendants les uns des autres. Une présentation est proposée dans la section suivante afin de décrire plus en détail ce processus.

## 1.2 Le fonctionnement de l'assurance santé

Les soins relevant du domaine de la santé sont nombreux. Il n'existe pas de classification normée des différents postes de soins. Une première partie sera consacrée à la description des différents postes ainsi qu'à celle des actes associés. Ensuite, une seconde partie décrira les termes techniques couramment utilisés. Enfin, un exemple illustré sera intégré afin de décrire le fonctionnement de remboursement d'un acte de soins.

### 1.2.1 Les différents postes de soins et les différentes garanties

Les actes de soins peuvent être hiérarchisés de différentes manières grâce aux postes et sous-postes de soins. Une liste non-exhaustive d'exemples est donnée dans la figure 1.4. Par exemple, les frais de pharmacie peuvent être intégrés dans le poste Soins médicaux courants ou bien considérés comme un poste de soins à part entière.



FIGURE 1.4: Divers actes, sous-postes et postes de soins.

L'**hospitalisation** regroupe notamment les frais de séjour, le forfait journalier hospitalier, les frais de chirurgie et d'anesthésie, les frais de transports, les frais liés à la maternité ainsi que les frais des options, par exemple pour une chambre particulière. Le forfait journalier hospitalier ne doit pas être confondu avec les frais de séjour. Il s'agit de la participation financière du patient aux frais d'hébergement et d'entretien entraînés par son hospitalisation. Depuis le 1er janvier 2018, il s'élève à 20€ par jour pour les séjours à l'hôpital et en clinique et à 15€ par jour pour les séjours dans le service psychiatrique d'un établissement de santé. Ce montant n'est pas remboursé par l'Assurance Maladie mais il est remboursé par les complémentaires santé.

Le poste **dentaire** regroupe les prestations relatives aux actes bucco-dentaires, notamment les soins dentaires (détartrage, traitement d'une carie, ...), les prothèses et implants dentaires, l'orthodontie et les actes de radiologie dentaires.

L'**optique** regroupe les dépenses en équipement tels que les montures, les verres et les lentilles qu'elles soient ou non remboursées.

Les **soins médicaux courants**, également appelés soins de médecine de ville, regroupent notamment les consultations chez un médecin généraliste ou un spécialiste, les actes délivrés par des auxiliaires médicaux, les analyses médicales telles que les prélèvements sanguins, l'imagerie médicale telle que les IRM et d'autres soins médicaux techniques tels que les actes d'anesthésie par exemple.

Les frais de **pharmacie** concernent les médicaments prescrits remboursés et non remboursés par la Sécurité sociale. On distingue plusieurs catégories de médicaments : la première concerne les médicaments hors nomenclature qui ne sont pas remboursés par la Sécurité sociale et les autres concernent les médicaments remboursés à 15%, 30%, 60% et 100%.

Les **aides auditives** regroupent l'ensemble des dispositifs médicaux visant à compenser une perte d'audition.

Les actes relatifs à ces différents postes de soins peuvent être totalement, partiellement ou pas du tout pris en charge par la Sécurité sociale et il en va de même pour les couvertures de complémentaire santé.

Chaque contrat de complémentaire santé détaille les montants pris en charge dans des tableaux appelés grilles de garanties, elles-mêmes présentées dans les Conditions Générales d'Utilisation (CGU). Il n'existe pas de format type concernant ces tableaux, les garanties sont formulées de manière hétérogène. Dans la continuité de la réforme du système de santé, un accord a été signé le 14 février 2019 entre le Centre technique des institutions de prévoyance, la Fédération nationale de la mutualité française, la Fédération française de l'assurance et l'Union nationale des organismes complémentaires d'Assurance Maladie, UNOCAM (2019). Cet accord inédit porte sur l'engagement des organismes assureurs en matière de lisibilité des contrats afin d'améliorer la compréhension des contrats ainsi que leur niveau de couverture. Ainsi, chaque contrat de complémentaire santé doit présenter ses garanties avec des intitulés identiques pour cinq grands postes et ne devra pas excéder plus de dix postes de garanties. Les cinq grands postes sont :

- ➡ Les aides auditives ;
- ➡ Le dentaire ;
- ➡ L'hospitalisation ;
- ➡ L'optique ;
- ➡ Les soins courants.

Les postes dentaire, optique et aides auditives devront faire une distinction entre les équipements 100% Santé et les équipements du "panier libre", conformément à la réforme 100% Santé. Quelques sous-postes ont été définis. Par exemple, le poste dentaire devra contenir un sous-poste "soins" et un second sous-poste "prothèses", le poste hospitalisation devra contenir un sous-poste "forfait journalier hospitalier" et le sous-poste "honoraires" et enfin le poste soins courants devra contenir les sous-postes "honoraires paramédicaux", "honoraires médicaux", "analyses et examens de laboratoire", "médicaments" et "matériel médical".

Dans le cadre de ce mémoire, huit postes de soins seront retenus. Les postes de soins sont divisés en sous-postes, il en a 27 au total, et chaque sous-poste est composé d'actes de soins. Par exemple, le sous-poste Monture appartient au poste Optique et à l'intérieur de ce poste se trouve l'acte Monture Panier libre. Le tableau 1.2 hiérarchisant les postes et sous-postes est présenté ci-dessous afin de définir le contexte de modélisation de la prochaine partie.

POSTE DE SOINS	SOUS-POSTE DE SOINS
Soins courants	Soins délivrés par des auxiliaires médicaux Analyses médicales (actes de biologie, prélèvements sanguins, ...) Radiologie (échographie, doppler, IRM, ...) Consultation chez un médecin généraliste Consultation chez un médecin spécialiste Autres soins (actes d'anesthésie, actes de sage femme, ...)
Dentaire	Prothèses dentaires Soins dentaires Acte de radiologie dentaire Orthodontie
Hospitalisation	Frais de séjour Forfait hospitalier Frais liés aux options (chambre particulière, télévision, ...) Transport Maternité Autres frais en lien avec le poste de soins hospitalisation
Optique	Monture Verres Lentilles remboursées par la Sécurité sociale Lentilles non remboursées par la Sécurité sociale
Pharmacie	Médicaments non remboursés par la Sécurité sociale Médicaments remboursés à 15% par la Sécurité sociale Médicaments remboursés à 30% par la Sécurité sociale Médicaments remboursés à 65% par la Sécurité sociale
Audiologie	Prothèses auditives
Cures thermales	Cures thermales
Prothèses	Prothèses, hors prothèses auditives et prothèses dentaires

TABLE 1.2: Hiérarchisation des sous-postes par postes de soins.

### 1.2.2 Le vocabulaire spécifique

Plusieurs termes techniques seront utilisés dans la suite de ce mémoire. Une présentation de ces termes est faite ci-dessous.

#### Dépense engagée

La dépense engagée, également appelée frais réels, désigne la totalité des frais de soins facturés à l'assuré. Une partie peut être prise en charge directement, via le tiers-payant, ou indirectement par la Sécurité sociale et une autre partie peut être remboursée par une complémentaire santé. La somme des remboursements de l'Assurance Maladie et de l'assurance santé complémentaire est nécessairement inférieure au montant des frais réels. La potentielle partie restante est appelée le reste à charge (RAC).

### **Base de remboursement de la Sécurité sociale (BRSS)**

La Base de remboursement de la Sécurité sociale est la base de calcul du montant des remboursements des frais de santé par la Sécurité sociale en France. Avant le 1er janvier 2006, la base de remboursement était appelée Tarif de Convention ou Tarif d'Autorité. Elle est fixée par convention entre la Caisse Nationale d'Assurance Maladie (CNAM) et les professionnels de santé. Son montant est inférieur ou égal aux frais réels. Pour chaque acte de soins, un montant propre est déterminé. Les bases de remboursement figurent sur le site [ameli.fr](http://ameli.fr).

### **Taux de remboursement de la Sécurité sociale**

Il s'agit du pourcentage de la base de remboursement pris en charge par la Sécurité sociale. Ainsi, le remboursement de la Sécurité sociale est égal au produit du taux de remboursement et de la base de remboursement.

### **Participation forfaitaire**

La participation forfaitaire de 1€ est un dispositif qui a été mis en place le 1er janvier 2005 afin de contribuer à réduire le déficit de l'Assurance Maladie mais aussi pour limiter les comportements non-responsables. Ce montant est déduit de tous les remboursements d'actes médicaux ou de consultations médicales effectués par la Sécurité sociale excepté pour les mineurs, les femmes enceintes à partir du sixième mois de grossesse et les bénéficiaires de la CSS. Les contrats responsables\* proposés par les complémentaires santé ne peuvent prendre en charge ce montant.

### **Dépassements d'honoraires**

Des tarifs sont établis chaque année par la Sécurité sociale afin de définir les montants de soins pour chaque acte médical. Les médecins conventionnés en secteur 1 appliquent ces tarifs, ils ne pratiquent pas de dépassements d'honoraires. A contrario, les médecins conventionnés en secteur 2 ou 3 peuvent appliquer des dépassements d'honoraires qui ne sont pas pris en charge par la Sécurité sociale. Les dépassements d'honoraires ne figurent pas dans la base de remboursement.

### **Ticket modérateur**

Le ticket modérateur est le montant des dépenses de santé qui reste à la charge de l'assuré après le remboursement de l'Assurance Maladie. Il ne prend pas en compte les dépassements d'honoraires, ni la participation forfaitaire. Il s'agit donc de la différence entre la base de remboursement et le montant réellement remboursé par la Sécurité sociale.

### **Forfait**

Un forfait désigne un montant fixe de remboursement. Le forfait maternité délivré par des complémentaires santé en est un exemple. Il s'agit d'un montant fixe permettant de faire face aux frais liés à la naissance d'un enfant qui peut être dépensé librement par les parents.

---

\*La définition d'un contrat responsable est présentée page 38.

## Reste à charge

Il s'agit des frais à la charge de l'assuré. Le reste à charge est composé de la participation forfaitaire et de la partie du ticket modérateur et des dépassements d'honoraires non pris en charge par les organismes complémentaires.

### 1.2.3 Le mécanisme de remboursement des frais de santé

Les remboursements proposés dans les contrats de complémentaires santé peuvent s'exprimer en pourcentage des frais réels, en pourcentage de la base de remboursement, en pourcentage du ticket modérateur ou bien en forfait. Ces derniers peuvent inclure ou non les remboursements de la Sécurité sociale. De plus, des plafonds et des franchises de remboursement peuvent également être définis.

Afin d'illustrer ces propos, il convient d'utiliser un exemple concret. Dans le cas d'une consultation de base chez un généraliste conventionné du secteur 2 adhérent à l'OPTAM\*, les frais réels (FR) pour l'assuré s'élèvent à 38€, la base de remboursement de la Sécurité sociale (BRSS) est de 25€ et le taux de prise en charge de la Sécurité sociale est de 70% pour le parcours de soins coordonnés<sup>†</sup>. Pour cet exemple, l'assuré est couvert par un contrat responsable de complémentaire santé. Ce dernier rembourse 125% de la base de remboursement de la Sécurité sociale y compris le remboursement de la Sécurité sociale et ne rembourse pas la participation forfaitaire (PF) de 1€. Les différents montants de remboursement sont décryptés ci-dessous et la figure 1.5 propose une synthétisation.

La prestation remboursée par la Sécurité sociale ( $R_{ss}$ ) est calculée en prenant en compte le taux de prise en charge et en enlevant la participation forfaitaire de 1€, elle s'élève à 16,50€.

$$R_{ss} = T_{R_{SS}} \times BRSS - PF = 70\% \times 25 - 1 = 16,50.$$

Le ticket modérateur ( $TM$ ) est la différence entre la base de remboursement de la Sécurité sociale et le remboursement de la Sécurité sociale, il s'élève à 7,50€.

$$TM = BRSS - R_{ss} = 25 - 16,50 = 7,50.$$

Le dépassement d'honoraires ( $Dep$ ) s'élève à 13€.

$$Dep = FR - BRSS = 38 - 25 = 13.$$

Le contrat de complémentaire santé garantit, quant à lui, un montant de 31,25€. L'assuré a donc un reste à charge ( $RAC$ ) de 6,75€.

$$Garantie_C = 125\% \times 25 = 31,25.$$

$$RAC = FR - Garantie_C = 38 - 31,25 = 6,75.$$

Afin de connaître le montant payé par la complémentaire santé, il faut enlever le montant remboursé par la Sécurité sociale ainsi que la participation forfaitaire de 1€. Ainsi, le montant remboursé par la complémentaire santé s'élève à 13,75€.

$$R_c = Garantie_C - R_{ss} - PF = 31,25 - 16,50 - 1 = 13,75.$$

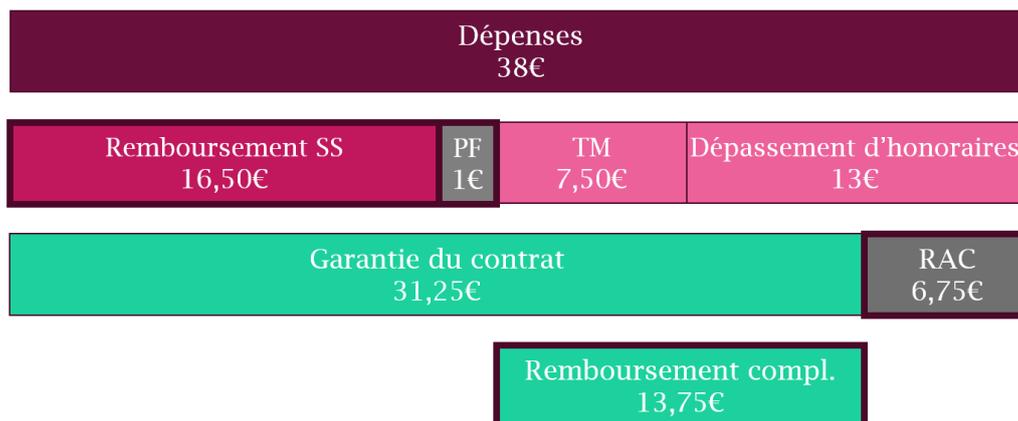


FIGURE 1.5: Exemple de remboursement.

Cette section a présenté en détail le vocabulaire ainsi que les mécanismes propres au domaine de l'assurance santé en France. Ces spécificités sont issues de plusieurs années de réformes et sont en perpétuelle évolution. La prochaine section est consacrée à la présentation de ces diverses réformes.

### 1.3 Les récentes réformes

Le secteur de l'assurance santé est caractérisé par de nombreuses réformes visant à améliorer son fonctionnement. Une liste non-exhaustive de ces récentes réformes est présentée ci-dessous.

#### 1.3.1 Des réformes encourageant la consommation responsable afin de réduire le déficit de l'Assurance Maladie

##### La loi Douste-Blazy (2004)

La loi n° 2004-810 du 13 août 2004 relative à l'Assurance Maladie, LÉGIFRANCE (2004), réforme et structure le parcours de soins, détermine la notion de médecin traitant, impose une franchise de 1€ et définit le dossier médical personnel (DMP).

Lorsque le processus du parcours de soins est respecté, l'individu obtient un remboursement maximal, à savoir un taux de remboursement de 70% pour la majeure partie des actes. Dans le cas contraire, l'individu reçoit une indemnisation moindre, 30%. Pour respecter le parcours de soins, il faut, au préalable, déclarer un médecin traitant. La notion de médecin traitant permet de renforcer le parcours de soins du patient afin qu'il puisse bénéficier d'un meilleur suivi et donc d'une meilleure prise en charge. Le médecin traitant doit être consulté en priorité, sauf en cas d'urgence. De plus, quelques exceptions au parcours de soins coordonnés existent. Par exemple, certaines spécialités comme les ophtalmologues et les gynécologues ne font pas partie du parcours de soins coordonnés.

\*L'Option Pratique Tarifaire Maîtrisée (OPTAM), qui a remplacé le Contrat d'Accès aux Soins (CAS) est un contrat signé entre l'Assurance Maladie et des médecins conventionnés exerçant (ou ayant la possibilité d'exercer) en secteur 2.

†Le parcours de soins coordonnés est introduit page 37.

Une participation forfaitaire de 1€ est imposée lors de consultations, d'analyses médicales ou encore d'examens de radiologie sauf pour certains cas particuliers comme une consultation par une infirmière, un chirurgien-dentiste, etc. De plus, certaines personnes sont exemptées de cette participation forfaitaire notamment les personnes bénéficiaires de la Complémentaire Santé Solidaire ou encore les femmes enceintes de plus de 6 mois.

### **Le contrat solidaire et le contrat responsable (2002-2006)**

Le contrat solidaire a pour objectif de faciliter l'accès aux soins quel que soit l'état de santé des assurés. Pour cela, le tarif d'un contrat solidaire ne doit dépendre ni des antécédents médicaux de l'assuré ni de son état de santé actuel. Quant au contrat responsable, il a pour but d'inciter l'assuré à modérer sa consommation en santé à travers des garanties limitées et réglementées. Il possède donc des garanties minimales mais aussi des garanties maximales. Parmi ces obligations, on retrouve tout d'abord l'interdiction de prise en charge de la participation forfaitaire de 1€. De plus, il y a obligation pour l'assureur de prendre intégralement en charge le ticket modérateur pour les consultations et actes médicaux du médecin traitant. Cette obligation encourage le patient à respecter le parcours de soins coordonnés. Une autre contrainte porte sur la limitation de la prise en charge des dépassements d'honoraires à hauteur de 100% de la base de remboursement pour les médecins conventionnés au secteur 2 non-signataires de l'OPTAM\*. Enfin, il y a une obligation de prise en charge intégrale du forfait journalier hospitalier et un encadrement strict des remboursements en optique.

Afin d'inciter le marché à proposer des contrats solidaires et responsables, l'Etat a baissé la taxe de solidarité annuelle (TSA) de ces derniers en la faisant passer de 20,27% à 13,27% et a également créé des avantages fiscaux et sociaux. De ce fait, aujourd'hui, la majeure partie des contrats complémentaires d'assurance santé sont des contrats solidaires et responsables.

### **Le contrat d'accès aux soins CAS et l'OPTAM (2016)**

L'Option Pratique Tarifaire Maîtrisée (OPTAM) et l'Option Pratique Tarifaire Maîtrisée Chirurgie et Obstétrique (OPTAM-CO) remplacent le dispositif du Contrat d'Accès aux Soins (CAS) depuis le 1er janvier 2017. Ces dispositifs sont issus de la convention médicale signée le 25 août 2016 par l'Assurance Maladie Obligatoire et trois syndicats de médecins. Ils ont pour but de réduire les dépassements d'honoraires. L'OPTAM concerne les médecins conventionnés au secteur 2 et l'OPTAM-CO concerne les chirurgiens et obstétriciens. Les médecins signataires s'engagent à modérer leurs tarifs en respectant le taux de dépassement d'honoraires moyen fixé, soit 100% du BRSS et à développer leur part d'activité sans dépassement d'honoraires. En contrepartie, le médecin obtient une prime, une aide financière pour s'installer dans les déserts médicaux et le remboursement de la Sécurité sociale est aligné sur le remboursement des médecins du secteur 1.

---

\*L'Option Pratique Tarifaire Maîtrisée (OPTAM), qui a remplacé le Contrat d'Accès aux Soins (CAS) est un contrat signé entre l'Assurance Maladie et des médecins conventionnés exerçant ( ou ayant la possibilité d'exercer) en secteur 2.

### 1.3.2 Des réformes dans le but d'augmenter la couverture sociale des français

#### L'ANI (2013)

Comme mentionné précédemment, l'Accord National Interprofessionnel, signé le 11 janvier 2013, MINISTÈRE DU TRAVAIL (2013), a instauré l'obligation pour les employeurs de proposer une complémentaire santé à leurs salariés. Cette généralisation de l'assurance santé complémentaire pour les salariés a permis de couvrir une plus grande partie de la population active.

De plus, l'ANI prévoit la portabilité des droits d'un salarié : elle garantit un maintien de droits pendant une durée égale à la durée du contrat et ce pour un maximum d'un an. Ce maintien de garanties est à titre gratuit, les salariés portant le coût de cette portabilité. En supplément, l'ANI prévoit des garanties de soins minimales à intégrer dans les contrats.

#### La loi Evin (2017)

L'article 4 de la loi n° 89-1009, LÉGIFRANCE (2017) a été modifié par un décret le 23 mars 2017 afin de lisser les hausses de tarifs. Ce décret modifie les règles de maintien de garanties santé des contrats collectifs obligatoires pour les anciens salariés d'une entreprise. Pendant la première année faisant suite au départ du salarié, la cotisation de ce dernier est la même que celle des salariés actifs. Cependant, elle peut augmenter au plus de 25% la seconde année et au plus de 50% la troisième année. Cette modification permet d'augmenter graduellement les hausses de cotisations des anciens salariés. En effet, la part auparavant financée par l'employeur est ensuite à charge de ces derniers. Ils font donc face à une hausse de leurs cotisations dès la première année de leur départ. Ainsi, cette loi leur fournit une protection sociale supplémentaire nécessaire.

#### La loi sur la résiliation infra-annuelle (2019)

La loi sur la résiliation infra-annuelle a été adoptée en juillet 2019 mais elle n'est pas encore en vigueur. Le décret n° 2019-733 fixe son entrée en vigueur au 1er décembre 2020 au plus tard, LÉGIFRANCE (2019b). Elle permettra à l'assuré qui a une durée de contrat supérieure à un an de résilier à tout moment de l'année son contrat individuel d'assurance santé. Cette mesure va entraîner une plus grande volatilité des clients et donc augmenter la concurrence sur le marché. Les assureurs devront à la fois faire face à une potentielle fuite de leurs clients et essayer d'attirer et de garder les clients plus volatiles qui ne vont pas hésiter à résilier leur contrat en cours d'année. Cette mesure aura très certainement un impact bénéfique pour les assurés car les assureurs vont leur proposer des couvertures à des tarifs plus attractifs ou bien de meilleure qualité si la concurrence sur le marché s'accroît.

#### La réforme 100% santé (2019)

Anciennement appelée « reste à charge zéro », elle a pour but de pouvoir offrir à tous des soins dentaires (prothèses dentaires), des remboursements d'optiques (lunettes de vue) ou encore en audiologie (prothèses auditives) remboursés à 100%, LÉGIFRANCE (2019a). Elle se déploie sur plusieurs années, depuis le 1er janvier 2019 jusqu'en 2021. Ces trois secteurs cibles sont caractérisés par des prix nettement supérieurs aux remboursements de la Sécurité sociale. En effet, ils sont fixés librement et cela entraîne une renonciation élevée aux soins.

Cette offre est destinée aux bénéficiaires de la Complémentaire Santé Solidaire et à tous les assurés d'un contrat responsable, ce qui représente la quasi-totalité des contrats du marché collectif. Les professionnels de santé sont tenus de proposer deux types de paniers à leurs clients : le panier "100% santé" et le panier libre.

Les offres incluses dans le panier 100% santé ont été choisies par des professionnels de la santé en collaboration avec l'Etat et l'Assurance Maladie mais aussi avec l'accord des complémentaires santé et des fabricants des dispositifs de santé. Les frais inclus dans le panier 100% santé seront intégralement remboursés par la Sécurité sociale et les complémentaires santé. Les assurés n'auront donc plus de reste à charge. A contrario, s'ils choisissent le panier libre, ils pourront avoir un reste à charge.

Ainsi, les réformes ont permis à la fois de réduire le déficit de l'Assurance Maladie, de mettre un accent sur la consommation responsable des soins et d'augmenter la protection sociale des Français.

La description du marché de l'assurance santé en France, notamment à travers le pilier public qu'est la Sécurité sociale et les piliers privés que sont les organismes de santé complémentaires, permet de mieux comprendre les enjeux ainsi que les réformes passées, en cours et à venir de ce secteur.

## 1.4 Le cadre et les enjeux de ce mémoire

Ce mémoire propose la construction d'un contrat de complémentaire santé collectif sous une contrainte de prix, à savoir un euro par jour. Il s'agit d'un contrat d'entrée de gamme destiné aux salariés dans l'impossibilité d'accéder à la Complémentaire Santé Solidaire (CSS) tout en étant dans l'incapacité de s'offrir une couverture complémentaire santé onéreuse. De plus, puisque les entreprises sont contraintes par l'Accord National Interprofessionnel (ANI) à financer au moins 50% des cotisations, ce type de contrat peu onéreux pourrait être proposé par des très petites et petites entreprises ayant une trésorerie limitée à adosser à la protection sociale de ses employés. Ainsi, les travaux menés ont pour objectif d'étendre la Complémentaire Santé Solidaire (CSS) afin de tendre vers une amélioration de la protection sociale en France tout en prenant en compte les limites financières de chacun.

Afin de tarifer au mieux ce nouveau produit, une approche innovante a été choisie : prendre en compte l'utilité des futurs assurés. En effet, aujourd'hui, les contrats de complémentaire santé collectifs sont issus d'une négociation entre les représentants des entreprises et les syndicats. Une balance entre les demandes des syndicats, voulant protéger au mieux les salariés, et les exigences des dirigeants est trouvée. Les assureurs complémentaires ont, quant à eux, un rôle de conseil. Ils connaissent les coûts de chaque garantie et évaluent les différentes calibrations de ces dernières. De plus, ils utilisent des *benchmarks*, des techniques de marketing afin d'observer, comparer et analyser les performances des autres sociétés d'assurance. Leurs analyses globales du marché leur permet de calibrer au mieux les garanties des contrats afin d'attirer les clients en proposant soit des tarifs plus attractifs, soit de meilleures garanties.

L'utilité des assurés est très souvent oubliée dans le processus de construction des offres. Ce mémoire propose au contraire de mettre en avant ce facteur. En effet, proposer un contrat adapté aux attentes des clients permettrait d'attirer plus de consommateurs en misant sur la personnalisation du produit par rapport aux attentes des clients. Afin de mesurer l'utilité des potentiels clients, un sondage sera créé et le modèle Cobb-Douglas sera utilisé pour résoudre un problème d'optimisation d'utilité sous contrainte de prix afin d'obtenir le contrat optimal.

En amont de cette optimisation, une étude et une modélisation des différents risques en santé sont nécessaires. En effet, la tarification du contrat dépend des niveaux de risque de chaque sous-poste de soins. De plus, les contrats de complémentaire santé intervenant en seconde position après la Sécurité sociale, deux modélisations doivent donc être réalisées en amont : une première modélisation portant sur la dépense engagée et une seconde traitant du remboursement de la Sécurité sociale.

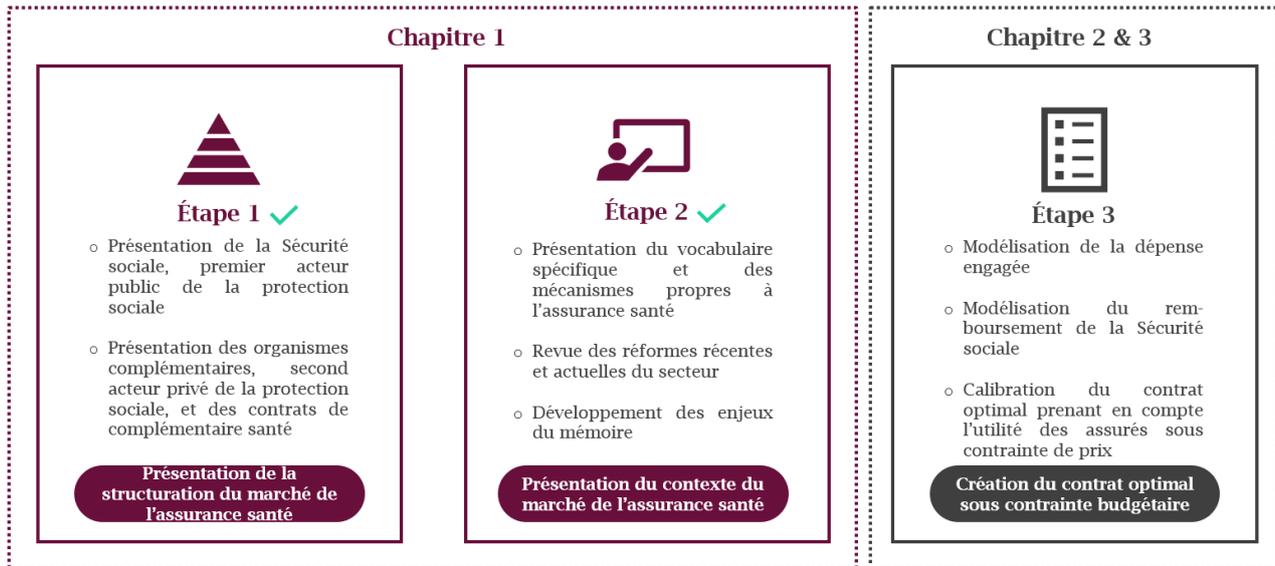


FIGURE 1.6: Schéma résumant les étapes du mémoire.



## Chapitre 2

# Modélisation de la dépense engagée et du remboursement de la Sécurité sociale

En amont de la création du contrat de complémentaire santé, les frais réels, autrement appelés la dépense engagée, doivent être analysés. Ensuite, ce sont les remboursements obligatoires qui doivent l'être. En effet, le chapitre précédent a présenté le fonctionnement de la protection sociale en santé en France et la Sécurité sociale constitue la première couche de cette protection tandis que les complémentaires santé, intervenant en deuxième position, viennent accroître cette dernière.



FIGURE 2.1: Décomposition de la dépense engagée.

L'étude sera faite à la maille sous-poste de soins, résultant d'un compromis entre application opérationnelle et modélisation fidèle. Plusieurs milliers d'actes étant référencés dans la Classification Commune des Actes Médicaux CCAM (*CCAM*), une modélisation à la maille acte aurait demandé des bases de données volumineuses ainsi qu'un temps de modélisation conséquent, tandis qu'une modélisation à la maille poste aurait pu être trop réductrice. Ainsi, 27 sous-postes de soins sont à analyser. Pour rappel, la hiérarchisation des sous-postes par postes de soins est présentée dans le chapitre 1 par le tableau 1.2.

Tout d'abord, une première partie théorique, concernant les outils statistiques utilisés par la suite dans le cadre des modélisations, est présentée au lecteur. Ensuite, une seconde partie pratique portant sur la modélisation de la dépense engagée est développée. Le risque sous-jacent à chaque sous-poste est modélisé par une approche coût-fréquence à l'aide de modèles linéaires généralisés. Enfin, une troisième partie est consacrée à la création d'un indicateur de remboursement de la Sécurité sociale.

## 2.1 Pré-requis théoriques sur les outils statistiques classiques

Les méthodes statistiques utilisées dans le cadre de ce mémoire peuvent être regroupées en deux parties :

- ➔ Les **algorithmes d'apprentissage supervisé**, pour lesquels la variable à prédire est connue et doit être expliquée en fonction de variables explicatives ;
- ➔ Les **algorithmes d'apprentissage non supervisé**, pour lesquels des observations sont données et leur lien doit être analysé.

Les modèles linéaires généralisés ainsi que les arbres de décisions appartiennent au premier type d'algorithmes. Le second type est, ici, représenté par l'algorithme des *k-means*, une méthode de partitionnement de données.

### 2.1.1 Les modèles linéaires généralisés, un outil d'apprentissage supervisé

Le modèle de tarification le plus utilisé aujourd'hui de par sa facilité d'analyse est le modèle linéaire généralisé. Il est commun de construire pour chaque garantie, deux modèles linéaires généralisés : un premier pour modéliser la fréquence et un second pour modéliser le coût moyen.

La description théorique à suivre est inspirée du cours de Modèles linéaires dispensé par Sophie Donnet à l'Université Paris Dauphine, DONNET (2019). Premièrement, le modèle linéaire gaussien est présenté. Ensuite, les modèles linéaires généralisés sont à leur tour décrits. De plus, les modèles de fréquence et de coût qui seront utilisés par la suite, sont également développés.

#### Le modèle linéaire gaussien

Un modèle linéaire gaussien consiste à étudier la liaison linéaire entre une variable  $Y$ , appelée variable réponse et une série de  $p$  variables explicatives  $X_1, \dots, X_p$ . Pour cela, des bases de données sont à disposition. Dans le cadre de ce mémoire,  $Y$  représentera soit la fréquence, soit le coût associé à chaque sous-poste de soins et les variables explicatives décriront l'assuré ainsi que son sinistre.

#### Notations :

Soient  $y_1, \dots, y_n$   $n$  observations indépendantes d'une variable quantitative. De plus, pour chaque observation  $i$ , les  $p$  variables explicatives non-aléatoires sont notées  $x_1^i, \dots, x_p^i$ . Le but est d' "expliquer"  $y_i$  comme une fonction linéaire de  $x_1^i, \dots, x_p^i$ , autrement dit, le but est de déterminer les "meilleurs" coefficients  $\beta_0, \dots, \beta_p$  ci-dessous :

$$\forall i = 1, \dots, n, \quad y_i = \beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i + e_i, \text{ où } e_i \text{ représente le terme d'erreur.} \quad (2.1)$$

Le choix des variables explicatives est important. Les variables ne doivent pas être corrélées entre elles, un seuil de corrélation de 0,7 sera utilisé dans la suite de ce mémoire. De plus, le nombre de variables doit être suffisant sans être trop important. Le cas échéant, une analyse en composantes principales peut être utilisée en amont afin de sélectionner les variables les plus pertinentes pour la modélisation.

Les notations suivantes seront utilisées par la suite :

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1^1 & \dots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_i^1 & \dots & x_i^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & \dots & x_n^p \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}, \quad \text{et} \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}. \quad (2.2)$$

Une version matricielle de l'équation 2.1 est :

$$y = X\beta + e, \quad (2.3)$$

où  $y$  et  $X$  sont observés et  $\beta$  est le paramètre inconnu que l'on cherche à déterminer.

Dans le cas particulier du modèle linéaire gaussien, les termes d'erreur  $e$  sont distribués selon la loi d'un vecteur gaussien de taille  $n$ , d'espérance nulle et de matrice de variance-covariance diagonale.

$$e \sim \mathcal{N}_n(0_{\mathbb{R}}^n, \sigma^2 I_n)$$

#### 4 postulats :

La loi sur les termes d'erreur implique quatre postulats concernant les observations pouvant être vérifiés par des outils graphiques et non par des tests statistiques :

- ❶ Les erreurs sont **centrées**. Des erreurs non centrées indiquent qu'un terme a été oublié dans le modèle ;
- ❷ Les erreurs sont de **variance constante**, il s'agit d'un modèle homoscédastique ;
- ❸ Les erreurs sont supposées **indépendantes** puisque les observations doivent résulter d'une expérience menée dans des conditions indépendantes ;
- ❹ Les erreurs sont supposées **gaussiennes**. Il s'agit d'un postulat fort qui peut être négligé dans le cas où le nombre d'observations est grand.

#### Estimation des paramètres par l'utilisation du principe des moindres carrés :

Le principe des moindres carrés consiste à chercher le  $\beta$  qui minimise la somme des carrés des résidus :

$$\hat{\beta}(y) = \arg \min_{\beta \in \mathbb{R}^n} \|y - X\beta\|^2 = \arg \min_{\beta \in \mathbb{R}^n} \sum_{i=1}^n \left( y_i - \sum_{k=1}^p \beta_k X_i^k \right)^2, \quad \text{où } \hat{\beta} \text{ est l'estimateur de } \beta. \quad (2.4)$$

**Théorème** Si  $X$  est de rang plein et  $\hat{\beta}$  vérifie l'équation 2.4 alors

$$\hat{\beta}(y) = (X'X)^{-1}X'y, \quad (2.5)$$

où  $X'$  est la transposée de  $X$ .

Estimation des paramètres par l'utilisation du maximum de vraisemblance :

Si  $y \sim \mathcal{N}_n(\mu, \Sigma)$ , alors  $y$  a pour densité

$$f_{\mu, \Sigma}(y) = \frac{1}{\sqrt{((2\pi)^n \det(\Sigma))}} e^{-\frac{1}{2}(y-\mu)' \Sigma^{-1}(y-\mu)}. \quad (2.6)$$

Si le quatrième postulat est vérifié, i.e.  $e$  suit une loi gaussienne, alors  $y \sim \mathcal{N}_n(X\beta, \sigma^2 I_n)$ . La log-vraisemblance des observations s'écrit alors :

$$\begin{aligned} l(y, \beta, \sigma^2) &= \log \left[ \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{1}{2}(y-X\beta)' \frac{1}{\sigma^2} I_n (y-X\beta)} \right] \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|y - X\beta\|^2. \end{aligned} \quad (2.7)$$

Ainsi, maximiser la vraisemblance en  $\beta$  revient à minimiser  $\|y - X\beta\|^2$ . Les deux méthodes ont donc des résultats qui coïncident dans le cas où les résidus sont supposés gaussiens.

Sélection de variables :

Certaines variables n'apportent pas d'explications supplémentaires sur la variable Y, il est donc préférable de les éliminer du modèle afin de ne pas augmenter la variabilité des estimateurs. Afin de définir le meilleur sous-ensemble de variables, il faut définir un critère. Plusieurs critères existent :

→ Le risque quadratique et le **critère Cp de Mallows** :

Le risque quadratique est principalement utilisé afin de montrer la convergence d'estimateurs paramétriques. Il peut également être utilisé afin de définir une distance entre deux modèles. Le risque quadratique entre un modèle  $m$  et le vrai modèle  $m^*$  est défini par la relation 2.8.

$$R(m, m^*) = \mathbb{E}[\|X^{(m^*)} \beta^{(m^*)} - X^{(m)} \hat{\beta}^{(m)}\|^2]. \quad (2.8)$$

Après réécriture, le risque quadratique peut être divisé en deux parties. La première partie est croissante en fonction du nombre de variables tandis que la seconde est décroissante en fonction de la dimension du modèle. Le but est de minimiser le risque quadratique. Le vrai modèle n'étant pas connu à l'avance, il faut réécrire le risque quadratique avec des données calculables à partir des données. L'expression simplifiée du risque quadratique est la suivante :

$$\frac{R(m, m^*)}{n(\sigma^*)^2} \approx \frac{(\tilde{\sigma}^{(m)})^2}{(\tilde{\sigma}^{(m_c)})^2} + \frac{2(|m| + 1)}{n} - 1, \quad (2.9)$$

où  $m$  représente le modèle testé,  $m^*$  représente le vrai modèle qu'il faut trouver parmi l'ensemble des modèles possibles,  $m_c$  représente le modèle complet,  $n$  est le nombre d'observations,  $|m|$  est le cardinal du modèle  $m$  et  $\tilde{\sigma}^{(m)}$  est l'écart-type associé au modèle  $m$ . Ainsi, minimiser le critère Cp de Mallows, défini par l'équation 2.10 équivaut à minimiser le risque quadratique.

$$C_p(m) = \frac{(\tilde{\sigma}^{(m)})^2}{(\tilde{\sigma}^{(m_c)})^2} + \frac{2(|m|)}{n}. \quad (2.10)$$

⇒ Le critère Akaike Information Criteria (AIC) :

Ce critère utilise une autre distance mais repose sur la même idée fondamentale que le critère précédent.

$$AIC(m) = \log((\tilde{\sigma}^{(m)})^2) + \frac{n + |m| + 1}{n - |m| - 3}. \quad (2.11)$$

⇒ Le critère Bayesian Information Criteria (BIC) :

Ce dernier critère est très ressemblant au critère précédent mais pénalise davantage le nombre de variables du modèle.

$$BIC(m) = \log((\tilde{\sigma}^m)^2) + \frac{\log(n)}{n} |m|. \quad (2.12)$$

Ces trois critères ont une forme particulière :  $Critere(m) = -f(Vraisemblance) + Penalisation(m)$  où  $f$  est en général la fonction identité ou logarithme. Ils doivent être minimisés. Un équilibre entre l'adéquation du modèle, il s'agit de la partie vraisemblance, et sa complexité, il s'agit de la partie pénalisation, doit être trouvé. Ainsi, ce compromis biais-variance s'explique par le fait qu'un modèle de grande dimension s'ajuste mieux mais entraîne une grande variabilité et donc de moins bonnes prédictions.

L'ensemble des sous-modèles à explorer est de taille conséquente ( $2^p$ ). Il existe trois procédures de sélection de variables connues qui permettent de limiter le nombre de modèles à tester :

- ⇒ La **méthode Forward** : l'initialisation se fait à partir du modèle sans variable. A la première itération, le modèle initial est comparé, à l'aide du critère choisi, aux modèles contenant une variable explicative. Le meilleur modèle, celui qui minimise le critère AIC par exemple, est retenu. Puis, à chaque nouvelle étape, une variable nouvelle est ajoutée au modèle jusqu'à ce que l'ajout d'une variable ne soit plus bénéfique, i.e. le critère AIC est maximisé par le modèle précédent.
- ⇒ La **méthode Backward** : il s'agit de la même stratégie que précédemment mais dans le sens contraire. L'initialisation se fait à partir du modèle complet et une variable est ôtée à chaque itération.
- ⇒ La **méthode Forward-Backward** : il s'agit d'une combinaison des deux méthodes précédentes. A chaque étape, une variable est enlevée et une variable est ajoutée. Ainsi, une variable jugée intéressante à une étape est réévaluée par la suite afin de s'assurer qu'elle l'est toujours. L'algorithme s'arrête lorsqu'aucune variable n'est ajoutée ou supprimée.

Les postulats des modèles linéaires gaussiens sont forts. Ils peuvent ne pas être vérifiés. Par exemple, le montant des dépenses engagées ne sera jamais négatif. Or, une variable gaussienne est à valeurs dans  $\mathbb{R}$ . Ainsi, les prédictions du modèle linéaire gaussien seront aberrantes. Un second exemple plus marquant concerne celui de la fréquence de sinistres : il s'agit d'un nombre entier positif. Là encore, l'hypothèse gaussienne n'est pas tenable et les prédictions seraient inexploitables. Une extension du modèle linéaire est le modèle linéaire généralisé, présenté ci-dessous.

### Les modèles linéaires généralisés

Le modèle linéaire généralisé est une extension de la régression linéaire, il permet au modèle linéaire d'être relié à la variable réponse via une fonction. Il s'agit de la donnée d'une loi de probabilité pour les  $Y_i$  et d'une fonction de lien, notée  $g$  telle que  $g(\mathbb{E}(Y_i)) = X_i'\beta$ . La fonction de lien permet ainsi d'obtenir une relation non-linéaire entre la variable à expliquer et les variables qualitatives.

La loi de distribution associée à la variable à expliquer doit appartenir à la famille exponentielle naturelle. C'est une famille de lois de probabilité s'écrivant sous forme exponentielle. Plus formellement,  $f$  appartient à la famille exponentielle naturelle si elle s'écrit sous la forme suivante :

$$f_Y(y) = e^{\frac{1}{\gamma(\phi)}(y\theta - b(\theta)) + c(y, \phi)}, \quad (2.13)$$

où  $c$  représente une fonction dérivable,  $b$  désigne une fonction trois fois dérivable de dérivée première inversible,  $\theta$  est le paramètre naturel et  $\phi$  est le paramètre de dispersion. Réaliser un modèle linéaire généralisé requiert donc le choix de deux éléments :

- ❶ **Le choix de la loi de probabilité** pour la variable à expliquer, ce choix est guidé par la nature du problème ;
- ❷ **Le choix de la fonction de lien  $g$**  utilisée.

Souvent, la fonction de lien est choisie comme la fonction qui transforme l'espérance  $\mathbb{E}(Y)$  en le paramètre naturel :  $g = (b')^{-1}$ . On nomme alors  $g$  fonction de lien canonique. Ce choix permet notamment d'assurer la convergence de l'algorithme vers le maximum de vraisemblance. En effet, les équations de vraisemblance n'ont pas toujours de solution explicite, sauf pour le cas du modèle linéaire gaussien. Il faut recourir à des algorithmes d'optimisation itératifs pour approcher la solution. Un algorithme souvent utilisé est l'algorithme de Newton-Raphson. Il a été prouvé que dans le cas de la fonction de lien canonique, cet algorithme converge vers le maximum de vraisemblance. Dans le cas particulier de la fonction de lien canonique, les formules suivantes sont donc appliquées :

$$g = (b')^{-1} \text{ et } \mathbb{E}(Y) = b'(\theta_i) = b'(x_i\beta). \quad (2.14)$$

Quatre lois de probabilité seront utilisées dans les modélisations, deux seront des lois de comptage pour les modèles de fréquence et deux autres des lois continues à valeurs positives pour les modèles de coût. Les lois qui ont été choisies sont celles couramment utilisées en tarification.

Deux exemples de lois de fréquence :

#### ⇒ La loi de Poisson

La fréquence est usuellement modélisée à l'aide d'une loi de Poisson. Elle appartient à la famille exponentielle, comme le montre l'équation 2.15 et a pour fonction de lien canonique  $g(u) = \log(u)$ , ainsi  $\mathbb{E}(Y_i) = e^{X_i'\beta}$ .

$$\forall k \in \mathbb{N}, f(k) = \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda + k \log(\lambda) - \log(k!)}, \text{ où } \lambda > 0 \text{ est le paramètre de la loi.} \quad (2.15)$$

Cependant, la loi de Poisson repose sur une hypothèse forte : l'espérance est égale à la variance. Or cette hypothèse est rarement vérifiée en pratique. Lorsqu'il y a une trop forte dispersion, une autre loi plus adaptée est utilisée. Il s'agit de la loi Binomiale négative. En effet, cette loi de fréquence est caractérisée par une variance plus forte que l'espérance et peut permettre une meilleure adéquation des données au modèle.

