

Collège des Ingénieurs
215 Boulevard Saint-Germain, 75007 Paris

Mémoire présenté devant le Collège des Ingénieurs pour l'admission à l'Institut des Actuaires

le 23 Juin 2016

Par : **Jean-Charles Samuelian**

**Titre : Création d'une nouvelle assurance
complémentaire santé : définition d'un produit sans
données historiques**

Confidentialité : Oui (2 ans)

Membres présents du jury de l'Institut des
Actuaires:

- Arnaud Cohen
- Florence Picard
- Michel Fromenteau

Membres présents du jury du Collège des
Ingénieurs:

- Frédéric de Courtois
- Eudes Charpentier
- Bruno Boulay

Directeur de mémoire : Michel Béra

Invité : Philippe Mahrer

Signature du candidat :



Collège des Ingénieurs
215 Boulevard Saint-Germain, 75007 Paris

Mémoire d'Actuariat - 2016

Création d'une nouvelle assurance complémentaire santé : définition d'un produit sans données historiques

Jean-Charles Samuelian

Mots-clés : Assurance Complémentaire Santé, Benchmark, Nouveau Produit, Itemisation, Scraping

Résumé

Le marché de l'Assurance Complémentaire Santé a vu sa réglementation évoluer au fil des années mais très peu de nouveaux acteurs innovants sont entrés dans le marché. La création d'une nouvelle Société d'Assurance implique la conformité à d'importantes contraintes réglementaires et la création de nouveaux produits.

La société Alan crée une nouvelle Assurance Complémentaire Santé 100% digitale orientée vers l'expérience utilisateur, la simplicité et la transparence. Comme toute nouvelle société, elle ne dispose pas de données historiques pour construire ses produits d'assurance. C'est l'objet de ce mémoire.

Pour construire un nouveau produit d'assurance complémentaire santé, nous partons d'abord à la chasse à la donnée. Nous construisons des robots appelés "scrapers" qui

vont collecter toutes les offres accessibles sur internet en faisant varier certains paramètres (intrinsèques et extrinsèques).

Une fois ces données collectées, nous les consolidons dans une grille standard et analysons. Nous testons aussi la cohérence des données en mesurant la linéarité de la fonction de prix. Nous nous apercevons que des incohérences subsistent toutefois dans le pricing de certains plans (coûts marginaux négatifs pour des augmentations strictes de remboursements...).

La démarche de ce mémoire repose sur la construction d'un produit de complémentaire santé sur la connaissance du prix marginal de chaque item. Pour cela, il est nécessaire de déconstruire le prix des mutuelles en comparant des offres proches en garanties.

Nous prenons deux hypothèses théoriques : 1. nous avons les données pour tous les plans qui puissent exister, et ceux-ci sont *cohérents*, c'est à dire que pour un plan avec certains paramètres intrinseques, ne correspond qu'un seul et unique prix, 2. le prix de chaque poste doit être indépendant de chaque autre poste.

Il est alors possible d'obtenir le prix de tout *item*, c'est à dire une addition de garantie à un point particulier, en procédant à la soustraction des prix d'un plan à un autre.

Nos données n'étant pas exhaustives, on ne pourra pas trouver, pour tout item, les plans correspondants et procédons donc par l'interpolation cubique monotone d'Hermite. Dans de nombreux cas, nous déduisons la valeur d'une groupe de garanties en faisant des pondérations des différentes garanties en termes de prix sur la base de données de frais réels moyens par acte et par fréquence.

Nous combinons ces solutions avec une approche inter-assureurs. Lorsque l'on compare les prix des différents assureurs, nous réalisons rapidement qu'ils ne sont pas en adéquation pour le même niveau de garantie, ce que nous recalibrons en repartant du panier minimal réglementaire.

Maintenant que nous avons calibré nos pricings, et que nous pouvons reconstruire chaque garantie indépendamment, il est possible de créer des produits originaux en additionnant le prix de chaque garantie. Trois étapes sont nécessaires: décomposer le produit en chacun de ses éléments, donner un prix à ces éléments, additionner les prix. Dans le cadre de ce mémoire et de la création de la Société Alan, nous appliquons notre solution à des produits utilisés par des startups et entreprises réelles qui nous ont partagé leurs plans d'assurance santé.

Nous comparons aussi notre méthode avec une méthode *bottom-up*. Pour se faire, nous avons donc fait appel à un cabinet de conseil en actuariat possédant un tarificateur basé sur la sinistralité. La comparaison du produit et de son prix nous offre la réalisation que les deux prix sont très similaires, et ainsi que la méthode développée dans ce mémoire est solide.

Néanmoins, la méthode présentée dans ce mémoire présente plusieurs limites que nous avons identifiées au cours de la construction du Produit d'Assurance Complémentaire Santé : notre comparaison est basée sur des nouveaux contrats alors que les contrats évoluent dans le temps entre les entreprises et leur assureur, les différents acteurs n'ont pas forcément la même stratégie pouvant rendre les comparaisons hasardeuses, enfin il est possible que les garanties ne soient pas entièrement indépendantes entre elles.

En conclusion, ce mémoire a été à la fois un exercice théorique et l'application concrète de la définition de notre nouveau produit d'assurance complémentaire santé. Il sera maintenant important de collecter notre propre historique de sinistralité pour le comparer à notre base d'utilisateurs.

Collège des Ingénieurs
215 Boulevard Saint-Germain, 75007 Paris

Actuarial Thesis - 2016

New health insurance: create an insurance product with no historical data

Jean-Charles Samuelian

Keywords: Health Insurance, Benchmark, New Insurance Product, Itemisation, Scraping

Abstract

The complementary health insurance market has seen its regulation evolve in the past 10 years, but there has been a surprisingly low number of innovative new actors entering the market.

The creation of a brand new insurance company implies conforming to stringent regulatory constraint, and creating new insurance products.

The company Alan creates a fully digital new complementary health insurance, with a focus on user experience, simplicity, and transparency. Like any new company, it doesn't possess any historical data to build its insurance products.

To build a new complementary health insurance product, we first gather some data, by building bots called *scrapers* that collect all the product offerings on the internet, while varying parameters (intrinsic and extrinsic).

Once these datapoints collected, we consolidate them in a standard grid, and analyse them. We also test the consistency of the data by measuring the linearity of the price function. We see some inconsistencies live through in some plans (net negative marginal costs for reimbursement increases).

This thesis relies on building a health insurance product on the knowledge of the marginal price of each item. For this, it's necessary to deconstruct the price of providers by comparing plans close in terms of guaranties.

We take two theoretical hypotheses: 1. We have the data for all the possible plans, and they are consistent, meaning that they only have one price per set of intrinsic parameters, 2. The price of each guaranty is independent of all other guaranties.

Then we can obtain the price of any item, meaning an additional coverage by subtracting the price of a plan to another.

Data not being exhaustive, we wouldn't be able to find for all items the corresponding plan, and therefore we are proceeding with Hermite's monotone cubic interpolation. In many cases, we deduct the value of a group of guaranties by weighting the prices on the basis of real cost by act and by frequency.

We combine these solutions with an inter-provider approach. When we compare the prices of the different providers, we quickly realize that they are not equated for the same coverage level. We recalibrate by going back to the regulatory minimum coverage basket.

Once we have calibrated our pricings, and that we can build each guaranty independently, it's possible to create original products by summing all of them. In the case of the creation of the company Alan, we apply our solution to products used by startups and companies who have shared their plans.

We also compare our method with a bottom-up approach. For this, we solicited a actuary consulting firm possessing a pricer based on historical claims. The comparison confirms that our prices are similar and that our approach is solid.

However, the method presented comports limits we identified in the building of the product: our comparison is based on new contracts' prices, although these evolve with time, the actors don't necessarily have the same strategy, making comparisons unwise, and it could be that guaranties are not completely independent.

To conclude, this thesis is at the same time a theoretical exercise and the concrete application of the definition of a new complementary health insurance product. It will be important to collect our own historical claims data to compare our model to our user base.

Remerciements :

Je tiens à remercier très sincèrement Florence Picard, Bruno Boulay et Philippe Mahrer de m'avoir donné l'opportunité de travailler sur ce sujet très lié à la nouvelle société que nous construisons.

Je remercie aussi vivement Michel Fromenteau et Michel Béra pour leur soutien en amont et tout au long de l'étude ainsi que pour tous les échanges très inspirants sur comment innover dans le milieu de l'assurance.

Enfin, j'exprime toute ma gratitude à l'équipe d'Alan avec qui c'est une immense joie et un honneur de créer l'Assurance Santé de demain et en particulier Charles Gorintin mon ami et cofondateur.

Sommaire :

<u>Introduction</u>	14
<u>1. Le Marché de l'Assurance Complémentaire Santé</u>	16
<u>1.1. Aperçu du Marché de l'Assurance Santé</u>	16
<u>1.2. Les différents types d'Acteurs</u>	17
<u>1.3. La structure de coût des acteurs</u>	22
<u>Les sinistres</u>	22
<u>Les frais de gestion des sinistres</u>	22
<u>Les frais d'acquisition</u>	23
<u>Les frais d'administration</u>	25
<u>Analyse du résultat</u>	25
<u>1.4. Evolution de la réglementation</u>	27
<u>Solvabilité II</u>	27
<u>Réglementation spécifique à la Santé en France</u>	34
<u>1.5. Contrats collectifs et individuels</u>	36
<u>1.6. Connexion à la Sécurité Sociale</u>	38
<u>1.7. La protection des données</u>	39
<u>1.8. Présentation d'Alan et de la vision de la Société</u>	41
<u>2. Collection des données</u>	44
<u>2.1 Introduction sur les données</u>	44
<u>2.2 Sources de données</u>	44
<u>2.3 Types de données</u>	45

<u>2.4 Difficulté d'avoir des sources disparates</u>	47
<u>2.5 Comment récupérer les données (Python + Selenium + Scrapy)</u>	51
<u>2.6 Blocage de l'IP</u>	54
<u>3. Analyse des données collectées</u>	57
<u>3.1 Des milliers de plans récupérés, des dizaines de fournisseurs</u>	57
<u>3.2 Dispersion des prix</u>	57
<u>3.3 Nomenclature des prix CNM</u>	58
<u>3.4 Constatations</u>	62
<u>4. Itemisation</u>	65
<u>4.1 Cas théorique</u>	65
<u>4.2 Interpolation monotone cubique</u>	67
<u>4.3 Cas des multiples vecteurs</u>	70
<u>4.4 Application</u>	71
<u>5. Complétion des nappes de prix par interpolation inter-assureur</u>	75
<u>5.1 Description du procédé</u>	75
<u>5.2 Application</u>	77
<u>6. Construction de produits originaux</u>	81
<u>6.1 Description du procédé</u>	81
<u>6.2 Benchmarking: Source de Données</u>	81
<u>6.3 Second benchmarking</u>	85
<u>6.4 Application à Alan</u>	87
<u>7. Limites</u>	91
<u>Conclusion</u>	94

<u>Bibliographie</u>	96
<u>Annexes</u>	98

Introduction

Le marché de l'Assurance Complémentaire Santé a vu sa réglementation évoluer au fil des années mais très peu de nouveaux acteurs innovants sont entrés dans le marché. La création d'une nouvelle Société d'Assurance implique la conformité à d'importantes contraintes réglementaires et à la création de nouveaux produits.

La société Alan répond à l'ambition de redéfinir la manière d'interagir des assurés français avec leur protection santé par la commercialisation et la distribution d'une assurance santé entièrement fondée sur la technologie et Internet via une interface utilisateur personnalisée.