### → La loi Binomiale négative

La loi Binomiale négative est une combinaison d'une loi de Poisson et d'une loi Gamma. Elle caractérise des données qui suivent une loi de Poisson de paramètre aléatoire. Par exemple, les données peuvent être assimilées à différentes populations qui sont caractérisées par des risques moyens différents. La distribution de la loi Binomiale négative définit la distribution d'une variable aléatoire discrète de la même manière qu'une distribution de Poisson mais autorise la variance à être différente de la moyenne. C'est pourquoi son utilisation est préférable lorsque la variance est fortement supérieure à l'espérance. Cette loi appartient également à la famille exponentielle et a la même fonction de lien  $g$  que la loi de Poisson :

$$\forall k \in \mathbb{N}, f(k) = \binom{k+r-1}{k} (1-p)^r p^k = e^{k \log(p) + r \log(1-p) + \log\left(\binom{k+r-1}{k}\right)}, \quad (2.16)$$

où  $r > 0$  et  $0 < p < 1$  sont les paramètres de la loi.

Deux exemples de lois de coût :

### → La loi Gamma

La loi Gamma possède une distribution asymétrique modulée par deux paramètres, un paramètre de forme et un paramètre d'échelle. Elle appartient également à la famille des lois exponentielles et a pour fonction de lien  $g(u) = \log(u)$ , ainsi  $\mathbb{E}(Y_i) = e^{X_i' \beta}$ .

$$\forall x \in \mathbb{R}_+, f(x) = \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha-1} \exp^{-\beta x} = e^{-\log(\Gamma(\alpha)) + \alpha \log(\beta) + (\alpha-1) \log(x) - \beta x}, \quad (2.17)$$

où  $\alpha > 0$  et  $\beta > 0$  sont les paramètres de la loi.

### → La loi Log-normale

La loi Log-normale ne fait pas partie des lois exponentielles. Cependant, une régression log-normale peut être obtenue en considérant une régression linéaire gaussienne appliquée sur le logarithme du coût. Pour passer des estimations du modèle log-normal aux prédictions du coût, un terme doit être rajouté. En effet, l'espérance de la loi Log-normale a un terme supplémentaire :  $e^{\frac{\sigma^2}{2}}$ .

$$Y \sim \mathcal{LN}(\mu, \sigma^2) \Leftrightarrow \log(Y) \sim \mathcal{N}(\mu, \sigma^2). \quad (2.18)$$

$$\forall i = 1, \dots, n, \log(Y_i) = X_i' \beta. \quad (2.19)$$

La différence entre la loi Gamma et la loi Log-normale réside dans le fait que la loi Gamma est plus sensible aux valeurs extrêmes. En effet, l'application de la fonction logarithme permet de diminuer l'importance des sinistres extrêmes.

En somme, cette partie a permis de présenter la théorie sur les modèles linéaires généralisés. Ces modèles sont des outils d'apprentissage supervisé, tout comme les arbres de régression.

### 2.1.2 Les arbres de régression, un cas particulier des arbres de décision

Les arbres de décision sont des méthodes couramment utilisées en apprentissage supervisé. Deux types d'arbres de décisions se distinguent, les arbres de classification et les arbres de régression. Ces derniers sont utilisés afin d'expliquer et de prédire les valeurs prises par une variable quantitative en fonction de variables explicatives. La description suivante, concernant l'algorithme présenté, est inspirée du cours de *Machine learning* dispensé par Patrice Bertrand à l'Université Paris Dauphine, BERTRAND (2020).

Construction d'un arbre de régression :

L'algorithme consiste à construire un arbre de décision. A l'initialisation, les individus représentent une unique classe. Ensuite, la classe est divisée en deux classes afin de former les deux groupes les plus distincts possibles. Pour cela, un critère est testé sur une des variables explicatives. Cette opération est répétée pour chaque classe jusqu'à ce que le critère d'arrêt soit vérifié. Le critère d'arrêt peut prendre plusieurs formes : la variance interclasse n'augmente plus de manière significative, le nombre minimal d'individus dans une classe est atteint, etc. A chaque classe est alors associée une valeur numérique pour la variable cible.

La variable testée à chaque noeud de l'arbre est celle qui réalise le meilleur partage. L'algorithme teste les différentes variables d'entrée possibles et sélectionne celle qui maximise le critère de division utilisé. Ce dernier est la maximisation de la réduction de déviance. La déviance d'un noeud  $D_{noeud}$  et la réduction de déviance d'un noeud  $R$  sont respectivement définies par les équations 2.20 et 2.21.

$$D_{noeud} = \frac{1}{Card(noeud)} \sum_{i \in noeud} (y_i - \bar{y}_i)^2, \text{ où } \bar{y}_i = \frac{1}{Card(noeud)} \sum_{i \in noeud} y_i. \quad (2.20)$$

$$R = D_{parent} - D_{fils_1} - D_{fils_2}. \quad (2.21)$$

En effet, le but est d'obtenir des classes telles que les individus au sein d'un même groupe soient les plus similaires possibles et que les groupes soient les plus dissimilaires possibles, i.e. la variance interclasse doit être maximisée.

L'arbre construit peut sur-ajuster les données : il peut être excessivement raffiné et donc conduire à un modèle de partitionnement très instable. Afin de limiter ce risque, l'arbre est élagué.

Elagage de l'arbre :

A nouveau, il faut trouver l'équilibre biais-variance. En effet, une situation de sur-ajustement est moins robuste pour les prévisions. Il s'agit de trouver un arbre optimal entre l'arbre trivial réduit à une unique classe et l'arbre maximal construit. La variance interclasse de chaque feuille de l'arbre est calculée et ce pour toute la séquence d'arbres possibles. A celle-ci est ajoutée une pénalisation due à la taille de l'arbre. Une validation croisée est utilisée et l'arbre retenu est celui qui minimise l'erreur quadratique pénalisée de la base de test.

Les avantages et limites des arbres de régression :

Les arbres de régression sont adaptés à la fois pour expliquer mais aussi prédire l'évolution d'une variable cible en fonction de variables explicatives. Ils offrent des **résultats simples, faciles à interpréter**. De plus, ils ne nécessitent pas d'hypothèses particulières sur les distributions des variables et incluent une sélection de variables "automatique" puisqu'une variable non-pertinente ne sera pas choisie par l'algorithme comme variable test.

Cependant, ces modèles sont instables, ils sont sensibles aux données présentées. En effet, l'élaboration de l'arbre se fait pas à pas et donc un optimum global peut ne pas être trouvé par l'algorithme. C'est pourquoi **l'algorithme est sensible aux fluctuations d'échantillons**. De plus, il convient de s'assurer du bon élagage de l'arbre afin d'éviter le sur-apprentissage des données. Les forêts aléatoires visent notamment à corriger cet inconvénient.

Les présentations faites sur les modèles linéaires généralisés et les arbres de régressions diffèrent totalement des algorithmes d'apprentissage non supervisés dont un algorithme est présenté ci-dessous, l'algorithme des *k-means*.

### 2.1.3 L'algorithme des *k-means*, un algorithme d'apprentissage non supervisé

L'apprentissage non supervisé correspond à l'observation de réalisations de  $n$  variables  $X_1, X_2, \dots, X_n$ , le but étant d'explorer ces données afin de comprendre leur répartition, leur loi, etc.

La description suivante, concernant l'algorithme présenté, est inspirée du cours de *Machine learning* dispensé par Patrice Bertrand à l'Université Paris Dauphine, BERTRAND (2020).

L'algorithme des *k-means* est une variante de l'algorithme des centres mobiles. Son objectif est de construire une partition en plusieurs classes telle que les classes soient les mieux séparées et les plus homogènes possibles. Il s'agit d'un algorithme d'apprentissage non supervisé basé sur l'affectation dynamique d'individus à des classes. Il requiert une représentation vectorielle des individus dans  $\mathbb{R}^p$  muni d'une métrique à définir.

Les distances usuelles :

Il convient de choisir au préalable une mesure de dissimilarité. Deux données qui se ressemblent auront une distance de dissimilarité réduite, tandis que deux objets différents auront une distance de séparation plus grande. Plusieurs types de distance existent. Les deux plus communément utilisées sont la **distance euclidienne**  $d_1$  et la **distance de Manhattan**  $d_2$ .

$$d_1(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \text{ et } d_2(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|}. \quad (2.22)$$

L'algorithme des centres mobiles :

A l'initialisation,  $k$  individus sont tirés au hasard et sont appelés centres. Ensuite, chaque individu est associé au centre le plus proche, au sens de la métrique utilisée. Ainsi,  $k$  classes sont obtenues. Le centre de gravité de chaque classe est calculé et devient le nouveau centre de la classe associée. Cette étape est répétée jusqu'à ce que le critère d'arrêt soit vérifié. Il existe plusieurs critères d'arrêt : le nombre maximal d'itérations est atteint, la partition reste la même ou encore lorsque la variance interclasse ne croît plus de manière significative. La variance interclasse est la variance entre les groupes tandis que la variance intraclasse est la variance à l'intérieur des groupes. La somme de ces deux variances est égale à la variance totale. Afin de séparer au mieux les individus, il faut maximiser la variance interclasse, autrement dit minimiser la variance intraclasse. En effet, la variance interclasse est majorée par la variance totale et croît de manière décroissante à chaque itération de l'algorithme. Ainsi, la convergence de l'algorithme est assurée. Cependant, la vitesse de convergence dépend de l'initialisation.

$$V_{intra} = \sum_{h=1}^k \frac{n_h}{n} \sum_{i_h} \frac{1}{n_h} (X_{i_h} - \bar{X}_h)^2 \text{ et } V_{inter} = \sum_{h=1}^k \frac{n_h}{n} (\bar{X}_h - \bar{X})^2, \quad (2.23)$$

où  $k$  est le nombre de classes et  $n_h$  est la taille de la classe  $h$ .

L'algorithme des  $k$ -means :

Il s'agit d'une modification de l'algorithme précédent. Dans le cas de l'algorithme des  $k$ -means, le centre de chaque classe est recalculé à chaque allocation d'un individu à une classe. Cela améliore l'efficacité de l'algorithme mais les résultats obtenus dépendent de l'ordre des individus dans le fichier. C'est pourquoi plusieurs initialisations sont nécessaires.

Les avantages et limites de la méthode des  $k$ -means :

Bien qu'il s'agisse d'une **méthode rapide pouvant traiter une volumétrie conséquente et s'adapter facilement, les résultats dépendent fortement de l'initialisation**. En effet, plusieurs initialisations peuvent conduire à plusieurs partitions très différentes dans leur composition mais très proches et quasi-optimales au sens du critère de maximisation de la variance interclasse. De plus, il convient de **choisir la métrique utilisée ainsi que le nombre de classes** et il n'existe pas de critères unique afin de déterminer ce nombre. En effet, le choix du nombre de classes peut s'avérer délicat. Un nombre élevé de classes peut conduire à un partitionnement trop fragmenté des données tandis qu'un nombre trop petit peut au contraire conduire à un partitionnement trop généraliste des données et donc ce dernier manquera de finesse. L'équilibre biais-variance doit être trouvé. Dans ce but, plusieurs méthodes existent afin de choisir le nombre de classes approprié.

Le choix du nombre de classes :

La méthode la plus utilisée afin de déterminer le nombre de classes optimal est **la méthode du coude**, appelée Elbow Method en anglais. Elle consiste à tracer l'évolution de l'inertie intraclasse en fonction du nombre de classes. Le nombre optimal de classes est le point représentant le coude : un nombre de classes supérieur ne serait pas judicieux car il ne réduirait pas suffisamment l'inertie interclasse tandis qu'un nombre inférieur de classes ne permettrait pas d'obtenir une classification assez fine.

Un second critère est **le coefficient de silhouette**, c'est un indicateur de la distance des classes. Ce critère mesure la distance de chaque point d'une classe aux autres classes. L'évolution du coefficient silhouette, étant la moyenne de tous les coefficients de silhouette de tous les points est représentée en fonction du nombre de classes. Le nombre optimal de classes est celui qui maximise ce coefficient.

Enfin, un dernier **critère graphique** peut être utilisé. Il s'agit du graphique représentant la classe associée à chaque point dans le plan des deux premiers axes principaux.

Une application de ces critères est présentée dans le cadre des retraitements de données figure 2.3.

Les outils statistiques présentés ci-dessus vont être à présent utilisés afin de modéliser la dépense engagée.

## 2.2 Modélisation de la dépense engagée

La première étape consiste à modéliser la dépense engagée de chaque sous-poste. Pour cela, les données d'une société d'assurance sont manipulées sous le logiciel de programmation R dans le cadre d'une modélisation coût-fréquence utilisant des modèles linéaires généralisés.

Les données sont, dans un premier temps, présentées puis retraitées afin de pouvoir réaliser les modèles de fréquence et les modèles de coût. Dans un second temps, les modèles sont détaillés et analysés. Il en résultera, dans une dernière partie, la présentation de ces analyses en termes de dépense engagée moyenne pour chacun des sous-poste de soins.

### 2.2.1 Présentation, analyses et retraitements des données

Avant de se lancer dans un modèle, il convient en amont de connaître les données utilisées mais également de les transformer au format le plus adapté à la modélisation. Cette section importante va permettre de les présenter et de les retraiter.

#### Présentation des données

Les données sont issues des systèmes de gestion d'une société d'assurance française. Elles concernent un portefeuille de contrats collectifs santé regroupant un total de 122 350 assurés et 210 554 bénéficiaires répartis sur l'ensemble du territoire français. La moyenne d'âge des bénéficiaires est de 35 ans et une répartition de femmes légèrement supérieure est observée (56%). Quatre conventions collectives nationales sont présentes dans les bases de données qui concernent les exercices d'assurance 2018 et 2019. De plus, il s'agit d'un portefeuille en expansion puisque le nombre d'assurés entre 2018 et 2019 a augmenté de 10%.

A chaque sous-poste est associée une base de données. Il y a donc un total de 27 bases de données. Chaque base de données constitue une matrice de 290 723 lignes et 24 colonnes. Chaque ligne représente la sinistralité d'un bénéficiaire pour l'exercice d'assurance correspondant et chaque colonne représente une variable décrivant soit l'assuré soit les sinistres qui lui sont associés. La liste des variables numériques et catégorielles ainsi que leur description et leurs valeurs sont respectivement présentées dans les tableaux A.9 et A.10 en Annexe 1.

Les variables numériques concernent pour la plupart des montants liés aux dépenses, par exemple, le montant de la dépense engagée, le montant du remboursement obligatoire de la Sécurité sociale, le montant de remboursement complémentaire ou bien le montant de reste à charge. Les autres variables numériques sont le nombre de sinistres sur l'exercice concerné, l'exposition de l'assuré sur l'exercice encouru et l'âge du bénéficiaire. Quant aux variables catégorielles, excepté la variable caractérisant le sous-poste de soins, elles caractérisent l'assuré : son sexe, son type, sa tranche d'âge, son foyer, son département, son collègue, la convention collective nationale à laquelle il est rattaché, le type de couverture complémentaire dont il bénéficie et l'exercice concerné.

Certaines variables ont été écartées de la modélisation. Étant donné que les bases ont été divisées en sous-postes, la variable "sousPoste" peut être supprimée. De plus, les variables numériques, qui ne concernent pas la dépense engagée, ont été supprimées. Ainsi la base de données exploitée par la suite contient 16 variables explicatives.

Des tests sur les bases de données ont été menés afin de vérifier qu'il n'y ait pas de valeur manquante ou aberrante. Par exemple, la concordance des montants entre les frais réels, le remboursement de la Sécurité sociale, le remboursement de la Complémentaire Santé et le reste à charge a été vérifiée.

### Analyses univariées et bivariées des données

Premièrement, le nombre de sinistres ainsi que la fréquence de sinistres de chaque sous-poste ont été calculés. Les résultats sont présentés dans le tableau 2.1.

SOUS-POSTE	VALEURS	NBR D'IND. SINISTRÉS	P. D'AVOIR AU MOINS UN SINISTRE	FRÉQUENCE MOYENNE DE SINISTRÉS	VALEURS (EN €)	COÛT MOYEN (EN €)
Auxiliaires médicaux	de 0 à 3 040	81 942	28,32%	3,46	de 1 à 1 043	14,88
Analyses médicales	de 0 à 133	107 601	37,11%	1,61	de 1 à 425	27,33
Autres soins courants	de 0 à 78	94 774	32,65%	0,85	de 1 à 7 799	33,4
Consult. Généraliste	de 0 à 83	190 255	65,53%	2,45	de 1 à 392	25,91
Consult. Spécialiste	de 0 à 410	173 618	59,94%	2,87	de 2 à 2 265	30,42
Radiologie (Soins courants)	de 0 à 30	94 521	32,54%	0,81	de 2 à 3 596	47,34
Prothèses dentaires	de 0 à 44	19 800	6,82%	0,16	de 19 à 13 403	555
Soins dentaires	de 0 à 55	87 738	30,21%	0,99	de 4 à 2 441	34,46
Radiologie (Dentaire)	de 0 à 55	116 531	13,68%	0,41	de 4 à 13 403	159,69
Orthodontie	de 0 à 7	11 570	4,01%	0,01	de 6 à 4301	430
Autres soins hospitaliers	de 0 à 213	8 494	2,83%	0,07	de 1 à 3 864	21,38
Frais optionnels	de 0 à 51	12 035	4,15%	0,06	de 2 à 3 080	211,29
Frais de séjour	de 0 à 264	53 202	18,30%	0,73	de 1 à 20 070	179,04
Forfait hospitalier	de 0 à 41	12 791	4,40%	0,08	20	20
Maternité	de 0 à 2	36	0,0001%	0,0000895	de 329 à 424	403,85
Transport	de 0 à 73	4 423	1,52%	0,029	de 2 à 8 955	132,52
Lentilles remb.	de 0 à 10	950	0,33%	0,0067	de 17 à 430	114
Lentilles non remb.	de 0 à 11	13 373	4,60%	0,1	de 12 à 968	89
Monture	de 0 à 4	61 450	21,14%	0,21	de 1 à 1 042	150
Verres	de 0 à 8	63 802	21,95%	0,44	de 3 à 1 484	152
Med. hors nomencl.	de 0 à 14	2 836	0,98%	0,02	de 2 à 405	31
Med. Remb. 15%	de 0 à 94	45 788	19,78%	0,64	de 1 à 229	2,88
Med. Remb. 30%	de 0 à 424	163 541	56,29%	5,9	de 1 à 664	2,38
Med. Remb. 65%	de 0 à 370	207 408	71,38%	14,22	de 1 à 1 748	6,22
Audiologie	de 0 à 7	2 762	0,95%	0,02	de 2 à 4 482	545
Cures thermales	de 0 à 63	3 227	1,11%	0,034	de 5 à 6 126	429
Prothèses	de 0 à 87	75 409	25,96%	0,68	de 0+ à 32 696	45,35

TABLE 2.1: Analyse de la fréquence et du coût moyen de chaque sous-poste.

Ce tableau montre que la fréquence et le coût moyen diffèrent d'un sous-poste à l'autre. Certains sous-postes sont caractérisés par des fréquences de sinistres élevées, c'est le cas pour les consultations chez le médecin généraliste, tandis que d'autres ont une fréquence plus faible, comme l'audiologie. La distribution du nombre de sinistres diffère également, le nombre de sinistres peut atteindre plusieurs centaines de sinistres comme ne pas dépasser dix. Il en va de même pour le coût. Certains sous-postes sont caractérisés par un coût moyen inférieur à la dizaine d'euros tandis que d'autres dépassent la centaine. Enfin, la combinaison fréquence - coût moyen est sensiblement différente d'un sous-poste à l'autre. Par exemple, ils peuvent être caractérisés à la fois par une fréquence faible et un coût faible et donc constituer un sous-poste peu onéreux ou bien le contraire et constituer un sous-poste conséquent en termes de coût.

Ces analyses montrent qu'une modélisation à la maille sous poste est justifiée. Une modélisation à la maille poste aurait été trop réductrice tandis qu'une modélisation à la maille acte aurait été complexe au niveau de la volumétrie de bases à analyser.

Pour chacune des variables présentes dans la base de données, des analyses univariées et bivariées ont été menées en amont de la modélisation. Par soucis de concision, toutes ces analyses concernant toutes les bases de données ne seront pas détaillées ici. Sont présentées ci-dessous les analyses les plus pertinentes.

Comme le montre la figure 2.2, certains âges sont sous-représentés dans les bases de données. Par exemple, les individus entre 18 et 25 ans, étant majoritairement des étudiants et adhérant donc à des mutuelles étudiantes, sont deux fois moins représentés ainsi que les personnes âgées de plus de 70 ans, majoritairement des retraités adhérant à des contrats de santé complémentaires individuels. En effet, comme expliqué dans le chapitre 1, les prix des complémentaires santé ne sont plus encadrés par la loi au delà des trois ans après le début de la retraite des anciens salariés. Il y a donc une très forte majorité de retraités qui souscrit des contrats individuels, souvent plus adaptés à leurs besoins.

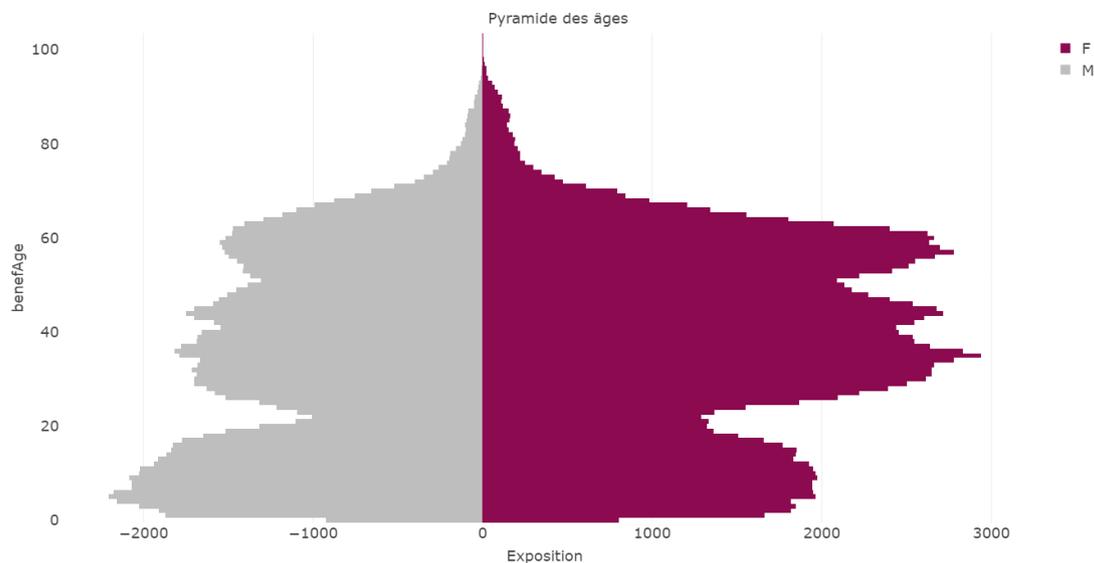


FIGURE 2.2: Répartition de l'exposition en fonction du sexe et de l'âge des assurés.

Un second facteur à analyser est celui de l'année de survenance du sinistre, 2018 ou bien 2019. Chaque base de données contient respectivement 138 795 et 151 927 lignes pour les années 2018 et 2019, soit une augmentation de 9,5% pour la seconde année. Outre ce déséquilibre d'effectif, certains sous-postes ont une consommation semblable d'une année à l'autre, d'autres croissante et certains décroissante. Ces observations faites sur la fréquence sont également vérifiées au niveau du coût moyen. La combinaison de ces deux facteurs entraîne une variation plus ou moins importante de la sinistralité d'une année à l'autre. Les variations de trois sous-postes de l'année 2018 à 2019 sont montrées ci-dessous dans le tableau 2.2 afin d'illustrer ces propos. Ainsi, ce premier facteur apparaît à la fois déterminant pour certains sous-postes et négligeable pour d'autres et ce pour la fréquence et/ou le coût moyen.

POSTE	SOUS-POSTE	VARIATION DE LA SINISTRALITÉ	VARIATION DE LA FRÉQUENCE	VARIATION DU COÛT MOUEN
Dentaire	Soins dentaires	0,58%	-1,21%	1,69%
Hospitalisation	Frais de séjour	-13,61%	3,79%	-18,63%
Audiologie	Audiologie	17,24%	17,03%	3,04%

TABLE 2.2: Variation de la sinistralité de différents sous-postes entre 2018 et 2019.

Les mêmes remarques peuvent être faites pour les autres variables telles que le type, le sexe, l'âge, la classe d'âge, le département, la structure du foyer, le collège, les caractéristiques de couverture et la convention collective de rattachement de l'assuré. Par exemple, la fréquence de soins en orthodontie est très sensible au type de l'assuré : les enfants sont les principaux consommateurs. A contrario, la variable "benefSexe" ne semble pas influencer la fréquence de consommation en prothèses auditives. Les différents modèles linéaires généralisés mis en place par la suite souligneront ces effets à travers des coefficients variables devant les modalités de ces variables.

Une fois ces premières analyses faites, il convient de retraiter les données afin de pouvoir ensuite construire les modèles de fréquence et de coût.

### Retraitements des données

Certaines variables ont dues être retraitées dans le but de pouvoir réaliser un modèle linéaire. Il s'agit des variables "benefDept" et "benefClasseAge" qui représentent respectivement le département de résidence et l'âge du bénéficiaire.

#### Retraitement de la variable "benefDept" :

Les variables explicatives d'un modèle linéaire généralisé doivent être numériques. Dans ce but, les variables catégorielles sont binarisées : une variable catégorielle à  $k$  modalités est remplacée par  $k$  variables indicatrices telles que la  $k$ -ième variable vaut 1 si l'individu a la modalité  $k$ , 0 sinon. Or la variable "benefDept" contient 104 modalités. Le nombre de variables créé par la binarisation est trop conséquent pour pouvoir réaliser un modèle linéaire généralisé. Elle doit donc être transformée. Il convient de la simplifier en regroupant les départements similaires ensemble. Or ces regroupements sont dépendants du poste de soins visé : en effet, les regroupements vont être formés à l'aide de variables externes et ces variables externes sont plus ou moins adaptées aux différents postes de soins. Les variables externes candidates ont été extraites de l'Institut National de la Statistique et des Etudes Economiques (INSEE). Elles ont été produites au niveau départemental et concernent l'année 2018. Voici la liste des variables extraites :

- ➔ Le nombre de médecins généralistes ;
- ➔ Le nombre d'ophtalmologues ;
- ➔ Le nombre de chirurgiens dentistes ;
- ➔ Le nombre d'audioprothésistes ;
- ➔ Le nombre de pharmaciens ;
- ➔ Le niveau de vie médian ;
- ➔ Le taux de chômage ;
- ➔ L'indice de vieillissement ;
- ➔ Le nombre d'établissements hospitaliers ;
- ➔ Le nombre de cures thermales ;
- ➔ Le nombre de laboratoires d'analyses ;
- ➔ Lentilles non remb.
- ➔ Le nombre de pharmacies ;
- ➔ Les taux bruts de mortalité ;
- ➔ Le taux de pauvreté ;
- ➔ Le nombre d'habitants.

Les variables traduisant un nombre d'équipements ont été modifiées à l'aide de la variable "Population" qui représente le nombre d'habitants par département afin d'obtenir la densité de ces équipements par département. Les autres variables sont économiques et fortement corrélées entre elles. Il faut donc être vigilant dans la sélection de ces variables et veiller à choisir les variables les plus adaptées à chaque sous-poste. Pour chaque poste de soins, la corrélation entre ces variables et la sinistralité est analysée. Si la variable est corrélée, elle est insérée dans le modèle. De plus, deux variables explicatives sont ajoutées : la fréquence de sinistres et le coût moyen des sinistres. Afin de limiter le nombre de retraitements, la création des classes a été faite au niveau de la dépense engagée. Ainsi, la sinistralité des différents départements est analysée en fonction des variables sélectionnées à travers un algorithme, l'algorithme des *k-means*, présenté dans la section 2.1.

Pour chaque poste, le nombre de classes a été déterminé en fonction des critères présentés précédemment (Le choix du nombre de classes), comme le montre la figure 2.3 pour le sous-poste Prothèses dentaires. Le premier graphique représente la méthode du coude et indique qu'un nombre de classes égal à 3 est optimal. Le second graphique en haut à droite représente la méthode silhouette et de même que le graphique précédent, il indique qu'un nombre de classes égal à 3 est optimal. Enfin, le dernier graphique trace les 3 classes formées dans le plan des deux axes principaux.

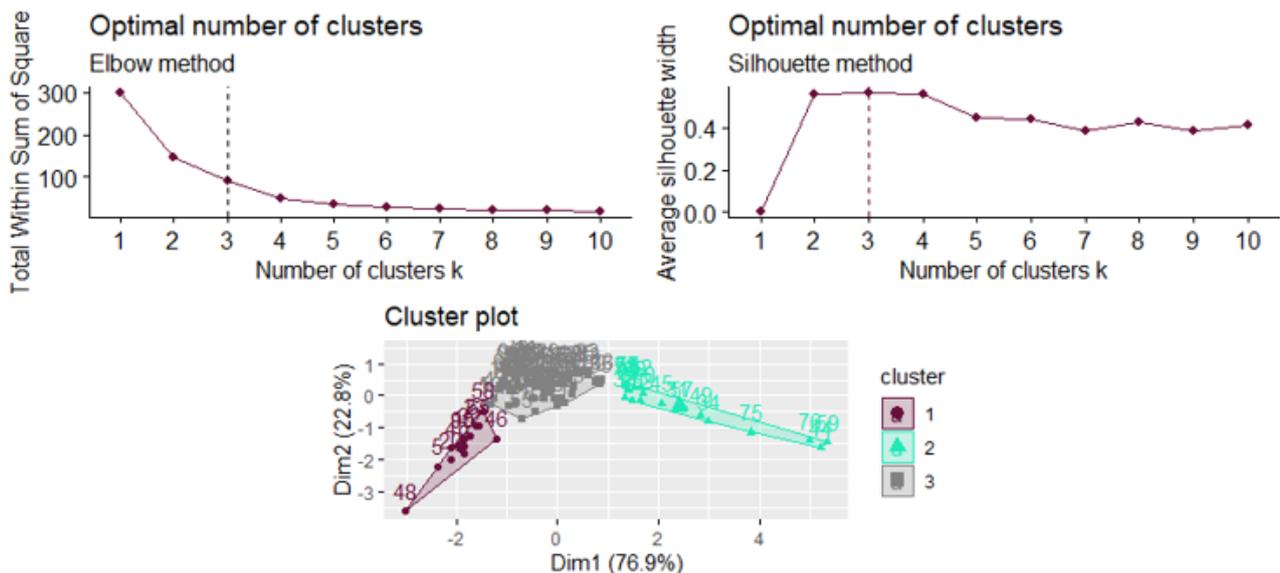


FIGURE 2.3: Représentation graphique des critères permettant de déterminer le nombre de classes optimales pour le sous-poste Prothèses dentaires.

Les départements ont été classés en 3 ou 4 classes pour chacun des sous-postes. Une représentation visuelle de la classification des départements de France métropolitaine pour ce sous-poste est donnée par la figure 2.4.

Ainsi, la simplification de la variable "benefDept" en quelques modalités choisies judicieusement permet de pouvoir binariser la variable sans créer un grand nombre de variables pour le modèle linéaire généralisé.

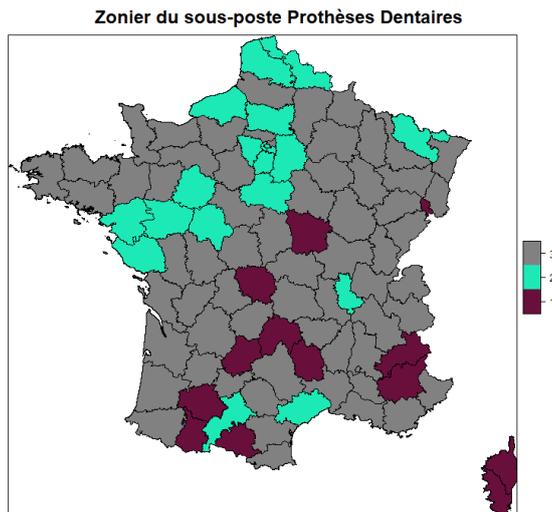


FIGURE 2.4: Cartographie de la classification obtenue pour le sous-poste Prothèses dentaires.

#### Retraitement de la variable “benefClasseAge” :

Afin de pouvoir comparer les montants de dépenses engagées en fonction des différents âges, il a d’abord fallu éliminer le biais dû à la composition du portefeuille, expliqué dans la section précédente. Ce biais a été éliminé en pondérant le montant des dépenses engagées de chaque âge par l’exposition de l’âge associé.

D’une base de modélisation à l’autre la répartition des sinistres en fonction de l’âge est très variable. Par exemple, les sinistres du sous-poste Audiologie concernent principalement des âges élevés tandis que les sinistres du sous-poste Orthodontie concernent principalement des âges faibles. Dans le cas du sous-poste Analyses Médicales, une surconsommation est observée pour les femmes entre 20 ans et 40 ans. Cette observation est très certainement due aux bilans sanguins réalisés dans le cadre de prescriptions contraceptives. Malgré ces distinctions spécifiques entre chaque sous-poste, une tendance exponentiellement croissante se dessine majoritairement : plus l’individu est âgé, plus le coût des actes de soins est élevé.

Ainsi la séparation de l’âge de manière uniforme en formant des classes de 10 en 10 est remise en question. De plus, utiliser la variable “benefAge” dans les modèles linéaires généralisés n’est pas possible. En effet, la dépense engagée ne suit une tendance linéaire due à l’âge. Un retraitement de la variable “benefClasseAge” est donc nécessaire. De même que pour le retraitement précédent, la variable “benefClasseAge” a été transformée en fonction de la dépense engagée. Les classes ont été créées afin de former des groupes d’individus ayant des caractéristiques de sinistralité homogènes à l’aide d’arbres de régression. Les arbres de régression sont présentés dans la partie 2.1 du mémoire.

Les individus sont regroupés successivement selon leurs caractéristiques, ici leur âge, jusqu’à l’obtention d’une unique catégorie. La dépense engagée est donc segmentée en fonction de l’âge des individus. A cet algorithme est ajoutée une méthode graphique de segmentation de la sinistralité par âge. Pour certains sous-postes, l’algorithme ne segmente pas assez les observations. C’est le cas pour les prothèses auditives par exemple. L’algorithme propose deux catégories ([0,69] et 70+) alors que visuellement, comme le montre la figure 2.5 une catégorie intermédiaire se dessine ([50,69]).

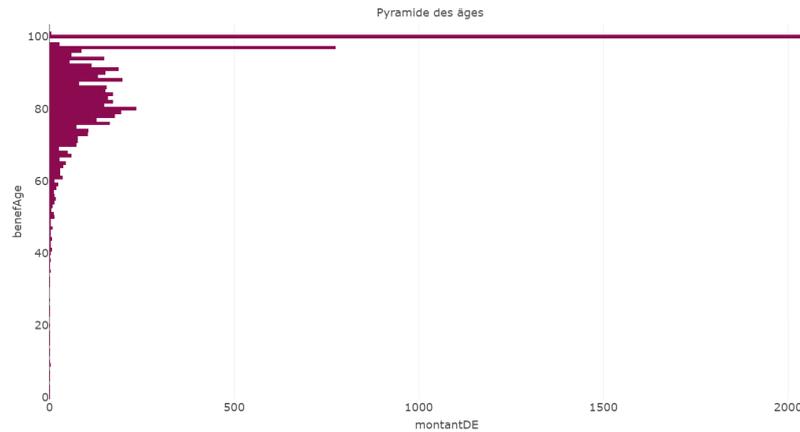


FIGURE 2.5: Variation de la dépense engagée en fonction de l'âge pour le sous-poste Audiologie.

Le tableau A.11 présenté en Annexe 2 résume les classifications choisies pour tous les sous-postes.

Ainsi les bases de modélisation ont été analysées puis transformées afin de pouvoir réaliser deux modèles linéaires généralisés par sous-poste : un premier modèle linéaire généralisé pour le coût et un second pour la fréquence.

### 2.2.2 La modélisation de la sinistralité de chaque sous-poste par une approche coût-fréquence

Dans cette partie, l'approche coût-fréquence utilisant des modèles linéaires généralisés est présentée pour chaque sous-poste excepté le sous-poste Maternité modélisé différemment puisqu'il ne contient pas assez d'observations afin d'obtenir un modèle de coût robuste. La modélisation a été inspirée par le cours de Statistique de l'assurance écrit par Arthur Carpentier, CARPENTIER (2010).

Tout d'abord, la théorie sur l'approche coût-fréquence est rappelée au lecteur. Avant de commencer la modélisation via des modèles linéaires généralisés, les sinistres extrêmes sont enlevés et modélisés à part. Ensuite, les étapes de modélisation sont développées dans un premier temps pour le modèle de fréquence et dans un second temps pour le modèle de coût.

#### L'approche coût-fréquence

La modélisation par l'approche coût-fréquence se fait de manière séparée : d'une part la fréquence de sinistres est modélisée et d'une autre part le coût moyen des sinistres est modélisé. L'approche repose sur le principe du modèle collectif.

#### Notations :

- \*  $N$  représente le nombre total de sinistres pour une année donnée, il s'agit d'une variable aléatoire à valeurs dans  $\mathbb{N}$  ;
- \*  $(X_i)_{i>0}$  où  $X_i$  représente la charge du sinistre  $i$ , il s'agit d'une suite de variables aléatoires continues à valeurs dans  $\mathbb{R}_+$  ;
- \*  $S = \begin{cases} \sum_{i=1}^N X_i & \text{si } N \geq 1 \\ 0 & \text{sinon} \end{cases}$  représente la charge annuelle des sinistres.

Hypothèses :

- \* La suite de variables aléatoires  $X$  est indépendante et identiquement distribuée. Ainsi, les montants de sinistres ont le même comportement aléatoire ;
- \* Pour tout  $i$  strictement positif, les variables  $N$  et  $X_i$  sont supposées indépendantes. Autrement dit, la sévérité n'a pas d'influence sur la fréquence ;

Grâce à ces hypothèses, il est possible d'obtenir une expression simple de l'espérance de la charge annuelle de sinistres.

$$\mathbb{E}(S) = \mathbb{E}\left[\mathbb{E}(S|N)\right] = \mathbb{E}\left[\mathbb{E}\left(\sum_{i=1}^N X_i|N\right)\right].$$

Or  $(X_i)_{i>0}$  est une suite de variables indépendantes identiquement distribuées et indépendante de  $N$ ,

$$\mathbb{E}(S) = \mathbb{E}\left[\sum_{i=1}^N \mathbb{E}(X_i|N)\right] = \mathbb{E}\left[\sum_{i=1}^N \mathbb{E}(X_i)\right] = \mathbb{E}\left[\sum_{i=1}^N \mathbb{E}(X_1)\right] = \mathbb{E}(N) \mathbb{E}(X_1).$$

Le principe de l'équité actuarielle consiste à équilibrer les engagements de l'assureur envers les assurés en imposant l'égalité des valeurs actuelles probables des flux futurs de l'assuré. La prime pure unique  $\Pi$  correspond donc à l'espérance de la variable  $S$ . En effet, il s'agit du coût moyen multiplié par le nombre de fois où l'acte est consommé.

La variable exposition des bases de données représente l'exposition du bénéficiaire sur l'exercice considéré. Il ne faut pas oublier de prendre ce facteur en compte dans la modélisation. En effet, la fréquence de sinistres entre une personne qui est exposée seulement une faible partie de l'année et une personne qui est exposée durant toute l'année ne peut être comparée. Il faut donc prendre en compte ce facteur dans la modélisation.

Ainsi, la fréquence de sinistres d'un assuré est la division entre le nombre moyen de sinistres et l'exposition de cet assuré. Tandis que le coût moyen est lui égal au coût total moyen des sinistres divisé par le nombre moyen de sinistres.

La prime s'écrit donc

$$\Pi = \underbrace{\frac{\text{Nombre moyen de sinistres}}{\text{Exposition}}}_{\text{la fréquence}} \times \underbrace{\frac{\text{Coût moyen des sinistres}}{\text{Nombre moyen de sinistres}}}_{\text{le coût moyen}}. \quad (2.24)$$

### La modélisation des sinistres extrêmes

Les sinistres extrêmes ont été retirés des bases. Ils ont été modélisés à l'aide d'une surprime, elle-même calculée à partir de la fréquence et du coût moyen de ces sinistres. En effet, les sinistres ayant des valeurs extrêmes peuvent avoir une très grande importance sur les résultats d'une régression linéaire. Les ôter de la modélisation permet de mieux ajuster les différentes lois aux données et d'obtenir de meilleurs résultats. Étant donné que les sinistres enlevés représentent une partie infinitésimale des bases de données, une modélisation de ces sinistres via un autre modèle linéaire généralisé est impossible. Une surprime a donc été modélisée à l'aide du coût moyen et de la fréquence de ces sinistres :

$$\text{Surprime}_i = \text{Coût moyen}_i \times \text{Freq}_i. \quad (2.25)$$

### La modélisation de la fréquence de sinistres

Cette section est consacrée aux modèles de fréquence. Elle présente la démarche suivie ainsi que les résultats obtenus par les différents modèles.

#### Lois de fréquence et prise en compte de l'exposition :

La fréquence est le nombre de sinistres divisé par l'exposition. Le nombre de sinistres est modélisé à l'aide de variables de comptage, comme la loi de Poisson et la loi Binomiale négative. Ces lois ont été présentées dans la section 2.1.

La loi de Poisson repose sur une hypothèse forte : l'espérance est égale à la variance. Or cette hypothèse est rarement vérifiée en pratique. Lorsqu'il y a une trop forte dispersion, une autre loi plus adaptée est utilisée. Il s'agit de la loi Binomiale négative. En effet, cette loi de fréquence est caractérisée par une variance plus forte que l'espérance et peut permettre une meilleure adéquation des données au modèle. C'est pourquoi son utilisation est préférable lorsque la variance est fortement supérieure à l'espérance.

Le nombre de sinistres d'un individu doit être pondéré par son exposition. En effet, la fréquence de sinistre d'un individu exposé à 50% n'est pas la même que celle d'un individu exposé à 100% sur un exercice donné. Afin de prendre en compte l'exposition, on utilise un offset dans le modèle de régression : le coefficient  $\beta$  devant la variable exposition à laquelle est appliquée la fonction logarithme vaut 1.

$$\begin{aligned} g(\mathbb{E}(Freq)) &= X' \beta \\ \Leftrightarrow \log(\mathbb{E}(\frac{NbSin}{Exposition})) &= X' \beta \\ \Leftrightarrow \log(\mathbb{E}(NbSin)) &= X' \beta + \log(Exposition) \\ \Leftrightarrow \mathbb{E}(NbSin) &= e^{X' \beta + \log(Exposition)}. \end{aligned}$$

#### Variables retenues pour la modélisation :

Avant de lancer les modèles, une analyse sur la corrélation des variables catégorielles a été menée. Elle a utilisé le V de Cramer, HARALD CRAMÉR (1946), qui est une normalisation de la statistique du test d'indépendance du  $\chi^2$  :

$$V = \sqrt{\frac{\chi^2}{n \cdot (\min(l, c) - 1)}} \quad (2.26)$$

où  $n$  est le nombre d'observations,  $l$  et  $c$  sont respectivement le nombre de modalités de la première et seconde variable. Il s'agit d'un nombre compris entre 0 et 1. Plus les variables sont indépendantes, plus  $V$  tend vers 0 et inversement, plus elles sont liées, plus  $V$  tend vers 1. Pour rappel, les variables explicatives d'un modèle linéaire ne doivent pas être corrélées, c'est pourquoi les variables fortement liées ont été enlevées de la modélisation. Dans ce but, une matrice résumant le V de Cramer pour chacune des variables explicatives de chaque sous-poste a été construite, comme le montre la figure 2.6 pour le sous-poste Prothèses. Pour tous les sous-postes, une forte corrélation entre les variables "contractFormula" et "couvFacultative" est observée ainsi qu'entre les variables couverture et "couvFacultative". Les variables "couverture" et "contractFormula" ont donc été retirées de la modélisation pour toutes les bases de données.

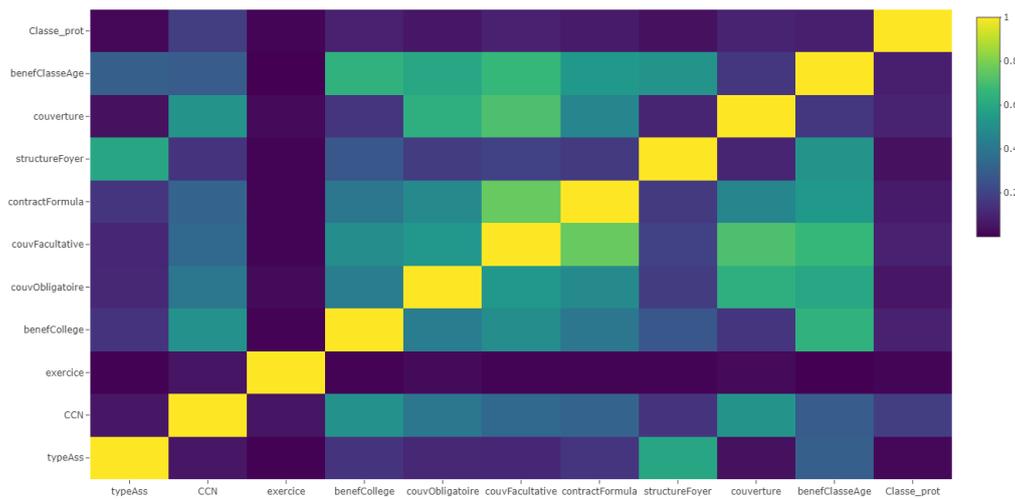


FIGURE 2.6: V de Cramer du sous-poste Prothèses.

De plus, la variable indiquant le sexe de l'assuré a été retirée de la modélisation. En effet, conformément à la directive européenne de 2004, il est interdit de discriminer par le sexe les assurés, EUR-LEX (2004).

Modélisation à l'aide des deux lois de fréquence et sélection du meilleur modèle :

Pour chaque sous-poste, les modèles de fréquence ont été réalisés sur une base d'apprentissage et analysés grâce à une base de test. Les critères AIC et BIC ainsi que la méthode Forward-Backward ont été utilisés afin de procéder à la sélection de variables des différents modèles de fréquence.

La loi de fréquence représentant le mieux la fréquence de chaque sous-poste a été choisie. Afin de faire ce choix, plusieurs indicateurs et plusieurs méthodes ont été combinées :

- ⇒ La **déviance** doit être minimisée. La déviance est l'écart entre les prédictions et les valeurs observées.
- ⇒ Le **RMSE** doit être minimisé. Il s'agit de la racine carrée de l'erreur quadratique moyenne, plus précisément il s'agit de la racine carrée de la somme des erreurs entre les valeurs observées et les prédictions.
- ⇒ Un **ratio** a été calculé : il s'agit de la somme des valeurs observées sur la somme des prédictions. S'il est supérieur à 1, cela signifie que le modèle sous-estime la fréquence et inversement s'il est inférieur à 1, il surestime la fréquence. Par mesure de prudence, un ratio supérieur à 1 est préféré à un ratio inférieur à 1 à écart égal.
- ⇒ Le **MAE** doit être minimisé. Il s'agit de la moyenne arithmétique des valeurs absolues des écarts.
- ⇒ Une **approche graphique** a été utilisée. La distribution empirique ainsi que la répartition empirique du nombre de sinistres ont été tracées. A celle-ci ont été comparées les distributions de la loi de Poisson et de la loi Binomiale négative.

La prise en compte de ces multiples indicateurs analysant un critère particulier permet ainsi d'obtenir une vision complète des prédictions faites par les modèles par rapport aux valeurs observées de la base de test.

Le tableau 2.3 et la figure 2.11 répertorient les différents indicateurs ainsi que les méthodes graphiques qui ont été employées à travers l'exemple du sous-poste Radiologie appartenant au poste Soins courants. Bien que le modèle utilisant la loi de Poisson minimise la déviance ainsi que le MAE, le modèle Binomiale négatif semble plus adapté. En effet, le modèle de Poisson surestime les faibles fréquences et sous-estime les grandes fréquences. C'est pourquoi, le modèle utilisant la loi Binomiale négative a été choisi pour modéliser la fréquence de ce sous-poste.

LOI DU MODÈLE	RMSE	RATIO	MAE	DÉVIANCE
Poisson	1,721	0,959	1,057	61 455,9
Binomiale négative	1,717	0,959	1,070	62 209,6

TABLE 2.3: Indicateurs des modèles de fréquence pour le sous-poste Radiologie (Soins courants).

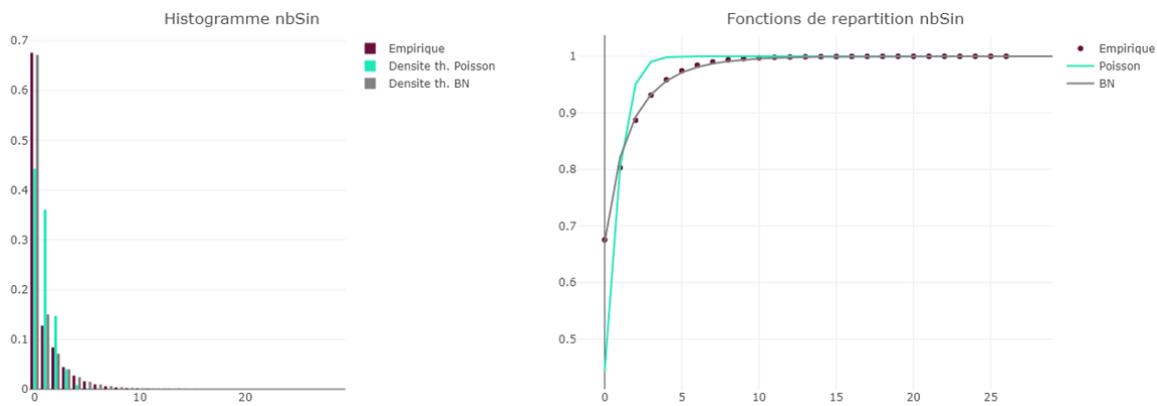


FIGURE 2.7: Adéquation graphique des modèles de fréquence pour le sous-poste Radiologie (Soins courants).

Cet exemple montre l'utilité d'utiliser plusieurs indicateurs variés. Bien que certains choix peuvent être plus complexes étant donné le nombre d'indicateurs, cela permet de prendre en compte différentes caractéristiques et donc de sélectionner le meilleur modèle.

Plus généralement, lorsque le nombre maximal de sinistres est élevé, la loi Binomiale négative semble être plus appropriée. Comme expliqué précédemment, elle s'adapte mieux en présence de surdispersion. A contrario, lorsque le nombre maximal de sinistres est faible, les deux distributions ont tendance à se superposer et des résultats similaires sont obtenus.

#### Pertinence des résultats du modèle retenu :

En premier lieu, afin d'analyser les résultats obtenus grâce au modèle, il convient d'analyser les résidus. En effet, il faut vérifier les hypothèses concernant les résidus décrites dans la partie théorique sur les modèles linéaires. Cependant, les résidus construits n'ont pas une variance constante. C'est pourquoi ils sont standardisés. Deux tests sont communément appliqués, le test des résidus de Pearson et le test des résidus de la déviance. Les statistiques de test, qui en dérivent, suivent une loi du  $\chi^2$  sous l'hypothèse de l'adéquation du modèle aux données.

Tout d'abord, le  $i$ -ème résidu de Pearson est un résidu normalisé, il est défini par l'équation 2.27 :

$$r_i^p = \frac{y_i - \hat{y}_i}{\sqrt{\mathbb{V}(\hat{y}_i)}}. \quad (2.27)$$

La figure 2.8 représente la répartition des résidus de Pearson en fonction de la fréquence prédite. Les résidus ne sont pas bien centrés. La première hypothèse sur les résidus n'est donc pas vérifiée. Le modèle a, en effet, tendance à sous-estimer la fréquence. De plus, les résidus ne forment pas une bande horizontale autour de zéro et ne semblent pas être distribués de manière aléatoire, l'hypothèse d'homoscédasticité ainsi que l'hypothèse d'indépendance ne sont donc pas non plus vérifiées.

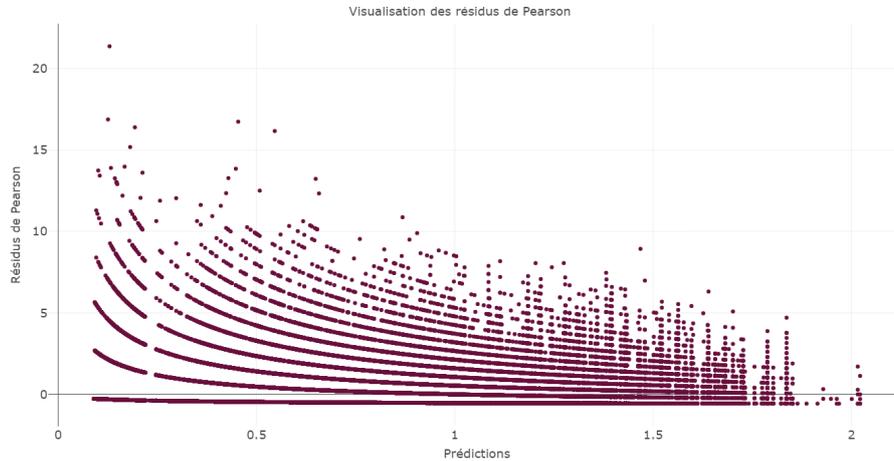


FIGURE 2.8: Graphe des résidus de Pearson en fonction des prédictions pour le modèle Binomial négatif du sous-poste Radiologie (Soins courants).

La statistique du test de Pearson est définie de la manière suivante :

$$\chi_p^2 = \sum_{i=1}^N (r_i^p)^2. \quad (2.28)$$

Sous l'hypothèse nulle de l'adéquation du modèle aux données, la statistique du test de Pearson suit une loi du  $\chi^2$  à  $(n - p - 1)$  degrés de liberté où  $n$  représente le nombre d'observations et  $p$  le nombre de variables. Le seuil de confiance choisi sera de 95% par la suite. Si la p-valeur est supérieure à 0,05 alors le modèle est adapté aux données.

Pour le cas du sous-poste Radiologie précédent, bien que graphiquement les hypothèses sur les résidus ne soient pas vérifiées, la p-valeur du test de Pearson est de 0,68, le modèle est donc validé.

Le second test, le test des résidus, est similaire au test précédent. Le  $i$ -ème résidu de déviance est défini par l'équation 2.29, où  $d_i$  représente la contribution de l'observation  $i$  à la déviance.

$$r_i^d = \sqrt{d_i} \times \text{signe}(y_i - \hat{y}_i). \quad (2.29)$$

Sous l'hypothèse nulle, la statistique  $\chi_d^2 = \sum_{i=1}^N (r_i^d)^2$  suit également une loi du  $\chi^2$  à  $(n - p - 1)$  degrés de liberté où  $n$  représente le nombre d'observations et  $p$  le nombre de variables. De même, si la p-valeur est supérieure à 0,05 alors le modèle est adapté aux données.

Cette fois-ci, la p-valeur atteint 1 pour le sous-poste Radiologie. Le modèle est donc une nouvelle fois jugé vérifié par un test statistique. Ainsi, les tests statistiques ne révèlent pas les constatations graphiques faites sur les résidus.

A contrario du sous-poste présenté, certains sous-postes n’obtiennent pas de p-valeurs suffisamment élevées afin de valider l’adéquation du modèle. Étant donné le nombre de modèles à réaliser, seulement deux lois de modélisation ont été utilisées, cependant d’autres lois plus complexes comme des lois à inflation de zéros aurait pu permettre une meilleure adéquation. Un bon candidat pour une loi *zero-inflated* aurait été le sous-poste Cures thermales. En effet, il est caractérisé par une très faible probabilité de survenance d’un sinistre (1,11%) et certains individus totalisent plusieurs dizaines de sinistres pour ce sous-poste (63 sinistres au maximum).

Le modèle obtenu pour le sous-poste Radiologie peut également être analysé à l’échelle des variables et de leurs modalités. Un exemple de graphe est montré par la figure 2.9. Comme constaté précédemment, les prédictions ont tendance à légèrement sous-estimer la fréquence de consommation. Cependant, dans l’ensemble, le modèle permet de bien représenter l’évolution de la fréquence de consommation des assurés en fonction de leur classe d’âge et de leur type d’assuré, et ce, même lorsque les modalités sont peu exposées dans la base.

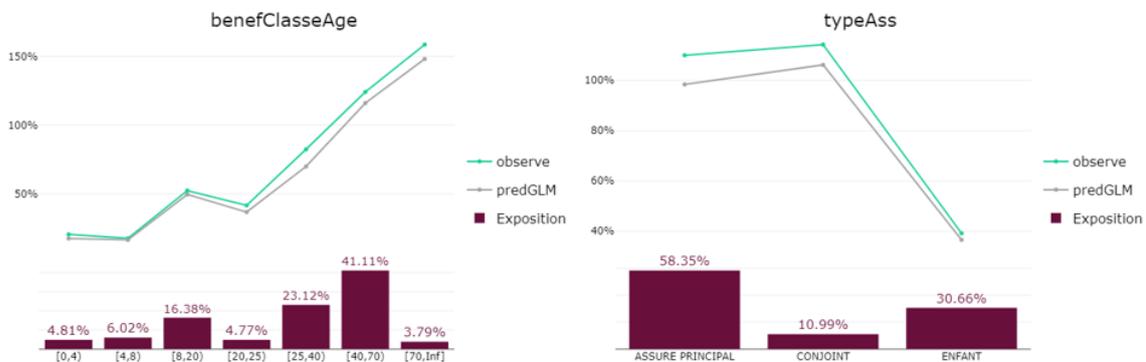


FIGURE 2.9: Evolution des prédictions et des observations en fonction des modalités des variables “structureFoyer” et “typeAss” pour le sous-poste Radiologie (Soins courants).

#### Utilisation du modèle retenu :

Une analyse de l’importance de chaque modalité peut être menée au travers des coefficients du modèle linéaire généralisé. Le tableau 2.4 répertorie les différents coefficients ainsi que la p-valeur issue du test de Student. Ce test permet de vérifier la non-nullité du coefficient. Si la p-valeur est supérieure à 0,05, alors l’hypothèse est vérifiée. De plus, la forte importance de la modalité est caractérisée par le symbole \*\*\*. L’intercept représente l’individu fictif de référence ayant les caractéristiques suivantes :

- ⇒ Il s’agit de l’assuré principal ;
- ⇒ Il a moins de 4 ans ;
- ⇒ Il s’agit d’un employé cadre ;
- ⇒ Il s’agit d’un individu en couple avec enfant(s) à charge ;
- ⇒ Il habite dans un département appartenant à la classe 1 ;
- ⇒ Il ne bénéficie pas d’une couverture complémentaire obligatoire ;
- ⇒ Il ne bénéficie pas d’une couverture complémentaire facultative ;
- ⇒ Il est rattaché à la CCN 1 ;
- ⇒ Il est rattaché à l’exercice de l’année 2018.

Cet individu fictif sert de référence. Si un individu a une modalité différente, le coefficient associé est ajouté afin de calculer la fréquence. Cette modalité différente peut avoir un impact plus ou moins important sur la fréquence prédite et il peut être positif ou négatif. Par exemple, si l'individu est rattaché à la deuxième convention collective nationale, toute caractéristique égale par ailleurs à l'individu de référence, alors sa fréquence prédite sera inférieure à ce dernier puisque le coefficient associé à la modalité est de  $-0,38$ .

VARIABLE	MODALITÉ	COEFFICIENT	P-VALEUR	IMPORTANCE DE LA MODALITÉ
Intercept		-2,04	< 2e-16	***
typeAss	Conjoint	-0,08	1,61e-06	***
	Enfant	-0,53	1,56e-08	***
benefClasseAge	[4,7]	-0,15	0,26	
	[8,19]	0,89	< 2e-16	***
	[20,24]	0,55	< 2e-16	***
	[25,39]	0,93	< 2e-16	***
	[40,69]	1,21	< 2e-16	***
	70+	1,30	< 2e-16	***
benefCollege	Ensemble du personnel	0,03	0,14	
	Employé non cadre	0,06	0,0086	**
structureFoyer	Couple sans enfant	0,008	0,66	
	Isole avec enfant	0,01	0,43	
	Isole sans enfant	-0,15	< 2e-16	***
benefClasseDept	2	-0,07	0,03	*
	3	0,02	0,56	
couvObligatoire	0	0,30	< 2e-16	***
	1	0,45	< 2e-16	***
	2	0,63	< 2e-16	***
couvFacultative	0	0,46	< 2e-16	***
	1	0,30	< 2e-16	***
	2	0,43	< 2e-16	***
CCN	2	-0,38	< 2e-16	***
	3	0,07	0,0044	**
	4	0,32	< 2e-16	***
exercice	2019	0,03	0,00146	**

TABLE 2.4: Coefficients du modèle de fréquence du sous-poste Radiologie (Soins courants).

Les coefficients des modalités des variables représentant la structure du foyer ainsi que le type de l'assuré ont été tracés dans la figure 2.10. L'évolution des coefficients est bien en adéquation avec les résultats de la figure 2.9 précédente. Cette fois-ci, les coefficients utilisent pour référence la première classe, à savoir la classe d'âge de 0 à 4 ans pour le graphique de gauche et l'assuré principal pour le graphique de droite. Par exemple, le pic de fréquence entre 8 et 20 ans est bien retrouvé sur le graphique de gauche et le coefficient associé à la modalité Enfant est élevé en valeur absolue puisque la fréquence de consommation est sensiblement inférieure pour les enfants.

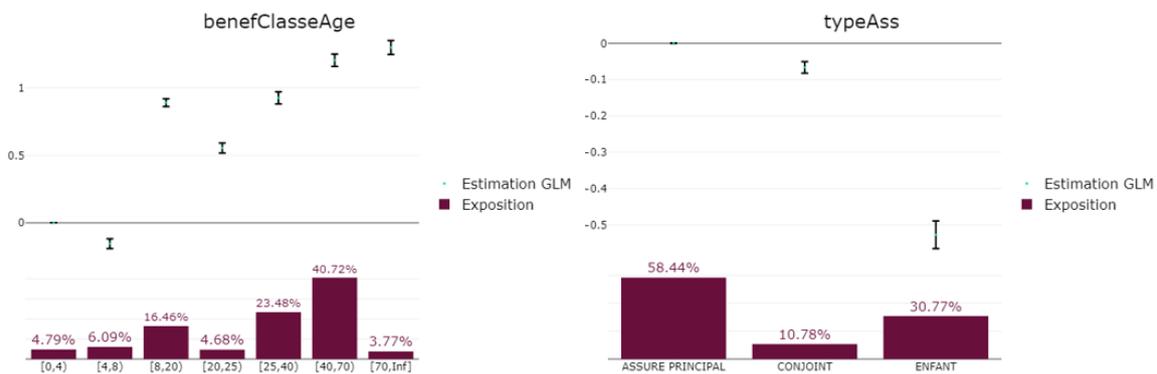


FIGURE 2.10: Évolution des coefficients en fonction des modalités des variables “benefClasseAge” et “typeAss” pour le sous-poste Radiologie (Soins courants).

Ainsi, les modèles linéaires généralisés permettent une étude de la fréquence de sinistres en fonction des caractéristiques des assurés. Comme expliqué lors des analyses univariées et bivariées des données, certains sous-postes sont caractérisés par différentes populations à risque. Par exemple, un enfant aura a priori un coût et une fréquence espérés plus élevés en orthodontie qu’un adulte. Ainsi, en fonction du sous-poste concerné, différents profils de risques peuvent être créés. Cette utilisation des modèles peut être adaptée à la tarification individuelle afin de créer un produit de complémentaire santé adapté à l’assuré bien que dans la suite de ce mémoire, les résultats des modèles seront exploités de manière collective.

Maintenant que la démarche utilisée pour les modèles de fréquence a été présentée, le coût doit l’être également.

### La modélisation du coût moyen des sinistres :

Une procédure similaire aux modèles de fréquence a été utilisée pour créer les différents modèles.

#### Modèles de coût et sélection du meilleur modèle :

Les sinistres extrêmes et les variables corrélées ont déjà été enlevés des bases de données. La seule modification à effectuer concerne la variable représentant la dépense engagée. Étant donné que le but est de prédire le coût moyen d’un sinistre, il faut diviser ce montant par le nombre de sinistres avant de commencer la modélisation.

Le coût est généralement modélisé par une loi Gamma, qui est une généralisation de la loi exponentielle, ou bien une loi Log-normale. Les deux modèles sont décrits dans le section 2.1. La différence principale entre la loi Gamma et la loi Log-normale réside dans le fait que la loi Gamma est plus sensible aux valeurs extrêmes. En effet, l’application de la fonction logarithme permet de diminuer l’importance des sinistres extrêmes.

Afin de comparer les deux modèles, les mêmes indicateurs ont été utilisés et un diagramme quantile-quantile a été ajouté pour la comparaison graphique des deux lois. L’exemple des modèles de coût pour le sous-poste Lentilles non remboursables par la Sécurité sociale est détaillé ci-dessous. La loi Gamma et la loi Log-normale permettent une bonne adéquation aux données. Cependant, la loi Log-normale est plus adaptée. Bien qu’il soit impossible de voir la différence en regardant le graphe des fonctions de répartition, le graphe des densités et le diagramme quantile-quantile ainsi que les métriques permettent

de voir cette différence. En effet, la loi Gamma sous-estime les sinistres ayant un coût inférieur à 100€ tandis que la loi Log-normale s'adapte mieux à ce niveau. De plus, les métriques utilisées dans le tableau 2.5 montrent de meilleurs résultats pour le modèle Log-normal. C'est pourquoi, la loi Log-normale a été retenue afin de modéliser le coût de ce sous-poste.

LOI DU MODÈLE	RMSE	RATIO	MAE	DÉVIANCE
Gamma	60,85	0,998	41,70	111 555,9
Log-normale	60,84	0,992	41,53	111 087,9

TABLE 2.5: Indicateurs des modèles de coût pour le sous-poste Lentilles non remboursables par la Sécurité sociale.

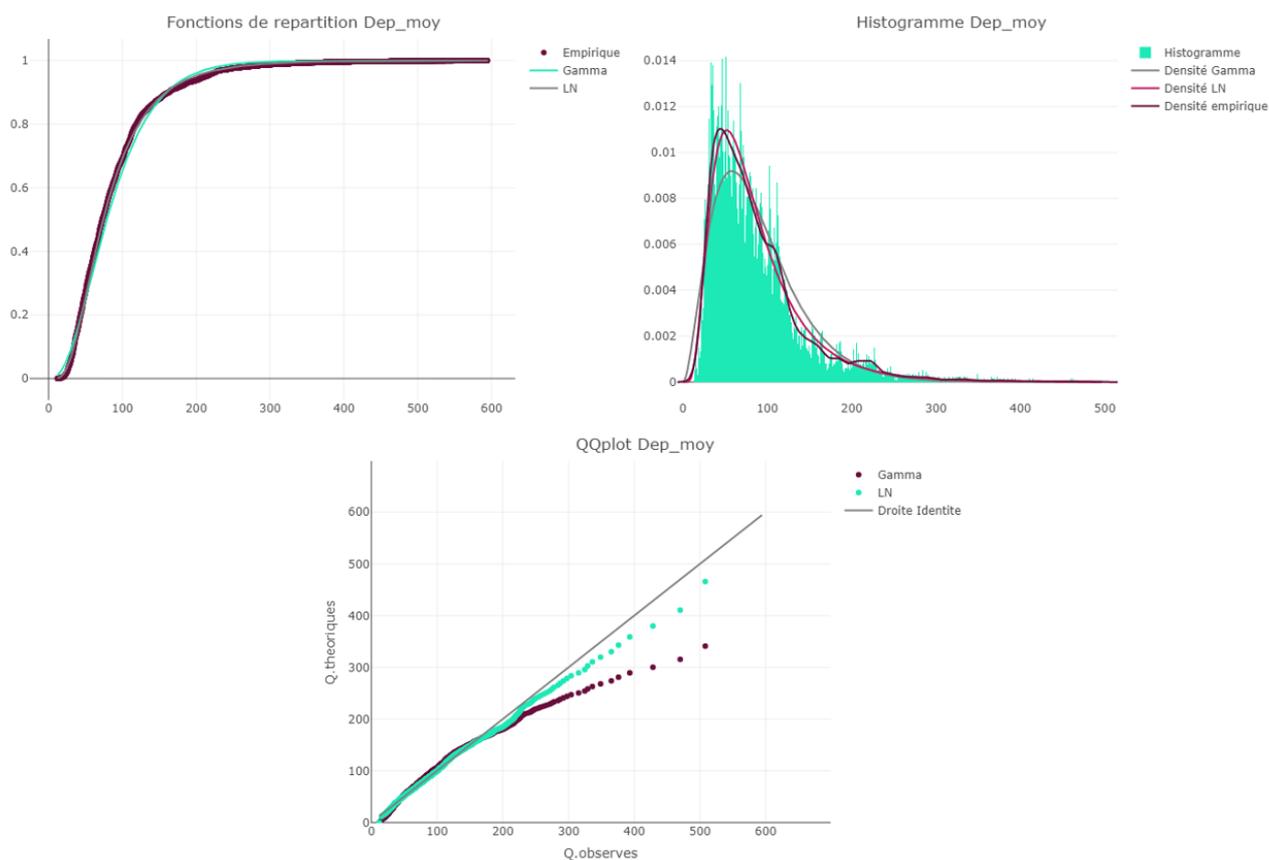


FIGURE 2.11: Adéquation graphique des modèles de coût pour le sous-poste Lentilles non remboursables par la Sécurité sociale.

#### Pertinence des résultats du modèle retenu :

Comme précédemment, afin de mesurer l'adéquation du modèle retenu, les tests de Pearson et de déviance ont été réalisés. Tous deux donnent des p-valeurs de 1, l'hypothèse de validité du modèle est donc vérifiée. De plus, le graphe 2.12, confirme cette fois-ci les hypothèses sur les résidus, à savoir l'hypothèse de centralité, l'hypothèse d'homoscédasticité et confirme l'absence de valeur aberrante qui aurait pu mener à un modèle moins performant.

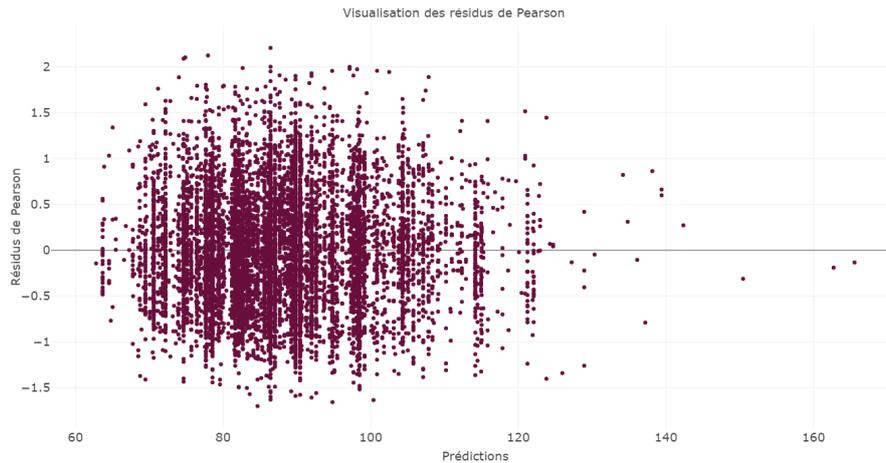


FIGURE 2.12: Graphe des résidus de Pearson en fonction des prédictions pour le modèle Log-normal du sous-poste Lentilles non remboursables par la Sécurité sociale.

De même que précédemment, l'adéquation pour chaque modalité de chaque variable a été analysée. Comme le montre la figure 2.13, quelque soit le collègue de l'assuré, le modèle s'ajuste bien mais sous-estime le coût tandis que pour la classe d'âge, le modèle surestime le coût pour les individus jeunes et sous-estime celui des individus plus âgés. Ces mauvaises adéquations sont expliquées par la faible proportion de ces populations dans la base de données comparée aux individus âgés entre 12 et 75 ans.

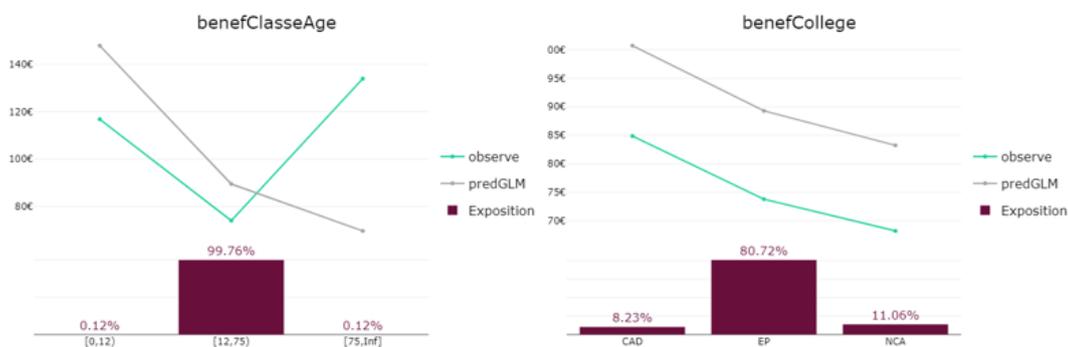


FIGURE 2.13: Évolution des observations et des prédictions en fonction des modalités des variables "benefClasseAge" et "benefCollege" du sous-poste Lentilles non remboursables par la Sécurité sociale.

#### Utilisation du modèle retenu :

Le tableau 2.6 répertorie les coefficients des modalités des variables retenues pour le modèle choisi et la figure 2.14 trace les coefficients des modalités des variables "benefClasseAge" et "benefCollege".

Le tableau montre que les variables ayant une forte importance sur le coût de ce sous-poste sont les variables caractérisant l'assuré, à savoir son type, sa classe d'âge, son collègue, la structure de son foyer ainsi que sa classe départementale. Comme le montre la figure 2.13, le coefficient associé aux personnes âgées de plus de 75 ans devrait être plus élevé, voire même supérieur à zéro.

VARIABLE	MODALITÉ	COEFFICIENT	P-VALEUR	IMPORTANCE DE LA MODALITÉ
Intercept		4,87	< 2e-16	***
typeAss	Conjoint	0,06	0,01	*
	Enfant	-0,11	2,23e-08	***
benefClasseAge	[12,74]	-0,61	8,79e-05	***
	75+	-0,10	0,0007	***
benefCollege	Ensemble du personnel	-0,13	2,62e-05	***
	Employé non cadre	-0,12	6,07e-06	***
structureFoyer	Couple sans enfant	0,07	0,01	*
	Isole avec enfant	-0,06	0,0007	***
	Isole sans enfant	-0,02	0,20	
benefClasseDept	2	0,10	5,95e-13	***
	3	0,10	6,06e-09	***
couvFacultative	0	-0,03	< 0,61	
	1	-0,01	< 0,75	
	2	0,21	< 2e-16	***
CCN	2	0,03	0,38	
	3	-0,05	0,13	
	4	0,08	0,006	**

TABLE 2.6: Coefficients du modèle de coût du sous-poste Lentilles non remboursables par la Sécurité sociale.

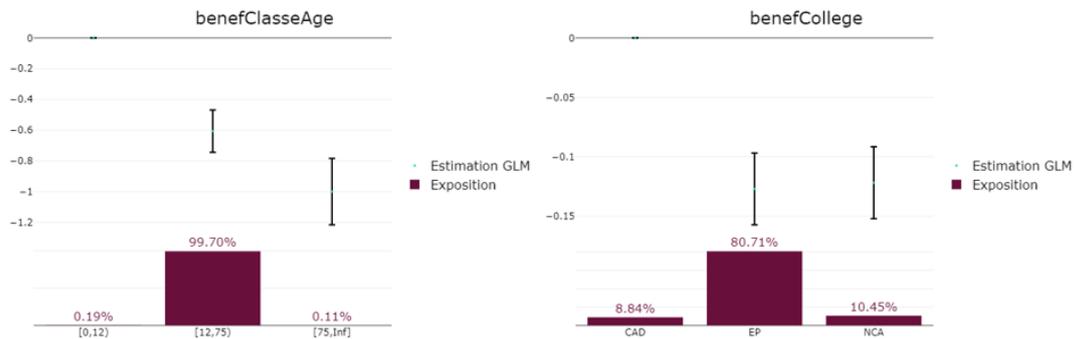


FIGURE 2.14: Évolution des coefficients en fonction des modalités des variables “benefClasseAge” et “benefCollege” du sous-poste Lentilles non remboursables par la Sécurité sociale.

A nouveau, cet exemple s’adapte bien aux lois proposées. Cependant, certains sous-postes sont composés d’actes variés caractérisés par des coûts différents. Certains, comme le sous-poste Auxiliaires médicaux sont caractérisés par des pics au niveau de leur histogramme comme le montre la figure 2.7. Plusieurs actes, notamment les actes infirmiers sont caractérisés par des frais réels inférieurs à 10€, par exemple, l’indemnité forfaitaire de déplacement lorsqu’une infirmière consulte à domicile s’élève à 2,50€. D’autres actes sont quant à eux, plus chers et oscillent autour de la vingtaine d’euros, par exemple une consultation chez un masseur-kinésithérapeute. Un second exemple est celui de l’orthodontie. Ce sous-poste est caractérisé par des coûts hétérogènes : une visite de contrôle coûtera quelques dizaines d’euros tandis que la pause d’un appareil en coûtera plusieurs centaines. Afin de mieux modéliser le coût de ces sous-postes, une modélisation à la maille acte de soins ou bien des méthodes de *machine learning* pourraient être utilisées. Par exemple, des algorithmes de classification pourraient permettre d’obtenir une meilleure segmentation des actes de soins et permettre ainsi d’obtenir de meilleurs résultats.

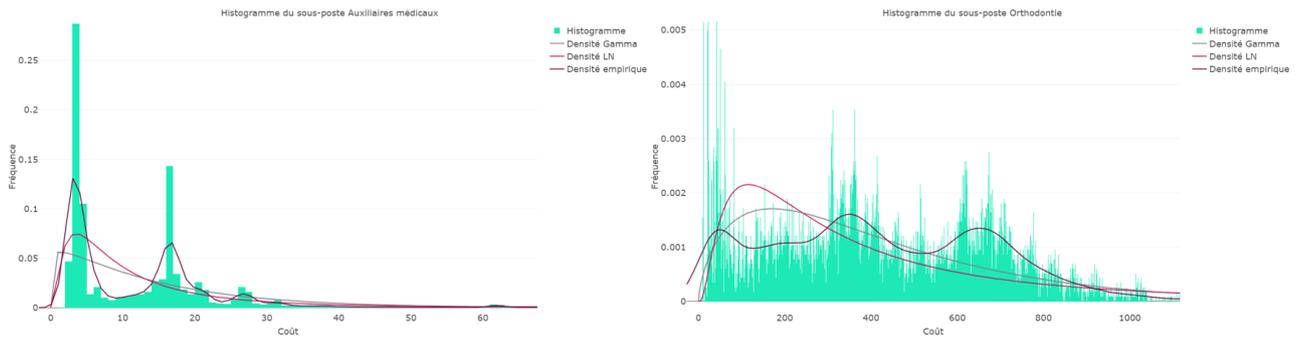


FIGURE 2.15: Densités empiriques et ajustement des lois pour les sous-postes Auxiliaires médicaux et Orthodontie.

Une description des résultats majeurs de chaque modèle de fréquence et chaque modèle de coût réalisé est présentée en Annexe 5. Ces descriptions ont pour but de résumer les facteurs les plus significatifs au niveau de la fréquence et du coût moyen de chacun des sous-postes.

Maintenant que la fréquence et le coût moyen ont été analysés séparément, il convient de regrouper ces résultats afin de former la sinistralité moyenne, tout en prenant en compte la surprime modélisée précédemment pour les sinistres extrêmes.

### 2.2.3 La sinistralité moyenne pour chaque sous-poste

Comme expliqué dans la section 2.3.1, la dépense moyenne associée à chaque sous-poste est le produit de la fréquence moyenne et du coût moyen de ce sous-poste. Les deux dernières quantités sont données par les modèles précédemment réalisés. De plus, les sinistres extrêmes ont été enlevés de la modélisation et modélisés par une surprime à ajouter à la quantité précédente. Ainsi la dépense moyenne engagée de chaque sous-poste est répertoriée dans le tableau 2.7.

Seuls les sous-postes Forfait hospitalier et Maternité ont été modélisés différemment. En effet, le forfait hospitalier, correspondant à la participation du patient aux frais liés à son hospitalisation, a un coût fixe de 20€, il n'y a donc qu'un modèle de fréquence réalisé pour ce sous-poste. Quant au forfait maternité, il représente le montant versé aux parents à la naissance de leur enfant par les contrats de complémentaire santé, lorsqu'un contrat est ouvert pour l'enfant, afin de faire face aux dépenses de santé futures liées à cet évènement. La base de données admet seulement 36 lignes de sinistres pour une base de données de 290 723 lignes, soit une fréquence d'environ 0,0001%. Un modèle linéaire généralisé ne peut être réalisé sur cette base de données. Parmi ces 36 lignes, une seule est caractérisée par un nombre de sinistres égal à 2 et une dépense engagée totale de 821,20€. De plus, le coût de chaque sinistre oscille entre 329€ et 424€, et est en moyenne égal à 403,85€. La dépense engagée de ce sous-poste sera représentée par le produit de la fréquence moyenne et du coût moyen. Il en résulte que ce poste a une dépense engagée moyenne de 0,04€.

L'association coût-fréquence mène à des résultats différents d'un sous-poste à l'autre. Par exemple, bien que le coût moyen des transports hospitaliers soit élevé (en moyenne 125€), la combinaison coût-fréquence amène à une sinistralité moyenne de 3,85€ étant donné la faible fréquence de consommation de ce sous-poste (0,029). En somme, ce sont les sous-postes caractérisés par une fréquence de consommation élevée ainsi qu'un coût moyen élevé qui forment des sous-postes conséquents en termes de sinistralité.

POSTE	SOUS-POSTE	FRÉQUENCE MOYENNE	COÛT MOYEN (EN €)	SURPRIME (EN €)	SINISTRALITÉ MOYENNE PAR SOUS-POSTE (EN €)	SINISTRALITÉ MOYENNE PAR POSTE (EN €)
Soins courants	Auxiliaires médicaux	3,11	14,78	4,05	50,02	312,43
	Analyses médicales	1,59	27,26	0,88	44,22	
	Autres soins	0,85	32,92	0,76	28,74	
	Consult. Généraliste	2,43	25,87	0,68	63,54	
	Consult. Spécialiste	2,72	30,34	4,28	86,80	
	Radiologie	0,82	47,29	0,32	39,10	
Dentaire	Prothèses dentaires	0,16	553,70	1,29	89,88	213,22
	Soins dentaires	0,99	32,57	0,51	32,75	
	Radiologie	0,39	116,29	11,82	57,17	
	Orthodontie	0,08	421,77	0,93	33,41	
Hospitalisation	Autres soins	0,07	19,73	0,09	1,49	140,04
	Frais optionnels	0,06	207,58	1,69	14,14	
	Frais de séjour	0,72	141,26	17,13	118,84	
	Forfait hospitalier	0,08	20	-	1,68	
	Maternité	-	-	-	0,04	
	Transport	0,03	125,30	0,22	3,85	
Optique	Lentilles remb.	0,01	115,12	0,004	0,76	107,89
	Lentilles non remb.	0,10	87,77	0,03	8,81	
	Monture	0,21	149,47	0,02	31,41	
	Verres	0,44	151,75	0,14	66,91	
Pharmacie	Med. hors nomencl.	0,02	31,10	0,001	0,62	98,83
	Med. Remb. 15%	0,59	2,82	0,14	1,80	
	Med. Remb. 30%	5,89	2,82	0,55	17,16	
	Med. Remb. 65%	14,23	5,44	1,83	79,24	
Audiologie	Audiologie	0,02	536,49	0,26	10,99	10,99
Cures thermales	Cures thermales	0,03	425,36	0,28	13,04	13,04
Prothèses	Prothèses	0,68	43,12	1,71	31,03	31,03

TABLE 2.7: Présentation des résultats des modélisations par sous-poste sur la dépense engagée.

Les sous-postes ayant la sinistralité la plus élevée sont :

- ❶ Les frais de séjour : ce sous-poste est caractérisé par un coût moyen élevé, une fréquence moyenne et une lourde queue de distribution (17,13€ de surprime) ;
- ❷ Les prothèses dentaires : bien que la fréquence soit faible pour ce sous-poste, le coût moyen est fortement élevé (554€) ;
- ❸ Les consultations chez le médecin spécialiste : a contrario des prothèses dentaires, c'est la fréquence de sinistres qui est élevée (2,62 en moyenne) ;
- ❹ Les médicaments remboursés à 65% par la Sécurité sociale, caractérisés par la fréquence de consommation la plus élevée parmi tous les sous-postes (14,23 en moyenne).

La figure 2.16 résume la répartition de la dépense moyenne engagée en soins de santé par poste. Les soins courants représentent plus d'un tiers de la consommation de soins en santé. Ils sont caractérisés par une fréquence élevée et un coût moyen oscillant entre 15€ et 50€. De plus, la consommation en médicaments a été modélisée à part mais elle aurait pu être assimilée à ce poste. Cette dernière représente à elle seule 11% de la consommation en santé et est principalement due à la consommation de médicaments remboursés à 65% par la Sécurité sociale, car considérés comme médicaments à service médical rendu\* majeur ou important. De plus, il ne faut pas oublier que les médicaments intégralement remboursés, car considérés comme irremplaçables, ne sont pas modélisés ici. La part consacrée aux médicaments est donc en réalité plus conséquente. En complément, le dentaire, l'hospitalisation et l'optique, tous trois caractérisés par des coûts élevés, représentent respectivement 23%, 15% et 12% de la consommation en santé. Ces cinq postes totalisent 95% de la dépense moyenne en santé. Les prothèses, les cures thermales et l'audiologie viennent compléter la dépense. Leur fréquence de sinistres faible contribue à limiter leur sinistralité. Au total, la dépense engagée moyenne s'élève à 927,47€ tout poste confondu.

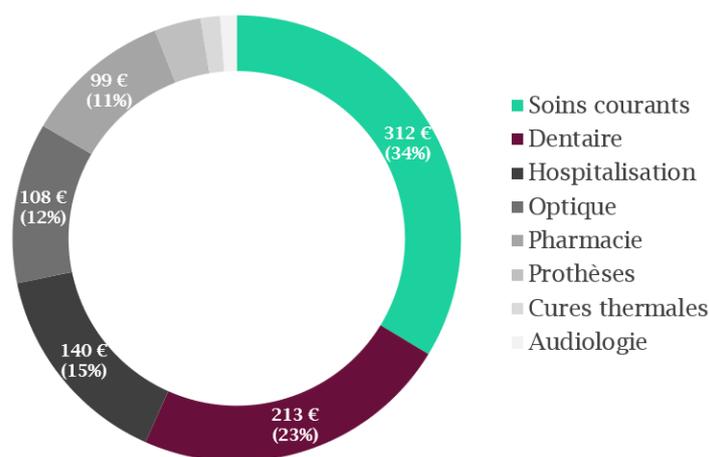


FIGURE 2.16: Décomposition de la dépense engagée en fonction des postes de soins.

Afin de pouvoir tarifier le contrat de complémentaire santé, la dépense engagée des différents sous-postes de soins a été analysée au travers de bases de données. Ces dernières ont été retraitées afin d'avoir un format adapté aux modèles linéaires généralisés. En effet, une approche coût-fréquence a été choisie pour décrypter la sinistralité. Les meilleurs modèles de coût et de fréquence ont été retenus pour la modéliser. Ils permettent d'analyser les risques sous-jacents en fonction des caractéristiques des individus et de déterminer la dépense engagée moyenne de chacun des sous-postes de soins mais également la distribution associée.

Tous les travaux effectués dans cette section concernent la dépense engagée moyenne. Ce sont les frais facturés pour les actes de soins réalisés. Ces frais se voient réduits par la Sécurité sociale au travers de l'Assurance Maladie avant l'intervention des complémentaires santé. La partie suivante est consacrée à la modélisation du remboursement accordé par la Sécurité sociale pour chacun des sous-postes de soins.

\*Le service médical rendu (SMR) est un critère permettant de classer les médicaments en fonction de leur utilité thérapeutique publié par la Haute Autorité de Santé (HAS).

## 2.3 Modélisation du remboursement obligatoire

Dans la continuité des étapes de modélisation proposées, cette partie est consacrée à la création d'un indicateur portant sur le remboursement de la Sécurité sociale à la maille sous-poste.

Plusieurs milliers d'actes sont référencés dans la Classification Commune des Actes Médicaux, CCAM (*CCAM*). De plus, chaque acte a des règles de remboursement spécifiques. La modélisation du remboursement de la Sécurité sociale à la maille acte demande donc un traitement conséquent. De plus, la modélisation précédente portant sur la dépense engagée a été réalisée à la maille sous-poste et l'optimisation se fera avec cette même maille. Ainsi, de même que pour la modélisation de la dépense engagée, la création d'un indicateur à la maille sous-poste est un bon équilibre entre adéquation des remboursements réellement délivrés par la Sécurité sociale et modélisation opérationnelle.

Le but est de créer un indicateur synthétique évaluant le niveau de remboursement délivré par la Sécurité sociale. Ce dernier représentera le pourcentage de remboursement dont bénéficie l'assuré pour un sous-poste donné. En effet, le but est d'estimer le montant moyen restant à payer après remboursement de la Sécurité sociale afin de connaître le montant moyen que devra payer la complémentaire santé suivant la garantie qu'elle proposera.