Le marché est caractérisé par un manque de transparence, les utilisateurs ne pouvant efficacement procéder à la comparaison des offres et ne les comprenant pas de par la manière dont elles sont exprimées.

Ce manque de transparence du marché s'accompagne de processus de souscription complexes et lourds impliquant la mise en œuvre de nombreuses étapes. Il est par conséquent extrêmement difficile pour un utilisateur particulier ou même un employeur de mesurer les coûts réels supportés et l'efficacité de leur couverture santé.

Ces problématiques constituent un enjeu d'autant plus important depuis la récente modification de la législation française imposant aux employeurs d'assurer la couverture santé de leurs salariés à compter de 2016 (Loi relative à la sécurisation de l'emploi du 14 juin 2013 transposant les principales mesures de l'accord national interprofessionnel du 11 janvier 2013).

C'est dans ce contexte que se pose la question de la création d'une nouvelle assurance

complémentaire santé. **Pour ce lancement, il est nécessaire de définir un nouveau produit d'Assurance qui doit combiner simplicité et transparence avec compétitivité et contraintes réglementaires, dans un contexte de start-up qui n'a pas de données historiques accessibles.** C'est le problème de ce mémoire.

Ce mémoire est organisé de la manière suivante. Dans un premier temps, nous rappelons le contexte réglementaire et marché dans le cadre desquels s'inscrit la démarche.

Puis, nous présentons les différentes étapes menant à la construction d'un nouveau produit. La première étant la construction d'outils qui permettent de collecter les données de base à la conception d'un produit nouveau. Une fois ces données collectées, on peut alors s'attaquer à leur analyse préliminaire et à leur indexation afin d'en tirer de premières conclusions et de préparer les étapes suivantes pouvant mener à la constitution d'un produit d'assurance complémentaire santé.

Nous présentons ensuite les modèles d'itémisation, et les cas théoriques associés, ainsi que de complétion des nappes de prix par interpolation. Ces modèles permettent de déterminer le coût marginal de chacune des lignes du produit d'assurance, et ainsi nous amènent à la partie suivante où nous reconstruisons notre propre produit d'assurance sur ces bases.

Enfin, nous confrontons ces résultats à des données de sinistralité obtenues à l'extérieur afin de confirmer nos modèles de construction dont nous énonçons les limites.

1. Le Marché de l'Assurance Complémentaire Santé

1.1. Aperçu du Marché de l'Assurance Santé

Le marché de l'assurance santé représente 33,9 milliards d'euros de cotisation en France en 2014¹ (hors sécurité sociale qui représente près de 200 milliards d'euros).

C'est un marché en croissance continue depuis 2000 : +5% par an en moyenne (euros courant) des cotisations collectées, du fait de la hausse cumulée du nombre d'individus couverts (+1 % par an en moyenne), de l'amélioration des garanties et des prix associés. La Santé représente 15% de l'activité totale du métier d'Assurance.

96% de la population française dispose d'une couverture santé complémentaire². En 2014, le marché était partagé entre trois types d'acteurs, les sociétés d'assurance, les institutions de prévoyance et les mutuelles soit 573 entreprises.

Ces acteurs gèrent des contrats complémentaires à la couverture de la Sécurité Sociale dont l'objet principal est le remboursement à l'assuré de ses soins médicaux : consultations de médecins, hospitalisations, achats de médicaments, optique, dentaire.

Les complémentaires santé financent 13,5% des dépenses de soins et de biens médicaux, contre 76,6% pour la Sécurité Sociale. Le solde étant majoritairement couvert par les foyers (8,5%).

¹ Source: Drees, Etudes & Résultats - Rapport Assurance Complémentaire Santé 2014

² Source: Haut conseil pour l'avenir de l'assurance maladie, La généralisation de la couverture complémentaire en santé (2013)

1.2. Les différents types d'Acteurs

Ce marché de la complémentaire santé est couvert par des organismes à la typologie variée. Il existe trois grands types d'acteurs qui se répartissent en :

- 453 mutuelles,
- 94 sociétés d'assurances, et
- 26 institutions de prévoyance.

Si leur mission est globalement la même : le remboursement de soins médicaux, leur structure juridique et de capital sont différentes. Nous les décrivons succinctement dans l'encadré ci-dessous :

Différences entre mutuelles, sociétés d'assurances et institutions de prévoyance

Mutuelles :

Les mutuelles sont des organismes à but non lucratif qui sont soumises au Code de la mutualité. Leur objet est donné à l'alinéa I de l'Article L111-1 du même Code. Elles sont immatriculées au Registre national des mutuelles.

Il s'agit d'un groupement ayant la capacité civile, le statut de la mutuelle relève du principe de l'autogestion. Elles poursuivent un but non lucratif, qui face au versement d'une cotisation, mènent des actions de prévoyance, de solidarité et d'entraide dans l'intérêt de leurs membres. Elles ne rétribuent aucun actionnaire et se financent majoritairement grâce aux cotisations de leurs membres.

Elles proposent uniquement des garanties d'assurances de personne.

Sociétés d'assurance :

Elles sont majoritairement à caractère commercial et soumises au code des assurances.

L'article L. 322-1 du Code des assurances définit leur forme, soit :

- sociétés anonymes (SA)
- sociétés d'assurances mutuelles (SAM)

Elles respectent des normes prudentielles de capital, provisions, solvabilité, placements financiers.

Les Sociétés Anonymes (SA) sont soumises au droit commun des sociétés commerciales.

Les sociétés d'assurances mutuelles sont à but non lucratif. Ces sociétés fonctionnent sans capital social. Elles sont constituées pour assurer les risques apportés par leurs sociétaires (article L.322-26-1 du Code des assurances).

En vertu du principe de spécialité, une entreprise d'assurance ne peut cumuler une activité d'assurance vie et non vie, sauf dans le cas des assurances maladies.

Institutions de Prévoyance (IP)

Les institutions de prévoyance sont soumises au Code de la Sécurité Sociale ou au code rural.

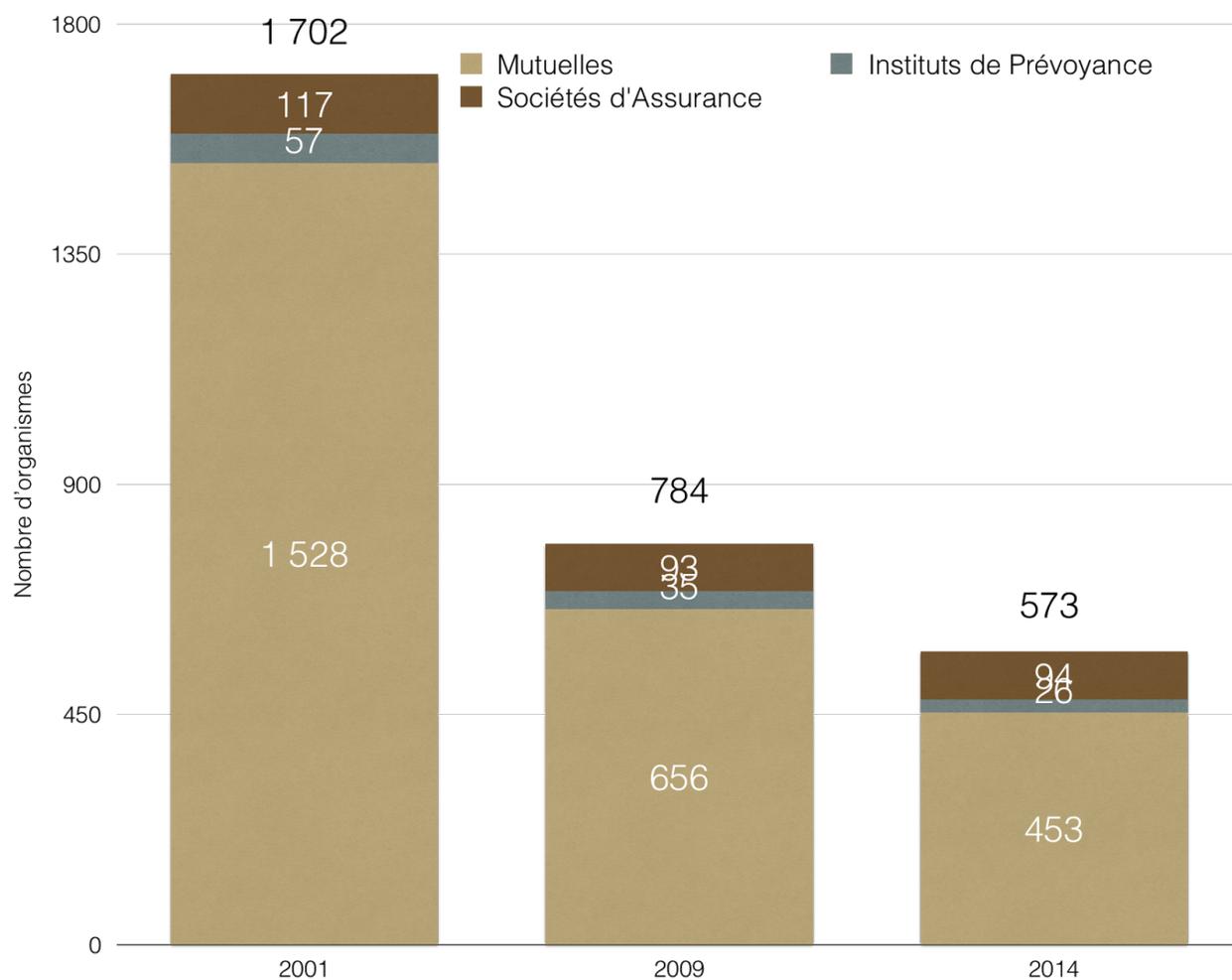
Elles sont positionnées dans la protection complémentaire d'entreprise au niveau prévoyance et retraite. Comme les mutuelles, elles ne peuvent exercer leur activité que dans le domaine des assurances de personnes.

Les IP sont créées par des partenaires sociaux, (syndicats professionnels d'une entreprise ou d'une branche) et sont dirigées par un conseil d'administration composé paritairement par des représentants des employeurs et représentants des salariés.

Les mutuelles sont les acteurs prédominants, malgré une consolidation continue depuis 2001 qui a vu leur nombre divisé par plus de trois. Le nombre d'Institution de Prévoyance a quant à lui été divisé par deux.

Ceci est principalement dû à l'évolution de la réglementation vers un marché unique européen, ayant un impact sur les contraintes réglementaires et la concurrence, ainsi qu'à un nombre historique trop important. La tendance de consolidation devrait continuer dans les années à venir, comme le montrent les différentes fusions en cours.