Une remarque importante est à spécifier : dans les bases de données, la participation forfaitaire a été associée au remboursement obligatoire. Ainsi le taux de remboursement qui va être modélisé intégrera cet élément, il sera donc légèrement supérieur au taux de remboursement de la Sécurité sociale. Étant donné que les contrats responsables ne peuvent prendre en charge cette participation forfaitaire, ce taux biaisé sera plus adapté à la modélisation du contrat de complémentaire santé.

La création de l'indicateur a pris en compte les caractéristiques de remboursement de chacun des sous-postes. Les taux de remboursement de chaque individu ont été analysés. Cette analyse a permis de répartir les sous-postes en trois catégories pour la modélisation de l'indicateur.

### Les sous-postes dotés d'un remboursement explicite

Dans un premier temps, il a été remarqué que le taux de remboursement de la Sécurité sociale est explicite pour certains sous-postes. Par exemple, les médicaments sont classés en sous-postes par pourcentage de prise en charge. A ces quatre sous-postes concernés viennent s'ajouter les sous-postes non couverts par la Sécurité sociale, il s'agit des lentilles non remboursées, du forfait hospitalier, des frais optionnels en hospitalisation et du forfait maternité. Ainsi, l'indicateur représentera fidèlement le remboursement réellement versé aux assurés par l'Assurance Maladie pour 8 des 27 sous-postes concernés. Le tableau 2.8 présente le taux de remboursement pour ces sous-postes.

POSTE	SOUS-POSTE	TAUX DE REMB. SS
Hospitalisation	Forfait hospitalier	0%
	Frais optionnels	0%
	Maternité	0%
Optique	Lentilles non remb.	0%
Pharmacie	Med. hors nomencl.	0%
	Med. remb. à 15% par la SS	15%
	Med. remb. à 30% par la SS	30%
	Med. remb. à 65% par la SS	65%

TABLE 2.8: Taux de remboursement du premier type de sous-poste.

### Les sous-postes représentés par le taux de remboursement moyen

Un indicateur intuitif du taux de remboursement de la Sécurité sociale est la moyenne des taux de remboursement pour chaque individu. Cette méthode simple permet de connaître le montant moyen de remboursement délivré par la Sécurité sociale, tout en prenant en compte la fréquence de chaque taux présent dans la base. Afin de valider cette méthode, il convient de s'assurer qu'il n'y ait pas une concentration de sinistres dont le taux de remboursement diffère significativement de cette moyenne. Dans le but de mesurer cette distorsion, la procédure suivante a été mise en place pour chaque sous-poste de soins :

- ❶ Les montants de dépense engagée ont été découpés en plusieurs catégories en fonction du pourcentage de remboursement de la Sécurité sociale à l'aide d'arbres de régression ;
- ❷ Pour chaque tranche de dépense engagée, le taux de couverture moyen de chaque sous-poste a été calculé ainsi que le poids de la tranche dans le portefeuille ;
- ❸ L'évolution des taux de couverture en fonction des tranches de dépense engagée a été analysée.

Deux sous-postes n'ont pas pu être divisés en tranches, ce sont les analyses médicales et les frais de transport hospitaliers puisqu'ils ont des remboursements homogènes en fonction de la dépense moyenne. Le taux moyen est donc un très bon indicateur pour ces deux sous-postes. Pour les autres, plusieurs tranches, plus ou moins nombreuses, ont été créées. L'algorithme crée un nombre élevé de tranches lorsque le taux de remboursement fluctue. De plus, le poids de la tranche est à analyser. En effet, une tranche ayant un poids élevé s'écartant de la moyenne a un impact plus important sur la précision de l'indicateur qu'une tranche ayant un poids négligeable. Comme signalé précédemment, les actes de soins sont nombreux et ont des spécificités de remboursement propres, certains peuvent être exclus des taux de remboursement usuels du sous-poste. S'ils représentent une faible proportion des actes, la conséquence est moindre. Mais s'ils représentent une plus grande proportion, ils peuvent montrer que l'indicateur moyen n'est pas adapté.

Pour certains sous-postes, comme la figure 2.17 de droite, une tendance linéaire est apparue entre la dépense engagée moyenne et le taux de couverture moyen, les sous-postes concernés sont traités dans la partie suivante. Pour les autres, les taux des différentes tranches oscillent et sont plus ou moins proches du taux de remboursement global, comme pour la figure de gauche. De plus, ce sont les tranches ayant un faible poids qui ont tendance à s'écarter de la moyenne. Les oscillations les plus fortes caractérisent les sous-postes ayant des actes de soins variés et montrent les limites de la modélisation par sous-postes. Cependant, pour la majorité des sous-postes concernés, le taux de remboursement par catégorie est semblable au taux de remboursement moyen.

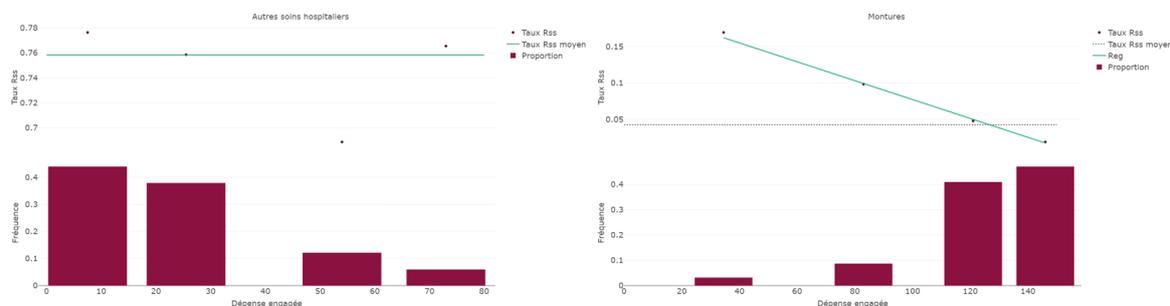


FIGURE 2.17: Visualisation graphique des taux de couverture des différentes tranches de dépense engagée en fonction du taux de couverture moyen pour deux sous-postes de soins.

Pour la plupart de ces sous-postes, le taux moyen de remboursement est un bon, voire très bon indicateur. Les résultats graphiques de ces analyses se trouvent en Annexe 2.1 et le tableau 2.9 résume les taux de remboursement de ces 12 sous-postes.

POSTE	SOUS-POSTE	TAUX DE REMB. SS
Soins courants	Auxiliaires médicaux	58,77%
	Analyses médicales	59,23%
	Consult. généraliste	60,99%
	Consult. spécialiste	53,34%
	Autres soins	61,13%
Dentaire	Soins dentaires	66,39%
	Orthodontie	30,20%
Hospitalisation	Frais de séjour	57,17%
	Transport	62,80%
	Autres soins	75,83%
Cures thermales	Cures thermales	40,07%
Prothèses	Prothèses	48,97%

TABLE 2.9: Taux de remboursement du second type de sous-poste.

### Les sous-postes représentés par une relation linéaire

7 sous-postes sont caractérisés par une décroissance du taux de couverture moyen en fonction de la tranche de dépense engagée. Les sous-postes concernés sont caractérisés par des coûts pouvant être élevés et un remboursement faible de la part de la Sécurité sociale lorsque ces coûts deviennent élevés. Par exemple, les montures ont un taux de prise en charge de 60% et une base de remboursement de 2,84€ pour les personnes majeures, soit un remboursement de 1,70€ par monture. Or les prix oscillent entre 1€ et 1 042€. Prendre comme indicateur la moyenne serait très mauvais pour ce cas de figure étant donné qu'il existe une relation linéaire entre le taux de remboursement et le coût moyen. Par exemple, si un individu achète une monture à 100€, son taux de remboursement sera de 1,7% tandis que s'il avait payé sa monture 300€, son taux de remboursement aurait été réduit de moitié (0,57%).

La méthode proposée ci-dessous a pour but d'éliminer le biais lié aux montants de dépense engagée. Pour cela, l'indicateur de remboursement est obtenu à l'aide d'une régression linéaire dépendant des tranches de dépense engagée ainsi que son poids. Pour les sous-postes concernés, le taux de remboursement majoré par un et minoré par zéro est obtenu grâce à la droite linéaire créée et au montant de dépense engagée moyen obtenu grâce aux modèles linéaires généralisés. Les résultats graphiques sont également détaillés en Annexe 2.2 et le tableau 2.10 répertorie les résultats numériques obtenus.

POSTE	SOUS-POSTE	RÉGRESSION	DÉPENSE ENGAGÉE MOYENNE (EN €)	TAUX DE REMB. SS
Soins courants	Radiologie	$y=-0,0001984x+0,8391$	39,10	66,46%
Dentaire	Prothèses dentaires	$y=-0,00001555x+0,2836$	89,88	28,22%
	Radiologie	$y=-0,001209x+0,8151$	57,17	74,60%
Optique	Lentilles remb.	$y=-0,0014132x+0,52108$	0,76	52,00%
	Monture	$y=-0,001298x+0,2069$	31,41	16,61%
	Verres	$y=-0,0003849x+0,1224$	66,91	9,66%
Audiologie	Audiologie	$y=-0,002008x+0,6588$	10,99	63,67%

TABLE 2.10: Taux de remboursement du troisième type de sous-poste.

### Intérêts et limites de la méthode utilisée

La méthode employée a pu être facilement déployée sur l'ensemble des sous-postes. **Il s'agit d'une solution pratique et opérationnelle. De plus, elle permet de comparer les niveaux de couverture des différents postes et sous-postes de soins.** Une telle comparaison n'aurait pas pu être possible sans l'indicateur étant donné la multitude de formulations de remboursement existantes. En effet, il est impossible de comparer les remboursements de deux actes s'ils ne sont pas exprimés de la même manière. L'indicateur permet donc de construire la figure 2.18.

En violet et en vert sont respectivement tracés les montants de dépense engagée moyens de chaque sous-poste avant et après le remboursement de l'Assurance Maladie. Avant l'intervention de la Sécurité sociale, huit sous-postes ont un montant de dépense engagée moyen supérieur à 60€, il s'agit des frais de séjour, des prothèses dentaires, des consultations chez le médecin spécialiste, des médicaments remboursés à 65% par la Sécurité sociale, des verres, des frais de séjour hospitaliers et des consultations chez le médecin généraliste. Ces sous-postes sont caractérisés par des coûts élevés et/ou une fréquence élevée. Après son intervention, trois sous-postes prédominent : les prothèses dentaires, les verres et les frais de séjour. Plus généralement, les soins les moins bien remboursés concernent ceux visés par la réforme 100% Santé en cours de déploiement. En effet, les équipements optiques (montures et verres) sont très peu remboursés par la Sécurité sociale. Ce sont principalement les mutuelles qui prennent en charge le reste à charge des assurés. Les prothèses auditives et dentaires, quant à elles, bénéficient d'une meilleure prise en charge par la Sécurité sociale mais leur coût étant plus élevé, le reste à charge reste conséquent pour les assurés. C'est pourquoi, la réforme 100% Santé porte sur ces trois sous-postes.

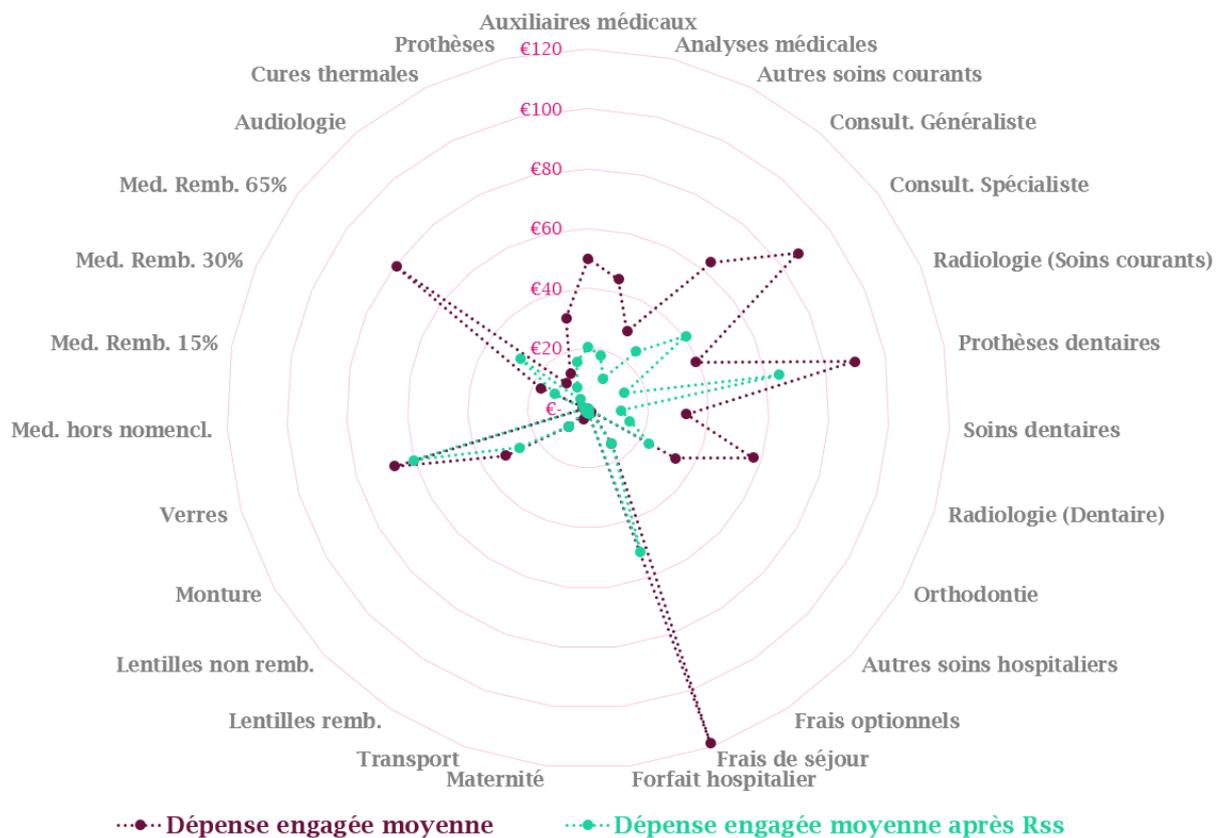


FIGURE 2.18: Dépense engagée des sous-postes de soins avant et après remboursement obligatoire.

Une analyse a également pu être réalisée à la maille poste de soins afin de mesurer l'évolution de la dépense engagée après remboursement de la Sécurité sociale. Tout d'abord, comme le montre la figure 2.19, l'ordre des postes en termes de dépense engagée moyenne reste le même avant et après remboursement de la Sécurité sociale excepté pour le poste Optique qui voit sa dépense engagée baisser de quelques euros à cause d'un faible remboursement au niveau des montures et verres optiques tandis que les autres postes sont pour la plupart divisés de moitié. Au total, la part des dépenses en santé passe de 927,47€ à 475,50€, soit une diminution de 49%.

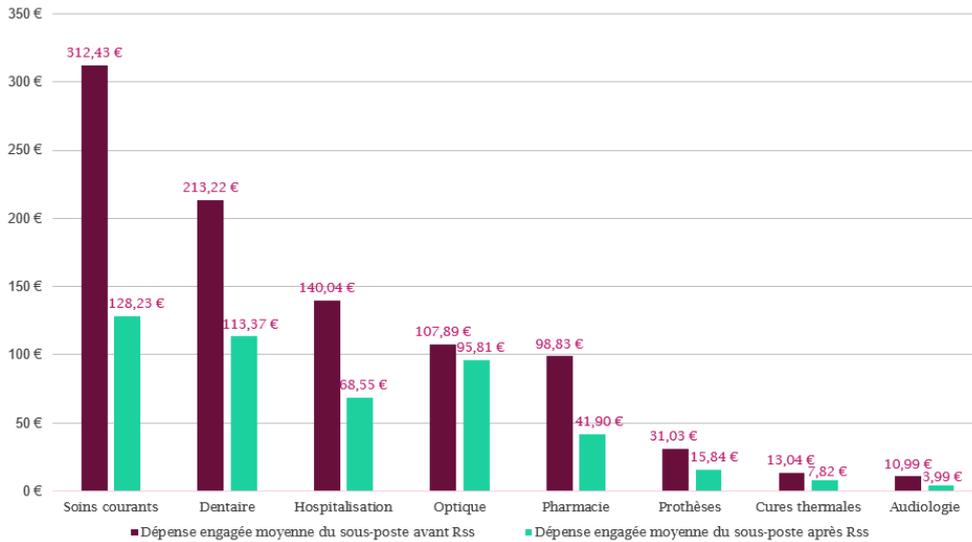


FIGURE 2.19: Dépense engagée des postes de soins avant et après remboursement obligatoire.

La répartition de la dépense engagée en fonction des différents postes de soins reste sensiblement la même que celle de la figure 2.16, exceptée pour les soins courants et les soins optiques. En effet, les soins courants sont bien remboursés et voient leur part diminuer à 27% tandis que, puisque les soins optiques sont très peu remboursés, leur part double et passe à 20% de la dépense engagée après remboursement de la Sécurité sociale.

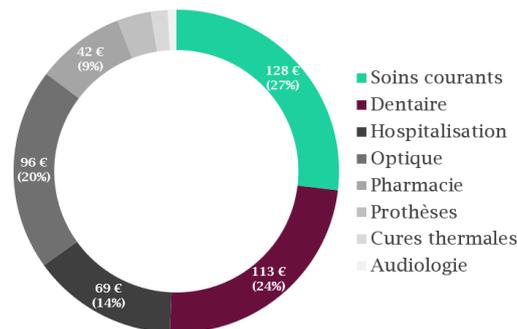


FIGURE 2.20: Décomposition de la dépense engagée après remboursement obligatoire en fonction des postes de soins.

La modélisation admet également des limites. Comme le montrent les résultats, **les sous-postes regroupent des actes variés remboursés à différents taux et l'utilisation d'un indicateur à la maille sous-poste ne permet pas de prendre en compte ces écarts.** Il s'agit de la limite principale de l'indicateur. Néanmoins, puisqu'il s'agit d'une moyenne de remboursement, l'indicateur permet de connaître l'espérance de la sinistralité d'un sous-poste de soins après remboursement de la Sécurité sociale et c'est cette quantité qui sera utilisée par la suite.

A l'aide d'outils statistiques, notamment grâce à l'utilisation de modèles linéaires généralisés, et ce dans le cadre d'une approche coût-fréquence, la dépense engagée a été modélisée pour chaque sous-poste de soins. Cette évaluation des différents risques en santé a permis d'analyser la consommation des individus selon leurs caractéristiques. Par la suite, la dépense engagée moyenne sera retenue afin de tarifier le contrat de complémentaire santé.

Le premier pilier de protection sociale en France étant la Sécurité sociale, son taux de remboursement a été modélisé à l'aide d'un indicateur adapté au type de sous-poste de soins présenté. Ainsi, le montant restant à charge, étant la différence entre la dépense engagée et le montant remboursé par la Sécurité sociale, est à présent connu pour chacun des sous-postes de soins.

L'objectif étant de construire un contrat de complémentaire santé, le chapitre prochain permettra de déterminer la prise en charge garantie par le contrat sur les montants restants à charge pour chaque sous-poste de soins. De plus, l'objectif principal du mémoire étant de construire le contrat de complémentaire santé optimal pour les assurés sous une contrainte de prix, leur utilité sera simulée et utilisée dans un problème de maximisation sous contrainte.

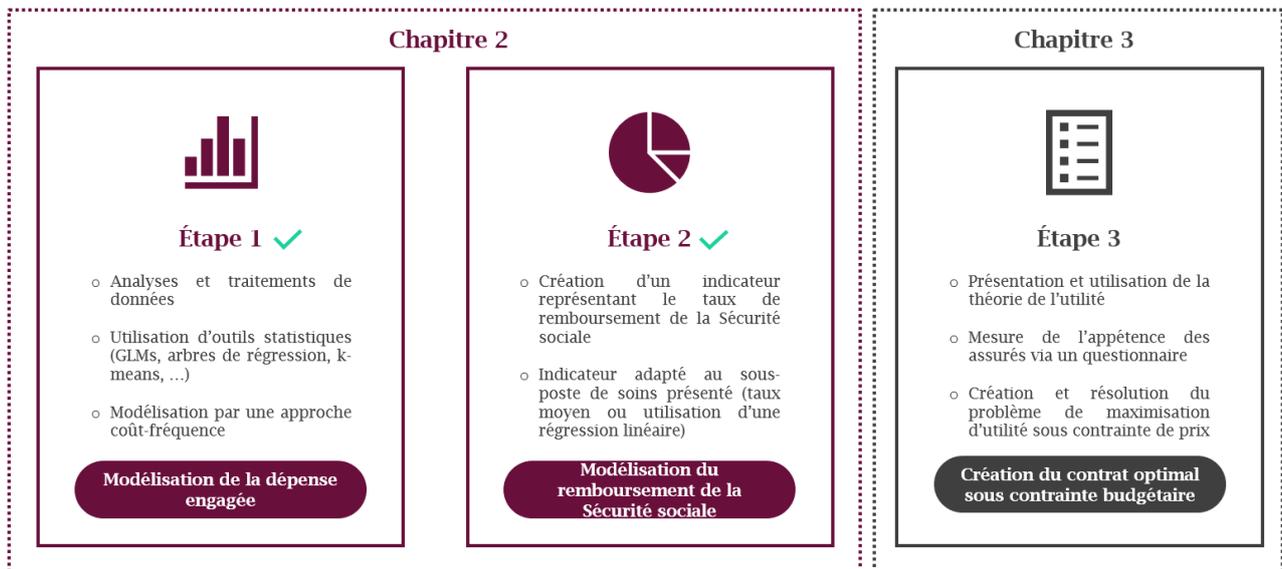


FIGURE 2.21: Schéma résumant les étapes du mémoire.



## Chapitre 3

# Création du contrat optimal maximisant l'espérance d'utilité des assurés sous contrainte budgétaire

Le premier chapitre de ce mémoire a présenté le fonctionnement de l'assurance santé en France : la Sécurité sociale intervient en premier afin de rembourser une partie de la dépense engagée des Français puis, la complémentaire santé intervient à son tour et prend en charge tout ou partie de la dépense engagée restante. Celui-ci a pour but de tarifier un contrat de complémentaire santé et il se basera sur les modélisations réalisées à la maille sous-poste de soins dans le précédent chapitre sur la dépense engagée et le remboursement de la Sécurité sociale.

Le contrat de complémentaire santé collectif d'entrée de gamme va être tarifé sous la contrainte budgétaire d'un euro par jour. La calibration de ses garanties aura pour objectif de maximiser l'espérance d'utilité des assurés afin de proposer le contrat optimal qui satisfera le plus les assurés sous cette limite de prix. Dans ce but, un algorithme d'optimisation d'utilité sous contrainte va être employé, il s'agit d'une adaptation du modèle Cobb-Douglas.

Tout d'abord, afin de comprendre les mécanismes et enjeux de la modélisation, les bases de la théorie de l'utilité ainsi que le modèle Cobb-Douglas vont être présentés. Grâce à ces bases théoriques, le modèle va pouvoir être utilisé pour résoudre la problématique posée dans ce mémoire dont l'objectif est de déterminer le contrat optimal maximisant l'espérance d'utilité des assurés tout en répondant à la contrainte budgétaire. Son utilisation va nécessiter la diffusion d'un questionnaire afin d'obtenir des informations concernant l'utilité des assurés pour les différents risques en santé et va déterminer la répartition optimale du budget disponible. Les garanties du contrat vont être calibrées de manière à égaliser les montants donnés par l'algorithme d'optimisation d'utilité. Pour cela, des espérances non-linéaires vont être calculées. Ensuite, le contrat optimal obtenu va être analysé sous plusieurs angles afin de pouvoir conclure sur l'intérêt des travaux menés. Enfin, une ouverture concernant l'adaptation de la méthodologie employée à la tarification individuelle en santé est présentée dans une ultime partie.

### 3.1 Pré-requis : la théorie de l'utilité et le modèle Cobb-Douglas

Afin de pouvoir comprendre le modèle d'optimisation mis en place, deux sujets sont ici abordés. Premièrement, les principes sur lesquels reposent la théorie de l'utilité sont détaillés pour pouvoir comprendre les mécanismes associés. Ensuite, le modèle Cobb-Douglas est à son tour présenté. Il est d'abord exprimé sous sa forme générale puis étudié sous sa forme simplifiée à deux dimensions.

#### 3.1.1 La théorie de l'utilité

La théorie de la décision repose sur le concept d'utilité. En économie, l'utilité est une mesure du bien-être ou de la satisfaction obtenue par la consommation, ou du moins l'obtention d'un certain nombre de biens et de services. Elle est liée à la notion de besoin.

La construction de fonctions d'utilité repose sur des axiomes fondamentaux. Ce sont des propriétés généralement admises qui permettent de cadrer la modélisation de ces dernières. Ces axiomes sont présentés ci-dessous. Dans la suite, l'ensemble des biens sera un ensemble fini noté  $\mathcal{L}$  et la relation de préférence sera notée  $\preceq$ .

**Axiome 1** Les préférences des individus sont **transitives** :

$$\forall L, L', L'' \in \mathcal{L}, L \preceq L' \text{ et } L' \preceq L'' \Rightarrow L \preceq L''. \quad (3.1)$$

**Axiome 2** Les préférences des individus sont **complètes** :

$$\forall L, L' \in \mathcal{L}, L \not\preceq L' \Rightarrow L \succ L'. \quad (3.2)$$

Les deux premiers axiomes définissent la rationalité des préférences des individus. Sans ces axiomes fondamentaux, aucune modélisation ne peut être faite. En effet, la notion de transitivité et de complétude sont nécessaires à tout début de raisonnement scientifique. Un troisième axiome vient compléter ces deux premiers.

**Axiome 3** Les préférences des individus sont **continues** :

$$\forall L, L', L'' \in \mathcal{L}, \{\alpha \in [0, 1], \alpha L + (1 - \alpha)L' \succeq L''\} \text{ et } \{\alpha \in [0, 1], \alpha L + (1 - \alpha)L' \preceq L''\} \text{ sont fermés.} \quad (3.3)$$

Le troisième axiome élimine les sauts d'utilité : un "petit" changement de probabilité n'entraîne pas de "gros" changements de préférences.

Ces trois premiers axiomes sont des conditions nécessaires afin de former une fonction d'utilité  $\mathcal{U}$ .

**Théorème de Debreu** Si  $\preceq$  satisfait les axiomes 1, 2 et 3 alors il existe une fonction d'utilité  $\mathcal{U} : \mathcal{L} \rightarrow \mathbb{R}$  continue qui représente les préférences, i.e.  $\forall L, L' \in \mathcal{L}, L \preceq L' \Leftrightarrow \mathcal{U}(L) \leq \mathcal{U}(L')$ .

La fonction d'utilité créée à partir de la relation de préférence  $\preceq$  n'est pas unique. Pour toute fonction  $f$  croissante, si  $\mathcal{U}$  représente  $\preceq$  alors  $f \circ \mathcal{U}$  représente également  $\preceq$ . L'utilité traduit une préférence ordinale et non cardinale. Une approche ordinale est plus facile à construire, les individus sont en général capables d'exprimer leurs préférences et de leur donner un ordre préférentiel mais il est en général difficile pour eux de quantifier leurs préférences.

Étant donné le problème posé dans ce mémoire, il est nécessaire de pouvoir quantifier l'appétence des individus pour chaque sous-poste proposé. Une construction cardinale de fonctions d'utilité aurait demandé plusieurs questions numériques dans le sondage et ce pour chacun des sous-postes. En plus de ce nombre de questions élevé, les sondés aurait dû être capables de mesurer leur utilité pour chaque sous-poste présenté, ce qui semble presque impossible. Un économiste américain, professeur à l'université de Havard, spécialiste en économie comportementale a publié plusieurs articles au sujet de la calibration des courbes d'utilité, RABIN (*Risk aversion and expected-utility theory : a calibration theorem*), dans lesquels il soulève les incohérences potentielles résultantes de la création d'une utilité cardinale. C'est pourquoi l'utilisation de fonctions d'utilité ordinales a été préférée.

Dans la littérature, plusieurs propriétés sur les préférences sont supposées vérifiées :

- ❶ Il est tout d'abord intuitif de supposer qu'une augmentation de la quantité d'un bien accroît l'utilité de l'individu. C'est pourquoi l'utilité est supposée **croissante**.
- ❷ Ensuite, il est coutume de supposer que les individus ne sont jamais rassasiés : toute quantité supplémentaire d'un bien augmente leur satisfaction, il s'agit de la propriété de non-satiété. Ainsi, l'utilité est supposée **strictement croissante**.
- ❸ En revanche, l'augmentation de l'utilité dépend de la quantité du bien initialement détenue. En effet, la première gorgée d'eau dans le désert procure une satisfaction élevée, la deuxième également mais un peu moins élevée et ainsi de suite jusqu'à ce que l'envie se tarisse. L'utilité marginale, la variation d'utilité qui résulte d'une modification d'une unité de quantité consommée d'un bien, est donc décroissante. La figure 3.1 représente ce phénomène, la variation de la quantité de bien X est supérieure à la variation de l'utilité associée. Ainsi, l'utilité est supposée **strictement concave**.

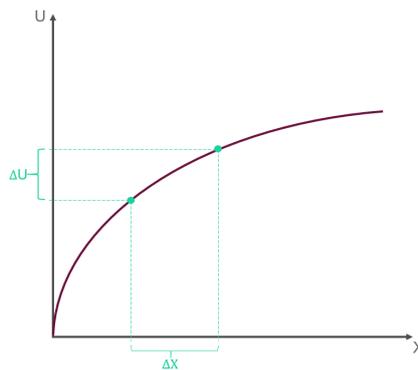


FIGURE 3.1: Représentation graphique d'une courbe d'utilité.

L'utilité est donc généralement représentée par une fonction strictement croissante et strictement concave ce qui traduit à la fois une satisfaction croissante et une utilité marginale décroissante.

Enfin, les axiomes ainsi que les propriétés des préférences des individus supposées vérifiées dans la suite de ce mémoire peuvent être discutées. En effet, il existe plusieurs paradoxes qui les remettent en question.

Ainsi, les fonctions d'utilité des individus sont généralement représentées par des courbes continues croissantes et concaves. Le modèle Cobb-Douglas utilise cette représentation, notamment grâce à la fonction logarithme  $x \mapsto \log(x)$ . Maintenant que les rappels concernant les bases de la théorie de l'utilité ont été présentés, le modèle d'optimisation utilisé se basant sur ces fondements va l'être également.

### 3.1.2 Présentation du modèle Cobb-Douglas

Ce modèle est un modèle d'optimisation sous contrainte permettant de maximiser une fonction donnée dans un espace de dimension  $N$ . Il est ici présenté puis décrypté sous sa forme simplifiée à deux dimensions afin de comprendre son mécanisme de fonctionnement.

#### Présentation du modèle

Le modèle Cobb-Douglas a été proposé en 1928 par un économiste américain Paul Douglas et un mathématicien américain Charles Cobb, CHARLES W. COBB AND PAUL H. DOUGLAS (1928). Le modèle a en premier lieu été utilisé en tant que modèle de production. Cependant, il est aussi adapté à la théorie de la décision et de l'utilité. En effet, utilisé en tant que modèle de production, il permet de trouver l'allocation de production de biens optimale étant donné une contrainte budgétaire. Utilisé en tant que modèle d'utilité, il permet de trouver l'allocation de biens optimale afin de maximiser l'utilité des individus sous contrainte budgétaire. La fonction générale du modèle Cobb-Douglas est la suivante :

$$\mathcal{Y}(x_1, \dots, x_n) = \prod_{i=1}^N x_i^{a_i}, \text{ où } 0 \leq a_i \leq 1 \text{ et } \sum_{i=1}^N a_i = 1, \quad (3.4)$$

avec l'indice  $i$  correspondant au  $i^{\text{ème}}$  bien de consommation et  $x_i$  représentant la quantité du  $i^{\text{ème}}$  bien.

La forme générale peut être linéarisée en appliquant la fonction logarithme :

$$\mathcal{Y}(x_1, \dots, x_n) = \sum_{i=1}^N a_i \log(x_i), \text{ où } 0 \leq a_i \leq 1 \text{ et } \sum_{i=1}^N a_i = 1. \quad (3.5)$$

La fonction  $\mathcal{Y}$  représente une combinaison linéaire de fonctions logarithmes qui représentent l'utilité de l'individu. A chaque bien  $i$  est associé un poids  $a_i$  qui quantifie l'importance du bien  $i$  par rapport aux autres biens.

#### Le modèle simplifié

Afin de simplifier les explications qui vont suivre, prenons une forme simple du modèle :

$$\mathcal{Y}(x, y) = x^a y^b, \text{ où } 0 \leq a \leq 1, 0 \leq b \leq 1, \text{ et } a + b = 1. \quad (3.6)$$

$x$  et  $y$  sont les quantités respectives d'un bien X et d'un bien Y et  $a$  et  $b$  sont des constantes définies dans le modèle. Le but est de maximiser  $\mathcal{Y}$ , qui représente l'utilité d'un individu. Or l'individu dispose d'une richesse finie  $\omega$ . Il faut donc déterminer les quantités  $x$  et  $y$  qui maximisent  $\mathcal{Y}$  sous la contrainte budgétaire.

Le concept de courbe d'indifférence permet de répondre au problème posé. Une courbe d'indifférence est l'ensemble des combinaisons de deux biens, ici le bien X et le bien Y, qui procurent à l'individu le même niveau de satisfaction, le même niveau d'utilité. Comme c'est le cas pour les fonctions d'utilité, les courbes d'indifférence satisfont des propriétés :

- ➡ **Les courbes d'indifférence ne se coupent pas.** Cette propriété découle de la transitivité des préférences de l'individu.
- ➡ **Plus la courbe d'indifférence est éloignée de l'origine, plus le niveau d'utilité associé est élevé.** En effet, plus la courbe est éloignée, plus les quantités des biens X et Y sont importantes et donc plus l'utilité de l'individu est grande.
- ➡ Les courbes d'indifférence ne sont pas constantes. Lorsque les biens ne sont ni substituables, ni complémentaires, étant donné l'hypothèse de non-saturation, **les courbes ont une pente négative et une forme convexe.** A utilité constante, l'individu doit substituer des biens X pour avoir plus de biens Y et inversement. Si les courbes avaient été concaves, il aurait été possible qu'un individu augmente à la fois sa quantité de biens X et sa quantité de biens Y tout en ayant la même utilité ce qui contredit les hypothèses précédentes.

Dans le cas du modèle Cobb-Douglas simplifié, la courbe d'indifférence a pour équation :

$$\mathcal{Y}(x, y) = c \iff y = \left( \frac{c}{x^a} \right)^{\frac{1}{1-a}}. \quad (3.7)$$

Il s'agit bien d'une fonction convexe décroissante. Tester plusieurs calibrations du paramètre  $a$  permet de comprendre l'importance de ce dernier. La figure 3.2 de gauche montre que le paramètre influe sur la convexité de la courbe d'indifférence, suivant la valeur de  $a$ , il faut plus ou moins de biens Y pour compenser une perte d'une unité du bien X. Cette quantité est appelée Taux Marginal de Substitution (TMS). Plus formellement, le Taux Marginal de Substitution est défini par :  $\frac{\Delta Y}{\Delta X}$ . Ce taux est toujours négatif. Ceci est la conséquence du fait que l'augmentation d'un des deux biens entraîne une diminution de l'autre bien. De plus, il est décroissant, ceci est la conséquence de la décroissance des utilités marginales. Par la suite les poids  $a_i$  représenteront l'importance des différents postes et sous-postes. Par exemple, plus un poste sera important, plus son poids sera élevé. Par conséquent, à augmentation égale de remboursement, le poste ayant un poids plus important aura un gain d'utilité supérieur à celui ayant un poids moins élevé.

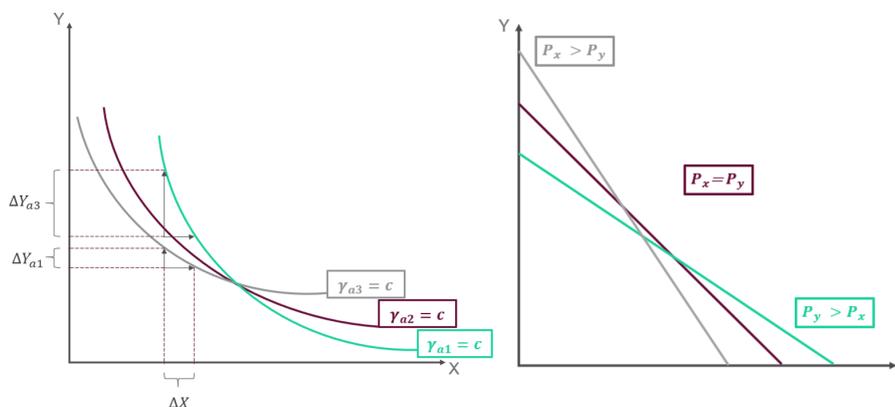


FIGURE 3.2: Variation des courbes d'utilité en fonction des paramètres  $a_1 < a_2 < a_3$  et variation des contraintes budgétaires en fonction du prix des biens.

La contrainte budgétaire de l'individu correspond à la principale limitation au comportement du consommateur. Il dispose d'une richesse  $\omega$  qu'il consacre en totalité à la consommation des biens  $X$  et  $Y$ . Le prix du bien  $X$  et le prix du bien  $Y$  sont respectivement notés  $P_x$  et  $P_y$ . Comme c'est le cas pour les poids  $a_i$ , les prix des biens influent sur l'utilité de l'individu. L'individu ne peut les contrôler mais ils pèsent dans son choix de panier optimal. Plus un bien coûte cher, plus une unité du bien réduira le budget restant pour acheter d'autres quantités. Ainsi, l'individu veut maximiser son utilité en égalisant le prix du panier de biens à sa richesse.

Le problème de maximisation pouvant se résoudre graphiquement et numériquement est le suivant :

$$\begin{aligned} \max_{x,y} x^a y^b. \\ \text{s.c. } x \cdot p_x + y \cdot p_y = \omega \end{aligned} \quad (3.8)$$

### Résolution graphique

Afin de résoudre le problème graphiquement, il suffit de tracer la contrainte budgétaire dans le plan ainsi que les différentes courbes d'utilité. La solution est donnée par la courbe d'indifférence qui est tangente à la contrainte budgétaire. L'allocation optimale est l'unique point de rencontre entre la courbe d'indifférence et la contrainte budgétaire. En effet, si la courbe d'indifférence est au-dessus de la contrainte budgétaire, l'allocation n'est pas possible, l'individu n'a pas assez d'argent pour pouvoir acheter les différents paniers. A contrario, si elle est en dessous, cela signifie qu'elle coupe à deux endroits différents la contrainte budgétaire et puisque la courbe est concave, l'individu pourrait acheter un panier moins cher qui lui donnerait la même utilité. Dans ce cas, toute la richesse de l'individu n'est pas utilisée, la contrainte budgétaire n'est pas saturée, il existe une meilleure allocation qui donne une meilleure utilité.

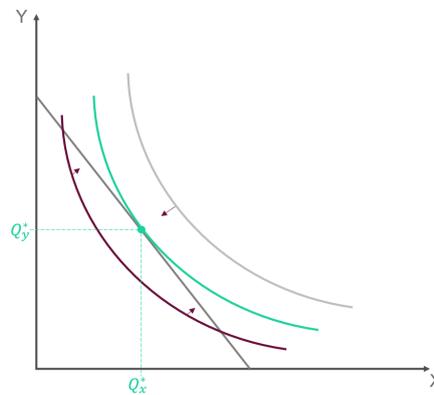


FIGURE 3.3: Résolution graphique du modèle Cobb-Douglas simplifié.

### Résolution numérique

Mathématiquement, la résolution du problème est plus simple lorsque la fonction logarithme est appliquée au problème. Le problème d'optimisation devient :

$$\begin{aligned} \max_{x,y} a \ln(x) + b \ln(y). \\ \text{s.c. } x \cdot p_x + y \cdot p_y = \omega \end{aligned} \quad (3.9)$$

L'utilisation du Lagrangien permet de résoudre aisément ce système. En effet, il s'agit d'un problème de maximisation d'une fonction strictement concave, strictement croissante et de classe  $\mathcal{C}^1$  sur  $\mathbb{R}_*^+$ , la méthode du Lagrangien peut donc être appliquée et elle permet d'obtenir un maximum global :

$$L(x, y, \lambda) = a \ln(x) + b \ln(y) - \lambda (x.p_x + y.p_y - \omega). \quad (3.10)$$

La fonction logarithme étant dérivable sur  $]0; +\infty[$ , il suffit de dériver  $L$  afin de trouver la solution optimale :

$$\begin{cases} \frac{\delta L(x,y,\lambda)}{\delta x} = 0 \Leftrightarrow x = \frac{a \omega}{P_x} \\ \frac{\delta L(x,y,\lambda)}{\delta y} = 0 \Leftrightarrow y = \frac{b \omega}{P_y} \\ \frac{\delta L(x,y,\lambda)}{\delta \lambda} = 0 \Leftrightarrow \lambda = \frac{1}{\omega}. \end{cases} \quad (3.11)$$

La solution trouvée dépend à la fois des poids d'importance des deux biens mais aussi de leur prix. Ces facteurs ont des effets contraires. Comme cela a pu être constaté graphiquement, plus un bien est coûteux, moins il sera préféré et plus un bien à un poids élevé et donc une importance élevée, plus il sera préféré.

Plus généralement, quelle que soit la dimension du modèle Cobb-Douglas l'existence d'un maximum global est vérifiée. En effet, la fonction à maximiser est la composition de fonctions puissances ayant pour exposant un nombre compris strictement entre 0 et 1. Ainsi, il s'agit de la composée de fonctions strictement concaves, strictement croissantes et de classe  $\mathcal{C}^1$  sur  $\mathbb{R}_*^+$ , qu'il faut maximiser sous une contrainte affine. Le théorème suivant énonce les conditions suffisantes afin d'obtenir un maximum global :

**Théorème** (Condition suffisante d'extremum global sous contrainte d'égalité affine) Si la contrainte d'égalité  $g$  est une fonction affine, si la fonction à maximiser  $f$  est une fonction concave et si  $x_0$  est un point stationnaire de  $f$  sur le domaine  $\mathcal{D} = \{x \mid g(x) = 0\}$  alors  $x_0$  est un maximum global de  $f$  sur  $\mathcal{D}$ .

### Atouts et limites du modèle

Ce modèle de maximisation d'utilité sous contrainte a été choisi pour ses atouts, à savoir sa **simplicité d'application**, sa **flexibilité** et son **interprétabilité**. En effet, il permet d'assurer une solution globale à l'optimisation et il peut être adapté à de grandes dimensions. De plus, des contraintes supplémentaires peuvent lui être ajoutées. Et enfin, il repose sur des mécanismes facilement compréhensibles.

Cependant, **la fonction d'utilité de l'agent est imposée**, il s'agit de la fonction logarithme. Comme expliqué dans la section précédente, la construction d'une fonction d'utilité cardinale demande une étude poussée et peut mener à des résultats non probants. Une seconde limite du modèle est la calibration de ses poids. Elle est très importante et **une mauvaise calibration entraînera un résultat erroné**.

Les bases théoriques concernant l'utilité et le modèle Cobb-Douglas étant posées, le modèle d'optimisation adapté à la problématique peut être à présent créé.

## 3.2 Description des étapes de modélisation

Sont détaillées ici les étapes clés permettant d'obtenir le contrat optimal répondant à la problématique posée dans ce mémoire. Le budget alloué au contrat de complémentaire santé va être divisé en fonction des différents sous-postes de soins en prenant pour critère l'utilité, la satisfaction, des assurés. Une fois cette décomposition réalisée, les garanties vont pouvoir être calculées afin de former la grille du contrat de complémentaire santé.

### 3.2.1 La décomposition de la prime commerciale d'un euro par jour

Le contrat est limité par une contrainte de prix. Il ne doit pas excéder 1€ par jour, soit 30€ par mois, soit 360€ par an. La prime pure correspond, pour les produits ayant un tarif équilibré, à l'espérance des sinistres. Il s'agit donc, dans ce cas précis, de l'espérance de remboursement de la complémentaire santé. La prime commerciale correspond, quant à elle, à la prime réellement payée par l'assuré. Pour l'obtenir, on ajoute à la prime pure les chargements de gestion et d'acquisition permettant de financer les coûts d'acquisition et d'administration supportés par l'assureur. On peut également inclure un chargement de sécurité. La modélisation de ces frais dépend donc de la structure de l'entreprise de l'assureur, c'est pourquoi, dans le cadre de ce mémoire, il a été décidé d'appliquer un pourcentage arbitraire à la prime pure afin d'obtenir la prime commerciale. Ainsi, la prime commerciale correspond à la prime pure majorée de 20%. Ce pourcentage correspond à ce qui est en moyenne observé sur le marché. La prime pure, correspondant à  $\frac{360}{1+20\%}$ , s'élève donc à 300€. La contrainte de prix du produit devient donc  $\sum_{i=1}^{27} RC_i = 300$  où  $RC_i$  représente le remboursement moyen du contrat de complémentaire santé pour le sous-poste  $i$  concerné.