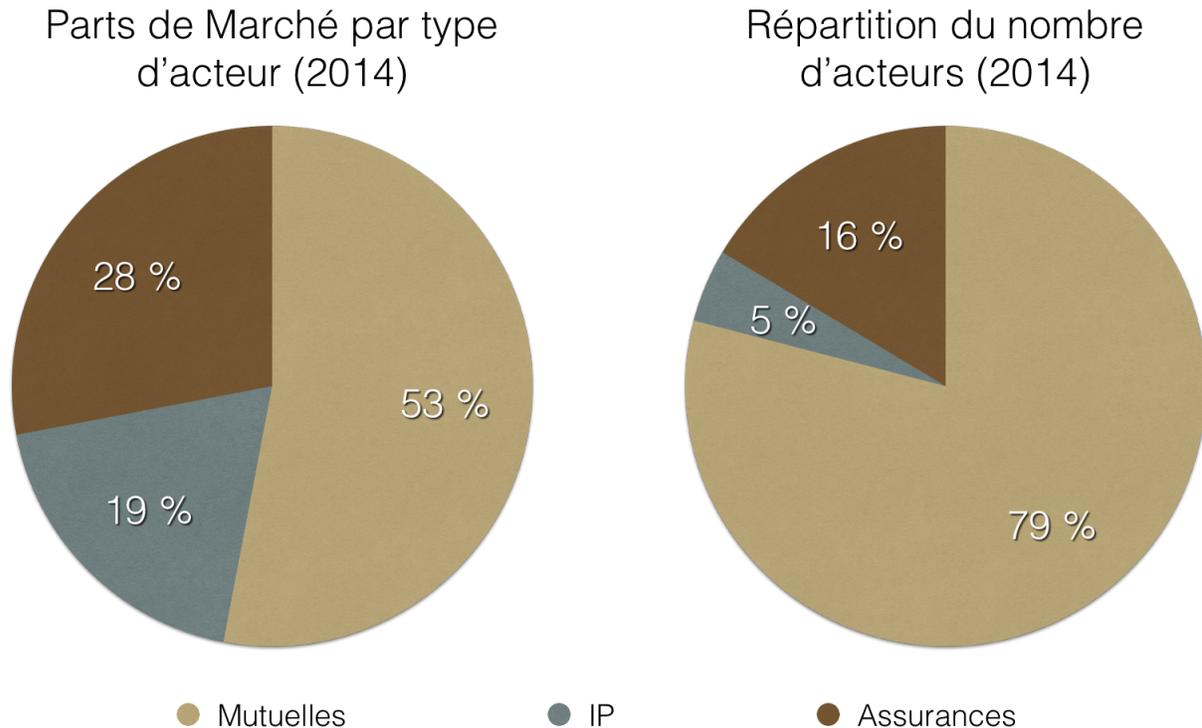
Evolution du nombre d'organismes de Complémentaire Santé depuis 2001



Sources : Drees, *Etudes & Résultats - Rapport Assurance Complémentaire Santé 2014*, *Rapport du Fonds CMU*

Les mutuelles sont clairement les acteurs les plus présents du marché en nombre mais ont proportionnellement moins de part de marché que les sociétés d'assurance, dont les parts de marché ont progressé de 19% en 2001 à 28% en 2014.

Elles ont 53% de part de marché alors qu'elles représentent 79% des organismes. Les IP présentent le meilleur ratio part de marché sur nombre d'acteurs avec seulement 26 organismes pour 19% de part de marché.



Sources : Drees, Etudes & Résultats - Rapport Assurance Complémentaire Santé 2014, Rapport du Fonds CMU

Les mutuelles se concentrent sur l'assurance complémentaire santé (84% de leur chiffre d'affaires). C'est aussi un des axes prioritaires des institutions de prévoyance (47% du chiffre d'affaires). En revanche, l'assurance complémentaire santé ne représente que 5% de l'activité des assurances qui sont beaucoup plus diversifiées.

1.3. La structure de coût des acteurs

Il est intéressant de se pencher sur les structures de coûts des différents acteurs de la complémentaire santé, et de les détailler. Alors que ses recettes consistent principalement des primes de ses clients, il existe quatre grandes catégories de coûts pour le fournisseur d'assurance complémentaire santé :

- Les sinistres
- Les frais de gestion des sinistres
- Les frais d'acquisition des clients
- Les frais d'administration

Les sinistres

Ce sont les remboursements que la société opère lorsque le souscripteur est sujet à un sinistre remboursé par les garanties du produit souscrit. Dans le cas présent, les frais de santé.

Ceux-ci constituent généralement la plus grande part des coûts pour la société, et pour cause, puisqu'ils sont l'objet de son activité.

Les frais de gestion des sinistres

Afin de pouvoir rembourser les sinistres, de recueillir les primes, et opérer un centre d'appel pour la gestion, la société d'assurance porte des coûts. Ils sont fréquemment

sous-traités en externe chez des sociétés spécialisées. Ces coûts sont les frais de gestion des sinistres.

Ils peuvent être exprimés de différentes manières lors des négociations avec un prestataire. Chaque façon de les exprimer comporte ses avantages.

La première manière d'expression est en terme de pourcentage des primes collectées, ce qui est rassurant souvent pour un assureur, mais qui n'a pas beaucoup de sens pour les produits haut de gamme, le coût de revient de la gestion étant souvent similaire entre un panier minimal de soins, et une couverture haut de gamme, malgré un différentiel de prix allant jusqu'à 5 fois plus cher.

La seconde façon de l'exprimer est en terme de prix par adhérent, mais encore une fois, ça ne correspond pas exactement au coût, à cause des conjoints et familles, bien que ce soit utile si l'on calcule les primes en termes d'adhérent.

Enfin, la troisième façon standard d'exprimer ces frais est en facturant par personne bénéficiaire. C'est la manière de facturation qu'a choisie Alan. Elle est intéressante car elle permet à la société de compter par personne qui compte pour les primes un chargement individuel, ce qui simplifie les projections.

Les frais d'acquisition

La société démarre sans souscripteurs. Ce qui veut dire qu'il faut qu'elle les acquière. Les frais d'acquisition de souscripteurs couvrent cette dépense. Elle peut se représenter sous plusieurs formes (y compris des frais internes pour l'acquisition comme ceux d'émission d'un contrat). Un exemple est sous la forme de commissions à payer à un courtier qui distribue les produits de la société. Le recours à un courtier est onéreux, mais il peut être bénéfique car il donne l'accès à une forte base de

souscripteurs potentiels. Un des avantages des courtiers est souvent qu'ils peuvent offrir des contrats d'assurance dans un domaine différent de celui dans lequel des souscripteurs possèdent déjà un contrat. Dans le cadre d'Alan, nous privilégions la distribution directe, ce qui veut dire que nous nous passons de courtiers en assurance.

Un autre exemple des frais qui peuvent compter comme frais d'acquisition sont les campagnes sur les réseaux sociaux, comme Twitter, Facebook, ou bien les campagnes AdWords avec Google. Le coût de mots-clés étant très élevé dans le cadre des assurances, cela n'est parfois pas le meilleur moyen d'obtenir des souscripteurs.

On peut aussi penser à des campagnes de publicité plus classiques dans la presse, à la télévision, au cinéma, ou sur des espaces aménagés (dans le métro, sur les aribus...).

Encore un exemple est le coût d'une équipe commerciale. Une équipe commerciale pourrait contacter les clients prospectifs et les solliciter pour qu'ils achètent les polices d'assurance. Le coût d'une telle équipe est aussi très élevé.

On peut aussi y compter les coûts de maintien d'un réseau de distribution par agences, qui est une force locale, principalement financée par des commissions sur primes ou encours. Ce modèle peut être remis en question avec l'apparition de nouvelles méthodes de distribution par internet.

Dans le cadre d'Alan, nous misons sur la supériorité des services offerts, ainsi que la forte probabilité de distribution organique par la presse. Cela nous permet de garder des frais d'acquisition très bas.

Les coûts d'acquisition peuvent par ailleurs être maîtrisés avec une intelligente utilisation de techniques de pilotage, qui permettent de cibler plus précisément les futurs souscripteurs.

Les frais d'administration

Les frais d'administration sont le coût restant de l'administration de la société. Il comporte en général les coûts qui ne sont pas comptabilisés dans les cas précédents. On peut y inclure les coûts du personnel d'administration, mais aussi souvent les coûts des systèmes d'information, ainsi que la gestion classique des contrats d'assurance (modification du contrat, changement d'adresses, envoi de la lettre annuelle...). Les équipes d'actuariat font aussi partie de ce coût.

Dans le cas d'Alan, ces coûts sont élevés proportionnellement en début d'exercice, car les excellents ingénieurs coûtent cher, mais ce coût est amorti rapidement.

Analyse du résultat

Le résultat net moyen sur l'ensemble des organismes est de 3,1% dans l'assurance complémentaire santé en 2014, les sociétés d'assurance enregistrant une marge nette moyenne de 3,7%.

Dans le marché, les marges sont supérieures pour les contrats individuels par rapport aux contrats collectifs. Ceci est dû à un historique d'évolution du marché que nous décrivons dans la partie suivante.

Les charges liées à l'acquisition d'un client représentent une part importante des ventes (jusqu'à 13%), les charges administratives 7%, la gestion des requêtes clients 4% et le coût de la sinistralité 80% en moyenne.

in %	Sociétés d'Assurance	Institutions de prévoyance	Mutuelles	Total
Sinistres	73 %	88 %	80 %	80 %
Frais de gestion des sinistres	5 %	5 %	4 %	4 %
Frais d'acquisition	13 %	5 %	6 %	8 %
Frais d'administration	6 %	5 %	9 %	7 %
Résultat technique	3,5%	-2,9%	0,6%	0,8%
Résultat net	3,7%	1,3%	1,3%	3,1%

Source: Drees, Etudes & Résultats - Rapport Assurance Complémentaire Santé 2014

Ce tableau est évidemment à analyser avec précaution car se base sur des données statistiques agrégées qui sont parfois peu représentatives, principalement au niveau des frais de gestion qui peuvent être noyés dans les sinistres. C'est pour cela que nous reconstruisons la structure de frais sur la base de notre business plan et non de moyennes marché dans le cadre de la construction d'un nouveau produit.

Le marché de la complémentaire santé dégage donc de faibles excédents, avec les sociétés d'assurance qui sont les plus profitables malgré des coûts de distribution très importants.

1.4. Evolution de la réglementation

Les organismes évoluant dans la complémentaire santé sont régulés. Nous nous concentrons particulièrement sur les SA d'assurance qui sont contrôlées par l'Autorité de contrôle prudentiel et de résolution (ACPR). Des règles similaires s'appliquent aux Mutuelles et aux Instituts de Prévoyance mais s'adaptent à leur structure de capital.

Solvabilité II

La Solvabilité se définit comme l'aptitude, pour une entreprise et plus précisément pour une Société d'Assurance, à faire face à ses engagements. Une entreprise peut être considérée comme insolvable dès lors que la valeur de ses engagements est supérieure à la valeur de son actif.

Historiquement, les Sociétés d'Assurances étaient régulées par la directive Solvabilité I. Sous Solvabilité I, les Sociétés d'Assurance étaient essentiellement tenues de respecter trois contraintes de bilan :

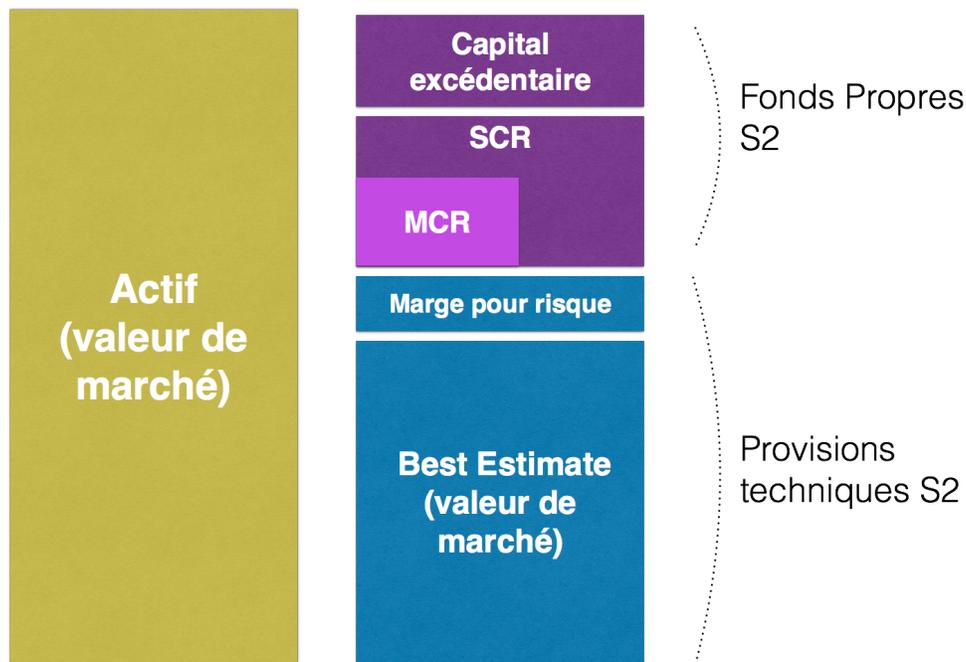
1. Les provisions techniques devaient être prudentes : hypothèses de calcul prudentes
2. Les actifs du bilan devaient permettre le règlement intégral des engagements :
 - a. en particulier, les engagements réglementés (provisions techniques et autres créances prioritaires) devaient être mis en face d'actifs sûrs, liquides, rentables, et
 - b. la base de dispersion (actifs en représentation diminués des créances sur réassureurs) devait satisfaire des exigences de dispersion et de composition du portefeuille

3. Les fonds propres devaient être supérieurs à un minimum appelé exigence de marge

Solvabilité I était simple, les exigences de marges étant facilement calculables et donc auditables. Cependant, l'exigence de marge ne prenait en compte qu'imparfaitement l'étendue des risques des assureurs.