De plus, le forfait hospitalier doit être modélisé différemment. Pour rappel, il s'agit de la participation du patient aux frais liés à son hospitalisation non remboursable par l'Assurance Maladie et prise en charge par les complémentaires santé. Son coût étant fixé légalement à 20€, l'utilité des assurés n'influencera pas la garantie associée du contrat de complémentaire santé. C'est pourquoi, le coût de ce sous-poste de soins est retiré en amont du modèle d'optimisation de la prime pure de 300€. La fréquence s'élevant à 0,08, le remboursement complémentaire moyen associé est donc de 1,68€. Le budget associé aux autres sous-postes de soins s'élève donc à 298,32€.

A présent que le budget disponible pour la tarification est calculé, il va être réparti en fonction des différents sous-postes de soins.

### 3.2.2 La décomposition du budget en fonction des sous-postes de soins

Le but est de construire le contrat qui maximise l'espérance d'utilité des assurés sous la contrainte budgétaire de 298,32€. Pour cela, le modèle Cobb-Douglas va être utilisé et un questionnaire a été diffusé afin de pouvoir calibrer les poids du modèle.

La segmentation du budget va à nouveau être faite à la maille sous-poste de soins et ce pour plusieurs raisons. Premièrement, les modélisations du chapitre précédent ont montré qu'il s'agissait d'un bon équilibre entre modélisation opérationnelle et ajustement face aux différents risques. Deuxièmement, cette maille permet d'obtenir une granularité suffisante au niveau de l'utilité des assurés. En effet, le questionnaire diffusé montre que cette maille permet de bien révéler les différences d'appétences tandis que s'il avait été réalisé à la maille acte de soins, les sondés auraient eu des difficultés à déterminer leurs appétences à une maille trop fine.

Ainsi, le problème d'optimisation à résoudre est le suivant :

$$\begin{aligned} \max \quad & \prod_{i=1}^{26} RC_i^{Poids_i} \\ \text{s.c.} \quad & \sum_{i=1}^{26} RC_i = 298,32, \end{aligned} \tag{3.12}$$

où  $Poids_i$  représente le poids associé au  $i^{ème}$  sous-poste de soins et  $RC_i$  représente le remboursement complémentaire moyen du contrat pour le sous-poste  $i$ . Les poids du modèle vont à présent être calibrés grâce à un questionnaire.

### La calibration des poids du modèle via le sondage

En amont de la résolution du problème d'optimisation, il convient de déterminer les poids du modèle ayant un rôle déterminant dans la construction du contrat. Dans ce but, un sondage a été créé et diffusé de manière anonyme afin de respecter le règlement général sur la protection des données, EUR-LEX (2016). Il contient une trentaine de questions. Chaque question propose au sondé une échelle d'importance pour un poste ou sous-poste donné, permettant ainsi de quantifier l'utilité apportée par ces derniers. En Annexe 6, sont proposés le lien d'accès au sondage et un rapport des réponses associées.

Le produit tarifé dans ce mémoire s'adressera à des personnes aux revenus modestes qui n'ont pas accès à la Complémentaire Santé Solidaire (CSS) mais qui ne peuvent pas se permettre de payer une complémentaire santé onéreuse. C'est pourquoi des questions caractérisant les sondés ont été introduites dans le questionnaire afin d'analyser le biais découlant des différences entre la population interrogée et la population cible du produit santé.

#### Caractéristiques des sondés :

Quelques questions afin de caractériser l'individu, notamment pour connaître son sexe, sa tranche d'âge, la composition de son foyer familial et sa profession ont été posées. Les résultats des 300 sondés sont présentés sous la forme de graphiques circulaires par la figure 3.4.

Une équirépartition est observée au niveau du sexe des sondés tandis qu'un léger biais, similaire à celui observé dans les bases de données, apparaît au niveau de l'âge. En effet, seulement 39 personnes de plus de 50 ans y ont répondu. De plus, il est à noter que les résultats obtenus dans le sondage seront pondérés afin de répliquer la structure des âges observée dans les bases de données ayant servies pour la tarification.

Un biais plus important au niveau de la catégorie socioprofessionnelle apparaît. La majeure partie des répondants est représentée par des cadres. Ce biais a été surveillé étant donné que le produit à tarifier ici s'adresse à des personnes n'ayant a priori pas ce statut. Dans la suite de l'analyse, les travailleurs indépendants ainsi que les chefs d'entreprise sont associés aux employés non cadres. Ce choix peut porter à discussion mais a été fait pour deux raisons. Tout d'abord, étant donné le faible nombre de réponses pour ces catégories, ils ne pouvaient pas créer une catégorie à part entière. De surcroît, ils ont été associés à la catégorie la plus proche au niveau de l'importance des postes de soins. C'est notamment cette catégorie de personnes, pesant pour 23,1% des réponses, qui est majoritairement visée par le produit à calibrer. Les étudiants et les retraités augmentent l'information portant sur l'évolution des appétences. Enfin, la composition du foyer du sondé a été répertoriée afin de pouvoir par la suite analyser les potentielles fluctuations d'appétence en fonction de ce critère.

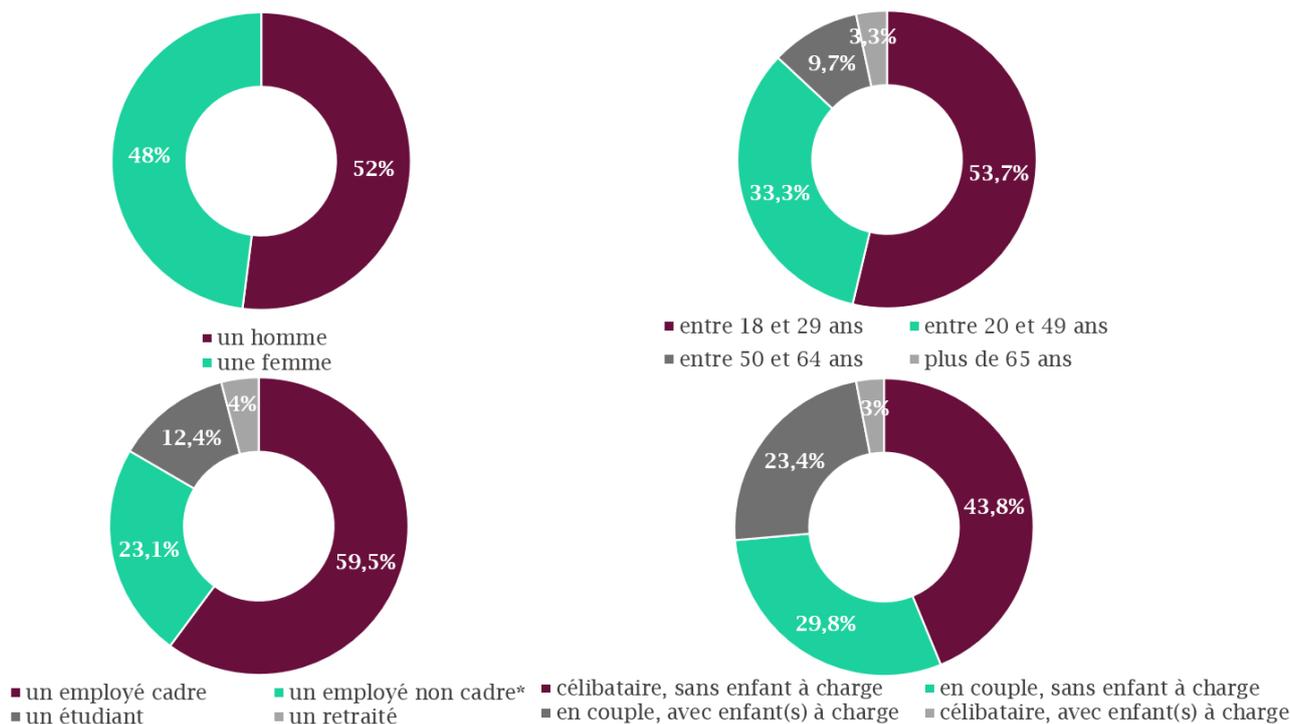


FIGURE 3.4: Répartition des sondés en fonction de leurs caractéristiques.

#### Présentation de la méthode utilisée :

La méthodologie exécutée afin de déterminer les poids du modèle Cobb-Douglas est résumée par la figure 3.5. Chaque sondé a estimé son importance pour un poste ou sous-poste de soins donné. Il s'agit de l'importance absolue de ce dernier, or l'objectif est de quantifier l'importance relative d'un poste ou sous-poste de soins par rapport aux autres. La première étape consiste donc à transformer ces importances absolues en importances relatives pour chaque sondé. Dans un second temps, le poids associé à chaque poste ou sous-poste de soins est calculé en faisant la moyenne des importances relatives de chaque sondé. Cette moyenne utilise une pondération en accord avec l'âge des individus dans les bases de données qui ont servies pour la tarification, à savoir :

- un poids de 40% est associé aux individus entre 18 et 29 ans ;
- un poids de 32% est associé aux individus entre 30 et 49 ans ;
- un poids de 21% est associé aux individus entre 50 et 64 ans ;
- un poids de 7% est associé aux individus de plus de 65 ans ;

Ainsi, cette pondération permet d'ajuster les réponses du sondage à l'âge des assurés présents dans les bases de données. Une segmentation en fonction de la catégorie socioprofessionnelle aurait limité la fiabilité des poids, il y aura donc un biais dû à la population visée par le sondage. Enfin, le modèle d'optimisation nécessitant une somme de poids égale à un, ces poids ont été normalisés.

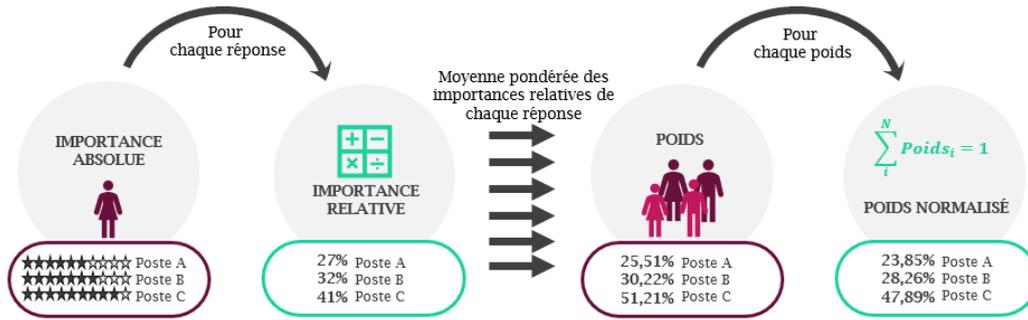


FIGURE 3.5: Schéma résumant les étapes permettant de calculer les poids des postes de soins.

Avant d’appliquer ces étapes à la maille sous-poste de soins, les importances ont été analysées à une maille moins fine, la maille poste de soins, afin de valider la pertinence du critère d’utilité.

Importances à la maille poste de soins :

La seconde partie du sondage concerne l’appétence des sondés pour les différents postes de soins proposés. Elle a pour objectif de vérifier que l’utilité est un facteur déterminant dans la construction d’un contrat optimal. En effet, si aucune différence de poids n’est observée au niveau des postes de soins, cela signifie que les individus sont indifférents entre ces derniers et donc qu’utiliser un modèle d’utilité n’est pas pertinent.

La description détaillée des étapes permettant de créer les poids à la maille poste de soins est présentée en Annexe 6. Elle est accompagnée d’interprétations sur les variations en fonction des caractéristiques des individus. Sont seulement présentés ici le résumé des étapes et de l’analyse des poids en fonction des différents postes de soins.

Tout d’abord, l’étude des importances absolues et relatives des sondés permet de déceler les variations d’importance entre les postes de soins pour les différents profils de sondés. De plus, les poids, étant obtenus grâce à une moyenne pondérée des importances relatives de chaque poste de soins, traduisent l’utilité des sondés. Ils permettent donc de comparer l’importance d’un poste de soins par rapport à l’ensemble des autres postes de soins. La figure 3.6 les répertorie.

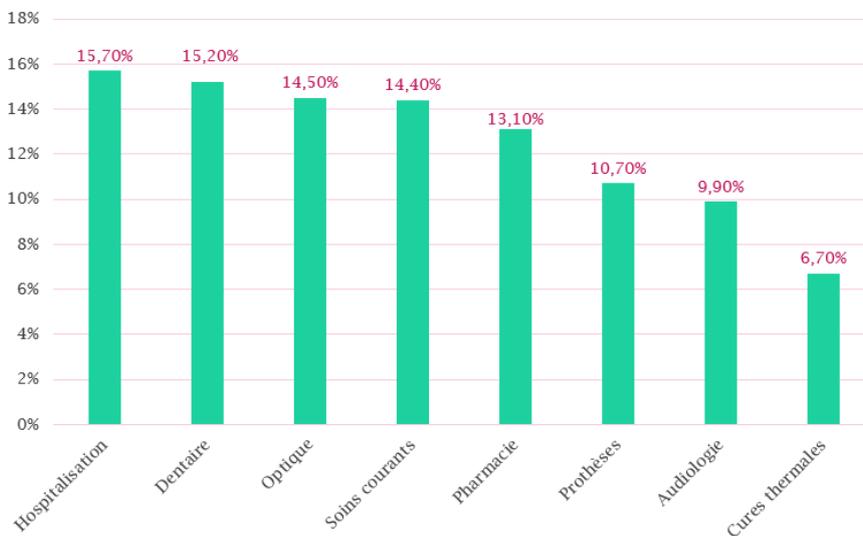


FIGURE 3.6: Poids à la maille poste de soins.

Les poids les plus élevés sont associés aux postes Hospitalisation et Dentaire, suivis par les postes Optique et Soins courants. Ce sont les postes de soins ayant une dépense engagée moyenne élevée. Ils sont complétés par le poste Pharmacie possédant un poids légèrement moins important. Enfin, l'Audiologie et les Prothèses ont un poids un tiers moins élevé et les cures thermales obtiennent un poids deux à trois fois moins élevé. Ainsi, les sondés ont accordé plus d'importance aux postes de soins totalisant une forte dépense engagée. La répartition des poids obtenus à l'aide du sondage est donc bien en cohérence avec l'importance en termes de dépense engagée de chaque poste de soins.

Une discussion est à mener à propos du dernier poste, les cures thermales. Bien qu'il obtienne une appétence moyenne de 3,697, une majorité de répondants (30%) lui a associé l'importance la plus faible et 80% des sondés lui ont accordé une importance faible à moyenne. De par sa structure de réponse, ce poste se détache des autres. En effet, les autres postes sont caractérisés par des réponses oscillant entre 5 et 10, voire uniformément réparties. Une étude approfondie concernant la nullité de ce paramètre devrait être réalisée. Cette étude demanderait des résultats comportementaux poussés dépendants du profil des individus afin d'obtenir l'influence au niveau du gain d'utilité d'une couverture plus ou moins avantageuse sur ce poste de soins. De plus, ce poste n'ayant pas une dépense engagée conséquente, son influence sur les résultats du contrat optimal restera limitée. C'est pourquoi, le poids du poste sera néanmoins égal à 0,067.

La division des poids à la maille poste ne sera pas utilisée dans le modèle d'optimisation étant donné que le problème sera basé sur une maille plus fine, la maille sous-poste. Cependant, cette section du sondage a permis d'analyser les appétences globales des sondés et de confirmer qu'ils accordent bien des niveaux d'importance variables et donc que les poids formés diffèrent. La prise en compte de l'utilité est donc bien justifiée.

Il convient de répéter cette division mais cette fois-ci à la maille sous-poste de soins afin de déterminer les poids du modèle d'optimisation.

#### Importances à la maille sous-poste de soins :

A nouveau, des différences d'importance sont observées entre les sous-postes de soins ainsi que des divergences en fonction du profil de risques des sondés. Tout comme la partie précédente, les résultats et interprétations détaillés sont présentés en Annexe 6.

26 poids ont été calculés grâce à la méthode présentée, la figure 3.7 les répertorie.



FIGURE 3.7: Poids à la maille sous-poste de soins.

Des caractéristiques similaires à celles faites précédemment sont retrouvées. Les sous-postes ayant les poids les plus élevés sont caractérisés par un fort reste à charge pour l'assuré. De plus, les sous-postes ayant de plus faibles poids sont majoritairement des sous-postes caractérisés par une consommation ciblée. Par exemple, les lentilles sont principalement consommées par des adultes actifs et, les appareils auditifs et les cures thermales sont principalement consommés par des retraités. La composition du portefeuille explique donc que ces dernières aient les poids les plus petits.

A l'intérieur des postes de soins, des disparités apparaissent. Par exemple, l'hospitalisation est le poste de soins qui a le plus grand poids alors que les frais optionnels et les transports hospitaliers sont parmi les sous-postes ayant le plus faible poids. De même, les verres ont un poids plus important que les montures et beaucoup plus important que les lentilles. Il y a donc une grande diversité d'appétence à l'intérieur des postes de soins. Ainsi, la prise en compte de l'utilité à la maille sous-poste de soins est à nouveau justifiée.

A présent, des intervalles de confiance portant sur les poids à la maille sous-poste de soins vont être calculés afin de pouvoir par la suite analyser les résultats de la modélisation par rapport à ces derniers.

#### Intervalles de confiance :

Le sondage a été réalisé sur un échantillon de 300 personnes et il présente un biais au niveau de la catégorie socioprofessionnelle. Dans le but de mesurer la marge d'erreur de chaque poids due à leur calibration sur un échantillon réduit, un intervalle de confiance au niveau 95% a été créé grâce à l'utilisation du théorème central limite. En effet, les poids ont été obtenus grâce aux calculs de moyennes empiriques. Le théorème central limite énonce que, pour un nombre d'observations  $n$  assez grand, la moyenne empirique de variables indépendantes et identiquement distribuées, de moyenne  $\mu$  et de variance  $\sigma$ , tend vers une loi gaussienne de même moyenne et de variance  $\frac{\sigma}{\sqrt{n}}$ . Pour chaque sous-poste de soins, le paramètre  $\sigma$  a été estimé à l'aide de la moyenne empirique  $\bar{X}$  par la formule suivante :

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (3.13)$$

Ainsi, le poids de chaque sous-poste de soins peut être encadré par une borne inférieure et supérieure. L'équation 3.14 donne la formule permettant de calculer un intervalle de confiance de niveau 95% :

$$\mathbb{P}\left(\bar{X} - q_{0,975} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + q_{0,975} \frac{\hat{\sigma}}{\sqrt{n}}\right) = 0,95 \quad (3.14)$$

où  $q_{0,975}$  est le quantile à 97,5% de la loi  $\mathcal{N}(0, 1)$ .

Les résultats des calculs sont répertoriés dans le tableau 3.1. Au niveau de confiance 95%, la marge d'erreur moyenne et la marge d'erreur maximale des poids sont respectivement de 0,4% et de 0,75%. Ces marges sont faibles pour deux raisons. Tout d'abord, le nombre de répondant est grand, en pratique le théorème central limite est appliqué à partir d'une trentaine d'observations. Ensuite, il apparaît que malgré la diversité des profils de sondés notamment au niveau professionnel, leurs appétences soient semblables : ils attribuent des niveaux d'importances différents en fonction des sous-postes de soins et pour chaque sous-poste de soins, mis à part quelques variations, leurs réponses sont semblables. Enfin, certains sous-postes de soins sont caractérisés par des marges d'erreur plus importantes car l'écart-type de l'échantillon est élevé pour ces derniers. Ce sont notamment l'orthodontie, le forfait maternité, l'audiologie, les prothèses, les frais optionnels en hospitalisation, les lentilles et les cures thermales. Ils sont consommés par des profils d'assurés précis tels que les enfants, les adultes ou bien les retraités, ce qui explique que les résultats des sondés varient plus fortement pour ces derniers.

POSTE	SOUS-POSTE	POIDS	BORNE INFÉRIEURE	BORNE SUPÉRIEURE	MARGE D'ERREUR
Soins courants	Auxiliaires médicaux	0,04003	0,03996	0,04009	0,32%
	Analyses médicales	0,04272	0,04266	0,04279	0,30%
	Autres soins	0,03893	0,03886	0,03899	0,33%
	Consult. Généraliste	0,04237	0,0423	0,04245	0,35%
	Consult. Spécialiste	0,04636	0,04629	0,04642	0,28%
	Radiologie	0,0419	0,04185	0,04196	0,26%
Dentaire	Prothèses dentaires	0,04448	0,0444	0,04456	0,36%
	Soins dentaires	0,04227	0,0422	0,04234	0,33%
	Radiologie	0,04063	0,04056	0,04069	0,32%
	Orthodontie	0,03787	0,03779	0,03796	0,45%
Hospitalisation	Autres soins	0,0448	0,04474	0,04487	0,29%
	Frais optionnels	0,0314	0,03132	0,03149	0,54%
	Frais de séjour	0,04618	0,04612	0,04624	0,26%
	Maternité	0,04122	0,04112	0,04131	0,46%
	Transport	0,03822	0,03816	0,03829	0,34%
Optique	Lentilles remb.	0,03396	0,03387	0,03406	0,56%
	Lentilles non remb.	0,03175	0,03166	0,03185	0,60%
	Monture	0,03813	0,03805	0,03821	0,42%
	Verres	0,04521	0,04514	0,04528	0,31%
Pharmacie	Med. Hors nomencl.	0,03645	0,03637	0,03652	0,41%
	Med. Remb. 15%	0,03791	0,03784	0,03797	0,34%
	Med. Remb. 30%	0,03791	0,03784	0,03797	0,34%
	Med. Remb. 65%	0,03791	0,03784	0,03797	0,34%
Audiologie	Audiologie	0,02985	0,02977	0,02993	0,53%
Cures thermales	Cures thermales	0,01994	0,01987	0,02002	0,75%
Prothèses	Prothèses	0,0316	0,03151	0,03168	0,54%

TABLE 3.1: Intervalle de confiance des poids calculés grâce au sondage au niveau de confiance 95%.

Ainsi, les poids calculés présentent une bonne stabilité. Les intervalles de confiance créés dans cette partie serviront par la suite pour analyser la sensibilité du contrat de complémentaire santé par rapport aux poids calculés.

En somme, le sondage a permis de recenser les importances de chaque poste et sous-poste de soins et des calculs ont été effectués afin d'obtenir les poids de chacun de ces sous-postes de soins. Ces poids, dont la somme est égale à 1, seront utilisés dans le modèle d'optimisation. De plus, des analyses portant sur la calibration de ces derniers ont été faites dans le but de mesurer la qualité des résultats obtenus.

Le modèle Cobb-Douglas est à présent paramétré. Sa résolution permettra d'obtenir la répartition optimale qui est un maximum global de la fonction à maximiser sous contrainte d'égalité (cf. 3.1.2) du budget disponible. Plus un sous-poste de soins aura un poids élevé, plus le budget qui lui sera accordé le sera. Une fois cette répartition connue, les garanties du contrat vont pouvoir être calculées.

### 3.2.3 Le calcul des garanties

Le modèle d'optimisation permet d'obtenir le montant de remboursement complémentaire moyen de chaque sous-poste de soins. Dans le cas d'un tarif équilibré, ce montant est la prime pure associée au sous-poste de soins en question. A présent, le but est de calibrer les garanties associées à chaque sous-poste de soins qui vont équilibrer ces montants de primes pures.

Le remboursement moyen d'une garantie donnée est majoré par le reste à charge de l'assuré : un assuré ne peut être remboursé plus que ce qu'il a déboursé. Ainsi, il est formulé comme suit :

$$RC_i = \mathbb{E}(\min(\text{Garantie}_i, DE_{\text{postSS}_i})), \quad (3.15)$$

avec  $\text{Garantie}_i$  représentant la garantie du contrat au titre du sous-poste de soins  $i$  et  $DE_{\text{postSS}_i}$  la dépense engagée après le remboursement de la Sécurité sociale.

Comme énoncé dans le premier chapitre, les grilles de garanties des contrats de complémentaire santé sont formées par différents types de remboursement. Les garanties peuvent être exprimées sous la forme d'un forfait, i.e. un montant fixe, ou bien sous la forme d'un pourcentage, par exemple un pourcentage de la dépense engagée, de la base de remboursement de la Sécurité sociale ou bien du ticket modérateur. Étant donné que les bases de remboursement et les tickets modérateurs varient d'un acte de soins à l'autre et que la modélisation a été faite à la maille sous-poste de soins, il est impossible d'exprimer les garanties sous ces formes, bien que la méthode employée peut tout à fait s'y adapter.

Deux modélisations seront donc utilisées par la suite : un pourcentage des frais réels et un forfait. Les remboursements de certains sous-postes de soins sont couramment exprimés sous la première forme, notamment les soins médicaux courants tandis que d'autres sont exprimés sous la forme d'un forfait. Enfin, suivant les contrats de complémentaire santé, les expressions peuvent différer. Les contrats de complémentaire santé des bases de données ont permis de statuer sur le type d'expression utilisé pour certains sous-postes comme les cures thermales qui peuvent être exprimées sous les deux formes. Les 9 sous-postes bénéficiant d'un remboursement exprimé sous la forme d'un forfait sont les suivants :

- |                         |                               |                         |
|-------------------------|-------------------------------|-------------------------|
| → Audiologie ;          | → Médicaments hors nomencl. ; | → Lentilles remb. ;     |
| → Cures thermales ;     | → Monture ;                   | → Lentilles non remb. ; |
| → Prothèses dentaires ; | → Verres ;                    | → Forfait maternité.    |

Maintenant que la distinction a été faite, les formules mathématiques des remboursements complémentaires moyens des sous-postes exprimés sous la forme d'un pourcentage de la dépense ou sous la forme d'un forfait sont présentées.

Calcul des garanties exprimées sous la forme d'un pourcentage de la dépense engagée :

$x_i$  représente le pourcentage de remboursement du contrat de complémentaire santé pour le sous-poste de soins  $i$ . Comme le montre l'équation 3.16, la garantie exprimée sous cette forme permet une simplification du calcul de l'espérance. Le raisonnement est donc basé sur la dépense engagée moyenne et le taux de remboursement moyen du sous-poste concerné.

$$RC_{i\%DE} = \mathbb{E}(\min(x_i \times DE_i, DE_{\text{postSS}_i})) = \min(x_i, 1 - R_{SS_i}) \times \mathbb{E}(DE_i) \quad (3.16)$$

où  $R_{i\%DE}$  représente le remboursement complémentaire moyen pour le sous-poste  $i$ ,  $DE_i$  et  $DE_{\text{postSS}_i}$  représentent respectivement la dépense engagée avant et après remboursement de la Sécurité sociale du sous-poste  $i$  et  $R_{SS_i}$  représente le taux de remboursement obligatoire du sous-poste  $i$ .

Ainsi, afin de trouver la garantie égalisant le remboursement complémentaire déterminé par l'algorithme d'optimisation, il faut que ce montant soit inférieur ou égal à la dépense engagée moyenne après remboursement de la Sécurité sociale. Cette contrainte doit donc être ajoutée à l'algorithme d'optimisation, sans quoi aucune grille de garanties ne pourra être construite.

Deux facteurs coexistants jouent sur le taux de remboursement du contrat de complémentaire santé :

- ➔ Plus la dépense engagée est élevée, moins ce taux l'est ;
- ➔ Plus le taux de remboursement de la Sécurité sociale est élevé, plus ce taux l'est.

Calcul des garanties exprimées sous la forme d'un forfait :

$Forfait_i$  représente le montant fixe de remboursement du contrat de complémentaire santé pour le sous-poste  $i$ . Le remboursement complémentaire moyen est donné par la formule ci-dessous :

$$RC_{i_{Forfait}} = \mathbb{E}(\min(Forfait_i, DE_{post\ SS_i})) = \mathbb{E}(\min(Forfait_i, (1 - R_{SS_i}) \times DE_i)) \quad (3.17)$$

où  $R_{i_{Forfait}}$  représente le remboursement complémentaire moyen pour le sous-poste  $i$ ,  $DE_i$  et  $DE_{post\ SS_i}$  représentent respectivement la dépense engagée avant et après remboursement de la Sécurité sociale du sous-poste  $i$  et  $R_{SS_i}$  représente le taux de remboursement obligatoire du sous-poste  $i$ .

Pour obtenir un forfait correspondant au remboursement complémentaire, il faut qu'à nouveau une contrainte soit ajoutée à l'algorithme d'optimisation : le remboursement complémentaire ne pourra excéder l'espérance de remboursement donnée par un forfait égalisant le montant maximum de la dépense engagée pour le sous-poste en question.

Étant donné que l'espérance de remboursement dépend ici d'une fonction non-linéaire, un raisonnement basé sur la dépense engagée moyenne n'est pas envisageable pour calibrer le forfait. L'espérance doit donc être calculée empiriquement en fonction du niveau du forfait. Pour cela, les dépenses engagées vont être simulées à l'aide des distributions propres à chaque sous-poste de soins. Afin d'obtenir une représentation fidèle des distributions, i.e. afin d'obtenir des résultats précis au centième près, 200 000 tirages aléatoires seront utilisés pour chaque sous-poste de soins concerné. Un taux de remboursement obligatoire correspondant à chaque montant de dépense engagée tiré aléatoirement est calculé : il s'agit soit du taux de remboursement moyen soit du taux de remboursement calculé à partir de la régression linéaire pour les sous-postes concernés (cf Chapitre 2 : Modélisation du remboursement de la Sécurité sociale). Ces valeurs permettent de calculer l'espérance de remboursement complémentaire empirique étant donné un montant du forfait pour chaque sous-poste de soins. La figure 3.13 résume graphiquement les étapes de ce calcul empirique d'espérance de remboursement complémentaire.

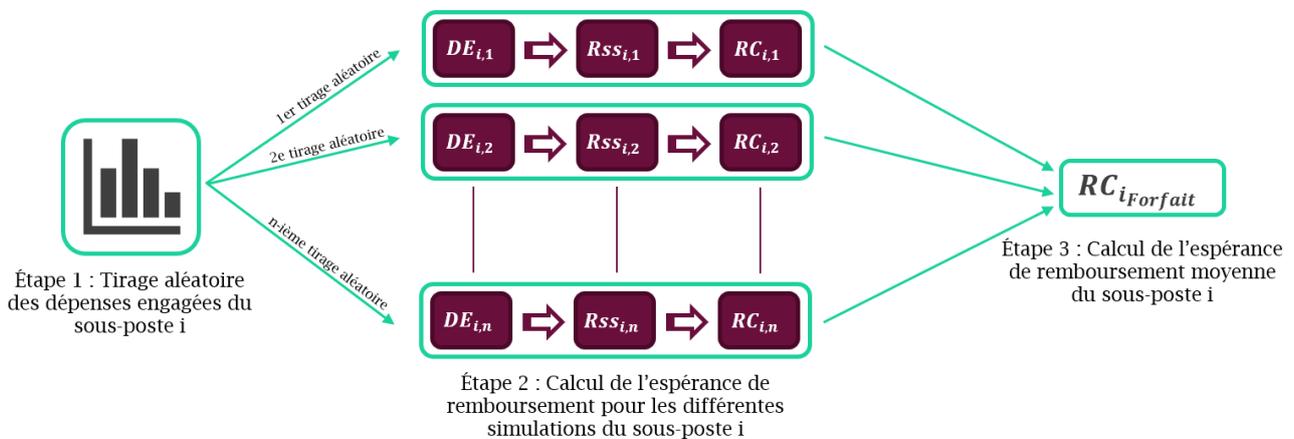


FIGURE 3.8: Schéma résumant les étapes permettant de calculer le remboursement complémentaire moyen d'une garantie exprimée sous la forme d'un forfait.

De même que pour le premier type de garantie, deux effets coexistent pour déterminer le forfait qui égalise le remboursement complémentaire correspondant au montant donné par l'algorithme d'optimisation :

- ➡ La distribution de la dépense engagée joue un rôle important. Si les montants sont hétérogènes et/ou élevés, alors le forfait sera moins élevé car le remboursement coûtera cher. Au contraire, s'ils sont homogènes et faibles, le forfait aura tendance à être élevé.
- ➡ De même que précédemment, plus le taux de remboursement de la Sécurité sociale est élevé, plus la garantie le sera également. De plus, pour les sous-postes modélisés par un taux non-linéaire, une corrélation entre le montant de la dépense engagée et le taux existe. Ainsi, si le montant de la dépense engagée est faible, alors le taux de remboursement sera légèrement supérieur et donc les petits sinistres favoriseront les garanties élevées et inversement pour les sinistres plus lourds.

Ainsi, suivant le type d'expression de garantie, le calcul mis en place diffère afin de passer du montant de remboursement complémentaire moyen à la garantie du contrat. Dans le cas d'un forfait, le calcul d'une espérance empirique est effectué. De plus, l'algorithme d'optimisation doit être enrichi de contraintes afin de pouvoir obtenir des garanties.

La stratégie proposée pour déterminer le contrat maximisant l'espérance d'utilité des assurés utilise un modèle d'optimisation sous contrainte, le modèle Cobb-Douglas. Il a été calibré grâce à un questionnaire qui a permis de confirmer l'importance de la prise en compte de ce critère. De plus, des contraintes vont lui être ajoutées afin de pouvoir calibrer des garanties à l'aide des résultats optimaux. Deux types de garantie seront modélisés : le forfait et le pourcentage de la dépense engagée. Le calcul de la première garantie va nécessiter des simulations stochastiques tandis que le deuxième peut être simplifié et nécessite seulement un calcul basé sur des espérances moyennes. Ainsi, les garanties obtenues dépendront de l'utilité des assurés mais également de la distribution de leur dépense engagée et du remboursement accordé par la Sécurité sociale. Le schéma ci-dessous résume les grandes étapes de modélisation à appliquer.

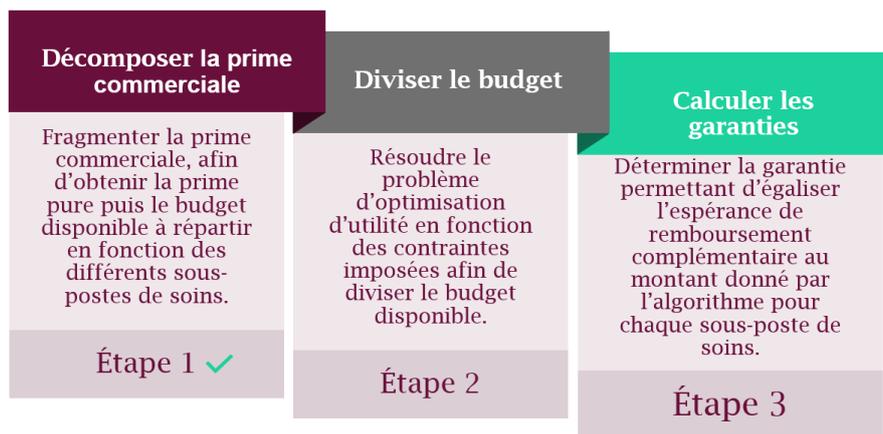


FIGURE 3.9: Description des étapes de modélisation.

Maintenant que la méthode mise en place a été expliquée en détail, ses résultats vont être analysés.

### 3.3 Construction du contrat optimal et analyse des résultats

Cette section a pour objectif de présenter les résultats obtenus suite à la modélisation mise en place, de justifier les correctifs apportés et d'analyser le modèle final sous plusieurs angles de vue. Tout d'abord, une analyse de la décomposition de la prime pure est présentée. Puis, une analyse par rapport au marché est réalisée afin de comparer les résultats obtenus. Ensuite, une analyse du reste à charge des assurés ainsi qu'une analyse globale de la décomposition de la dépense engagée sont faites. Enfin, une synthèse des travaux est réalisée dans une dernière partie.

#### 3.3.1 Premiers résultats

L'algorithme d'optimisation multidimensionnelle sous contrainte retourne les primes pures associées à chaque sous-postes de soins. Une fois ces montants connus, les garanties, exprimées soit sous la forme d'un forfait soit sous la forme d'un pourcentage de la dépense engagée, sont calculées. Les résultats des étapes de modélisation sont présentés dans le tableau 3.2.

POSTE	SOUS-POSTE	PRIME PURE MAXIMALE	PRIME PURE	GARANTIE	GARANTIE AJUSTÉE	PRIME PURE AJUSTÉE
Soins courants	Auxiliaires médicaux	20,62	17,54	35,04%	35%	17,51
	Analyses médicales	18,03	<b>18,03</b>	40,84%	40%	17,69
	Autres soins	11,17	<b>11,17</b>	38,94%	40%	11,17
	Consult. Généraliste	24,79	18,50	29,13%	30%	19,06
	Consult. Spécialiste	40,50	20,34	23,42%	25%	21,70
	Radiologie	13,11	<b>13,11</b>	33,53%	35%	13,11
Dentaire	Prothèses dentaires	64,52	19,53	308,87	310	19,60
	Soins dentaires	11,01	<b>11,01</b>	33,63%	35%	11,01
	Radiologie	14,52	<b>14,52</b>	25,43%	25%	14,29
	Orthodontie	23,32	16,58	49,65%	50%	16,71
Hospitalisation	Autres soins	0,36	<b>0,36</b>	24,23%	25%	0,36
	Frais optionnels	14,14	13,62	96,30%	95%	13,43
	Frais de séjour	50,90	20,17	17,02%	15%	17,83
	Forfait hospitalier	-	-	-	20	1,68
	Maternité	0,04	<b>0,04</b>	425	425	0,04
Transport	1,43	<b>1,43</b>	37,14%	35%	1,35	
Optique	Lentilles	0,37	<b>0,37</b>	146,60	145	0,37
	Lentilles non remb.	8,81	<b>8,81</b>	660,38	660	7,64
	Monture	26,19	16,77	80,71	80	16,61
	Verres	60,44	19,84	90,26	90	19,69
Pharmacie	Med. hors nomencl.	0,62	<b>0,62</b>	166,94	165	0,62
	Med. Remb. 15%	1,53	<b>1,53</b>	80%	85%	1,53
	Med. Remb. 30%	12,01	<b>12,01</b>	70%	70%	12,01
	Med. Remb. 65%	27,73	16,74	21,13%	20%	15,85
Audiologie	Audiologie	3,99	<b>3,99</b>	1.071,80	1.070	3,99
Cures thermales	Cures thermales	7,82	<b>7,82</b>	1.906,39	1.905	7,82
Prothèses	Prothèses	15,84	13,87	44,74%	45%	13,96

TABLE 3.2: Grille de garanties du contrat de complémentaire santé créé.

Pour rappel, les forfaits sont des montants annuels. Il s'agit du remboursement maximal garanti annuellement pour un risque. Le montant versé annuellement par le contrat de complémentaire santé est donc le minimum entre ce plafond et la dépense engagée. De plus, quelques soins tels que les soins optiques sont réglementés par des contraintes numériques dans le cadre des contrats responsables, par exemple le remboursement en optique est limité à une paire de lunettes tous les deux ans pour les personnes majeures sauf en cas d'évolution de la vue conformément au décret n° 2014-1374 du 18 novembre 2014, LÉGIFRANCE (2014). De même, le forfait en audiologie est fixé par oreille et limité à deux prothèses par an. Enfin, le forfait maternité est accordé sous les conditions précédemment citées, à savoir l'ouverture d'un contrat au nom de l'enfant. Ces spécificités seront appliquées au contrat et elles ont bien été tarifées avec des contrats utilisant les mêmes règles, la tarification est donc bien adaptée.

La solution de l'algorithme d'optimisation sature les sous-postes ayant la plus faible dépense engagée après le remboursement de la Sécurité sociale. Cette constatation rejoint les remarques faites dans la partie précédente : un individu préférera un remboursement diversifié portant sur l'ensemble des garanties, cependant, puisque le coût de chaque garantie varie et que le budget est limité, les remboursements associés aux sous-postes ayant un coût élevé seront inférieurs à ceux associés aux sous-postes ayant un coût moindre. De plus, pour les contraintes non saturées, l'utilité des assurés est prise en compte au travers des poids du modèle calculés grâce au questionnaire. C'est pourquoi, les sous-postes ayant un poids plus élevé auront un remboursement plus important à coût égal. Ainsi, l'algorithme d'optimisation a permis de diviser le budget disponible en fonction des différents sous-postes de soins.

Ensuite, les garanties permettant d'égaliser le montant des primes pures de chaque sous-poste de soins ont été calculées. Elles sont exprimées soit sous la forme d'un forfait, soit sous la forme d'un pourcentage de remboursement de la dépense engagée. Lorsque la garantie est saturée pour le deuxième type de remboursement, la somme du taux de remboursement obligatoire et du taux de remboursement complémentaire est égale à 1, ainsi les assurés ont en moyenne un reste à charge nul. Et il en est de même pour les forfaits ayant une contrainte saturée.

Enfin, qu'elles soient exprimées en forfait ou en pourcentage, les garanties ont été arrondies afin de pouvoir être présentables aux clients, elles figurent dans la sixième colonne du tableau. Par conséquent, la prime pure, qui est la somme des espérances de remboursement complémentaire pour tous les sous-postes de soins est modifiée et s'élève à 296,63€ au lieu de 300€ et donc le taux de 20% pour les chargements pourra être complété par cette marge de sécurité supplémentaire s'élevant à 1,4%.

La grille de garanties construite répond ainsi à la problématique de ce mémoire qui était de prendre en compte l'utilité des assurés afin de tarifer le contrat optimal étant donné la contrainte budgétaire. Elle a été obtenue grâce à des calculs techniques et actuariels. La partie suivante a pour vocation d'y ajouter un point de vue pratique, un point de vue métier, qui permettra d'ajuster le contrat au marché français de l'assurance santé. En effet, certaines garanties doivent être revues afin d'être en adéquation avec les garanties proposées sur le marché.

### 3.3.2 Revue des critères portant sur les garanties

L'algorithme d'optimisation repose sur des calculs mathématiques et ne permet pas de prendre en compte des facteurs importants. Cependant, il est modulable et des contraintes peuvent lui être rajoutées afin d'éliminer ses incohérences.

Tout d’abord, la méthode employée ne permet pas de prendre en compte la concurrence. Les garanties exprimées sous la forme d’un forfait sont uniquement calculées de manière à maximiser une espérance de remboursement sous une contrainte et donc pour certains sous-postes de soins, le montant obtenu diverge totalement des remboursements proposés sur le marché. Par exemple, le remboursement des cures thermales est de 1 905€, ce montant est nettement supérieur aux garanties proposées par le marché. De tels remboursements pourraient entraîner deux effets bien connus en théorie comportementale : premièrement les individus se sachant bien remboursés vont augmenter leur consommation, il s’agit de l’aléa moral, et deuxièmement le contrat est susceptible d’attirer les “mauvais risques”, des individus ayant besoin d’un fort remboursement, il s’agit du phénomène de sélection adverse. Il faut donc reconnecter ces tarifs à ce qui est pratiqué par le marché. De plus, le contrat vendu est un contrat d’entrée de gamme, est-il justifié qu’il offre un remboursement nettement supérieur au marché pour les cures thermales ? pour les médicaments non remboursés par la Sécurité sociale ? pour les lentilles non remboursées par la Sécurité sociale ?

Ainsi, les contraintes imposées sur ces forfaits doivent être revues. Des montants ont été choisis en cohérence avec les grilles de garanties des contrats de complémentaire santé présents dans les bases de données. Dans un contexte de commercialisation du produit, les dirigeants et les syndicats des entreprises souscrivant le contrat auraient fait ce choix en étant accompagnés par l’assureur qui dispose d’une vision complète du marché. Le choix fait dans ce mémoire pour les contraintes repose sur une analyse d’experts, incluant une étude de la distribution des dépenses engagées et le fait que le contrat soit un contrat d’entrée de gamme. Il va permettre d’obtenir un contrat optimal suivant ces critères. Le tableau 3.3 répertorie les forfaits maximums retenus. Ces contraintes sont bien évidemment modifiables. Enfin, elles auront des conséquences limitées sur les résultats de la modélisation étant donné le faible poids de la dépense engagée pour ces sous-postes de soins.

POSTE	SOUS-POSTE	GARANTIE MAXIMALE (EN €)	PRIME PURE MAXIMALE
Optique	Lentilles non remb.	200	6,77
Pharmacie	Med. hors nomencl.	100	0,54
Audiologie	Audiologie	1 000	3,75
Cures thermales	Cures thermales	500	4,81

TABLE 3.3: Contraintes pesant sur les forfaits du contrat.

Une seconde modification est à faire au niveau des contraintes, elle porte sur les remboursements de médicaments. Les médicaments sont classés par la Haute Autorité de Santé (HAS) suivant leur service médical rendu\*. Plus ils sont jugés comme étant des médicaments importants, plus leur remboursement obligatoire est élevé. L’algorithme attribue un remboursement complet pour les médicaments remboursés à 15% et à 30% tandis qu’un reste à charge strictement positif est observé pour les médicaments remboursés à 65% à cause du coût élevé de ces derniers. Cette décroissance de remboursement n’est pas cohérente, elle doit être modifiée : l’algorithme doit être contraint à délivrer un remboursement croissant en fonction du taux de remboursement obligatoire.

Les étapes permettant de construire la grille de garanties ont été répétées mais cette fois-ci les contraintes supplémentaires données ci-dessus ont été ajoutées afin de combiner optimisation d’utilité et ajustement face au marché. Les tableaux 3.4 et 3.5 détaillent les résultats obtenus.

---

\*Le service médical rendu (SMR) est un critère permettant de classer les médicaments en fonction de leur utilité thérapeutique publié par la Haute Autorité de Santé (HAS).

POSTE	SOUS-POSTE	PRIME PURE MAXIMALE	PRIME PURE	GARANTIE (EN €)	GARANTIE AJUSTÉE (EN €)	PRIME PURE AJUSTÉE	RAC (EN €)
Dentaire	Prothèses dentaires	64,52	18,92	298,20	300	19,03	45,49
Hospitalisation	Forfait hospitalier	-	-	-	20	1,68	-
	Maternité	0,04	<b>0,04</b>	425	425	0,04	-
Optique	Lentilles	0,37	<b>0,37</b>	146,60	145	0,37	-
	Lentilles non remb.	6,77	<b>6,77</b>	200	200	6,77	2,04
	Monture	26,19	16,27	78,44	80	16,61	9,58
	Verres	60,44	19,23	87,82	90	19,69	40,75
Pharmacie	Med hors nomencl.	0,54	<b>0,54</b>	100	100	0,54	0,08
Audiologie	Audiologie	3,75	<b>3,75</b>	1 000	1 000	3,74	0,25
Cures thermales	Cures thermales	4,81	<b>4,81</b>	500	500	4,81	3,00

TABLE 3.4: Grille de garanties des sous-postes exprimés sous la forme d'un forfait.

POSTE	SOUS-POSTE	PRIME PURE MAXIMALE	PRIME PURE	GARANTIE	GARANTIE AJUSTÉE	PRIME PURE AJUSTÉE	TAUX DE REMB. MOYEN	RAC (EN €)
Soins courants	Auxiliaires médicaux	20,62	17,00	34,03%	35%	17,51	94%	3,11
	Analyses médicales	18,03	17,98	40,64%	40%	17,69	99%	0,34
	Autres soins	11,17	<b>11,17</b>	38,94%	40%	11,17	100%	-
	Consult. Généraliste	24,79	18,15	28,53%	30%	19,06	91%	5,73
	Consult. Spécialiste	40,50	19,67	22,62%	25%	21,70	78%	18,80
	Radiologie	13,11	<b>13,11</b>	33,53%	35%	13,11	100%	-
Dentaire	Soins dentaires	11,01	<b>11,01</b>	33,63%	35%	11,01	100%	-
	Radiologie	14,52	<b>14,52</b>	25,43%	25%	14,29	100%	0,23
	Orthodontie	23,32	16,13	48,25%	50%	16,71	80%	6,61
Hospitalisation	Autres soins	0,36	<b>0,36</b>	24,22%	25%	0,36	100%	-
	Frais optionnels	14,14	13,22	93,49%	95%	13,43	95%	0,71
	Frais de séjour	50,90	19,46	16,42%	15%	17,83	72%	33,07
	Transport	1,43	<b>1,43</b>	37,14%	35%	1,35	98%	0,09
Pharmacie	Med. Remb. 15%	1,53	<b>1,53</b>	85%	85%	1,53	100%	-
	Med. Remb. 30%	12,01	<b>12,01</b>	70%	70%	12,01	100%	-
	Med. Remb. 65%	27,73	27,72	34,96%	35%	27,73	100%	-
Prothèses	Prothèses	15,84	13,15	42,34%	40%	12,41	89%	3,42

TABLE 3.5: Grille de garanties des sous-postes exprimés sous la forme d'un pourcentage de la dépense engagée.

Les contraintes faites sur les forfaits réduisent de 1,8% le budget qui leur est consacré tandis que la contrainte faite sur les médicaments implique une hausse de 3,7% pour les médicaments remboursés à 65%. Afin d'équilibrer la décomposition du budget, les sous-postes les moins appréciés et les plus importants en termes de dépenses sont imputés. Ce sont notamment les consultations chez le médecin généraliste et spécialiste, les verres, les montures, les frais de séjour et les prothèses dentaires. Ces variations étant mineures, les garanties ajustées restent inchangées pour la majorité des sous-postes de soins. Ainsi, la nouvelle grille de garanties obtenue permet d'obtenir une prime technique de 302,18€ et donc une marge pour les chargements légèrement inférieure à 20%, 19,1%. Ce delta ne doit pas être négligé. Dans le cas d'une mise en pratique des travaux, un ajustement à la marge d'une des garanties peut être décidé afin d'obtenir un tarif respectant la contrainte sur les chargements.

Ainsi, la prime pure a été divisée en tenant compte de l'importance des sous-postes et de la distribution des dépenses engagées grâce à des contraintes supplémentaires afin d'obtenir l'espérance de remboursement correspondant à chaque sous-poste de soins tout en respectant la contrainte budgétaire.

Cette division permet de créer la grille de garanties optimale du contrat de complémentaire santé. De plus, de nouvelles contraintes ont été ajoutées dans le but d'obtenir un modèle cohérent en adéquation avec les pratiques de marché. A présent, l'importance de la calibration des poids dans la création du contrat va être analysée.

### 3.3.3 Étude de la sensibilité des résultats par rapport aux poids du modèle

Tout d'abord, les poids du modèle ont été calibrés grâce à un sondage effectué sur un échantillon réduit de personnes, c'est pourquoi des intervalles de confiance ont été construits. Il convient à présent d'étudier l'impact de ces fluctuations sur le contrat créé. Ensuite, dans un second temps, d'autres variations plus importantes vont être appliquées aux poids du modèle afin d'en mesurer les conséquences.

#### Sensibilité des résultats par rapport aux fluctuations d'échantillonnage du sondage

Le contrat de complémentaire santé créé repose en partie sur les poids déterminés grâce au sondage. Afin de mesurer la fiabilité de ces derniers, des intervalles de confiance ont été calculés au niveau de confiance 95% dans une partie précédente (Intervalles de confiance). Les marges d'erreur associées aux poids sont faibles, cependant elles peuvent avoir un impact important sur la calibration du contrat. Cette section a pour objectif d'étudier la sensibilité des résultats par rapport à la fluctuation des réponses au sondage.

Dans ce but, les poids du modèle ont été modifiés un par un. Pour chaque sous-poste de soins, le poids extrême supérieur et le poids extrême inférieur ont été testés. Puisqu'il s'agit d'un algorithme d'optimisation nécessitant une somme de poids égale à un, les autres poids ont à chaque fois été ajustés uniformément. Par exemple, si le poids testé a été augmenté de  $x\%$  alors les autres poids ont été diminués uniformément d'un total de  $x\%$ . Ainsi, la sensibilité étudiée se concentre sur la variation unilatérale d'une garantie.

Les variations de poids n'ont pas d'impact sur les variables saturées. En effet, la faible variation des poids ne permet pas de modifier la répartition déterminée par l'algorithme pour ces sous-postes. Ainsi, les variations ont seulement des impacts sur les variables non saturées. Le tableau 3.8 résume les résultats obtenus pour celles-ci.

POSTE	SOUS-POSTE	RÉSULTATS		VARIATIONS À LA BAISSÉ / À LA HAUSSE				VARIATION DE PRIME PURE
		PRIME PURE	PRIME PURE AJUSTÉE	PRIME PURE	GARANTIE	GARANTIE AJUSTÉE	PRIME PURE AJUSTÉE	
Soins courants	Auxiliaires médicaux	17,00	17,51	16,97 / 17,67	33,93% / 35,34%	35% / 35%	17,51 / 17,51	- / -
	Analyses médicales	17,98	17,69	17,95 / 18,03	40,64% / 40,84%	40% / 40%	17,69 / 17,69	- / -
	Consult. Généraliste	18,15	19,06	18,11 / 19,02	28,53% / 29,93%	30% / 30%	19,06 / 19,06	- / -
	Consult. Spécialiste	19,67	21,70	19,25 / 19,72	22,22% / 22,72%	20% / 25%	17,36 / 21,70	-4,34 / -
Dentaire	Prothèses dentaires	18,93	19,03	18,88 / 18,99	298,02 / 298,43	300 / 300	19,03 / 19,03	- / -
	Orthodontie	16,13	16,71	16,07 / 16,50	48,15% / 49,35%	50% / 50%	16,71 / 16,71	- / -
Hospitalisation	Frais optionnels	13,22	13,43	13,16 / 13,31	93,09% / 94,09%	95% / 95%	13,43 / 13,43	- / -
	Frais de séjour	19,46	17,83	19,06 / 19,51	16,01% / 16,42%	15% / 15%	17,83 / 17,83	- / -
Optique	Monture	16,27	16,61	16,11 / 16,41	77,69 / 79,20	80 / 80	16,61 / 16,61	- / -
	Verres	19,23	19,69	19,16 / 19,27	87,56 / 87,91	90 / 90	19,69 / 19,69	- / -
Prothèses	Prothèses	13,15	12,41	13,06 / 13,23	42,04% / 42,64%	40% / 45%	12,41 / 13,96	- / 1,55

TABLE 3.6: Sensibilité des résultats en fonction de la variation des poids due aux fluctuations d'échantillonnage.

Le budget déterminé par l'algorithme varie de quelques centièmes à la hausse ou à la baisse suivant la variation appliquée sur le poids du sous-poste de soins. Ces changements entraînent par conséquent des fluctuations au niveau des garanties. Cependant, puisque ces dernières sont arrondies, les effets s'annulent dans la quasi-totalité des cas.

Ainsi, grâce à un nombre de réponses élevé au sondage, une faible variance au niveau des réponses des sondés pour chaque sous-poste de soins et un ajustement commercial des garanties, la fluctuation due à l'échantillonnage a un impact quasi-nul sur le contrat créé. A présent des modifications plus importantes vont être faites pour mesurer leur impact dans la création du contrat.

### Sensibilité des résultats par rapport aux poids

La section précédente a analysé les conséquences de faibles variations portant sur les poids du modèle afin de vérifier la qualité des résultats du sondage sur les travaux menés. A présent, un autre point de vue est pris : dans le cas d'une application concrète des travaux sur un corps de métier précis, les résultats obtenus au sondage peuvent sensiblement changer. Dans le but d'illustrer les conséquences possibles de telles variations, deux exemples de chocs sont présentés.

Premier choc : Quels auraient été les résultats si le sous-poste ayant le plus grand poids avait eu le plus petit poids ?

POSTE	SOUS-POSTE	POIDS : 4,635%		NOUVEAU POIDS : 1,994%			VARIATION DE PRIME PURE
		PRIME PURE	PRIME PURE AJUSTÉE	PRIME PURE	GARANTIE	GARANTIE AJUSTÉE	
Soins courants	Consult. spécialiste	19,67	21,70	8,45	9,71%	10%	-13,02

TABLE 3.7: Variation à la baisse du poids du sous-poste Consultations chez le spécialiste.

Second choc : Quels auraient été les résultats si les montures avaient eu le même poids que les verres ?

POSTE	SOUS-POSTE	POIDS : 3,813%		NOUVEAU POIDS : 4,521%			VARIATION DE PRIME PURE
		PRIME PURE	PRIME PURE AJUSTÉE	PRIME PURE	GARANTIE (EN €)	GARANTIE AJUSTÉE (EN €)	
Optique	Monture	16,27	16,61	18,87	92,02	90	1,90

TABLE 3.8: Variation à la hausse du poids du sous-poste Monture.

Ces deux chocs illustratifs permettent de révéler l'importance des poids dans la calibration du contrat. En effet, le premier choc témoigne d'une variation unilatérale équivalente à plus de 4% de la prime pure et le second choc fait augmenter la garantie de plus de 12%.

De plus, il convient de rappeler que les chocs appliqués ici, dans le but d'étudier la sensibilité par rapport aux poids, concernent la variation d'un poids par rapport aux autres. Ainsi, si plusieurs variations de poids sont appliquées conjointement, alors un contrat sensiblement différent sera créé. La calibration des poids est donc déterminante dans la création du contrat.

Cette section a permis d'analyser l'impact des poids dans la construction du contrat optimal. A présent, le contrat créé va être analysé sous plusieurs angles afin d'acquérir une vision complète des travaux réalisés.

### 3.3.4 Analyses du contrat créé

Les sections précédentes ont permis de construire le contrat de complémentaire santé optimal répondant à la problématique posée. A présent, dans le but de l'analyser, quatre visions sont choisies :

- ❶ Une vision assurantielle analysant la composition de la prime pure ;
- ❷ Une vision concurrentielle analysant le contrat par rapport au marché ;
- ❸ Une vision portant sur le reste à charge des assurés ;
- ❹ Une vision globale analysant l'ensemble de la sinistralité.

#### Vision assurantielle : analyse de la composition de la prime pure

Cette première vision va permettre d'analyser le contrat obtenu grâce aux différents mécanismes entrant en jeu dans sa construction.

La répartition de la prime pure entre les différents postes et sous-postes de soins, présentée par la figure 3.10, est semblable à la distribution de la dépense engagée après remboursement de la Sécurité sociale : plus un poste ou un sous-poste de soins a une dépense engagée élevée après le remboursement obligatoire, plus la prime pure associée l'est également. Cette observation est expliquée par deux effets. Premièrement, les sous-postes ayant un faible reste à charge voient leur contrainte saturée et donc la prime pure qui leur est associée l'est également puisque le remboursement est maximal. Deuxièmement, les sondés ont accordé des importances plus fortes aux sous-postes pour lesquels ils ont un fort reste à charge. L'algorithme leur attribue donc des budgets supérieurs.

La contrainte imposée sur les remboursements de médicaments entraîne une prime pure légèrement plus élevée pour le sous-poste de médicaments remboursés à 65%. Ensuite, ce sont les consultations chez le spécialiste et le généraliste, les verres, les prothèses dentaires et les frais de séjour qui ont les primes pures les plus élevées. L'ordre de ces sous-postes diffère de celui de leur dépense engagée moyenne après remboursement de la Sécurité sociale. Ceci est dû à la prise en compte de leur poids d'importance dans le calcul d'optimisation. A l'inverse de ces sous-postes de soins, les sous-postes ayant un faible montant de remboursement complémentaire voient leur contrainte saturée. Ce sont par exemple l'audiologie et les cures thermales.

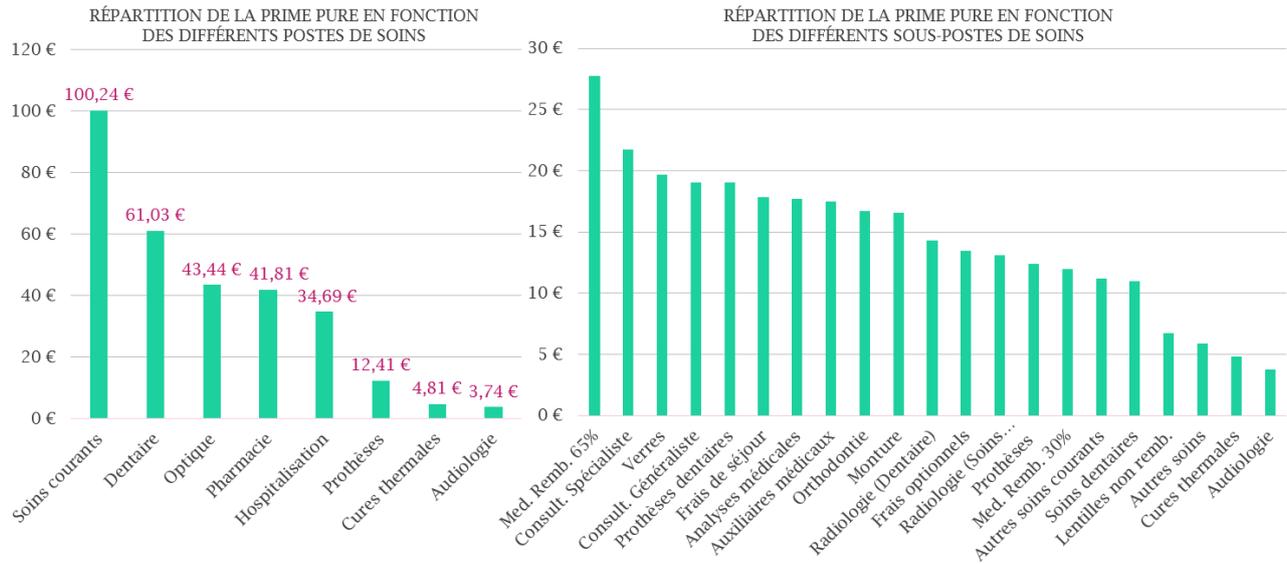


FIGURE 3.10: Répartition de la prime pure en fonction des différents postes et sous-postes de soins.

A présent que la composition de la prime pure du contrat a été analysée, une analyse concurrentielle va permettre de challenger le contrat par rapport à d'autres contrats vendus sur le marché.

### Vision concurrentielle : analyse par rapport aux pratiques de place

Le contrat optimal créé a été comparé aux quatre contrats de complémentaire santé collectifs présents dans les bases de données. Dans ce but, deux indicateurs ont été calculés sur des échantillons des bases de données  $Ratio_{RO}$  et  $Ratio_{RC}$  :

$$Ratio_{RO} = \frac{\sum RO}{\sum DE} \text{ et } Ratio_{RC} = \frac{\sum RC}{\sum DE}. \quad (3.18)$$

Ils permettent d'obtenir le taux de remboursement obligatoire et le taux de remboursement complémentaire moyens d'un sous-poste de soins. La figure 3.11 propose une restitution graphique de la somme de ces taux pour les quatre contrats collectifs présents dans les bases de données et le contrat de complémentaire créé. De plus, la dépense engagée moyenne après remboursement de la Sécurité sociale a été insérée sous la forme d'un histogramme afin de pouvoir facilement identifier les sous-postes ayant un fort reste à charge avant le remboursement de la complémentaire santé.

Tout d'abord, une remarque importante doit être faite. Les contrats présents dans la base de données ont des tarifs commerciaux plus élevés, pouvant aller de 400€ à 700€. Pour certains contrats, les tarifs sont fixes pour tous les employés tandis que pour d'autres, ils sont indexés sur le salaire. Ainsi, si les tarifs des contrats avaient été les mêmes, il aurait été facile de voir si le contrat créé diverge au niveau des taux de remboursement par rapport aux contrats présents dans les bases de données, ce qui aurait permis de mesurer directement les impacts de la modélisation effectuée.

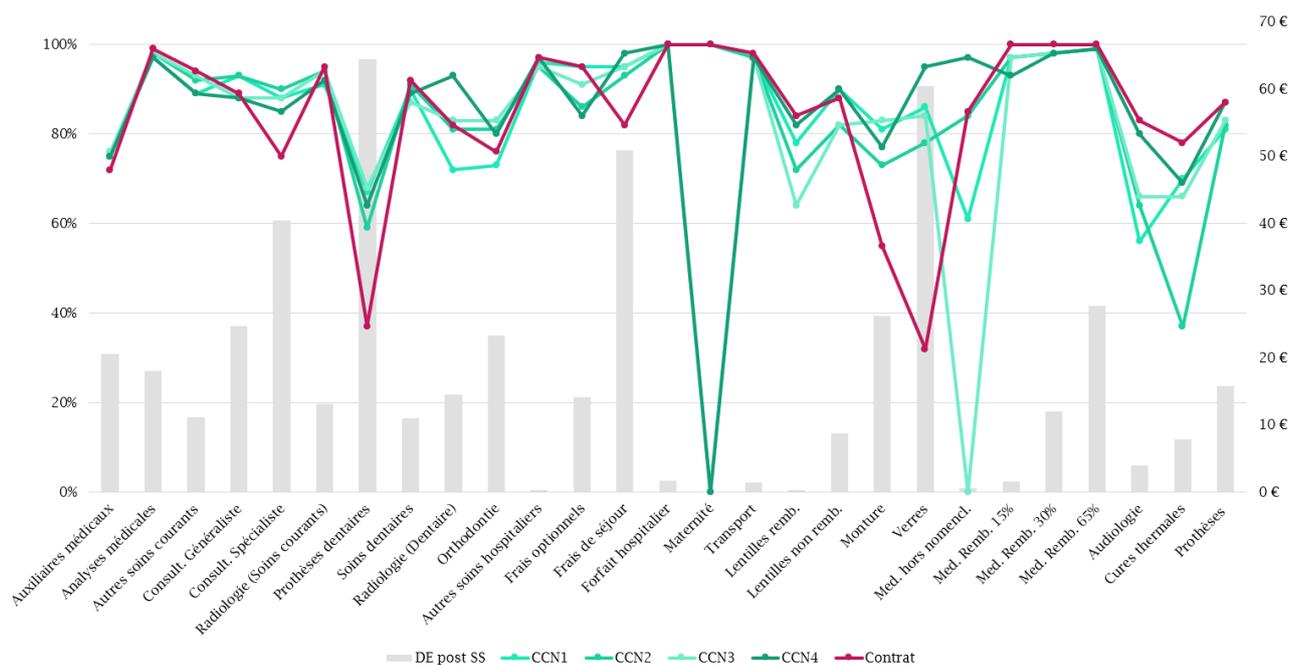


FIGURE 3.11: Analyse des remboursements du contrat par rapport aux contrats de complémentaire santé présents dans les bases de données.

La différence de budget peut notamment être observée au niveau des sous-postes ayant la plus forte dépense engagée après remboursement de la Sécurité sociale, à savoir les prothèses dentaires, les montures et verres optiques, les frais de séjour et les consultations chez le spécialiste. Le contrat créé offre un taux de remboursement nettement inférieur aux autres contrats pour ces quatre sous-postes puisque leur remboursement coûte plus cher.

Les écarts de remboursement entre le contrat optimal et ces quatre contrats collectifs peuvent être quantifiés en utilisant les montants moyens de la dépense engagée de chaque sous-poste. Une variation de remboursement portant sur un sous-poste ayant une dépense engagée faible impliquera une différence moins élevée en termes de remboursement qu'une variation portant sur une dépense engagée élevée. Au total, les écarts de remboursement entre le contrat optimal et la moyenne des contrats collectifs totalisent une différence de 103,27€.

D'autre part, les remboursements sont semblables d'une convention collective à l'autre excepté pour quelques sous-postes de soins tels que le forfait maternité, les médicaments hors nomenclature et les cures thermales. Ce phénomène révèle la présence d'une concurrence forte sur le marché poussant les assureurs à tarifier leurs contrats en fonction des contrats vendus sur le marché. Les quelques sous-postes divergeant sont des leviers permettant de se différencier du marché.

Certains sous-postes sont caractérisés par des remboursements plus faibles, ce sont notamment les soins dentaires, les équipements optiques et les prothèses auditives qui sont concernés par la réforme 100% Santé tandis que les soins en milieu hospitalier bénéficient d'une prise en charge importante.

Exceptées les divergences du contrat créé au niveau des remboursements ayant un coût élevé, quelques variations apparaissent. Le contrat se démarque au niveau des garanties en audiologie et en cures thermales. Il s'agit de garanties ayant des contraintes saturées à un maximum fixé arbitrairement et pouvant être modifiées. Néanmoins, il convient de rappeler que le forfait fixé pour les cures thermales s'élève à 500€ et que le quantile à 50% de la dépense engagée de ce sous-poste s'élève à 1 034€. Ainsi,

bien qu'il soit élevé par rapport aux contrats présentés, son forfait ne l'est pas comparé à la distribution de la dépense engagée. Le même raisonnement peut être appliqué aux prothèses auditives. De plus, le forfait maternité et le forfait pour les médicaments non remboursés ne sont pas remboursés par tous les contrats. Ce sont des garanties qui peuvent influencer les clients et elles doivent être pilotées avec discernement par les assureurs et les entreprises.

Mis à part ces variations, le contrat semble converger avec les garanties mises en place par les autres assureurs. A nouveau, de par la différence de prime, la comparaison est biaisée et donc il est difficile de juger de la déviance du contrat créé par rapport à ceux vendus sur le marché.

A présent que le contrat obtenu a été analysé, la prochaine vision va se porter sur le reste à charge des assurés.

### Vision du reste à charge : analyse du point de vue de l'assuré

Le paragraphe suivant porte sur le reste à charge moyen des assurés. Puisqu'il s'agit d'une espérance, il ne permet pas de représenter le reste à charge d'un individu qui peut sensiblement différer.

Le reste à charge moyen des assurés s'élève à 173,32€. Comme le montre la figure 3.12, trois sous-postes cumulent 65% de ce montant, ce sont les prothèses dentaires, les verres optiques et les frais de séjour. Plus généralement, ce sont les sous-postes ayant une dépense engagée élevée, caractérisés par un faible remboursement obligatoire et une densité hétérogène qui forment le reste à charge. De plus, plusieurs sous-postes ayant une faible importance pour les assurés ont un reste à charge faible mais strictement positif bien que leur dépense engagée ne soit pas élevée, ce sont par exemple les prothèses et les frais optionnels en milieu hospitalier. Cette observation prouve une nouvelle fois que l'algorithme d'optimisation prend bien en compte le facteur qu'est l'utilité.

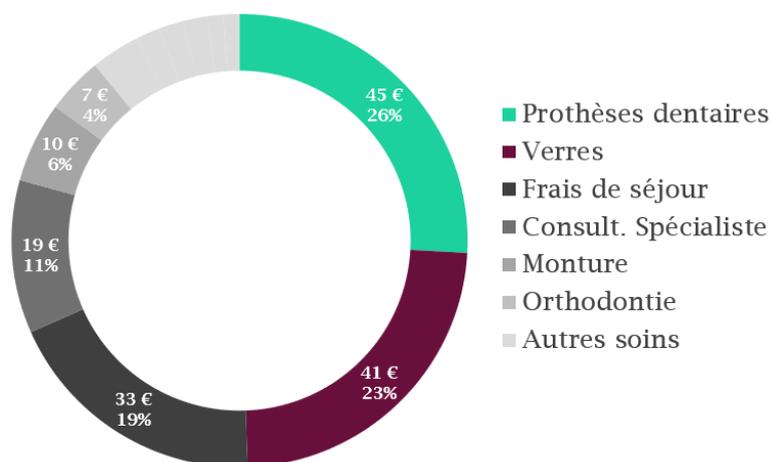


FIGURE 3.12: Répartition du reste à charge moyen des assurés.

### Vision globale : analyse d'ensemble de la sinistralité

Une ultime analyse regroupant toutes les modélisations réalisées durant le mémoire va permettre de décomposer la dépense engagée moyenne des individus couverts par le contrat de complémentaire santé créé. La figure 3.13 représente à gauche la structure des dépenses engagées à la maille poste de soins et à droite celle à la maille sous-poste de soins.

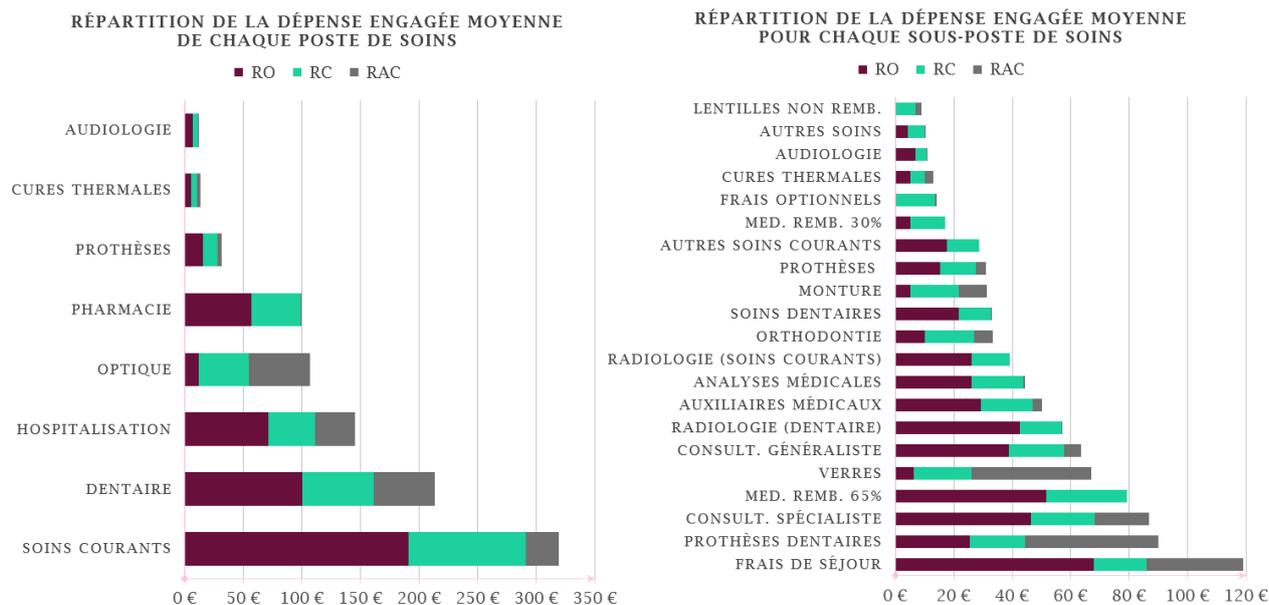


FIGURE 3.13: Décomposition de la dépense engagée des différents postes et sous-postes de soins.

Quelque soit le montant de la dépense engagée moyenne des postes de soins, la Sécurité sociale représente un poids important dans le remboursement oscillant entre 40% et 60% excepté pour un seul poste, l'optique. Ce dernier est caractérisé par une très faible part de remboursement obligatoire. A ce premier palier de remboursement vient s'ajouter le remboursement complémentaire. Il est plus ou moins important selon l'appétence des assurés, le montant et la distribution de la dépense engagée du poste après remboursement obligatoire. Ces deux facteurs de remboursement entraînent un reste à charge plus ou moins élevé pour les différents postes de soins. C'est pourquoi les restes à charge optiques et dentaires sont les plus élevés, notamment à cause des verres optiques et des prothèses dentaires.

Les mêmes constatations peuvent être faites à une maille plus fine. A nouveau, ce sont les verres et les montures qui ont les remboursements obligatoires les plus faibles suivis par les prothèses dentaires. De même, le remboursement complémentaire varie en fonction des éléments cités précédemment. C'est pourquoi, les restes à charge sont majoritairement associés aux sous-postes chers, peu remboursés et pour lesquels les assurés portent une faible importance.

Ainsi, le modèle a permis d'obtenir une répartition équilibrée des remboursements du contrat tout en incluant l'utilité des assurés, la distribution des dépenses engagées, les taux de remboursement obligatoire et les facteurs concurrentiels.

### 3.3.5 Synthèse des travaux

#### Enseignements

La prime pure a été décomposée en fonction des différents sous-postes de soins en utilisant pour critère l'utilité, c'est-à-dire la satisfaction des potentiels assurés. Ce critère bien souvent oublié dans les négociations commerciales entre les entreprises et les assureurs ainsi que dans les négociations salariales entre les syndicats et les représentants patronaux, a été ici mis en valeur à l'aide d'un modèle faisant appel à la théorie de l'utilité. Ensuite, une fois la répartition obtenue, les garanties ont pu être calculées,

de manière à respecter l'équilibre actuariel, grâce aux modélisations préliminaires effectuées dans le chapitre précédent. De plus, un point de vue métier a été mis en avant dans la calibration des garanties afin qu'elles soient en adéquation avec les produits vendus sur la place. Ainsi, le contrat a été tarifé en prenant en compte à la fois l'utilité des assurés, la distribution de la dépense engagée de chaque sous-poste de soins ainsi que le taux de remboursement obligatoire associé et également les pratiques du marché.

De plus, les analyses faites ont permis de justifier la prise en compte de l'utilité dans le processus de tarification à plusieurs reprises. Ce sont d'abord les réponses au sondage qui ont confirmé l'existence de variations d'appétences entre les différents soins en santé que ce soit à la maille poste ou à la maille sous-poste de soins. Par la suite, l'analyse de quatre conventions collectives nationales a également fait ressortir les similitudes entre les contrats proposés sur le marché dues à une intensité concurrentielle élevée. Enfin, l'étude de sensibilité menée a confirmé l'importance du facteur d'utilité dans la construction du contrat.

La grille de garanties obtenue a également été analysée sous plusieurs angles afin d'obtenir une vision complète du contrat créé. La première analyse a permis d'étudier la répartition du budget du contrat et de confirmer que les appétences des assurés ont bien été prises en compte dans sa construction. La seconde analyse a comparé le contrat à des produits vendus sur le marché, et a confirmé que le contrat créé reste compétitif sur des garanties ayant un coût faible à modéré et qu'il existe bien une convergence des garanties proposées sur le marché. Enfin, les deux dernières analyses ont permis de conclure les travaux effectués en étudiant le reste à charge des assurés et en résumant la répartition des frais de santé en fonction des différents postes et sous-postes de soins.

En somme, ces travaux mémorialistes ont à la fois permis de répondre à la problématique posée et de construire une démarche actuarielle innovante en partant d'un nouveau point de vue. Un cheminement inversé, par rapport à la tarification classique, a été mis en place : partir d'une contrainte budgétaire et déterminer les garanties grâce à un problème d'optimisation sous contrainte qui divise le budget sous critère de maximisation d'utilité.

Les travaux effectués peuvent être continués afin d'obtenir un produit commercialisable, les étapes à suivre vont être présentées.

### **Continuité**

Une liste des éléments d'attention est présentée dans le cadre de la continuation des travaux menés afin de pouvoir commercialiser le produit.

#### Créer une grille de garanties détaillée en optique :

Les remboursements sur les verres optiques sont exprimés sous la forme d'un forfait et dépendent du type de verres (verre simple/verre complexe), ainsi que de la dioptrie de l'assuré. Plus les verres sont complexes, plus leur coût est élevé, c'est pourquoi les remboursements complémentaires doivent être adaptés. La construction d'une grille de garanties optique nécessite donc des données plus fines répertoriant les caractéristiques des verres consommés.

#### Obtenir un contrat responsable :

Le contrat tarifé étant un contrat d'entrée de gamme, il doit nécessairement respecter les normes imposées sur les contrats responsables bénéficiant d'avantages fiscaux non négligeables.

Bien que la prise en compte de la non-coverage des franchises sur les soins courants et les médicaments ait été incluse durant la tarification, grâce à une inclusion de la participation forfaitaire à la modélisation du remboursement de la Sécurité sociale, certains points n'ont pas pu être modélisés. Par exemple, la prise en charge du ticket modérateur pour une majorité des sous-postes nécessite d'obtenir des données supplémentaires. De plus, des contraintes de remboursement minimales et maximales existent pour certaines garanties. Par exemple, les verres optiques doivent respecter des montants de remboursement en fonction de la dioptrie. Ces existences peuvent être incluses dans l'algorithme d'optimisation modulable mais nécessitent davantage de données.

#### Inclure les remboursements accordés pour les paniers 100% Santé :

Les données étant issues des années d'exercice 2018 et 2019, la sinistralité résultant des adaptations réglementaires liées à la réforme 100% Santé n'a pas pu être étudiée. En effet, cette récente réforme touche un point crucial bien connu en assurance : les changements comportementaux. Une étude menée par Sia Partners en 2019, SIA PARTNERS (2019), démontre l'importance de ces effets non-linéaires qui ne peuvent être décelés par des modèles classiques tels que les modèles linéaires généralisés et doivent être étudiés avec attention. De plus, la réforme étant récente, les données pouvant mesurer son impact se font encore rares, c'est pourquoi la modélisation n'a pas pu être faite dans ce mémoire.

Ainsi, la modélisation réalisée peut être complétée pour une mise en place pleinement opérationnelle. Néanmoins, les avantages et les limites de cette dernière peuvent dès à présent être analysés.

### **Atouts et limites**

L'apport principal de cette étude est l'approche utilisée afin de créer le contrat respectant la contrainte de prix qui repose sur la théorie de l'utilité. Elle permet également d'obtenir une grille de garanties, et ce quelles que soient les circonstances, grâce à l'utilisation d'un modèle simple, facilement interprétable et pouvant être aisément adapté. En effet, la méthodologie réalisée est modulable au niveau de la contrainte budgétaire, au niveau de la maille de modélisation, au niveau de l'utilité, au niveau de l'expression des garanties et au niveau de leur calibration.

Un second point fort de l'étude réalisée est la maille utilisée. Bien qu'elle ait demandé un nombre conséquent de modèles, elle a permis d'obtenir des résultats plus fins qu'une analyse à la maille poste de soins et ce à plusieurs degrés : la sinistralité, le remboursement obligatoire et l'appétence des assurés divergent fortement d'un sous-poste de soins à l'autre. Ainsi, l'utilisation de cette maille a contribué à définir un modèle tarifaire plus robuste.

De plus, grâce à une expertise métier, des contraintes supplémentaires ont été ajoutées au modèle d'optimisation afin de prendre en compte des effets comportementaux tels que l'aléa moral. Ces contraintes ont été choisies avec une vision globale du marché en tenant compte de la gamme du contrat tarifé, bien que d'autres contraintes auraient pu être choisies. Ces ajouts témoignent de la flexibilité de la méthode déployée.

Plus généralement, les résultats obtenus dépendent des contraintes imposées, des poids calculés grâce au sondage et des modélisations faites précédemment. Effectivement, la modélisation repose sur un unique jeu de données et l'approche coût-fréquence utilisant des modèles linéaires généralisés a montré ses limites. Des techniques de tarification plus avancées telles que des méthodes utilisant du *machine learning* aurait pu permettre une meilleure adéquation des risques aux modèles. Par exemple, ces techniques pourraient permettre d'obtenir une meilleure segmentation des actes de soins ou bien de prendre en compte les effets comportementaux non-linéaires dans la tarification.

De plus, les résultats obtenus ne peuvent être comparés aux contrats vendus dans le cadre de la Complémentaire Santé Solidaire (CSS) puisque les assurés ayant accès à cette couverture bénéficient d'avantages, comme la non-facturation de dépassements d'honoraires, qui biaisent la tarification effectuée. Cependant, la comparaison du contrat créé par rapport aux contrats collectifs, présents dans les bases de données et ayant des tarifs commerciaux supérieurs, a permis de montrer que les différences de couverture sont fortement visibles au niveau des sous-postes de soins ayant un fort reste à charge et au niveau de certains sous-postes qui jouent le rôle de leviers commerciaux.

Outre ce fait, une contrainte est le manque de données. Sans ce support, un tarif commercial complet ne peut être obtenu. En effet, comme le montre la section précédente, les travaux peuvent être complétés mais pour cela, des données plus exhaustives sont nécessaires.

L'application concrète de ces travaux à la construction ou à la revue de contrats de complémentaire santé permettra de dépasser les limites de manque de données, de mettre en contexte les contraintes choisies et d'obtenir une meilleure calibration des poids du modèle. De plus, la contrainte tarifaire deviendra une variable endogène à déterminer. Ainsi, les salariés seront, dans un premier temps, sondés afin de connaître leurs besoins mais également le budget qu'ils peuvent allouer à leur protection sociale. Dans un second temps, conseillés par les actuaires et s'appuyant sur les résultats du questionnaire, ce sont les représentants patronaux et les syndicats qui décideront de la prime commerciale et des contraintes imposées sur les garanties. Enfin, une fois ces décisions prises, le contrat optimal répondant à ces critères pourra être construit en suivant les étapes réalisées dans ce mémoire.

Ainsi, les travaux menés ont permis de construire une alternative à la tarification usuelle. Cette méthode de tarification inversée a été mise en place afin de construire un contrat de complémentaire santé collectif. Une ouverture concernant l'application de cette méthode aux contrats individuels est à présent proposée.

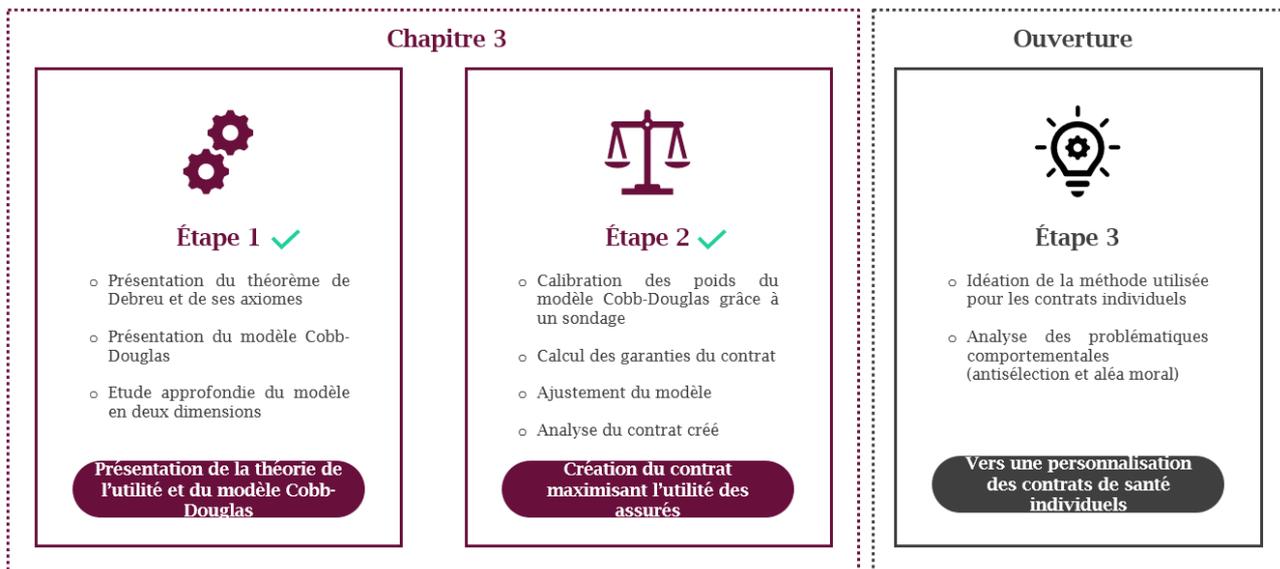


FIGURE 3.14: Schéma résumant les étapes du mémoire.

### 3.4 Ouverture : vers une tarification individuelle personnalisée ?

Les travaux menés dans le cadre de ce mémoire ont eu pour but d'insérer le critère d'utilité dans la tarification d'un contrat de complémentaire santé et ce au travers d'un contrat d'entrée de gamme à 1€ par jour. La prise en compte des appétences des potentiels assurés permet, en réalité, d'ajouter un nouvel axe d'optimisation à la tarification en santé et ce pour un plus grand champ d'application, à savoir tous les contrats de complémentaire santé qu'ils soient collectifs ou individuels.

Concernant les contrats de complémentaires santé collectifs, les travaux menés ont permis de développer une méthodologie modulable prenant en compte ces aspects dans la tarification. Il en résulte une plus-value commerciale permettant d'obtenir une meilleure cohérence entre les produits commercialisés et la demande. Quant aux contrats de complémentaires santé individuels, une réflexion portant sur la mutualisation des contrats est proposée ci-dessous notamment au niveau des enjeux et des limites de cette approche. En effet, il s'agit d'un sujet d'actualité qui vient diviser le monde de l'assurance. Ce dernier est construit sur le principe de mutualisation et, aujourd'hui, de plus en plus d'assureurs proposent des contrats segmentés selon le profil de risques des assurés voire même des contrats sur-mesures, personnalisés par rapport à leurs besoins et proposant des garanties supplémentaires facultatives qui complètent les couvertures des contrats collectifs.

#### 3.4.1 L'approche

La méthodologie présentée ici est basée sur les besoins des assurés. L'idée est de **proposer un contrat individuel sur-mesure adapté aux besoins du client**.

Premièrement, **la prime commerciale peut être adaptée au budget de ce dernier** grâce à la technique de tarification inversée mise en place tout au long de ce mémoire. Certains clients vont viser des produits similaires à des produits d'entrée de gamme, comme la tarification proposée ici, ou bien des contrats hauts de gamme caractérisés par de meilleurs remboursements et donc des primes plus élevées.

Deuxièmement, **l'individu pourrait choisir d'accentuer ses garanties de remboursement sur les postes, sous-postes ou actes de soins qu'il souhaite** grâce à la prise en compte de son utilité. En effet, les actes de soins sont multiples et variés et chaque assuré a des besoins différents en termes de santé. L'idée serait de modéliser une répartition de base des poids du modèle pour les différents actes de soins et de les déformer suivant les souhaits du client.

Connaissant la contrainte budgétaire et l'appétence du client, les assureurs pourraient tarifier le contrat de complémentaire santé en fonction de ses caractéristiques. L'approche de modélisation coût-fréquence via des modèles linéaires généralisés permet de moduler aisément les tarifs en segmentant les assurés en profils types. En effet, comme le montrent les remarques faites en Annexe 5, les risques associés aux différents profils d'assurés peuvent varier significativement. C'est pourquoi une **segmentation des tarifs en prenant en compte à la fois les caractéristiques de l'assuré, son budget et ses appétences permettrait une meilleure adéquation de la prime pure** et par conséquent de la prime commerciale.

Ainsi, apporter une nouvelle dimension à la segmentation actuelle restreinte des produits sur le marché grâce à la prise en compte de l'utilité des assurés et y ajouter une liberté au niveau du montant de la prime commerciale serait un atout majeur pour les assureurs complémentaires.

### 3.4.2 Quels enjeux ?

Le marché de l'assurance santé est caractérisé par un **nombre limité de types de contrats**. En général, un assureur santé va proposer trois à cinq gammes de contrats. De plus, lorsqu'une personne souhaite être bien remboursée au niveau de quelques soins, elle est contrainte de prendre un contrat de gamme élevée. L'idée serait de faire évoluer les pratiques courantes de marché en proposant aux clients des contrats personnalisés à leur profil de risques.