C'est pourquoi la Réglementation Solvabilité II est entrée en vigueur le 1er janvier 2016 après trois ans de transition entre les deux régimes. Elle crée un régime de solvabilité, harmonisé au niveau de l'Union Européenne, qui se veut mieux adapté aux risques encourus par les assureurs.

A la différence du bilan économique de Solvabilité I où les postes du bilan sont évalués prudemment, en accord avec les principes de comptabilité générale français, le bilan Solvabilité II est économique.



Bilan Economique Solvabilité 2

En effet, l'actif est inscrit au bilan à sa valeur de marché et les provisions techniques sont calculées selon une vision "*Best Estimate*", contrairement aux provisions Solvabilité I déterminées à partir d'hypothèses prudentes. La marge pour risque s'ajoute au Best Estimate des passifs, permettant ainsi d'intégrer une marge de prudence dans le calcul des provisions techniques, liés à plusieurs éléments tels que le coût du capital à immobiliser pour faire face à des risques non couvrables.

Cette nouvelle approche Solvabilité est constituée en 3 piliers qui dépassent les exigences seulement quantitatives : le pilier 1 recouvre les exigences quantitatives de capitaux propres, le pilier 2 recouvre les exigences relatives à la gouvernance et au management des risques encourus par les assureurs et le pilier 3 concerne les exigences en matière de reporting et de transparence.

Les nouvelles Sociétés d'Assurance qui sont au coeur de ce mémoire doivent donc se conformer aux exigences issues de la réglementation « Solvabilité II » et résultant de la transposition en droit français (i) de la directive 2009/138/CE du Parlement européen et du Conseil du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice ainsi que (ii) du règlement délégué (UE) 2015/35 de la Commission du 10 octobre 2014 complétant la directive 2009/138/CE (la « Réglementation Solvabilité II »), par l'ordonnance n°2015-378 du 2 avril 2015 et le décret n°2015-513 du 7 mai 2015.

Pilier 1 : Exigences Quantitatives

Le premier Pilier est celui des exigences quantitatives. Avant de nous intéresser aux exigences en capital proprement dites, il est intéressant d'aborder le mode de comptabilisation et d'évaluation du passif.

L'idée générale est d'avoir une approche économique du bilan, ou comptabilisation "market consistent" des différents postes, contrairement à une approche de valeur historique. Cette approche est proche de la réglementation bancaire Bâle II et s'appuie sur les règles éditées par l'International Accounting Standards Board (IASB).

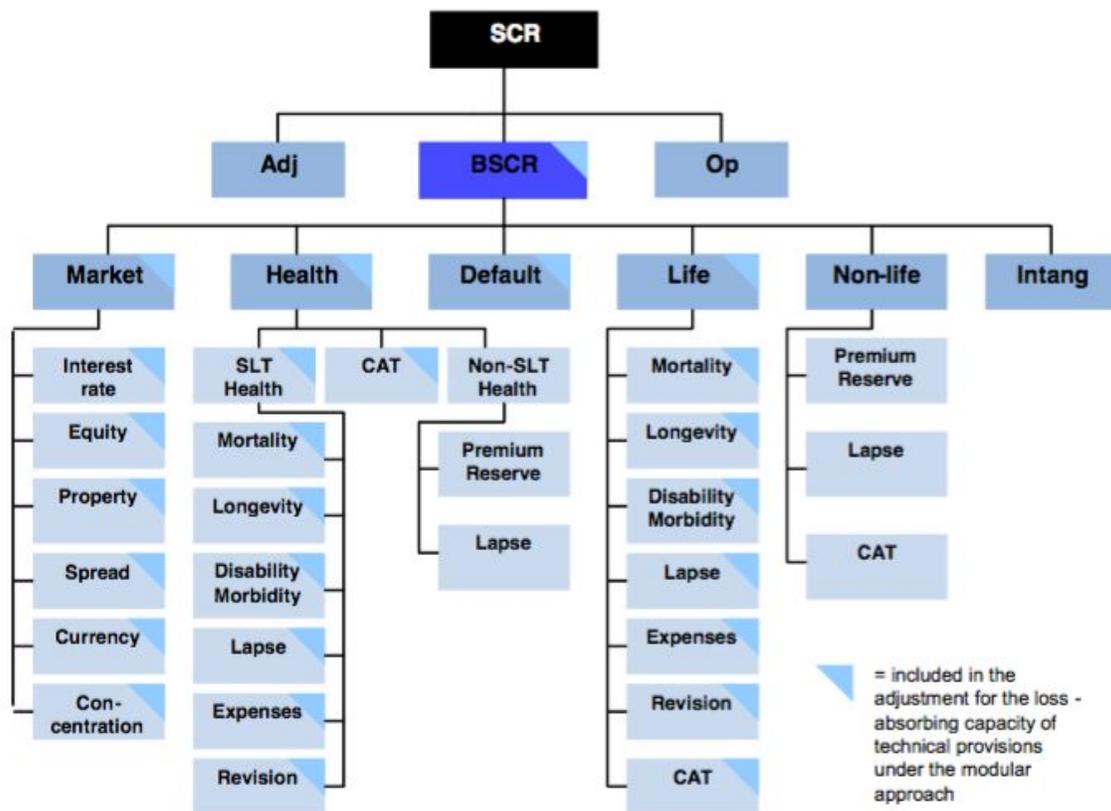
Les provisions techniques doivent être calculées comme l'espérance des engagements futurs actualisés, augmentée d'une marge pour risque.

Les exigences en capital sont déterminées par deux seuils :

- le MCR (Minimum Capital Requirement) représente le seuil en-deça duquel l'assureur se voit retirer son agrément

- le SCR (Solvency Capital Requirement) est le montant minimal de fonds propres en-deça duquel il encourt une intervention du régulateur (rappel à l'ordre et / ou sanctions).

Le SCR peut être déterminé par une formule dite “standard” qui est particulièrement pertinente dans le cas d’une nouvelle Société d’Assurance.



Structure Générale de la Formule Standard

Source : Préparation à Solvabilité II - Traduction partielle du document de l’EIOPA sur les hypothèses sous-jacentes à la formule standard pour le calcul du SCR, ACPR, août 2014

Le SCR est déterminé de sorte que la probabilité de ruine de la compagnie (l'événement de ruine correspondant à une valeur de marché des fonds propres négative), à horizon d'un an, soit inférieure ou égale à 0,5%. Ce chiffre correspond statistiquement à un événement de ruine d'une société tous les 200 ans, en supposant que les ruines des compagnies soient indépendantes et identiquement distribuées (hypothèse qui pourrait être remise en question).

Les risques que doit couvrir le SCR et qui sont décrits dans la formule Standard sont multiples : marché, souscription, crédit, opérationnel. Nous ne détaillons pas dans ce mémoire le calcul de ces différents éléments du SCR qui sont pourtant aussi nécessaires à la création d'une nouvelle assurance. Nous nous concentrerons ici sur la création d'un nouveau produit d'assurance dont les caractéristiques et la sinistralité auront un impact sur le SCR Santé.

Le MCR représente 25 à 45 % du SCR.

Pilier 2 : exigences qualitatives

Le deuxième pilier de la directive Solvabilité II concerne la gouvernance de la Société d'Assurance et sa surveillance par l'ACPR.

Les conditions de gouvernance sont définies dans les articles 41 à 48 de la directive :

- Exigences en termes d'honorabilité et de compétence des dirigeants ;
- La Société d'Assurance doit avoir un service de gestion des risques ;

- La Société d'Assurance doit conduire régulièrement une analyse du risque et de solvabilité, de sorte que ces évaluations soient intégrées dans les décisions stratégiques et communicables sur demande à l'ACPR ;
- La Société d'Assurance doit avoir un service de contrôle interne ;
- La Société d'Assurance doit avoir un service d'audit interne ;
- La Société d'Assurance doit avoir une fonction actuarielle ;
- La sous-traitance doit être gérée de façon appropriée.

Ces conditions sont surveillées par l'autorité de contrôle, l'ACPR en France, qui a les pouvoirs de mettre en place les contrôles nécessaires et de mener des stress tests.

Pilier 3 : Reporting prudentiel et communication

Le reporting prudentiel, à destination de l'ACPR en France, ou de chaque superviseur national, est encadré.

Cette partie de la directive prévoit aussi la publication annuelle d'un rapport dit de solvabilité et de situation financière à l'attention du public. En cas de non-respect des ratios de SCR ou de MCR, des dispositions particulières sont à prendre en compte dans la communication.

L'ensemble des acteurs européens doit renseigner les mêmes modèles, permettant une meilleure transparence du marché et une comparabilité simplifiée entre les acteurs.

Ces informations doivent être fournies périodiquement par le biais de deux types de document, à transmettre par l'assureur :

- Des rapports qui donnent une vision descriptive de la politique prudentielle mise en place par l'assureur : le SFCR (Solvency Financial and Condition Report) qui est public, et le RSR (Regulator Supervisory Report) uniquement dédié au superviseur et plus complet
- Des reportings quantitatifs appelés QRT qui sont des tableaux de bord définis par l'Autorité européenne des assurances et des pensions professionnelles (AEAPP ou EIOPA). Ces reportings couvrent les principaux domaines d'activité d'un assureur : Gestion d'actifs, Provisions techniques, Fonds propres, Bilan, Programme de réassurance...

L'EIOPA publie tous les modèles.

Réglementation spécifique à la Santé en France

En sus des règles européennes pour les sociétés d'assurance, la Santé est un secteur avec des réglementations supplémentaires.

En particulier, l'Assurance Complémentaire Santé est un marché où la réglementation évolue régulièrement. Récemment la législation a évolué quant aux contrats responsables et l'introduction d'une nouvelle obligation pour les entreprises issue de l'Accord National Interprofessionnel (ANI) du 11 janvier 2013. Cet accord a été transposé dans la Loi du 14 juin 2013 relative à la sécurisation de l'emploi.

La qualification de contrat santé en contrat « responsable » est dispositif existant depuis 2004. Il permet aux cotisations versées dans le cadre du contrat santé collectif, de bénéficier d'un traitement fiscal et social avantageux. La contribution du taux de taxe sur les conventions d'assurance (TCA) est réduite à 7% du montant des cotisations

dans le cadre du contrat responsable, contre 14% pour les contrats non responsables. De plus, la contribution de l'employeur bénéficie d'une exonération de charges sociales. Cette exonération est plafonnée à 6 % du plafond annuel de la Sécurité sociale (PASS) et à 1,5 % du salaire brut annuel par salarié ainsi que d'une déduction fiscale pour les salariés.