De surcroît, ce marché étant caractérisé par un **nombre élevé d'acteurs** et donc une **forte concurrence**, ce changement structurel permettrait de plus facilement attirer les clients. La loi portant sur la résiliation infra-annuelle, devant s'appliquer fin 2020, contribuera également à accroître l'aspect concurrentiel du marché. Un terme venant du marketing caractérise ce type de marché : les océans rouges, caractérisés par une intensité concurrentielle "sanglante".

L'idée serait de créer un océan bleu, un marché nouveau non affecté par la concurrence. Le terme océan bleu est, quant à lui, une analogie à l'immensité du potentiel non exploré. La prise en compte de l'utilité de la demande du marché de l'assurance santé permettrait de **conquérir une réelle demande encore non satisfaite** par les produits proposés sur le marché actuel.

Enfin, ces contrats individualisés aux besoins des assurés s'ancreraient parfaitement dans l'évolution de la société actuelle. En effet, le développement des sciences contribue aux avancées en médecine et mène au concept de médecine personnalisée. Cette dernière consiste à adapter les traitements en fonction des caractéristiques du patient à l'aide du développement du numérique, notamment du big data, et de l'intelligence artificielle. Un exemple illustrant ces avancées est celui des objets connectés permettent de recenser des informations sur la santé des individus. Ainsi, le contrat de complémentaire sur-mesure viendrait **compléter le concept de médecine personnalisée**.

Maintenant que les enjeux ont été présentés, il convient d'aborder les difficultés techniques de mise en pratique, notamment dues à deux phénomènes comportementaux bien connus, l'antisélection et l'aléa moral.

### 3.4.3 Difficultés de mise en pratique

Les contraintes de cette nouvelle tarification ne sont pas dues à la liberté portant sur le montant de la prime mais elles sont causées par la personnalisation des garanties par rapport aux appétences des assurés venant rompre le principe même de mutualisation des risques. Le premier risque comportemental qu'est l'antisélection est causé par une personnalisation du contrat tandis que le second risque, l'aléa moral, est causé par l'adaptation des garanties aux appétences des assurés.

#### **Perte de la mutualisation due à l'antisélection et solutions potentielles**

Le problème majeur lié à l'individualisation du contrat en fonction des besoins de l'assuré est l'antisélection. Il s'agit d'un problème bien connu en théorie comportementale : l'assuré dispose de plus d'informations sur lui-même qu'en à l'assureur quand il souscrit le contrat. Cette asymétrie d'information résulte, dans le cas de contrats individualisés, en une sélection de risques dite négative.

En effet, les individus vont demander à être bien couverts sur les soins qu'ils consomment le plus. Ainsi, leurs dépenses seront supérieures pour ce type de soins par rapport à celles prédites par un modèle de tarification standard. Par exemple, si un individu demande à être bien remboursé en optique, il semble peu vraisemblable que sa fréquence annuelle de consommation en monture soit égale à celle des bases de données (0,21). De plus, ce phénomène concerne tous les individus couverts. Ainsi, la mutualisation des contrats ne permet pas d'annuler ce biais.

Afin de déterminer un tarif adéquat au profil de risque des assurés, il convient d'adapter le processus de tarification en disposant de bases de données recensant la consommation d'individus de type risqué. Les données de tarification seraient enrichies au fil des années grâce aux années de développement des assureurs et permettraient une meilleure adéquation des tarifs proposés. Cependant, il faut veiller à rester compétitif par rapport au marché sinon les assurés, notamment les "bons" risques, pourraient résilier leur contrat et souscrire ailleurs.

De plus, afin de limiter le phénomène d'antisélection, plusieurs techniques peuvent être envisagées :

- ➡ Inclure un délai d'attente, un **délai de carence** d'assurance complémentaire santé : ce délai permet de se prémunir contre les assurés anticipant de fortes dépenses à court terme.
- ➡ Étant donné que les contrats de complémentaires santé sont des contrats annuels renouvelables, la mise en place de **clauses de type bonus/malus** y est adaptée. En assurance automobile, lorsqu'un assuré a un accident, il voit son bonus diminuer et donc sa prime commerciale augmenter. Le même mécanisme peut être appliqué en santé : si un individu a une forte consommation en soins de santé, il peut voir son tarif augmenter l'année suivante. Cependant, contrairement à l'assurance automobile, le risque encouru est ici un risque physiologique soulevant à la fois des problèmes déontologiques, de par la responsabilité sociale des assureurs complémentaires, et moraux : est-il éthique qu'un individu atteint d'une maladie entraînant des frais de santé élevés voit ses tarifs de santé augmenter fortement ?

### Obstacles dus à l'aléa moral et solutions potentielles

Un second obstacle comportemental est l'aléa moral. Il s'agit d'une constatation empirique selon laquelle l'individu est affecté par l'assurance : l'individu se sentant bien couvert, diminue sa vigilance et/ou augmente ses dépenses. Ce phénomène comportemental ne doit pas être confondu avec l'antisélection qui résulte d'un fait immuable, par exemple d'une maladie.

En effet, étant donné que l'individu va se sentir protégé par un contrat adapté à sa demande, il se peut qu'il diminue légèrement sa vigilance en termes de santé et/ou qu'il augmente sa consommation en soins de santé. Il s'agit de l'aléa moral ex post.

Afin de limiter ce risque, trois méthodes sont principalement utilisées :

- ➡ **Développer des campagnes de prévention** : afin de sensibiliser les assurés aux risques potentiels en santé ;
- ➡ **Le remboursement partiel** : l'assureur complémentaire ne rembourse pas intégralement les frais mais une partie seulement ;
- ➡ **La franchise absolue** : il s'agit du montant des frais payé par l'assuré en contrepartie d'une prise en charge des frais, après déduction de la franchise, par l'assureur complémentaire.

Ces deux dernières méthodes ont pour impact un reste à charge pour l'assuré l'incitant à avoir une consommation responsable et raisonnée.

La personnalisation des contrats de santé prenant en compte l'utilité des assurés, entraînant un risque élevé d'antisélection et d'aléa moral, réduit ainsi l'efficacité du marché. Cette diminution peut être limitée grâce aux solutions potentielles citées précédemment mais elle ne peut être totalement absorbée.

Par conséquent, afin d'éviter ces effets, un nombre restreint de contrats doit être créé. Ce nombre fini de contrats permettra aux assurés de choisir le contrat le plus adapté à leurs besoins : un individu de type "risqué" choisira naturellement un contrat haut de gamme et inversement pour un individu de type "non risqué". C'est ce qui est actuellement proposé sur le marché.

Une autre marque différenciante serait de créer ce nombre de contrats limité grâce à deux critères. Tout d'abord, les appétences des assurés doivent être prises en compte, l'idée serait de proposer un nombre restreint de contrats qui satisfasse le plus les clients. De plus, leur pouvoir d'achat doit également être considéré. En effet, ils ont des besoins et des budgets différents. Le croisement de ces deux variables permettrait ainsi d'obtenir un panel de contrat, à mi-chemin entre mutualisation et individualisation, en adéquation avec la demande de marché et il en résulterait un réel avantage concurrentiel par rapport aux produits actuellement vendus sur le marché.

En somme, les travaux menés dans ce mémoire dans le cadre de la tarification de contrats collectifs pourraient également apporter un gain au niveau de la tarification de contrats individuels. Cependant, ils doivent être adaptés aux risques comportementaux sous-jacents à la tarification individuelle et un équilibre entre mutualisation et individualisation doit être trouvé.



# Conclusion

Au cœur des débats politiques depuis plusieurs décennies, la protection sociale reste aujourd'hui encore un levier majeur et l'année 2020, ayant connu une crise sanitaire sans précédent, devrait en accroître davantage l'importance. Elle se définit par la mise en place de programmes publics ou privés, permettant d'instaurer des régimes obligatoires ou volontaires, dotés de sources de financement et de modes de fonctionnement propres. Bien qu'une forte disparité persiste encore au niveau de la protection sociale en santé entre les différents États du monde, une tendance de couverture croissante se dessine. C'est dans cette lignée que la France, reposant sur une structure unique inspirée par les modèles bismarckiens et beveridgiens, constituée par la Sécurité sociale et les organismes complémentaires, et composée de programmes collectifs et individuels, continue de réformer le système de santé existant.

Ce mémoire a voulu s'inscrire dans ce cadre en proposant la construction d'un contrat de complémentaire santé collectif répondant à une contrainte de prix forte, à savoir un euro par jour, dans le but de créer une protection sociale complémentaire s'inscrivant dans la continuité de la Complémentaire Santé Solidaire mise en place par l'État. A cette première cible, un objectif supplémentaire a été ajouté : renouer l'offre à la demande en construisant le contrat visant à satisfaire au mieux les clients, i.e. trouver le contrat qui va maximiser l'espérance d'utilité des assurés, et ainsi placer le client au cœur du processus de tarification.

Afin de répondre à ces enjeux, une double approche a été mise en place. Dans un premier temps, des techniques tarifaires usuelles, basées sur l'exploitation de données et l'utilisation d'outils à la fois mathématiques, actuariels et statistiques, ont permis d'adapter le processus de tarification au secteur d'assurance étudié, marqué par l'intervention de plusieurs acteurs. Ensuite, cette première démarche a été complétée par une méthode nouvelle : contrairement à une approche classique, le processus de tarification a été inversé. En effet, ce ne sont pas les engagements qui ont permis de déterminer le coût du produit mais la contrainte budgétaire qui a influencé la calibration des garanties optimales. Cette fois-ci, les techniques actuarielles permettant d'évaluer les engagements ont été complétées par des outils venant de la théorie de la décision et par un sondage recençant les préférences des français en santé. Enfin, le contrat a également été construit de manière à respecter les contraintes métiers spécifiques à la structure et à la réglementation du secteur de l'assurance santé. C'est ainsi que la complétude des méthodes utilisées a permis de former une réponse actuarielle adaptée à la problématique posée.

Les études menées sur les différentes étapes permettant la construction du contrat d'entrée de gamme qui répond aux objectifs de ce mémoire, ainsi que les diverses analyses réalisées sur ce dernier ont permis de dévoiler les facteurs impactant sa calibration. Plus particulièrement, les trois éléments déterminants sont la distribution des dépenses engagées avant et après remboursement obligatoire, les appétences des assurés et les contraintes concurrentielles. Ainsi, deux effets concourants majeurs interviennent dans la construction du contrat : les garanties apportant une grande utilité pour les assurés sont les plus protectrices et celles ayant un fort coût sont les moins élevées. De plus, malgré la limite de budget imposée, le contrat reste compétitif par rapport à d'autres produits vendus sur le

marché au niveau des garanties ayant un coût faible à modéré puisque la différence de couverture, due à une variation de prix, va principalement intervenir au niveau des garanties ayant un coût important. Enfin, l'étude réalisée peut être complétée afin d'obtenir un produit commercialisable, mais pour cela, des données plus exhaustives sont nécessaires.

En réalité, les travaux réalisés s'inscrivent dans une plus grande sphère et peuvent être utilisés pour répondre à des problématiques plus larges. En effet, la vision macroéconomique de départ, voulant aboutir à une optimisation tarifaire de l'offre par rapport à la demande et motivée par un double enjeu, à savoir augmenter l'efficacité du marché dans le but de tendre vers une amélioration de la protection sociale et apporter une plus-value commerciale aux acteurs en forte concurrence sur le secteur, a mené à la mise en place d'une méthode aisément modulable, facilement analysable et pouvant être commodément enrichie. De par ces atouts, la méthode créée pourrait être naturellement transposée afin de traiter d'autres problèmes tout en permettant de prendre en compte les contraintes métiers auxquelles font face les actuaires. Par exemple, une première étape serait d'intégrer la contrainte budgétaire comme une variable à part entière du modèle ou bien d'analyser les variations de préférences sur différents panels d'assurés. Quelle serait la variation de la satisfaction des clients si la contrainte budgétaire était modifiée ? Ou encore, quel serait le contrat calibré si les appétences des clients étaient modifiées ? Ensuite, les travaux mis en place pourraient également être utilisés pour tarifier d'autres types de contrat, comme par exemple des contrats de complémentaire santé individuels ou des contrats d'assurance non-vie. Dans ce cas, comme l'a montré l'ouverture proposée sur la tarification individuelle, les travaux réalisés doivent être utilisés à bon escient par les assureurs compte tenu des effets comportementaux sous-jacents tels que l'antisélection et l'aléa moral qui pourraient nuire à la mutualisation des risques, pilier central de l'assurance.

En somme, en plus de proposer une étude actuarielle complète sur la décomposition de la sinistralité en santé à travers la création d'un contrat sous une contrainte de prix, ce mémoire a développé une nouvelle approche, permettant de repenser les prismes, qui pourrait venir enrichir les pratiques tarifaires existantes, sans pour autant impacter la rentabilité des contrats, grâce à l'exploitation d'un nouvel axe d'optimisation, l'utilité des assurés. La poursuite des travaux pourrait permettre de mesurer l'impact d'une telle méthode et ce pour plusieurs vecteurs d'application.

# Bibliographie

- BERTRAND, P. (2020). Cours de Machine Learning. Université Paris Dauphine.
- CARPENTIER, A. (2010). Cours de Statistique de l'Assurance. Université de Rennes 1 et Université de Montréal.
- CCAM. CCAM. URL : <https://www.ameli.fr/accueil-de-la-ccam/index.php>.
- CHARLES W. COBB AND PAUL H. DOUGLAS (1928). A theory of production. Amherst College - University of Chicago.
- COUR DES COMPTES (2019). La situation financière de la Sécurité sociale en 2018 : un solde proche de l'équilibre, des efforts de maîtrise des dépenses à poursuivre. URL : <https://www.ccomptes.fr/system/files/2019-07/20190620-rapport-situation-financiere-securite-sociale-2018.pdf>.
- DIRECTION DE LA SÉCURITÉ SOCIALE (2019). Loi n° 2004-810 du 13 août 2004 relative à l'Assurance Maladie. URL : <https://www.securite-sociale.fr/files/live/sites/SSFR/files/medias/DSS/2019/CHIFFRES%20CLES%202019.pdf>.
- DONNET, S. (2019). Cours de Modèles Linéaires. Université Paris Dauphine.
- DREES (2017). Personnes âgées dépendantes : les dépenses de prise en charge pourraient doubler en part de PIB d'ici à 2060. URL : <https://drees.solidarites-sante.gouv.fr/IMG/pdf/er1032.pdf>.
- EUR-LEX (2004). Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services. URL : <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32004L0113>.
- EUR-LEX (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). URL : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32016R0679>.
- HARALD CRAMÉR (1946). Mathematical methods of statistics. Princeton University Press.
- LÉGIFRANCE (2004). Les chiffres clés de la Sécurité sociale 2018. URL : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000625158>.
- LÉGIFRANCE (2014). Décret n° 2014-1374 du 14 novembre 2014 relatif au contenu des contrats d'assurance maladie complémentaire bénéficiant d'aides sociales et fiscales. URL : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000029777871/>.
- LÉGIFRANCE (2017). Décret n° 2017-372 du 21 mars 2017 relatif à l'application de l'article 4 de la loi jno89-1009 du 31 décembre 1989 renforçant les garanties offertes aux personnes assurées contre certains risques. URL : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000034258687&categorieLien=id>.
- LÉGIFRANCE (2019a). Décret n° 2019-21 du 11 janvier 2019 visant à garantir un accès sans reste à charge à certains équipements d'optique, aides auditives et soins prothétiques dentaires. URL : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000037995163&categorieLien=id>.

- LÉGIFRANCE (2019b). Loi n° 2019-733 du 14 juillet 2019 relative au droit de résiliation sans frais de contrats de complémentaire santé (1). URL : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038766567&categorieLien=id>.
- MINISTÈRE DU TRAVAIL (2013). Accord du 11 janvier 2013 relatif à la compétitivité des entreprises. URL : [https://www.journal-officiel.gouv.fr/publications/bocc/pdf/2013/0009/boc\\_20130009\\_0000\\_0029.pdf](https://www.journal-officiel.gouv.fr/publications/bocc/pdf/2013/0009/boc_20130009_0000_0029.pdf).
- RABIN, M. Risk aversion and expected-utility theory : a calibration theorem.
- SIA PARTNERS (2019). 100% Santé - Comment anticiper l'évolution des comportements. URL : [https://www.sia-partners.com/system/files/document\\_download/file/2020-06/Etude%20juin%202019%20-%20Analyse%20Comportementale.pdf](https://www.sia-partners.com/system/files/document_download/file/2020-06/Etude%20juin%202019%20-%20Analyse%20Comportementale.pdf).
- UNOCAM (2019). Engagement pour la lisibilité des garanties de complémentaire santé. URL : <https://unocam.fr/nos-actions/lisibilite-des-garanties/>.

# Annexes

## A.1 Annexe 1 : Variables présentes dans les bases de données

NOM DE VARIABLE	DESCRIPTION DE LA VARIABLE	VALEURS
benefAge	Age du bénéficiaire au moment des soins	Entières, de 0 à 103
exposition	Exposition du contrat sur l'exercice encouru	Réelles, entre 0 et 1
montantDE	Montant de dépense engagée / des frais réels	Réelles positives
montantRO	Montant du remboursement obligatoire / de la Sécurité sociale	Réelles positives
montantRC	Montant du remboursement complémentaire	Réelles positives
montantRC_base	Montant du remboursement complémentaire lié au niveau de couverture de base du contrat	Réelles positives
montantRC_option	Montant du remboursement complémentaire lié au niveau de couverture optionnel du contrat	Réelles positives
montantRC_obligatoire	Montant du remboursement complémentaire lié au(x) niveau(x) de couverture obligatoire(s) du contrat	Réelles positives
montantRC_facultatif	Montant du remboursement complémentaire lié au(x) niveau(x) de couverture facultatifs(s) du contrat	Réelles positives
montantRAC	Montant du reste à charge	Réelles positives
nbSin	Nombre de sinistres sur l'exercice considéré	Entières positives

TABLE A.9: Présentation des variables numériques.

NOM DE VARIABLE	DESCRIPTION DE LA VARIABLE	VALEURS
exercice	Exercice d'assurance	2018 2019
typeAss	Qualité du bénéficiaire	ASSURE PRINCIPAL CONJOINT ENFANT
benefSexe	Sexe du bénéficiaire	F M
benefClasseAge	Tranche d'âge du bénéficiaire au moment des soins	[0,20[ [20,30[ [30,40[ [40,50[ [50,60[ [60,70[ [70,80[ 80+
structureFoyer	Structure du foyer familial du bénéficiaire	coupleAvecEnfant coupleSansEnfant isoleAvecEnfant isoleSansEnfant
benefDept	Département de résidence du bénéficiaire	<i>Tous les départements français</i>
benefCollege	Collège du bénéficiaire	Cadres Non-cadres Ensemble personnel
CCN	Convention collective à laquelle est rattaché le bénéficiaire	CCN1 CCN2 CCN3 CCN4
couverture	Niveau de couverture du bénéficiaire	0 ( niveau de base ) 1 ( 1ère option ) 2 ( 2ème option )
couvObligatoire	Niveau de la couverture obligatoire du bénéficiaire	- 99 (pas de couverture obligatoire) 0 ( niveau de base ) 1 ( 1ère option ) 2 ( 2ème option )
couvFacultative	Niveau de la couverture facultative du bénéficiaire	- 99 (pas de couverture facultative) 0 ( niveau de base ) 1 ( 1ère option ) 2 ( 2ème option )
contractFormula	Caractère du contrat souscrit par l'ouvrant droit (obligatoire ou facultatif + niveau)	Oblig Fac+1 Fac+2 Fac+3
sousPoste	Sous-poste de soins	<i>28 modalités</i>

TABLE A.10: Présentation des variables catégorielles.

Quelques précisions sur les variables catégorielles sont données ci-dessous :

- ⇒ Les modalités de la variable `benefDept` correspondent aux numéros de départements français. Il est à noter que la modalité 20 représente les deux départements de Corse et que les départements d'outre-mer sont numérotés de 971 à 978 et la modalité 980 représente l'ensemble des autres collectivités et territoires d'outre-mer.
- ⇒ La couverture correspond au niveau de contrat souscrit, le niveau de base correspond au niveau d'entrée de gamme tandis que le niveau 1 et le niveau 2 représentent respectivement les niveaux de gamme moyen et haut.
- ⇒ La modalité `Oblig` de la variable `contratFormula` représente la couverture obligatoire proposée par l'entreprise tandis que les autres modalités représentent le nombre de niveaux supplémentaires souscrits par l'assuré par rapport à la couverture obligatoire proposée par l'entreprise.

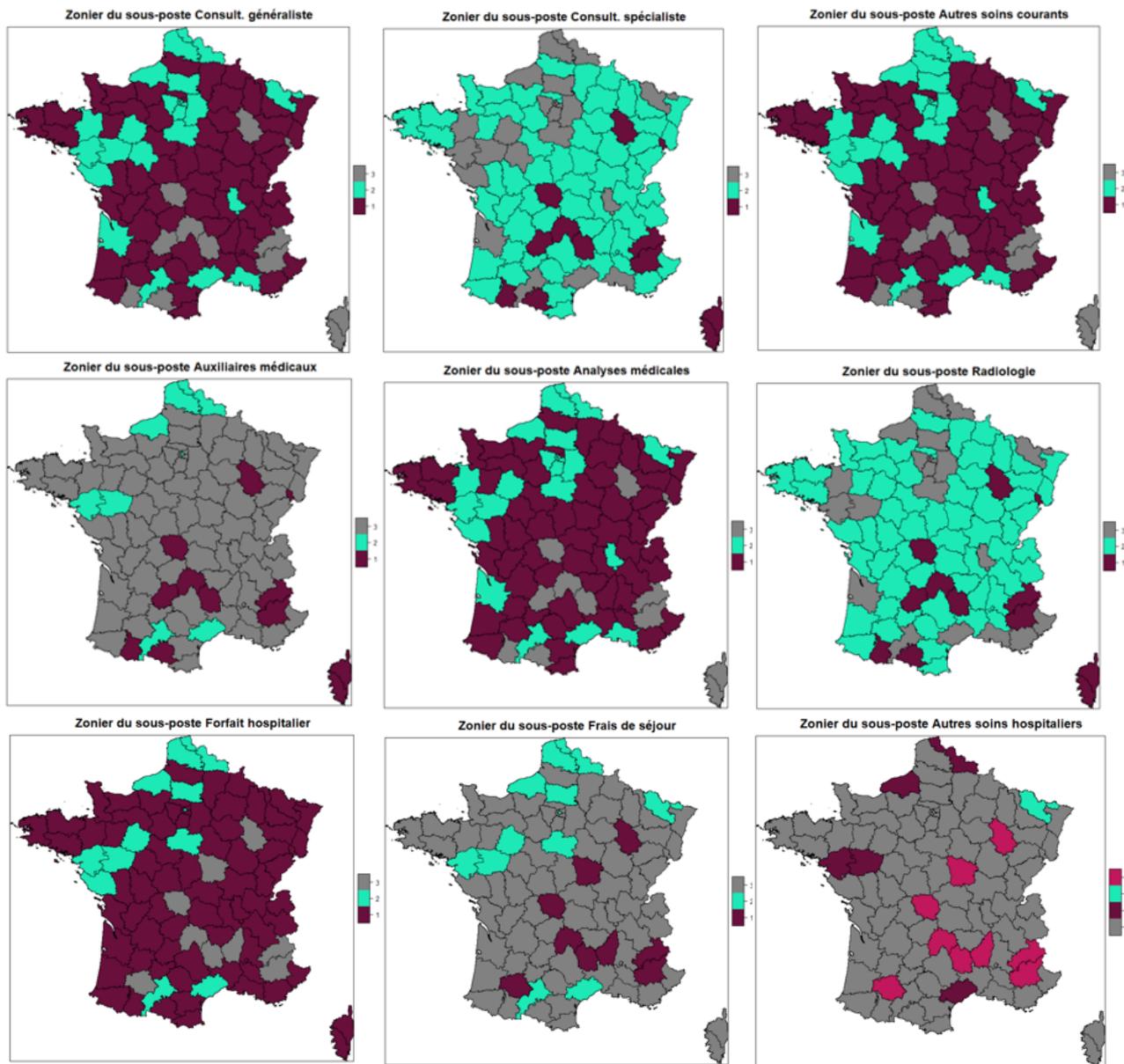
Les valeurs numériques, quant à elles, changent d'une base de données à l'autre. Par exemple, certains sous-postes sont caractérisés par des montants élevés de frais réels tandis que d'autres sont au contraire caractérisés par des montants plus faibles.

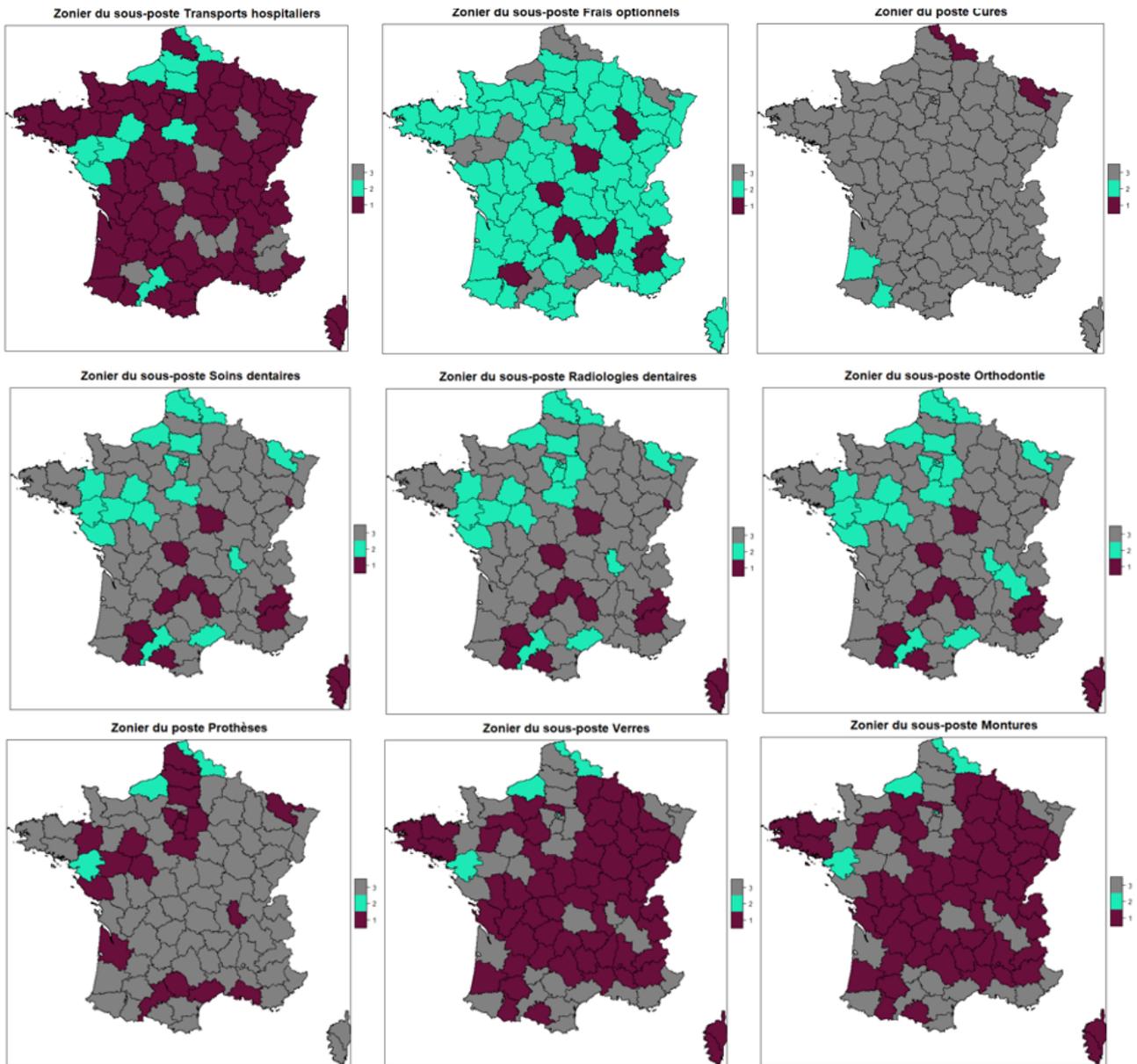
## A.2 Annexe 2 : Classes d'âge créées

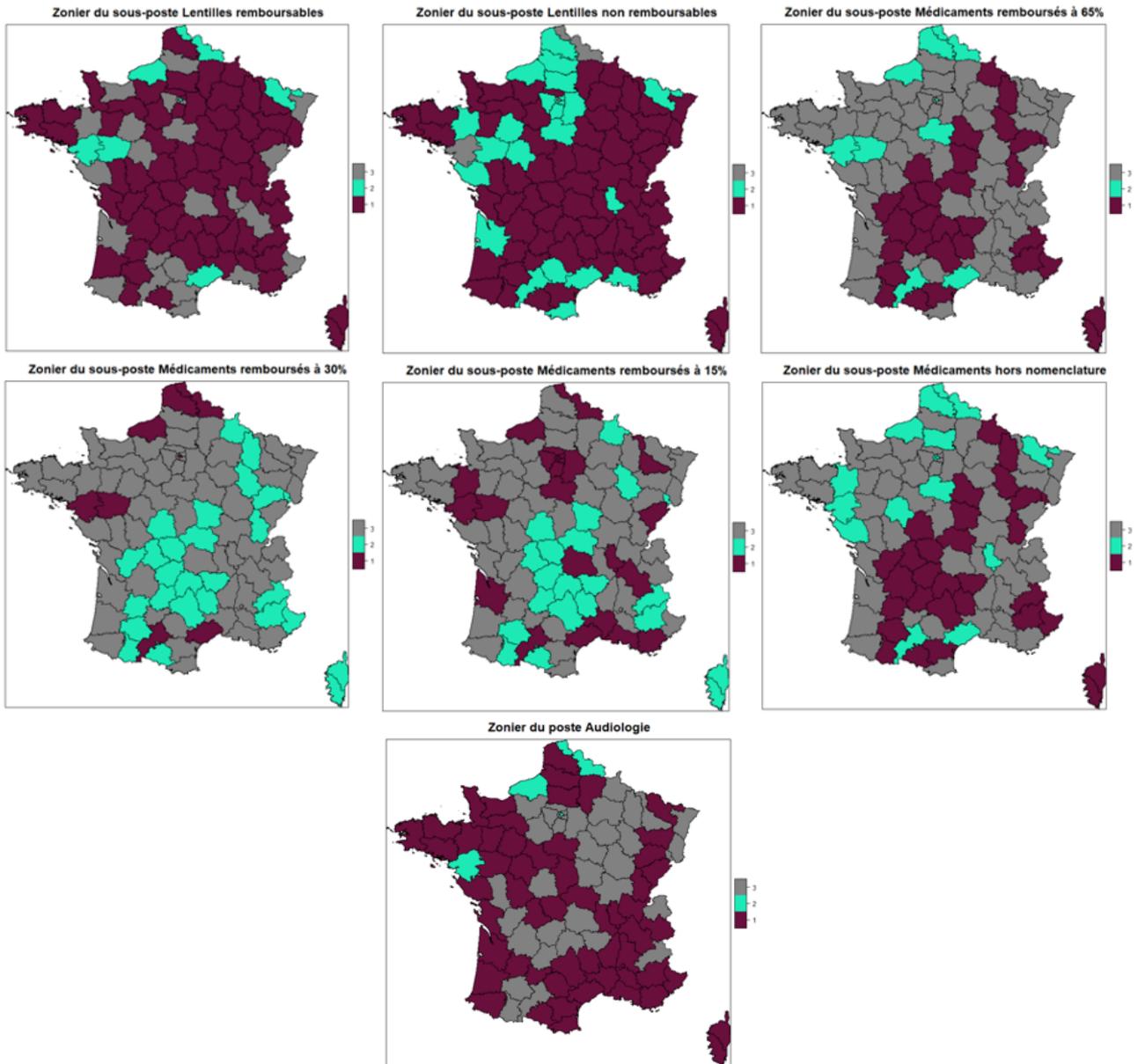
POSTE	SOUS POSTE	TRANCHES D'ÂGE
Soins courants	Auxiliaires médicaux	[0,9],[10,19],[20,79],80+
	Analyses médicales	[0,9],[10,29],[30,44],[45,79],80+
	Radiologie	[0,3],[4,7],[8,19],[20,24],[25,39],[40,69],70+
	Consult. généraliste	[0,9],[10,19],[20,59],60+
	Consult. spécialiste	[0,9],[10,19],[20,69],70+
	Autres soins	[0,5],[6,19],[20,59],60+
Dentaire	Prothèses dentaires	[0,16],[17,43],44+
	Soins dentaires	[0,6],[7,28],29+
	Radiologies dentaires	[0,5],[6,19],20+
	Orthodontie	[0,7],[8,19],20+
Hospitalisation	Frais de séjour	[0,2],[3,67],68+
	Forfait hospitalier	[0,2],[3,58],59+
	Frais optionnels	[0,24],[25,72],73+
	Transport	[0,69],70+
	Autres frais	[0,74],75+
Optique	Monture	[0,9],[10,41],[42,74],75+
	Verres	[0,44],[45,74],75+
	Lentilles remb.	[0,11],[12,76],77+
	Lentilles non remb.	[0,11],[12,74],75+
Pharmacie	Médicaments hors nomenclature	[0,3],[4,15],[16,54],55+
	Médicaments remb. à 15%	[0,19],[20,62],63+
	Médicaments remb. à 30%	[0,58],59+
	Médicaments remb. à 65%	[0,3],[4,51],52+
Audiologie	Prothèses auditives	[0,49],[50,69],70+
Cures thermales	Cures thermales	[0,55],56+
Prothèses	Prothèses	[0,59],60+

TABLE A.11: Tranches d'âge créées pour chaque sous-poste.

### A.3 Annexe 3 : Classes départementales créées







## A.4 Annexe 4 : Résultats des modèles linéaires généralisés

Les écrits ci-dessous présentent de manière succincte les résultats des modèles de fréquence et des modèles de coût réalisés sur l'ensemble des sous-postes de soins.

### Les soins courants

#### Les consultations chez le médecin généraliste : Modèle Binomiale négative - Gamma

- ➡ La variable la plus significative pour le modèle de fréquence est la tranche d'âge. La consommation est la plus élevée pour les assurés de moins de 10 ans et la plus faible pour les assurés entre 20 et 59 ans.
- ➡ La variable la plus significative pour le modèle de coût est à nouveau la tranche d'âge. Cette fois-ci, au contraire, ce sont les assurés de moins de 10 ans qui ont un coût moyen légèrement inférieur.

Les consultations chez le médecin généraliste étant encadrées en fonction du secteur d'appartenance du médecin, le coût moyen associé est difficilement mesurable en fonction des caractéristiques des assurés. Cependant, le modèle de coût montre une légère baisse du coût moyen pour les consultations des moins de 10 ans, sûrement due au secteur de convention des médecins pratiquant la consultation. Cette tranche d'âge a, en effet, une fréquence élevée et donc les patients nécessitent un suivi régulier opéré par le même médecin, souvent le médecin traitant. C'est pourquoi les parents ont tendance à choisir un médecin conventionné du secteur 1 pour lequel les tarifs sont réglementés.

#### Les consultations chez le médecin spécialiste : Modèle Binomiale négative - Gamma

- ➡ A nouveau, un pic de consommation est observé chez les moins de 10 ans. De plus, les assurés appartenant à la classe départementale 1 ont une fréquence de consommation légèrement plus élevée. Enfin, la fréquence de consommation en consultations chez le spécialiste a diminué en 2019 par rapport à 2018.
- ➡ Deux variables ont une importance significative pour le modèle de coût. Premièrement, le coût moyen est croissant en fonction de l'âge et deuxièmement, il est plus élevé pour les assurés appartenant à la seconde classe départementale.

Ce sont les enfants de moins de 10 ans, ayant besoin d'un suivi particulier durant le début de leur vie, qui totalisent la plus forte consommation en terme de consultations chez le médecin spécialiste. Et ce sont les seniors qui totalisent le coût moyen le plus élevé. De plus, des disparités à l'échelle départementale se dessinent au niveau de la fréquence et du coût moyen.

#### Les auxiliaires médicaux : Modèle Binomiale négative - Gamma

- ➡ A nouveau, le facteur âge joue sur la fréquence. Elle est importante pour les 10-19 ans et les plus de 80 ans. De plus, elle dépend positivement du niveau de couverture complémentaire des assurés et est plus élevée chez les cadres. Enfin, une différence de consommation apparaît au niveau géographique : les assurés rattachés aux départements de la classe 3 ont une consommation plus faible.
- ➡ Le facteur le plus important pour expliquer les variations de coût moyen est l'âge : le coût moyen est moins important pour les moins de 20 ans.

Les auxiliaires médicaux regroupent plusieurs professionnels médicaux (les infirmiers, les masseurs kinésithérapeutes, les podologues, les orthophonistes, ...). Deux types de consommateurs se dessinent : les 10-19 ans et les plus de 80 ans. Ils ne concernent a priori pas les mêmes professionnels médicaux. Par exemple, un orthophoniste interviendra en majorité chez les 10-19 ans tandis que les plus de 80 ans auront des visites d'infirmiers plus régulières. C'est pourquoi, des différences de coût moyen peuvent être observées au niveau des différentes tranches d'âge.

#### Les analyses médicales : Modèle Poisson - Gamma

- ⇒ La variable la plus significative au niveau de la fréquence est l'âge. Les individus ayant entre 30 et 44 ans et plus de 80 ans ont une consommation plus importante.
- ⇒ Deux facteurs principaux se dessinent au niveau du coût moyen. Le premier concerne l'âge, le coût moyen des analyses médicales est supérieur chez les 30-44 ans. Le second facteur est géographique, les assurés associés à la troisième classe départementale ont un coût moyen supérieur.

Les analyses médicales sont principalement consommées par deux types d'assurés. Le premier concerne les personnes de plus de 80 ans qui vont avoir des surveillances plus accrues au niveau de leur bilan sanguin. La seconde concerne les femmes entre 30 et 44 ans et plus largement les femmes prenant des contraceptifs hormonaux ou enceintes qui ont besoin de faire des bilans réguliers.

#### Les autres soins courants : Modèle Poisson - Lognormal

- ⇒ La fréquence de consommation est moins élevée pour les 6-19 ans et plus élevée chez les plus de 60 ans. De plus, elle est croissante en fonction du niveau de couverture complémentaire de l'assuré.
- ⇒ Le facteur le plus important au niveau du coût moyen est le suivant : le coût moyen des plus de 60 ans est plus élevé.

Ce sous-poste de soins est caractérisé par une dépense engagée plus élevée chez les plus de 60 ans due à une fréquence et un coût moyen plus importants.

#### La radiologie : Modèle Binomiale négative - Gamma

- ⇒ La fréquence de consommation en radiologie est croissante en fonction de l'âge avec un pic plus important au niveau des 8-19 ans.
- ⇒ Le coût moyen en radiologie oscille en fonction de l'âge. Il se dessine une légère tendance décroissante en fonction de ce facteur.

Ainsi, le facteur le plus important concernant la dépense engagée en radiologie est celui de l'âge.

Les sous-postes présentés ci-dessus sont tous spécifiques et ont leurs propres caractéristiques. L'âge apparaît comme le facteur le plus important afin d'expliquer l'évolution de la dépense engagée. Chaque sous-poste est plus ou moins consommé selon la tranche d'âge choisie. De plus, le coût des actes est corrélé à la tranche d'âge du patient.

## L'hospitalisation

### Les frais de séjour : Modèle Poisson - Gamma

- ⇒ Le premier facteur significatif est le niveau de couverture complémentaire. Plus il est élevé, plus la fréquence est importante. De plus, la fréquence est moins élevée chez les 3-68 ans. Enfin, les assurés rattachés à la première classe départementale ont une fréquence légèrement supérieure.
- ⇒ Les deux seuls facteurs significatifs vérifiant le test de non-nullité sont l'âge et l'exercice. Le coût moyen est plus élevé chez les nourrissons (les moins de 3 ans) et moins élevé pour les 3-68 ans et il est plus élevé en 2018 qu'en 2019.

La dépense engagée en frais de séjour dépend fortement de l'âge des patients. Ce sont principalement les nourrissons et les personnes âgées qui vont bénéficier de soins hospitaliers et les nourrissons sont caractérisés par des coûts hospitaliers supérieurs. De plus, le facteur géographique et le niveau de couverture complémentaire influent sur la fréquence.

### Le forfait hospitalier : Modèle Binomiale négative

Il s'agit d'un modèle Binomial négatif pour la fréquence. Le facteur le plus significatif est l'âge. La fréquence en frais de séjour est plus élevée chez les enfants de moins de 3 ans et elle est la moins élevée pour les 3-58 ans. Le second facteur significatif est le niveau de couverture obligatoire et facultatif. Plus ils sont élevés, plus la fréquence l'est également.

Le coût du forfait hospitalier est de 20€, ce montant est fixé réglementairement. Ainsi, un seul modèle linéaire a été réalisé et il montre que l'âge et le niveau de couverture complémentaire influent sur la fréquence.

### Les autres soins : Modèle Poisson - Gamma

- ⇒ Plusieurs facteurs interviennent dans le modèle de coût. Le plus important est le facteur géographique. La troisième classe départementale a une fréquence beaucoup plus élevée tandis que la première l'est légèrement moins comparée aux deux autres. Ensuite, un second facteur important est l'âge : les plus de 75 ans ont une plus forte fréquence. Enfin, un troisième facteur important est le niveau de couverture complémentaire, plus il est élevé, plus la fréquence l'est également.
- ⇒ Le coût moyen dépend aussi fortement du facteur géographique : le coût moyen associé aux assurés de la classe départementale 3 est beaucoup moins élevé que celui des trois autres classes. De plus, les niveaux de couverture obligatoire élevés ont un coût moyen plus faible.

En résumé, deux facteurs prépondérants expliquent la dépense engagée : le premier est un facteur géographique, jouant à la fois sur la fréquence et le coût, et le second est le facteur d'âge, impactant la fréquence.

### Les transports : Modèle Binomiale négative - Lognormal

- ⇒ Deux facteurs importants peuvent être retenus : la fréquence est plus élevée chez les assurés âgés de plus de 70 ans et elle dépend du département de l'assuré. Les assurés rattachés à la classe départementale 3 ont une fréquence beaucoup plus élevée.
- ⇒ A nouveau, deux facteurs peuvent être retenus mais cette fois-ci pour le modèle de coût : le type de l'assuré et sa classe départementale. Les enfants ont un coût moyen plus élevé et la troisième classe départementale a un coût moyen moins élevé.

Il n'est pas étonnant que le facteur géographique joue un rôle important pour ce sous-poste étant donné qu'il s'agit de transports de patients. Il en est de même pour l'âge des assurés.

Les frais optionnels : Modèle Poisson - Gamma

- ⇒ La fréquence de consommation en frais optionnels admet trois caractéristiques majeures. Premièrement, la consommation est croissante avec l'âge des assurés. Deuxièmement, la fréquence est plus élevée chez les contrats de complémentaire hauts de gamme. Enfin, les cadres ont une fréquence également plus importante.
- ⇒ Une seule variable vérifie l'hypothèse de non-nullité : la tranche d'âge de l'assuré. Plus l'assuré est âgé, plus le coût moyen associé à ces frais optionnels est élevé.

Ce sous-poste caractérisé par un taux de remboursement nul de la part de la Sécurité sociale est principalement consommé par des assurés ayant un niveau de vie et/ou couverture élevé. De plus, le facteur âge corrélé positivement pèse également dans sa consommation.

Ainsi, les soins liés à l'hospitalisation sont caractérisés par une forte dépendance par rapport à l'âge. Ce sont principalement les personnes âgées mais aussi les nourrissons qui vont avoir une forte consommation en frais hospitaliers. De plus, les autres frais tels que les transports et frais optionnels sont principalement dépendant de l'âge mais aussi du niveau de couverture complémentaire et du niveau de vie des patients. Enfin, le facteur géographique joue également sur les différents sous-postes.

## **Le dentaire**

Les soins dentaires : Modèle Poisson-Gamma

- ⇒ La variable la plus significative pour le modèle de fréquence est la tranche d'âge de l'assuré : la fréquence est croissante en fonction de l'âge. Deux autres variables sont également significatives, ce sont les niveaux de couverture obligatoires et facultatifs. Plus le niveau de couverture est élevé, plus la fréquence l'est aussi.
- ⇒ Peu de variables sont sélectionnées pour le modèle de coût. Seules deux variables vérifient le test de non-nullité. Ce sont les variables représentant la convention collective nationale de l'assuré et sa tranche d'âge. Le coût est plus élevé pour les conventions collectives numéro 1 et 2 et il augmente en fonction de l'âge.

La fréquence ainsi que le coût des soins dentaires augmentent avec l'âge. De plus, les personnes bénéficiant d'une couverture importante ont une consommation en moyenne plus élevée.

Les prothèses dentaires : Modèle Poisson-Gamma

- ⇒ La fréquence de consommation en prothèses dentaires chez les enfants est beaucoup moins élevée que celle des adultes. Elle est également plus importante chez les cadres ainsi que chez les assurés bénéficiant d'une haute couverture complémentaire.
- ⇒ Peu de variables sont sélectionnées dans le modèle de coût. Seules deux variables apparaissent significatives. Ce sont la tranche d'âge et la classe du département. Le coût des prothèses dentaires est croissant avec l'âge et la première classe départementale a un coût moyen plus élevé que la troisième.