La législation a fait évoluer et renforce les conditions que doivent remplir les contrats collectifs de santé pour être qualifiés de responsables.

Les contrats complémentaires santé doivent désormais respecter de nouveaux planchers et plafonds de prise en charge sur certaines garanties. Les planchers visent à servir de minima de garantie dans le cadre de la généralisation de la complémentaire santé mise en œuvre à compter du 1er janvier 2016. Les plafonds visent à fixer des limites aux garanties accordées dans le cadre des contrats collectifs et, ce faisant, limiter l'inflation des coûts médicaux qui portent préjudice à l'accès aux soins pour le plus grand nombre.

Avant l'implémentation effective de l'ANI, 76 % des salariés français bénéficiaient d'une complémentaire santé financée, en partie, par leur employeur. A compter du 1er janvier 2016, toutes les entreprises seront obligées de mettre en place une complémentaire santé pour l'ensemble de leurs salariés qu'ils doivent financer à 50% au minimum. Ils peuvent la couvrir jusqu'à 100%.

Les impacts de la généralisation de la complémentaire santé et de l'instauration de ces planchers et limites vont avoir un impact de marché. Nous pouvons estimer que cela affectera le taux d'équipement des TPE/PME, notamment moins équipées que les grandes entreprises avant la mise en œuvre de la réforme.

1.5. Contrats collectifs et individuels

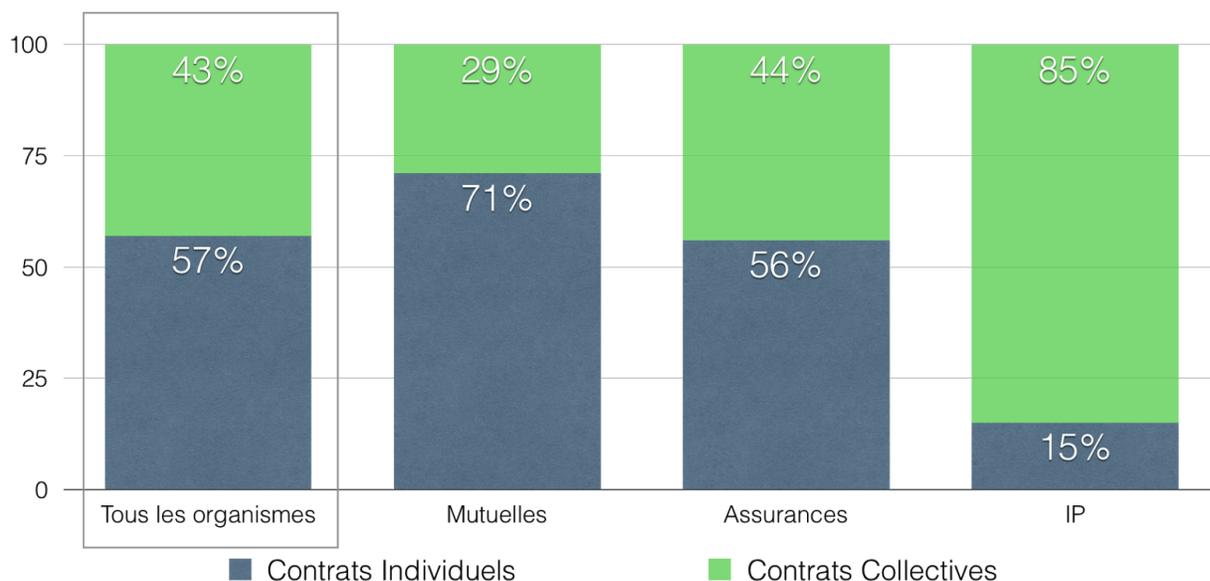
Le marché se divise en deux grands types de contrats : les contrats collectifs et les contrats individuels. Historiquement, ces deux types de contrat ont été gérés de manière différente par les acteurs du marché. Des produits plus standards dans le cas des contrats individuels et plus de personnalisation pour les contrats collectifs, surtout pour les sociétés de taille importante.

Nous pensons que le contrat responsable et l'évolution du marché tend vers une standardisation de ces contrats et une convergence entre individuel et responsable, surtout pour les petites entreprises.

Les contrats collectifs représentent 43 % du marché en 2014. L'assurance individuelle, souscrite directement par les particuliers, représentant le solde.

Historiquement, les institutions de prévoyance commercialisaient essentiellement des contrats collectifs (85 % de leurs cotisations en santé en 2014), tandis que les mutuelles proposaient surtout des contrats individuels (71 %).

Parts contrats collectifs et individuels (2014)



Source : Drees, *Etudes & Résultats - Rapport Assurance Complémentaire Santé 2014*

Ces agrégats sont en train de changer profondément depuis le 1er janvier 2016. Les tendances vont vers un inversement avec plus de 60% de contrats collectifs et moins 40% de contrats individuels.

Dans le cadre de ce mémoire, nous ne différencierons pas particulièrement les deux types de contrat, car nous pensons qu'ils peuvent être approchés de la même manière avec une vision standard du marché, à quelques nuances contractuelles près.

Historiquement les contrats individuels étaient plus "profitables" avec un ratio sinistres sur primes (S/P) de 75%, contre 86% pour les contrats collectifs. Ces derniers étant très souvent personnalisés et négociés par les grands groupes. C'est pourquoi les IP qui travaillent principalement avec les grands groupes reversent en moyenne la plus grande part des cotisations en prestations (88%).

1.6. Connexion à la Sécurité Sociale

Tout nouvel acteur se positionnant sur le marché doit s'interconnecter aux systèmes de la Sécurité Sociale, appelés Noémie. En effet, le rôle de complémentaire santé implique de connaître les flux de remboursements de la Sécurité Sociale pour déterminer quand elle doit rembourser ses souscripteurs.

La télétransmission représente environ 90% des flux de remboursement d'une complémentaire santé et il est donc nécessaire de mettre en place les bons outils de gestion sur ce sujet.

En effet, les processus de gestion des contrats complémentaires de frais de santé sont soumis à des contraintes opérationnelles fortes :

- Il s'agit de processus fortement industrialisés liant des parties prenantes variées : dirigeants et salariés des entreprises clientes, Sécurité Sociale, réseaux de tiers payant et professionnels de santé, organismes bancaires, etc.
- Des économies d'échelles jouent fortement dans l'objectif de minimiser les frais de gestion imputés sur les contrats d'assurance
- Des acteurs spécialisés se positionnent de longue date sur ce marché comme "délégataires de gestion" avec l'ensemble des technologies et des compétences requises ; ils sont rompus à fonctionner selon des accords de niveaux de services formalisés.

Dans ces conditions, il semble naturel pour un nouvel acteur d'opter pour l'externalisation des services de prestation de gestion des contrats.

Les fonctions gérées par une société d'assurance sont les suivantes :

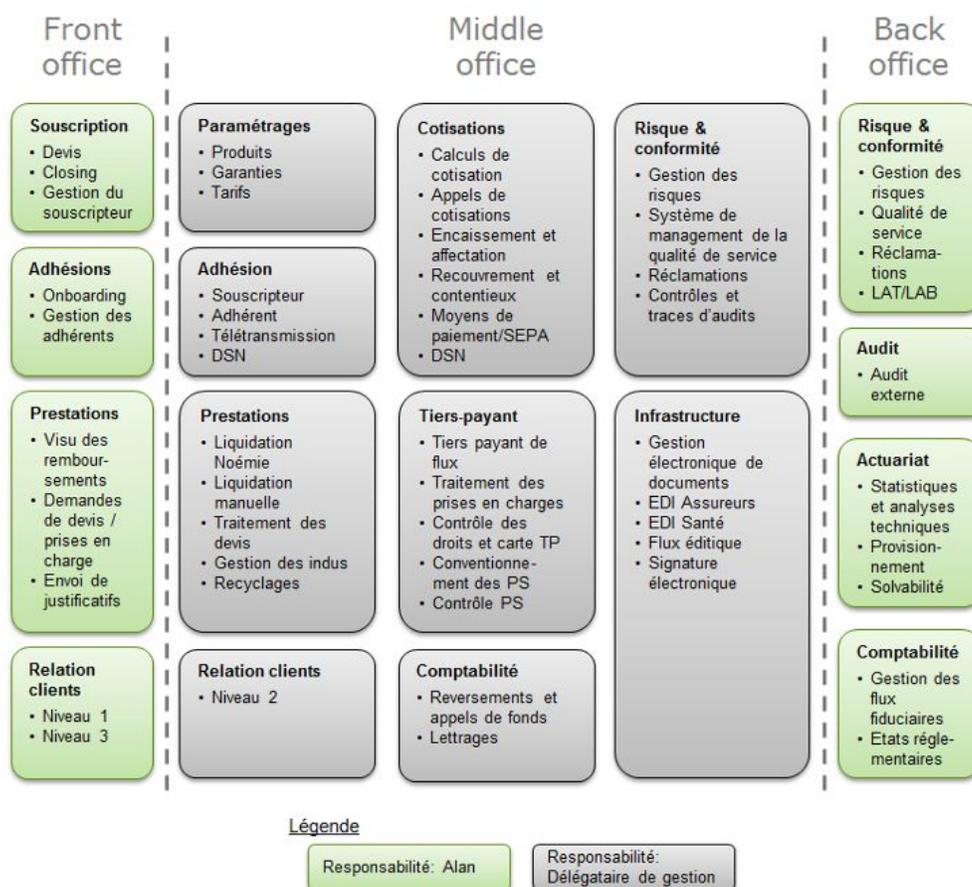


Schéma fonctionnel de responsabilité pour la Société Alan

L'appel à la sous-traitance permet donc une entrée plus rapide de l'Assureur sur le marché et la connexion directe aux outils de la Sécurité Sociale.

1.7. La protection des données

Le métier d'Assurance est réglementé quant à l'usage de ses données et tout nouvel acteur doit être très vigilant quant la mise en place de son infrastructure de données.

Les traitements de données à caractère personnel doivent se conformer aux dispositions légales et réglementaires applicables en matière de protection des données à caractère personnel et en particulier la loi 78-17 du 6 janvier 1978 modifiée

relative à l'Informatique, aux Fichiers et aux Libertés.

En tant qu'organisme d'assurance, les sociétés peuvent se soumettre à l'ensemble des dispositions du Pack de conformité ASSURANCE publié par la CNIL.

Au titre de la conservation des données, et en conformité avec le Pack de conformité ASSURANCE, il faut prévoir des stratégies de conservation des données pour les Clients et les Prospects.

Pour les clients :

- Conservation des dossiers de souscription et d'adhésion des contrats d'assurance Frais de santé : 5 ans à compter de la passation du contrat d'assurance, puis les dossiers et les données qu'ils contiennent sont traités en archivage définitif ;
- Conservation des données produites dans le cadre de la gestion et de l'exécution des contrats d'assurance Frais de santé : 5 ans à compter de leur collecte, puis les données sont traitées en archivage définitif ;
- Dans les deux cas, pour des motifs de sécurité, les données de santé sont conservées 2 ans à compter de leur collecte, puis les données de santé sont traitées en archivage intermédiaire pour une durée de 3 ans, puis traitées en archivage définitif. Les données de santé sont traitées et conservées conformément au Code de bonne conduite annexé à la Convention AERAS.

Pour les prospects : 3 ans à compter de leur collecte, puis suppression.

Au titre du traitement des données personnelles, la Société peut se mettre dans le cadre de différentes normes du pack conformité d'ASSURANCE :

- De la passation, de la gestion et de l'exécution des contrats d'assurance – Norme simplifiée n° 16,
- De la gestion commerciale des clients et des prospects – Norme simplifiée n° 56,
- De la collecte du NIR (Numéro d'inscription au répertoire) et de la consultation du RNIPP (Registre National d'identification des personnes physiques) – Autorisation Unique n° AU 32
- De la lutte contre la fraude – Autorisation unique n° AU 39.

Les Sociétés d'Assurances doivent donc souscrire à un engagement de conformité auprès de la CNIL à l'ensemble de ces normes et autorisations uniques et en particulier mettre en place :

Une information éclairée, préalable et réitérée des assurés au titre du traitement de leurs données personnelles (i) au sein des documents contractuels d'adhésion et de souscription aux garanties et (ii) au sein des espaces digitaux mis à leur disposition par des rubriques dédiées et aisément accessibles.

Pour les données de santé à caractère personnel qui relève de l'article L. 1111-8 du Code de la Santé publique, les Sociétés d'Assurance doivent faire appel au service d'hébergement d'un hébergeur agréé de données de santé (HADS).

1.8. Présentation d'Alan et de la vision de la Société

Ce mémoire s'inscrit dans la constitution d'une nouvelle société d'assurance, Alan, dont Jean-Charles Samuelian est co-fondateur et Président Directeur Général.

La création de la société Alan répond à l'ambition de redéfinir la manière d'interagir des assurés français avec leur protection santé par la commercialisation et la distribution

d'une assurance santé entièrement fondée sur la technologie et Internet via une interface utilisateur personnalisée.

Nous avons constaté des carences dans les connaissances du grand public concernant l'assurance santé principalement sur le rapport qualité/prix des services rendus. Ces carences sont dues principalement au manque de transparence caractérisant le marché de la protection de la santé, les utilisateurs ne pouvant efficacement procéder à la comparaison des offres.

Ce manque de transparence du marché s'accompagne de processus de souscription complexes et lourds impliquant la mise en œuvre de nombreuses étapes. Il est par conséquent extrêmement difficile pour un utilisateur particulier ou même un employeur de mesurer les coûts réels supportés et l'efficacité de leur couverture santé.

Ces problématiques constituent un enjeu d'autant plus important depuis la récente modification de la législation française imposant aux employeurs d'assurer la couverture santé de leurs salariés à compter de 2016 (Loi relative à la sécurisation de l'emploi du 14 juin 2013 transposant les principales mesures de l'accord national interprofessionnel du 11 janvier 2013).

C'est dans ce contexte que la Société se propose d'apporter ses solutions et a sollicité son agrément en tant que société d'assurance auprès de l'Autorité de Contrôle Prudentiel et de Résolution (l'« ACPR »).

Alan offrira une solution de couverture santé innovante et 100% en ligne (souscription via internet uniquement) dont l'interface simplifiée et adaptable sur tout type de support (Web, Android, iOS) permettra de souscrire en quelques minutes sur la base d'un questionnaire souscripteur simple. Une interface de suivi des coûts et de leurs impacts sur les finances des assurés et des employeurs sera également mise à disposition.

La plateforme de la Société a également vocation à proposer dans les mois qui suivront l'agrément de la Société, plusieurs outils pour les particuliers assurés destinés à simplifier leurs démarches parmi lesquels un historique des prescriptions, un outil de prise de rendez-vous avec les médecins, un comparateur de prix, une base de données des symptômes / outils de diagnostic médical ainsi qu'un service de téléconseil.

La Société a une très forte expertise dans le développement de la relation clients "front-end", et des interfaces utilisateur. Elle développe en interne tous les éléments de cette plate-forme qu'elle liera aux outils de gestion des sinistres d'un acteur spécialisé du back-office qui agira en qualité de prestataire.

2. Collection des données

2.1 Introduction sur les données

L'assurance santé pour les entreprises s'est étendue fortement avec les Accords Nationaux Interprofessionnels (ANI), qui nécessitent des entreprises qu'elles fournissent une couverture complémentaire santé pour leurs salariés. Les établissements habilités à fournir cette assurance ont réagi en étoffant leur offre, et notamment en la proposant sur Internet.

On peut notamment trouver sur Internet des comparateurs des prix de différentes mutuelles, instituts de prévoyance ou assurances.

Dans le cas de la société Alan et sa vision de standardisation et de simplification des produits, nous pouvons considérer que les remboursements seront comparables pour des plans individuels et pour des plans collectifs, et nous pourrions donc augmenter notre base de données de certains plans individuels.

2.2 Sources de données

Pour notre problème, nous avons décidé d'utiliser le site assurancepro.com, car c'est l'un des seuls sites permettant d'avoir un devis en temps réel, ce qui est essentiel pour pouvoir récupérer les prix et les plans correspondants.

Ce site est jumelé avec lecomparateurassurance.com, ce qui nous permet de récupérer des prix et plans individuels sur une offre plus grande, avec bien plus d'assureurs.

Nous avons aussi compilé les offres sur les sites lesforets.com, mutuelle.fr, assurland.com, dans des cas individuels.

D'autres sites ont été essayés comme la Mutuelle-conseil.com qui n'avait aucune mutuelle à proposer, ainsi que Mon-comparateur.fr qui n'avait pas de prix disponibles sur internet.

Nous jugeons qu'avec ces sites nous couvrons une large part des offres accessibles sur internet pour la Complétoire Santé.

2.3 Types de données

Les données sont disponibles sur les sites comparateurs, mais elles ne sont disponibles qu'au compte-gouttes, en sélectionnant un choix spécifique de champs correspondant au profil de l'entreprise ou de l'individu, les paramètres que l'on nommera *intrinsèques*. Ces paramètres intrinsèques font varier les prix pour un plan donné. Il existe aussi d'autres paramètres tout aussi importants, qui font alors uniquement varier les plans, mais qui conservent l'unicité de prix pour un plan donné. Nous nous référerons à ces paramètres sous le nom de paramètres *extrinsèques*. Ce sont des paramètres de choix que les futurs souscripteurs peuvent ajuster et qui ne correspondent pas à leur condition personnelle.

Les paramètres intrinsèques que nous avons dans notre cas entreprise sont :

- la convention collective
- le code postal de l'entreprise
- le nombre d'employés
- le nombre de cadres

- l'âge moyen des cadres
- l'âge moyen des non-cadres

Dans notre cas, par simplicité, nous considérerons uniquement les variations d'âge moyen des cadres comme paramètre intrinsèque. La convention collective sera fixée sur Syntec, qui correspond à la plupart des startups, notre cible initiale. Le code postal de l'entreprise sera fixé sur Paris, et le nombre d'employés sera égal au nombre de cadres, avec 10 employés-cadres.

Pour le nombre d'employés, nous avons fait des calculs de sensibilité sur le nombre d'employés cadres, et les offres ne variaient pas en prix pour les entreprises de moins de 50 personnes.

Pour la variation géographique, les différents fournisseurs avaient un coefficient multiplicateur (allant de 0.9 à 0.95) pour les villes hors métropoles.

Nous avons décidé de fixer ces paramètres car ils correspondent le plus à notre cible initiale.

Les paramètres extrinsèques correspondent à la qualité de la couverture offerte par un certain plan sur les quatre postes principaux : soins courants, hospitalisation, optique, et dentaire.

Nous avons une approche exhaustive, il est donc nécessaire et intéressant de récolter les résultats de prix pour chaque combinaison de paramètres.

Un des derniers paramètres est le nombre de personnes à charge par contrat. Pour celui-ci, nous avons décidé de nous baser sur une approche historique macro de la

sinistralité que nous avons obtenu sur des données agrégées d'un portefeuille, et de mesurer le coût supplémentaire d'un bénéficiaire, ou d'un enfant.

Nous avons donc basé nos développements sur des technologies adaptées pour obtenir les prix et caractéristiques de chacun des plans disponibles pour un jeu de paramètres intrinsèques donné.

2.4 Difficulté d'avoir des sources disparates

Il n'existe pas de source ni de code spécifique pour la présentation d'un produit de complémentaire santé : les nomenclatures sont différentes en fonction du prestataire. Cela crée des difficultés pour les souscripteurs prospectifs, pour pouvoir comparer les offres entre elle, d'une manière équilibrée.

Une seconde difficulté réside dans le fait que certaines garanties sont exprimées différemment selon le contrat.

Certaines garanties sont exprimées sous forme de pourcentage de la base de remboursement de la Sécurité Sociale (BR), parfois contenant le remboursement de celle-ci, parfois sans, donc ajoutant un terme de BR auquel se soustrait le ticket modérateur (TM).

Le Ticket Modérateur (TM) est la partie des dépenses de santé qui reste à la charge de l'utilisateur après le remboursement de l'Assurance Maladie, hors prise en compte de l'Assurance Complémentaire Santé. La Complémentaire Santé peut le prendre en charge tout ou partie.

Son taux peut varier en fonction des actes et médicaments, de la situation, et du respect ou non du parcours de soins coordonnés.

D'autres garanties, comme les frais de séjour hospitaliers sont exprimées sous la forme de pourcentage des frais réels (FR). Certaines garanties sont enfin proposées sous la forme de pourcentage du plafond mensuel de la sécurité sociale, ou enfin sous forme de plafond en euros.

Le plafond de la Sécurité sociale (PSS) est un outil de calcul pour certaines cotisations sociales (assurance vieillesse, régimes complémentaires de retraite, ...), certaines prestations de Sécurité sociale (notamment Assedic).

En 2016, le Plafond Mensuel de la Sécurité Sociale est de 3 218 euros.

Il est donc nécessaire de choisir une unique façon de représenter une garantie entre les différentes sources de données. Nous avons décidé de nous baser sur des informations partagées par beaucoup d'acteurs, que nous pouvons retrouver sur certains sites internet de comparaison d'assurance.