Le sous-poste Prothèses dentaires voit sa fréquence de consommation ainsi que son coût moyen augmenter avec l'âge. En effet, les prothèses dentaires définitives apparaissent plus nombreuses chez les adultes et plus coûteuses que les prothèses temporaires faites sur les enfants qui n'ont pas encore atteint la fin de leur croissance osseuse. De plus, la fréquence de consommation dépend du niveau de vie et de couverture de l'assuré. Enfin, le coût moyen est plus ou moins élevé selon les départements.

#### L'orthodontie : Modèle Binomiale négative - Lognormal

- ⇒ La variable la plus importante pour le modèle de fréquence est la tranche d'âge de l'assuré : ce sont les 8-19 ans qui ont la fréquence de consommation la plus élevée en orthodontie. De plus, les classes départementales 2 et 3 ont une fréquence de consommation plus élevée. Enfin, la fréquence est croissante en fonction du niveau de couverture complémentaire.
- ⇒ Le coût moyen est également plus élevé pour les assurés entre 6 et 20 ans. Il dépend également positivement du niveau de couverture ainsi que du collège de l'assuré, les cadres ont un coût moyen plus important.

L'orthodontie est caractérisée par une consommation forte chez les 8-19 ans. C'est à cette tranche d'âge que les traitements dentaires importants sont réalisés. C'est pourquoi leur fréquence mais aussi leur coût moyen sont plus élevés. De plus, les assurés adaptent leur consommation en fonction de leur niveau de vie et de leur niveau de couverture complémentaire.

#### Les radiologies dentaires : Modèle Binomiale négative - Lognormal

- ⇒ La fréquence de consommation en radiologies dentaires est la plus élevée entre 6 et 20 ans et elle augmente en fonction du niveau de couverture complémentaire.
- ⇒ De même, le coût moyen est plus élevé chez les 6-20 ans. Il est également plus important chez les cadres ainsi que chez les assurés bénéficiant d'un contrat complémentaire haut de gamme.

Les radiologies dentaires sont des actes principalement consommés par les 6-20 ans. C'est à cet âge-là que le développement osseux est surveillé. De plus, la consommation en radiologies dentaires est corrélée avec les actes d'orthodontie, c'est pourquoi un pic de consommation est observé à ces âges-là. Enfin, le coût moyen dépend également du niveau de vie et de couverture des assurés.

Les sous-postes dentaires peuvent être regroupés en deux catégories avec d'une part les soins et prothèses dentaires et d'une autre part l'orthodontie et les radiologies dentaires. Le premier regroupement est caractérisé par une augmentation de la consommation en fonction de l'âge tandis que le second a une tranche d'âge plus ciblée. De plus, les sous-postes dentaires sont caractérisés par une consommation dépendant fortement du niveau de vie et du niveau de couverture complémentaire des assurés qui adaptent leur consommation. En effet, ces sous-postes sont parmi les plus importants en terme de dépense engagée, c'est pourquoi aujourd'hui encore le renoncement aux soins reste élevé en dentaire.

## **L'optique**

#### Les verres : Modèle Poisson-Gamma

- ⇒ La fréquence de consommation est plus élevée chez les 45-74 ans bien qu'elle soit importante pour les 10-44 ans et les plus de 75 ans.

- ➡ Le coût des verres est croissant en fonction de l'âge. Cette même information est analysée au niveau du type de l'assuré : les enfants ont un coût moyen plus faible.

Sans surprise, la dépense engagée dépend fortement de l'âge et son coût croît avec l'âge. En effet, la vue se dégrade avec l'âge et les personnes plus âgées nécessitent des verres complexes, notamment à cause de la presbytie. De plus, ce trouble de la vision intervient entre 45 et 60 ans, c'est pourquoi, puisque leur vision change plus vite, certains patients peuvent avoir une fréquence de consommation plus élevée à ces âges.

#### Les montures : Modèle Poisson-Gamma

- ➡ A nouveau, la fréquence de consommation en monture est plus élevée chez les 42-74 ans. De plus, la fréquence est plus importante en 2018 qu'en 2019. Enfin, elle est plus supérieure pour les niveaux de couverture élevés.
- ➡ Le coût est croissant en fonction de l'âge. Il est également dépendant des classes de département créées et est plus élevé pour la deuxième classe. Les salariés non-cadres ainsi que les personnes assurées par des niveaux de couverture de base ont un coût moyen plus faible. Enfin, le coût moyen est plus élevé en 2019 qu'en 2018.

En 2019, la consommation de monture a vu son coût augmenter et par conséquent sa fréquence baisser. De plus, sa consommation est plus élevée chez les 42-74 ans, notamment à cause des problèmes de vue évoqués dans la section précédente, ainsi que chez les personnes ayant un niveau de couverture complémentaire élevé. En effet, certaines montures peuvent avoir un coût plutôt élevé et ce sont principalement des patients bien couverts et/ou plutôt aisés qui les achètent.

#### Les lentilles remboursables : Modèle Binomiale négative - Gamma

- ➡ La fréquence de consommation en lentilles remboursables par la Sécurité sociale dépend fortement de l'âge, elle est plus importante chez les 12-76 ans. Elle dépend également fortement de la convention collective nationale à laquelle est rattaché l'assuré.
- ➡ Le modèle de coût est basé sur un nombre limité d'observations, 950. Pour ce modèle, peu de variables ont été retenues et un nombre limité de facteur vérifie le test de non-nullité. La seule caractéristique retenue est la suivante : le coût est plus élevé chez les personnes assurées par un niveau de couverture de base.

Il s'agit d'un sous-poste ayant une fréquence de consommation faible. Il est en majeure partie consommé par des individus âgés entre 12 et 76 ans et est dépendant de la convention collective nationale.

#### Les lentilles non remboursables : Modèle Poisson - Lognormal

- ➡ La variable la plus importante est la classe d'âge de l'assuré. Les individus ayant entre 12 et 74 ans ont une plus forte consommation. De plus, les individus couverts par un contrat de complémentaire santé avec des niveaux élevés de garanties ont une plus grande fréquence de consommation.
- ➡ Le coût moyen est plus élevé chez les enfants de moins de 12 ans. De plus, il est plus élevé chez les cadres et les individus souscrivant une couverture facultative à niveau de garanties élevé.

Ce sous-poste est très sensible à l'âge et sa consommation est plus élevée chez les assurés ayant souscrit à des niveaux de couverture supplémentaires.

Le poste Optique peut être scindé en deux groupes avec d'une part les montures et verres optiques et d'une autre part les lentilles. Ces deux groupes ont un pic de consommation dépendant de l'âge, respectivement 42-74 ans et 12-74 ans. Ils sont également sensibles aux niveaux de couverture des assurés qui adaptent leur consommation en fonction de ce paramètre.

## La pharmacie

### Les médicaments hors nomenclature : Modèle Poisson - Gamma

- ➡ La consommation en médicaments dépend fortement de la convention collective nationale à laquelle est rattaché l'assuré et de son niveau de couverture complémentaire. En effet, l'assuré adapte sa consommation en fonction du niveau de remboursement offert par son contrat de complémentaire étant donné que la Sécurité sociale n'intervient pas pour cette catégorie de médicaments. De plus, la fréquence dépend du département de l'assuré ainsi que de son âge : la classe départementale 2 a une plus grande fréquence que les autres et les 16-54 ans consomment plus de médicaments non remboursés. Enfin, la fréquence de consommation est plus élevée en 2018 qu'en 2019.
- ➡ Les variables significatives retenues sont similaires à celle du modèle de fréquence. Ce sont la convention collective nationale, la tranche d'âge et l'année de consommation. Le coût moyen est plus élevé chez les individus âgés de plus de 16 ans et il est également plus élevé en 2019 qu'en 2018.

La consommation en médicaments hors nomenclature a vu son coût augmenter et par conséquent sa fréquence baisser en 2019. De plus, les assurés adaptent leur consommation en fonction de leur niveau de couverture complémentaire étant donné que c'est leur seule protection sociale pour ce type de médicaments.

### Les médicaments remboursés à 15% par la Sécurité sociale : Modèle Poisson - Gamma

- ➡ La fréquence de consommation en médicaments remboursés à 15% dépend principalement du niveau de couverture complémentaire ainsi que de la convention collective nationale à laquelle l'assuré est rattaché. De plus, ces médicaments sont majoritairement consommés par les moins de 20 ans et les plus de 63 ans.
- ➡ Peu de variables ont été retenues dans le modèle de coût, seulement une vérifie le test de non-nullité. Il s'agit de la tranche d'âge de l'assuré. Le coût moyen chez les moins de 20 ans est inférieur à celui des plus de 20 ans.

Les médicaments remboursés à 15%, jugés médicaments à service médical rendu\* faible, ont une consommation dépendante du niveau de couverture complémentaire et de l'âge des assurés.

---

\*Le service médical rendu (SMR) est un critère permettant de classer les médicaments en fonction de leur utilité thérapeutique publié par la Haute Autorité de Santé (HAS).

Les médicaments remboursés à 30% par la Sécurité sociale : Modèle Binomiale négative - Lognormal

- ⇒ A nouveau, la fréquence dépend du niveau de couverture complémentaire ainsi que de la convention collective de rattachement. De plus, les individus âgés de plus de 59 ans ont une consommation plus élevée. Les cadres ont également une fréquence plus importante. Enfin, les assurés appartenant aux départements de la classe 3 ont une fréquence inférieure aux autres départements.
- ⇒ Peu de modalités vérifient l'hypothèse de non-nullité. La variable la plus pertinente concerne l'âge des assurés. Les assurés ayant plus de 59 ans ont un coût moyen légèrement plus important.

Les médicaments remboursés à 30%, jugés comme ayant un service médical rendu\* modéré, voient leur consommation principalement caractérisée par l'âge des assurés ainsi que par leur niveau de couverture complémentaire.

Les médicaments remboursés à 65% par la Sécurité sociale : Modèle Binomiale négative - Lognormal

- ⇒ La variable la plus caractéristique est celle de l'âge des assurés. Les enfants de moins de 4 ans ont une fréquence de consommation plus élevée, bien qu'elle soit également importante pour les plus de 52 ans. De plus, la fréquence de consommation est croissante en fonction du niveau de couverture obligatoire. Enfin, les assurés de la classe départementale 3 ont une fréquence moins élevée.
- ⇒ De faibles variations sont observées en fonction des modalités des différentes variables. Cependant, une variable principale peut caractériser le coût moyen en médicaments remboursés à 65%, il s'agit de la tranche d'âge. Ce dernier est plus élevé pour les moins de 4 ans ainsi que pour les plus de 52 ans.

La consommation des médicaments remboursés à 65% par la Sécurité sociale, jugés comme ayant un service médical rendu\* majeur ou important, est fortement dépendante de l'âge des assurés. Ils sont notamment consommés par les nourrissons et les enfants de moins de 4 ans ainsi que par les personnes de plus de 52 ans.

Pour rappel, étant donné que les données viennent d'une société d'assurance, la consommation en médicaments intégralement remboursés par la Sécurité sociale ne peut être analysée ici.

Ainsi, la consommation en médicaments admet des facteurs explicatifs différents selon le niveau de remboursement de la Sécurité sociale. Pour ceux ayant un remboursement faible voire nul, le niveau de couverture complémentaire impacte fortement la fréquence. De plus, la fréquence et le coût moyen sont caractérisés par une plus forte consommation aux âges faibles et aux âges élevés.

**L'audiologie** : Modèle Binomiale négative - Gamma

- ⇒ La principale variable explicative pour le modèle de fréquence est l'âge. En effet, plus l'assuré est âgé, plus la fréquence de consommation moyenne associée est élevée. De plus, les assurés bénéficiant d'un contrat de couverture haut de gamme ont une plus grande fréquence de consommation. Enfin, les assurés cadres ainsi que les assurés rattachés à la classe départementale 2 ont une fréquence de consommation plus élevée et la fréquence est plus élevée en 2019.

---

\*Le service médical rendu (SMR) est un critère permettant de classer les médicaments en fonction de leur utilité thérapeutique publié par la Haute Autorité de Santé (HAS).

- ➡ La principale variable explicative du modèle de coût est la variable indiquant le niveau de couverture facultative. Le coût moyen est croissant en fonction de ce niveau. De plus, le coût moyen en 2019 est supérieur à celui de 2018.

L'audiologie est ainsi caractérisée par une augmentation de la fréquence aux âges élevés. De plus, la fréquence de consommation est dépendante du niveau de couverture complémentaire et du niveau de vie de l'assuré. Enfin, le coût moyen en prothèses auditives est en augmentation croissante.

#### **Les cures thermales : Modèle Poisson - Gamma**

- ➡ La fréquence dépend de plusieurs facteurs. Premièrement, les enfants sont de faibles consommateurs de cures thermales. Deuxièmement, les employés non-cadres consomment plus de cures thermales. De plus, les assurés de la classe départementale 1 regroupant la Moselle et le Nord, ont une fréquence plus élevée. Enfin, les assurés en couple sans enfant ont une fréquence également plus élevée.
- ➡ Seules deux variables apparaissent significatives dans l'analyse du modèle de coût. Ce sont les variables caractérisant le niveau de couverture obligatoire et le collège de l'assuré. Plus le niveau de couverture obligatoire est élevé, plus le coût moyen est élevé et les cadres ont un coût moyen en cures thermales plus élevé.

Les cures thermales sont majoritairement fréquentées par des personnes âgées. Il semblerait également que les non-cadres aient à la fois une fréquence de consommation plus élevée et un coût moyen moins élevé que les cadres.

#### **Les prothèses : Modèle Binomiale négative - Gamma**

- ➡ Deux facteurs sont significatifs concernant la fréquence de consommation en prothèses. Le premier concerne l'âge des assurés : les assurés de plus de 60 ans ont une fréquence de consommation plus importante. Le second concerne le niveau de couverture complémentaire : la fréquence est corrélée positivement avec le niveau de couverture.
- ➡ Trois caractéristiques majeures peuvent être citées concernant le coût moyen. Il est moins élevé chez les enfants, et plus élevé pour les plus de 60 ans. En effet, les matériaux utilisés sont différents : la prothèse d'un enfant va a priori être changée régulièrement étant donné sa croissance tandis que celle d'une personne adulte devra être résistante dans la durée. De plus, le coût moyen des prothèses est plus élevé chez les cadres.

Les prothèses voient leur consommation augmenter en fonction de l'âge des assurés. De plus, leur fréquence ainsi que leur coût dépendent du niveau de vie et du niveau de couverture de l'assuré.

Cette revue de l'ensemble des variables significatives pour chaque modèle de fréquence et chaque modèle de coût permet de mieux comprendre les spécificités liées à chaque risque de santé sous-jacent. Il apparaît que la variable la plus importante est en général l'âge. D'autres facteurs sont également prépondérants, ce sont le niveau de couverture complémentaire santé et le niveau de vie des assurés. Par exemple, ces facteurs sont importants en optique et en dentaire. Enfin, des disparités apparaissent au niveau départemental à la fois sur la fréquence et le coût moyen.

Grâce à ces analyses, il peut être possible d'individualiser un tarif selon les caractéristiques de l'assuré.

## A.5 Annexe 5 : Indicateur du remboursement de la Sécurité sociale

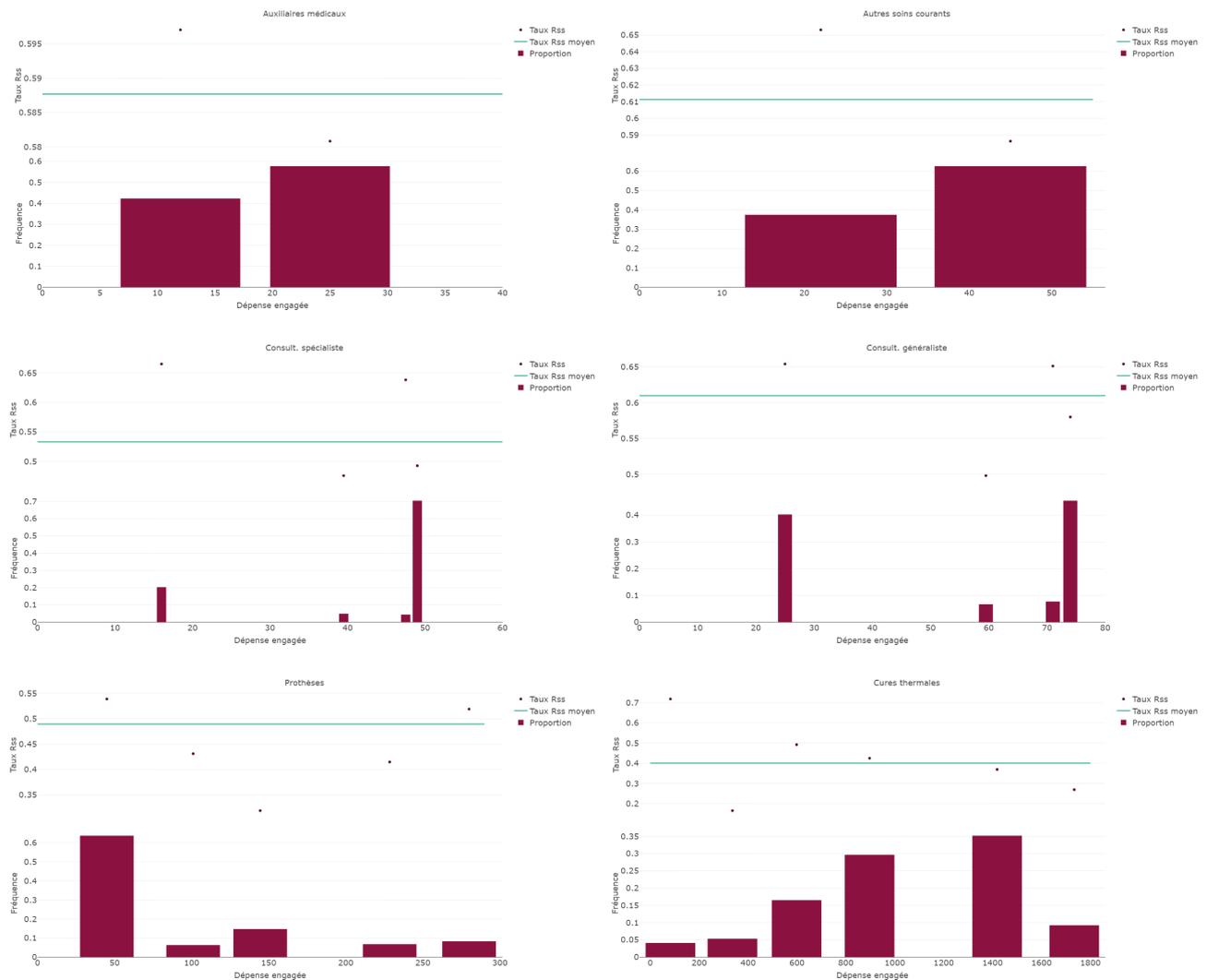
### A.5.1 Les sous-postes représentés par le taux de remboursement moyen

Deux sous-postes n'ont pas pu être divisés en tranches, ce sont les analyses médicales et les frais de transport hospitaliers puisqu'ils ont des remboursements homogènes en fonction de la dépense moyenne.

D'autres postes ont peu de catégories et un taux par catégorie très proche du taux moyen, les soins délivrés par des auxiliaires médicaux et les autres soins courants en sont de parfaits exemples.

Certains postes ont une segmentation par catégorie plus fine et des taux sensiblement plus différents d'un montant de dépenses engagées à l'autre. Ce phénomène caractérise les sous-postes ayant des actes de soins variés et montre les limites de la modélisation par sous-postes, il s'agit par exemple des cures thermales et des équipements prothétiques.

Ainsi, pour la majorité des sous-postes concernés, le taux de remboursement par catégorie est semblable au taux de remboursement moyen.



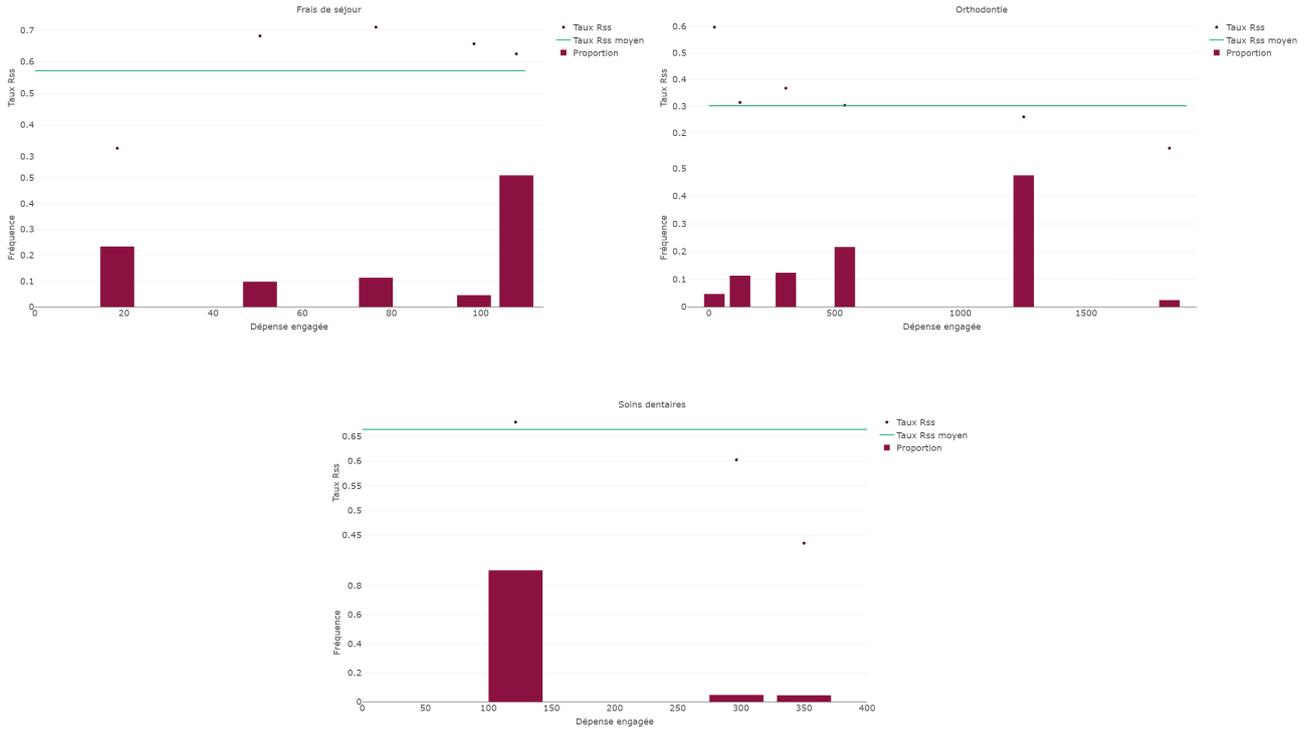


FIGURE A.17: Visualisation du taux de remboursement moyen en fonction des tranches de dépenses.

### A.5.2 Les sous-postes représentés par une relation linéaire

Les arbres de régressions ont créé 3 à 6 tranches suivant les sous-postes. Une droite de régression a été tracée en prenant en compte les poids de chacune des tranches. Le coefficient de détermination, noté  $R^2$ , définit par l'équation A.19, permet de mesurer la qualité de prédiction des régressions créées :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (\text{A.19})$$

où  $n$  représente le nombre d'observations,  $y_i$  la valeur observée,  $\hat{y}_i$  la valeur prédite et  $\bar{y}$  la valeur moyenne.

Le coefficient de détermination, représentant la part de variance expliquée par la régression linéaire, est compris entre 0 et 1. Plus il se rapproche de 1, plus il explique une part importante de la variance. Il est supérieur à 0,8 pour tous les sous-postes et atteint même 0,99 pour certains sous-postes. La qualité des différentes régressions est donc bonne, voire très bonne.

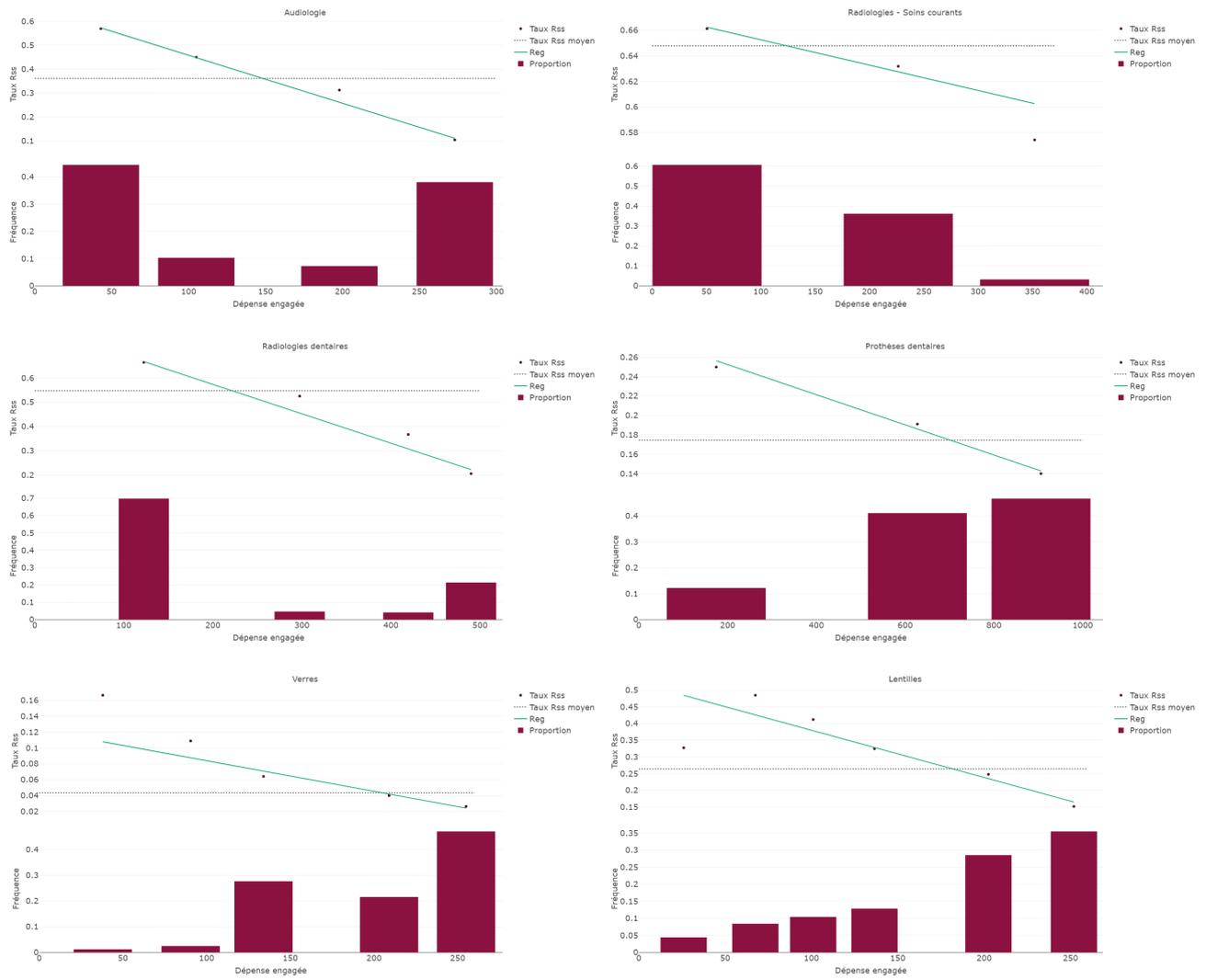


FIGURE A.18: Visualisation du taux de remboursement moyen en fonction des tranches de dépenses.

## A.6 Annexe 6 : Le sondage

### Le sondage

Le sondage, accessible via le lien <https://siapartners2.typeform.com/to/ANZ8jjYT>, est structuré de la manière suivante :

- ❶ Questions portant sur l'assuré (sexe, âge, composition du foyer familial, catégorie socioprofessionnelle) ;
- ❷ Questions portant sur les importances du sondé en santé à la maille poste de soins ;
- ❸ Questions portant sur les importances du sondé en santé à la maille sous-poste de soins.

### Le rapport du sondage

Un rapport résumant les réponses des sondés est disponible via le lien suivant <https://siapartners2.typeform.com/report/ANZ8jjYT/5NUvx4QrEpe2mz82>.

### Analyse des résultats du sondage à la maille poste

#### Importances absolues des postes de soins :

L'étude des importances absolues présentées par la figure A.19 permet de déceler les variations d'importance entre les différents postes de soins pour les différents profils de sondés.

Dans l'ensemble, les femmes accordent des degrés d'importance plus élevés aux différents postes de soins que les hommes. Trois postes en particulier sont concernés par ce phénomène, ce sont les cures thermales, les soins courants et la pharmacie. En effet, ces postes sont caractérisés par une consommation supérieure chez les femmes par rapport à celle des hommes.

Une tendance croissante des importances absolues en fonction de la tranche d'âge se détache pour la majorité des postes. Certains sont plus marqués, ce sont notamment l'audiologie et les cures thermales. En effet, les problèmes auditifs concernent majoritairement des personnes âgées de plus de 70 ans et les cures thermales sont principalement fréquentées par des personnes retraitées. A contrario, trois postes dérogent à cette analyse, ce sont les soins courants, l'optique et la pharmacie. La pharmacie est caractérisée par une décroissance de l'importance absolue en fonction de l'âge tandis que l'optique et les soins courants sont peu sensibles à la tranche d'âge du sondé. Enfin, les prothèses ont une importance supérieure pour les individus âgés entre 50 et 64 ans alors que c'est le contraire pour les soins optiques.

A nouveau, des effets similaires sont analysés au niveau de la catégorie socioprofessionnelle : l'importance absolue accordée par les retraitées est en général supérieure à la moyenne et celle des étudiants inférieure. Mis à part cette singularité, la catégorie socioprofessionnelle ne semble pas influencer les postes Soins courants, Hospitalisation, Optique, Dentaire et Pharmacie. Elle est cependant influente pour les cures thermales, les prothèses et l'audiologie : les non-cadres accordent une importance plus élevée pour ces postes. Deux explications possibles sont le coût moyen élevé pour l'audiologie et la consommation moins élevée en soins prothétiques et en cures thermales pour les cadres.

Enfin, l'analyse des résultats obtenus en fonction de la composition du foyer familial est difficilement interprétable puisqu'aucune tendance ne s'en détache.

Ci-dessus ont été analysées les appétences moyennes de chaque poste de soins en fonction des caractéristiques des sondés. Or, le but est d'obtenir la répartition de chaque poste. Pour cela, les réponses de chaque sondé ont été pondérées afin d'obtenir le pourcentage d'importance de chaque poste de soins pour chaque sondé.

#### Importances relatives des postes de soins :

Les poids obtenus sont issus de la moyenne pondérée des importances relatives de chaque poste de soins. Ils permettent donc de comparer l'importance d'un poste de soins par rapport à l'ensemble des autres postes de soins. Des analyses complémentaires peuvent être faites à celles présentées sur les importances absolues grâce à la figure A.20.

Tout d'abord, le poste Soins courants a une importance relative décroissante en fonction de l'âge : plus un individu devient âgé, moins ce poste a un poids lourd dans sa consommation en santé. Tandis que les postes Audiologie et Cures thermales voient, au contraire, leur importance relative augmenter avec l'âge.

A nouveau, les employés non-cadres accordent une importance relative plus élevée pour les cures thermales. L'hospitalisation ainsi que la Pharmacie ont un poids plus élevé dans la consommation en santé chez les étudiants. Enfin, l'Optique et le Dentaire ont un poids plus élevé dans la consommation en santé chez les cadres. En effet, comme vu précédemment, ce sont des postes ayant un coût moyen élevé et une consommation moindre chez les individus moins aisés.

Enfin, les prothèses ont une importance relative plus élevée chez les individus âgés entre 50 et 64 ans et particulièrement chez les hommes, ce qui est cohérent avec la sinistralité observée dans les bases de données.

### **Analyse des résultats du sondage à la maille sous-poste**

Une description détaillée est présentée ci-dessous par regroupement par poste de soins.

#### Le poste Dentaire :

L'importance pour l'orthodontie diminue fortement en fonction de la tranche d'âge. En effet, les étudiants et les sondés ayant des enfants à charge accordent une plus forte importance à ce poste. Ces constatations sont expliquées par une consommation principalement observée chez les enfants et forment le poids le plus faible du poste dentaire.

Les soins dentaires ont le poids le plus important du poste dentaire. Ce poids diminue en fonction de la tranche d'âge tandis que le contraire est observé pour les prothèses dentaires. Comme vu précédemment, ces sous-postes ont une consommation croissante en fonction de l'âge, il est donc normal de retrouver cette croissance pour les prothèses dentaires et étonnant de retrouver cette décroissance pour les soins dentaires.

De plus, les radiologies dentaires sont majoritairement consommées par les personnes de moins de 20 ans et donc leur poids devrait être décroissant en fonction de la tranche d'âge. Or il ne semble pas dépendre de la tranche d'âge visée et un pic est atteint pour les individus âgés entre 50 et 64 ans.

### Le poste Optique :

L'importance pour les lentilles remboursables et non remboursables par la Sécurité sociale diminue fortement pour les sondés de plus de 50 ans, qui consomment moins ce type de produit tandis que l'effet inverse est observé pour les montures et verres. De plus, les cadres accordent une importance plus élevée aux montures que les non-cadres.

Le poids associé aux verres est nettement supérieur aux autres sous-postes optiques tandis que ceux associés aux lentilles sont parmi les plus faibles poids de sous-postes.

### Le poste Pharmacie :

Étant donné que les taux de remboursement de la Sécurité sociale des médicaments ne sont pas toujours connus par tous, seules deux questions ont été posées dans le sondage, elles concernent :

- ⇒ l'importance pour les médicaments non remboursés par la Sécurité sociale ;
- ⇒ l'importance pour les médicaments remboursés partiellement par la Sécurité sociale.

Ainsi, les médicaments remboursés partiellement auront le même poids qu'ils soient remboursés à 15%, 30% ou 65%. De plus, les résultats montrent que l'importance absolue des médicaments remboursables et non remboursables est sensiblement la même, avec respectivement 6,72 et 6,95. Les sondés ont donc estimé une appétence très légèrement supérieure pour les médicaments remboursables. Ces derniers sont remboursables car ils sont estimés comme étant des médicaments à service médical rendu\* plus ou moins élevé. La différence d'appétence peut être remise en cause et mériterait une analyse plus poussée. Elle est cependant prise en compte par la suite et mène aux poids 3,65% et 3,79%.

### Le poste Soins courants :

Les importances pour les consultations chez le médecin généraliste et spécialiste apparaissent invariables en fonction des caractéristiques des sondés. En termes d'importance relative, elle est supérieure chez les étudiants.

Les analyses médicales ainsi que les auxiliaires médicaux ont, quant à elles, une importance supérieure chez les retraités qui consomment plus ce type de soins. Cependant, une importance égale entre les hommes et les femmes est observée au niveau du sous-poste Analyses médicales alors qu'une consommation supérieure est observée pour les femmes dans les bases de données.

Enfin, l'importance des sous-postes Radiologie et Autres soins courants diminue en fonction de l'âge. Les consultations chez le médecin spécialiste ont le poids le plus important et c'est bien ce sous-poste en question qui a une des plus fortes dépenses engagées. Ensuite, ce sont les analyses médicales et les consultations chez le médecin généraliste. Enfin, les autres soins ont un poids plus faible et ce sous-poste a bien la dépense engagée la plus faible du poste.

### Le poste Hospitalisation :

Ce poste a une importance plus élevée chez les personnes âgées. C'est pourquoi, une importance plus élevée est retrouvée pour les plus de 50 ans sur les frais de séjour et les autres soins hospitaliers.

Les frais optionnels ainsi que les frais de transport ont, quant à eux, une importance plus faible chez les étudiants et plus forte chez les retraités.

---

\*Le service médical rendu (SMR) est un critère permettant de classer les médicaments en fonction de leur utilité thérapeutique publié par la Haute Autorité de Santé (HAS).

Le sous-poste Maternité a une importance moindre chez les retraités et plus élevée chez les 18-50 ans. En effet, ce sous-poste particulier concerne les jeunes parents, c'est pourquoi, ces caractéristiques sont retrouvées au niveau de l'importance du sous-poste.

Les frais sur les soins dispensés à l'hôpital ont un poids plus élevé. De plus, étant donné que les bases de données sont formées à plus de 70% d'individus âgés de moins de 50 ans, le sous-poste Maternité a un poids plutôt élevé et les frais optionnels et les frais de transport ont un poids moins élevé.

#### Les postes Audiologie, Cures thermales et Prothèses :

L'audiologie est caractérisée par une importance croissante en fonction de l'âge, ce qui est cohérent avec les analyses faites sur les bases de données.

Les cures thermales ont une importance et un poids supérieur chez les non-cadres et les retraités ainsi que chez les femmes. A nouveau, ces remarques sont en phase avec ce qui a été constaté dans les bases de données.

Les prothèses sont caractérisées par un pic d'importance au niveau des 50-64 ans. Elles ont également une importance et un poids supérieur chez les hommes et inférieur chez les cadres.

En somme, que ce soit pour les postes ou les sous-postes, les importances absolues et relatives varient plus ou moins en fonction des caractéristiques des sondés et résultent en des poids différents. Plusieurs remarques ont pu être formulées à partir des résultats et la plupart coïncide avec les résultats précédemment présentés lors des analyses des bases de données.

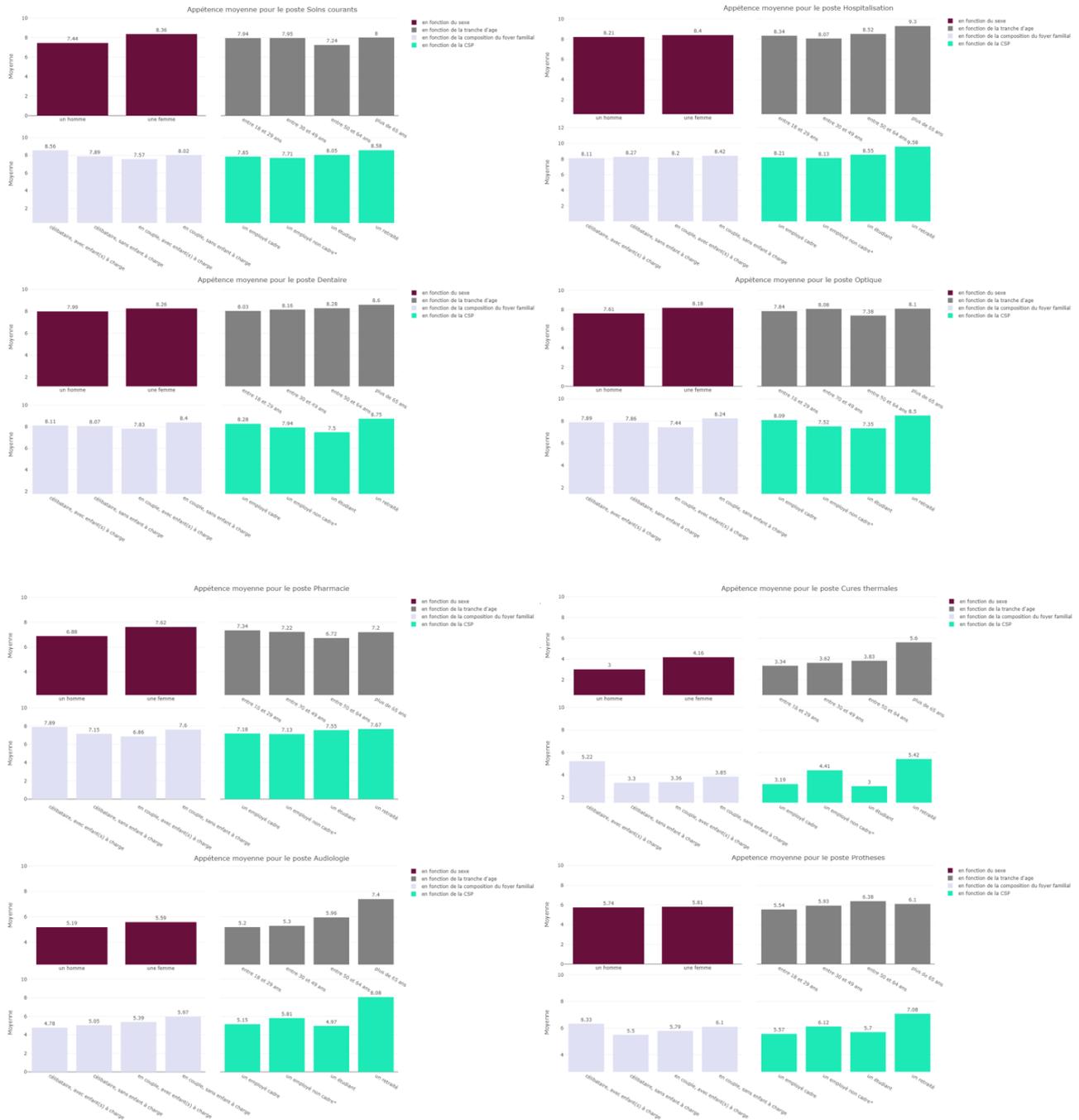


FIGURE A.19: Analyses graphiques des importances absolues des différents postes selon les caractéristiques des sondés.

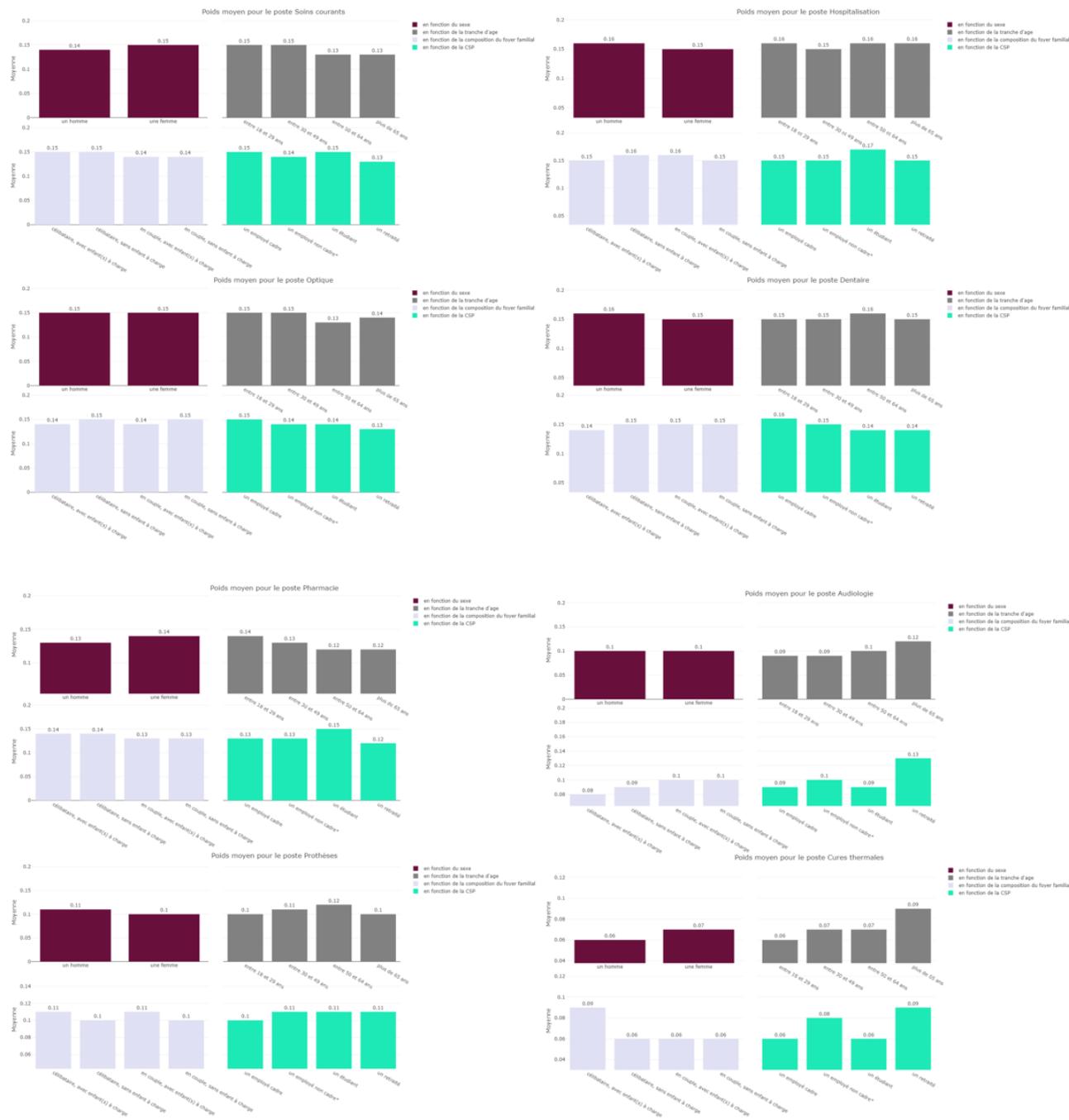


FIGURE A.20: Analyses graphiques des poids des différents postes selon les caractéristiques des sondés.