Nous avons ainsi choisi de représenter les garanties sous la forme suivante:

SOINS MEDICAUX (SS incluse)	Type de présentation
Consultation Généraliste (médecins adhérents au CAS)	% BR
Consultation Généraliste (médecins non adhérents au CAS)	% BR
Consultation Spécialiste (médecins adhérents au CAS)	% BR
Consultation Spécialiste (médecins non adhérents au CAS)	% BR
Consultation Psychiatrie et assimilé (médecins adhérents au CAS)	% BR
Consultation Psychiatrie et assimilé (médecins non adhérents au CAS)	% BR
Infirmiers / Auxiliaires médicaux	% BR
Analyses	% BR
Radiologie (médecins adhérents au CAS)	% BR

Radiologie (médecins non adhérents au CAS)	% BR
Pharmacie	% BR

HOSPITALISATION (SS incluse)

Frais de séjour / Etablissements conventionnés ou non conventionnés	% BR
Honoraires médecins adhérents au CAS	% BR
Honoraires médecins non adhérents au CAS	% BR
Forfait journalier	FR
Transport	% BR
Chambre particulière	€/jour <i>limitée à 30 jours</i>
Lit accompagnant	€/jour <i>limitée à 30 jours</i>
Maternité / Honoraires d'accouchement (médecins adhérents au CAS)	% BR
Maternité / Honoraires d'accouchement (médecins non adhérents au CAS)	% BR
Maternité / Chambre particulière	€/jour <i>limitée à 30 jours</i>

OPTIQUE (SS incluse)

Monture + 2 verres simples <i>1 paire tous les deux ans pour les adultes sans prescription médicale</i>	€ <i>Dont monture maximum 150€</i>
Monture + 1 verre simple + 1 verre complexe <i>1 paire tous les deux ans pour les adultes sans prescription médicale</i>	€ <i>Dont monture maximum 150€</i>
Monture + 2 verres complexes <i>1 paire tous les deux ans pour les adultes sans prescription médicale</i>	€ <i>Dont monture maximum 150€</i>
Monture + 1 verre simple + 1 verre très complexe <i>1 paire tous les deux ans pour les adultes sans prescription médicale</i>	€ <i>Dont monture maximum 150€</i>
Monture + 1 verre complexe + 1 verre très complexe <i>1 paire tous les deux ans pour les adultes sans prescription médicale</i>	€ <i>Dont monture maximum 150€</i>
Monture + 2 verres très complexes <i>1 paire tous les deux ans pour les adultes sans prescription médicale</i>	€ <i>Dont monture maximum 150€</i>
Lentilles acceptées	% BR+ €/an

Lentilles refusées	€/an
Chirurgie Myopie	€/œil/an

FRAIS DENTAIRES (SS incluse)	
Consultation dentiste	% BR
Orthodontie acceptée	% BR
Prothèses dentaires prises en charge ou non par la Sécurité Sociale (y compris Inlay/Onlay)	% BR
Implant dentaire	€/an
Détartrage bi-annuel	% BR

AUTRES (SS incluse)	
Audioprothèses	% BR + €/an
Cures	% BR
Prime de naissance	€/an
Contraception	€/an
Traitement anti-Tabac	€/an
Visite annuelle du sport	€/an
Vaccins internationaux	€/an
Prothèses	% BR
Appareillage	% BR
Actes de prévention pris en charge par la Sécurité Sociale	% BR

Cette présentation nous permet de normaliser les informations qui sont prises des différentes sources de données et de les traduire en termes similaires afin de pouvoir comparer.

Nous avons fait ce choix précis car il correspond à la logique adoptée dans la plupart des décrets, notamment dans le décret dit de “contrat responsable” (Décret n°

2014-1374 du 18 novembre 2014 relatif au contenu des contrats d'assurance maladie complémentaire bénéficiant d'aides fiscales et sociales).

2.5 Comment récupérer les données (Python + Selenium + Scrapy)

Un *scraper* est un programme informatique permettant de naviguer sur internet automatiquement, et de récupérer des données sur certains sites internet, en appliquant des transformations et en déchiffrant les données reçues des sites rencontrés.

Le programme est codé dans le langage Python, adapté à la manipulation des données. Les sites à *scraper* étant dynamiques en Javascript, il nous était impossible d'utiliser uniquement la technologie de Scrapy³, une librairie qui permet de construire facilement des *scrapers*.

Scrapy est un outil gratuit et open-source de "web crawling", écrit en Python. Il s'agit d'une brique de base du crawling (un crawler ou robot d'indexation est un logiciel qui explore automatiquement le Web afin de l'indexer) qui est utilisé par de nombreuses applications permettant de récupérer des données d'internet.

Scrapy est particulièrement populaire car simple, productif, rapide, extensible, portable, Open Source et robuste.

Il ne fonctionne qu'avec des pages statiques.

Il nous a fallu combiner cette technologie avec le *framework* Selenium⁴, originalement construit pour tester des programmes automatiquement en conditions réelles, mais qui a la capacité de pouvoir ouvrir un navigateur web, dans notre cas Firefox, et d'envoyer des actions comme les ferait un utilisateur réel.

³ <http://scrapy.org/>

⁴ <http://www.seleniumhq.org/>

Selenium est un outil de test automatisé pour les applications web.

Sa grande flexibilité permet d'automatiser les tâches pour les navigateurs web, même dans le cas de pages dynamiques.

Le programme remplit donc en temps réel les champs de paramètres, navigue, puis sélectionne de manière aléatoire les plans qui s'affichent à l'écran. Le programme inclut de ne jamais sélectionner des doublons. A chaque itération, quatre plans sont sélectionnés (la limite du site), et tous les champs sont stockés.

Vous recherchez une assurance

Quelle assurance vous intéresse ?

Civilité Monsieur Madame Mademoiselle

Nom

Prénom

Adresse du siège

Code postal du siège

Ville du siège

Adresse email

Téléphone direct

Si vous le souhaitez, L'AssurancePro.com peut vous faire parvenir par message électronique les meilleures offres de ses partenaires, exclusives et personnalisées. Vous pourrez évidemment vous désabonner à tout moment.

- Oui, je souhaite recevoir des propositions exclusives par email.
- Non, je ne souhaite pas recevoir de propositions.

▶ VALIDEZ

Les informations qui vous sont demandées par PIXFO responsable de traitement sont indispensables pour la prise en compte de votre demande de devis sans lesquelles elle ne pourra pas être traitée. Elles

Première étape de flux d'informations à remplir

Votre société
Vos besoins
90%

Convention collective

Nombre de cadres à couvrir

Nombre de non-cadres à couvrir

Age moyen des cadres

1486: Convention collective nationale applicable au pers

10

0

30

Niveaux de besoin

Les réglettes sont positionnées sur les niveaux de besoins minimum que vous pouvez proposer à vos salariés conformément au panier de soins ANI (Accord National Interprofessionnel). Vous avez bien sûr la possibilité de choisir des niveaux plus élevés si vous le souhaitez.

Soins médicaux

Soins hospitaliers

Soins optiques

Soins dentaires

★
★★
★★★
★★★★
★★★★★

▶ VALIDEZ

La deuxième étape dans le flux d'informations à remplir

Pour le stockage, nous différencions les plans par leurs noms et les paramètres intrinsèques afin d'éliminer les doublons.

2.6 Blocage de l'IP

Notre *scraper* a passé plusieurs jours à récupérer des prix sur la plateforme. Cela représente des milliers d'appels sur le serveur de lassurancepro.com, qui a fini par se rendre compte du fait que notre programme scrapait leur site. Ils ont donc bloqué notre IP, et nous nous sommes retrouvés devant la page suivante:

I'm a teapot



Heureusement, c'est une problématique classique du recouvrement de données sur Internet, et en utilisant un VPN (réseau privé virtuel), nous avons pu continuer notre activité en masquant notre adresse IP.

Un VPN permet d'accéder à des ordinateurs distants comme si l'on était connecté au réseau local.

Un VPN dispose généralement d'une passerelle permettant d'accéder à l'extérieur, ce qui permet de changer l'adresse IP source apparente de ses connexions. Cela rend plus difficile l'identification et la localisation approximative de l'ordinateur émetteur par le fournisseur de service.

Le scraping est une activité légale, c'est la récupération de données publiques et disponibles. Elle est très utilisée sur Internet. Google a par exemple construit son premier moteur de recherche à partir de données scrapées.

Pour se protéger contre certains abus, les sites peuvent parfois bloquer une adresse IP d'où viennent les requêtes qui leur semblent suspectes. Cela est aussi légal, et permet de se défendre contre des activités qui seraient malicieuses, comme une

attaque par déni de service, qui inonde un site de requêtes fausses, si bien que le site ne pourra pas servir les vraies requêtes. Ce n'est pas le cas dans le cadre d'un *scrapeur* qui est un outil classique.

3. Analyse des données collectées

3.1 Des milliers de plans récupérés, des dizaines de fournisseurs

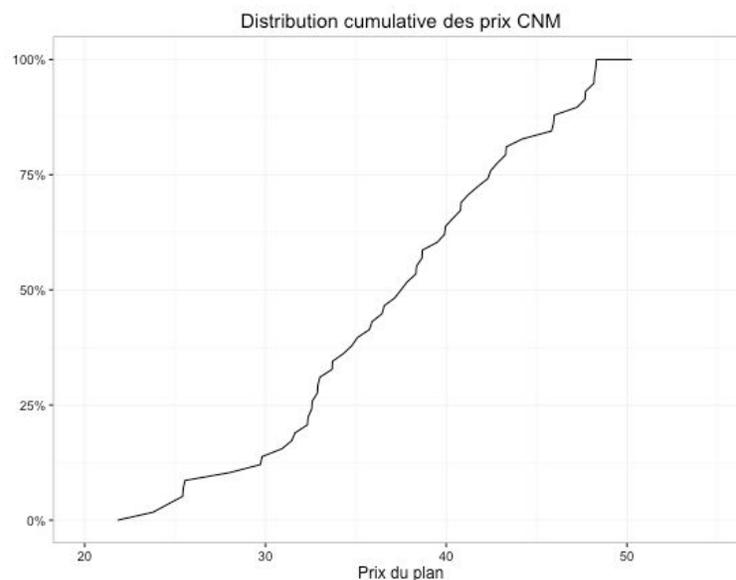
Nous avons pu récupérer plus de 3000 plans pour notre analyse, ce qui nous permet de développer nos méthodes de pricing.

Sur notre outil, nous avons récupéré des plans provenant de dizaines de différentes mutuelles, compagnies d'assurance, courtiers et institutions de prévoyance, LMP, Carrefour, A Comme Assur, AllSecur, Samassur, Ampli, Mutualia, Alptis, Smatis, Miltis, April, CNM, Adrea, Thélem, Mutuelle Générale, Cocoon, et AG2R.

CNM propose des plans très modulables, donc très adaptés à notre exercice. Nous avons pu collecter 58 plans différents de CNM pour des paramètres intrinsèques fixes.

3.2 Dispersion des prix

Afin d'illustrer la distribution des prix en ayant suffisamment de granularité pour une unique compagnie, nous représentons la distribution cumulative des prix pour CNM dans le graphe ci-dessous.



Nous pouvons observer sur le graphique ci-dessus que les prix des plans CNM vont de 25 euros par mois à 47 euros par mois, et que la dispersion est relativement régulière, avec légèrement plus de densité sur les premiers prix, ce qui indique que la gamme est plutôt diversifiée, avec une offre qui nous permettra d’opérer nos calculs.

3.3 Nomenclature des prix CNM

Les sociétés utilisent parfois une nomenclature particulière pour leur plans afin de les indexer et qu’il soit plus simple de les suivre. Les indexations se font de manière différente selon le fournisseur. Ces indices sont intéressants pour notre exercice, car ils permettent de comparer des prix avec moins de variables.

Par exemple, CNM utilise quatre indices, allant de 1 à 6, pour ses offres, 1 correspondant au panier minimal. Chaque index correspond à un poste différent :

- Soins courants
- Hospitalisation

- Dentaire
- Optique

Ainsi, pour 24,59€, le plan 1121 aura la couverture minimale dans quasiment tous les postes, mais une couverture légèrement supérieure en dentaire.

SOINS MEDICAUX (SS incluse)	
Consultation Généraliste (médecins adhérents au CAS)	100% BR
Consultation Généraliste (médecins non adhérents au CAS)	100% BR
Consultation Spécialiste (médecins adhérents au CAS)	100% BR
Consultation Spécialiste (médecins non adhérents au CAS)	100% BR
Consultation Psychiatrie et assimilé (médecins adhérents au CAS)	100% BR
Consultation Psychiatrie et assimilé (médecins non adhérents au CAS)	100% BR
Infirmiers / Auxiliaires médicaux	100% BR
Analyses	100% BR
Radiologie (médecins adhérents au CAS)	100% BR
Radiologie (médecins non adhérents au CAS)	100% BR
Pharmacie	100% BR
HOSPITALISATION (SS incluse)	
Frais de séjour / Etablissements conventionnés ou non conventionnés	100% BR
Honoraires médecins adhérents au CAS	150% BR
Honoraires médecins non adhérents au CAS	125% BR
Forfait journalier	FR
Transport	100% BR

Chambre particulière	40€/jour limitée à 60 jours
Lit accompagnant	40€/jour limitée à 15 jours
Maternité / Honoraires d'accouchement (médecins adhérents au CAS)	150% BR
Maternité / Honoraires d'accouchement (médecins non adhérents au CAS)	125% BR
Maternité / Chambre particulière	40€/jour limitée à 60 jours
OPTIQUE (SS incluse)	
Monture + 2 verres simples 1 paire tous les deux ans pour les adultes sans prescription médicale	100€ Dont monture maximum 150€
Monture + 1 verre simple + 1 verre complexe 1 paire tous les deux ans pour les adultes sans prescription médicale	150€ Dont monture maximum 150€
Monture + 2 verres complexes 1 paire tous les deux ans pour les adultes sans prescription médicale	200€ Dont monture maximum 150€
Monture + 1 verre simple + 1 verre très complexe 1 paire tous les deux ans pour les adultes sans prescription médicale	150€ Dont monture maximum 150€
Monture + 1 verre complexe + 1 verre très complexe 1 paire tous les deux ans pour les adultes sans prescription médicale	200€ Dont monture maximum 150€

Monture + 2 verres très complexes	200€
1 paire tous les deux ans pour les adultes sans prescription médicale	Dont monture maximum 150€
Lentilles acceptées	100% BR+ 100€/an
Lentilles refusées	100€/an
Chirurgie Myopie	100€/œil/an
FRAIS DENTAIRES (SS incluse)	
Consultation dentiste	100% BR
Orthodontie acceptée	125% BR
Prothèses dentaires prises en charge ou non par la Sécurité Sociale (y compris Inlay/Onlay)	125% BR
Implant dentaire	-
Détartrage bi-annuel	100% BR
AUTRES (SS incluse)	
Audioprothèses	100% BR
Cures	100% BR
Prime de naissance	100€/an
Contraception	100€/an
Traitement anti-Tabac	30€/an
Visite annuelle du sport	30€/an
Vaccins internationaux	30€/an
Prothèses	100% BR
Appareillage	100% BR
Actes de prévention pris en charge par la Sécurité Sociale	100% BR

- Tests de cohérence des données, linéarité de la fonction de prix

Afin de s'assurer que les données collectées ont du sens, il est essentiel de pratiquer des tests sur les différents plans collectés afin de s'assurer que la même option aura le même prix dans deux plans différents.

Pour faire ces vérifications, nous pouvons formaliser la proposition ci-dessus (utilisons la notation P pour prix) :

$$P(plan1 + \Delta plan) - P(plan1) = P(plan2 + \Delta plan) - P(plan2) = P(\Delta plan) \quad (1)$$

En d'autres termes, nous devons vérifier que la fonction prix d'un plan est une fonction distributive sur l'opérateur d'addition entre les espaces vectoriels de plans et de prix, c'est-à-dire qu'elle respecte l'opération d'addition (et de soustraction).

Nous pouvons écrire du code nous permettant de vérifier cette supposition. La nomenclature utilisée par CNM est particulièrement pratique dans notre cas, car elle nous permet de comparer des plans de manière équivalente. Il faut par exemple vérifier que la différence entre le plan 1112 - plan 1111 est équivalent à la différence plan 1122 - plan 1121, que l'on pourra noter $\Delta plan_{...1}$. La notation a du sens car elle ne dépend pas des caractéristiques des sous-postes non concernés.

3.4 Constatations

Des incohérences subsistent toutefois dans le pricing de certains plans.

Nous avons pu déceler dans les prix des coûts marginaux négatifs pour des augmentations strictes de remboursements. Ainsi, $\Delta plan_{,3..}$ est négatif, alors que les remboursements sont meilleurs.

Afin d'interpréter la raison pour ce coût marginal négatif, nous avons émis quatre hypothèses:

- Marketing

Parfois, pour obtenir une plus grande part de marché sur l'assurance santé, les providers vendent certaines gammes de produit moins cher en discomptant leurs prix. Il est alors possible de trouver un produit étonnamment peu cher par rapport au reste de la gamme de l'offre, l'opportunité de faire "une bonne affaire" ayant un impact psychologique sur la vente.

- Equilibrage de portefeuille

Les portefeuilles d'assurance sont basés sur la mutualisation du risque. Il est nécessaire d'équilibrer le portefeuille d'assurance en ayant des adhérents de profils différents. En réduisant les prix de certains plans, on peut alors raffermir le nombre d'adhérents sur une gamme de produits.

- Erreurs d'interpolation

Comme nous le verrons dans la suite, il est possible que les interpolations faites par les fournisseurs d'assurance soient non monotones. Dans ce cas, il est possible de trouver des incohérences dans les plans. On explique plus en détail dans la partie interpolation de Hermite comment éviter que les interpolations ne soient pas monotones et créent ces irrégularités.

On dit qu'une interpolation est monotone, si pour une garantie en tous points supérieure, correspond un prix supérieur.

- Négligence

Une autre solution est potentiellement que les plans n'aient tout simplement pas été revus et seraient donc seulement une erreur humaine dans les prix.

4. Itemisation

La démarche de ce mémoire repose sur la construction d'un produit de complémentaire santé sur la connaissance du prix marginal de chaque item. Pour cela, il est nécessaire de déconstruire le prix des mutuelles en comparant des offres proches en garanties.

4.1 Cas théorique

Le cas théorique est relativement simple. Supposons que nous avons les données pour tous les plans qui puissent exister, et qu'ils sont *cohérents*, c'est à dire que pour un plan avec certains paramètres intrinseques, ne correspond qu'un seul et unique prix.

Cela équivaut à avoir un tarificateur parfait, prenant en entrée les garanties, et donnant en sortie un prix.

Il nous faut aussi faire une autre hypothèse importante, qui est vérifiée dans le cas pratique. Le prix de chaque poste doit être indépendant de chaque autre poste. Comme nous l'avons discuté plus haut, il est nécessaire que la fonction de prix soit linéaire.

Si l'on considère que notre produit est un vecteur de dimension $1 \times N$, où N est le nombre de garanties proposées, on peut alors écrire, dans l'espace où les vecteurs de base e_i sont générateurs et que l'on considère que l'on peut définir notre produit

sous la forme $X = \sum_{i=1}^N x_i e_i$:

$$P(X) = P\left(\sum_{i=1}^N x_i e_i\right) = \sum_{i=1}^N P(x_i e_i)$$

Il est alors possible d'obtenir le prix de tout *item*, c'est à dire une addition de garantie à un point particulier, en procédant à la soustraction des prix d'un plan à un autre. Notre item, qui porte sur la garantie i , pourra s'écrire de la façon suivante:

$$\Delta X = X - X' = (x_i - x'_i)e_i \quad (2)$$

Nos données n'étant pas exhaustives, on ne pourra pas trouver, pour tout item, les plans X et X' correspondants. Il est donc nécessaire de procéder à de l'interpolation.

Selon la quantité de données disponible, nous pourrons avoir une capacité de décomposer le prix colinéairement à un vecteur de garantie ou non.

Le premier cas est plus simple: nous pouvons choisir deux plans connus X et X' qui n'ont pour différence qu'un taux de remboursement pour une garantie particulière décrite dans l'équation (2), où notre item $y \in [x_i, x'_i]$.

Dans ce cas, nous pouvons *interpoler* le prix de notre garantie. Les méthodes d'interpolation peuvent être multiples, la plus simple étant linéaire:

$$P(y) \approx (y - x_i)/(x'_i - x_i) (P(x'_i) - P(x_i))$$

La supposition que les prix soient linéaires n'est pas vérifiée en pratique. En effet, une couverture à 100% de la base de remboursement de la sécurité sociale ne vaudra pas la moitié de la couverture à 200% de la base de remboursement de la sécurité sociale (BR). Il est intuitif qu'elle représentera un prix plus élevé que la moitié. On dit alors que l'interpolation est *sous-modulaire*. On peut rapidement prouver cette caractéristique :

$$\forall x \in \mathbb{R}^+, \max(x, 100\%) = 1/2 \times \max(2x, 200\%) \geq 1/2 \times \max(x, 200\%)$$

La première égalité vient du fait que:

$$\forall x, y \in \mathbb{R}^+, \max(x, y) = 1/2 \times \max(2x, 2y)$$

L'inégalité sort de:

$$\forall x, y \in \mathbb{R}^+, \max(2x, y) \geq \max(x, y)$$

Ce qui prouve l'inégalité $\forall x \in \mathbb{R}^+, \max(x, 100\%) \geq 1/2 \times \max(x, 200\%)$.

Il faut donc utiliser une manière d'interpoler qui sera sous-modulaire.

4.2 Interpolation monotone cubique

Un candidat naturel est l'interpolation cubique monotone d'Hermite. Il permet d'être plus précis que l'interpolation linéaire, tout en conservant la monotonie de notre interpolation. La monotonie de l'interpolation est indispensable, car sinon, nous tombons dans l'erreur de CNM, et il devient plus avantageux d'acheter certains produits mieux couvrants, à un prix inférieur.

Nous avons un certain nombre d'observations, qui forment des correspondances $x_k \rightarrow P(x_k) = y_k$, sur un seul poste. Nous ne réécrivons pas l'indice du poste à chaque fois, car cela compliquerait trop la notation. Nous souhaitons recréer la fonction P' extrapolée. La méthode classique est la méthode de Fritsch-Carlson:

1. Calcul de la pente des droites sécantes (les mêmes que l'interpolation linéaire):

$$\Delta_k = \frac{y_{k+1} - y_k}{x_{k+1} - x_k}$$

pour $k = 1, \dots, n - 1$, les points que nous avons observé.

2. Initialisation des tangentes comme la moyenne de la pente des sécantes :

$$m_k = \frac{\Delta_{k-1} + \Delta_k}{2} \text{ si } \Delta_{k-1} \text{ et } \Delta_k \text{ sont de même signe}$$

$$m_k = 0 \text{ sinon}$$

pour $k = 2, \dots, n - 1$. Ces pentes pourront être mises à jour dans les prochaines étapes. Pour les bords, nous utilisons qu'un seul côté :

$$m_1 = \Delta_1 \text{ et } m_n = \Delta_{n-1}$$

3. Pour $k = 1, \dots, n - 1$, si $\Delta_k = 0$ (donc si deux points consécutifs sont à la même hauteur), alors fixons $m_k = m_{k+1} = 0$, puisque l'interpolante doit être plate entre ces deux points. On peut alors ignorer l'étape 4 pour ces k .
4. Soit $\alpha_k = m_k / \Delta_k$ et $\beta_k = m_{k+1} / \Delta_k$. Pour forcer la monotonie, au moins une des conditions suivantes doit être vérifiée :

La fonction
$$\phi(\alpha, \beta) = \alpha - \frac{(2\alpha + \beta - 3)^2}{3(\alpha + \beta - 2)}$$
 doit être strictement positive;

$\alpha + 2\beta - 3 \leq 0$; ou $2\alpha + \beta - 3 \leq 0$.

Une façon simple de forcer la monotonie est de fixer $m_k = \min(\alpha_k, 3)\Delta_k$ et $m_{k+1} = \min(\beta_k, 3)\Delta_k$.

Un seul passage de l'algorithme est nécessaire.

Une fois que cet algorithme a eu lieu, nous pouvons utiliser l'interpolation cubique de Hermite.

Nous avons évalué les sécantes, que nous pouvons alors utiliser dans l'algorithme de construction de l'interpolation d'Hermite.

Nous forçons :

$$p_0 = x_0$$

car cela nous permet d'avoir notre interpolation exacte aux bords.

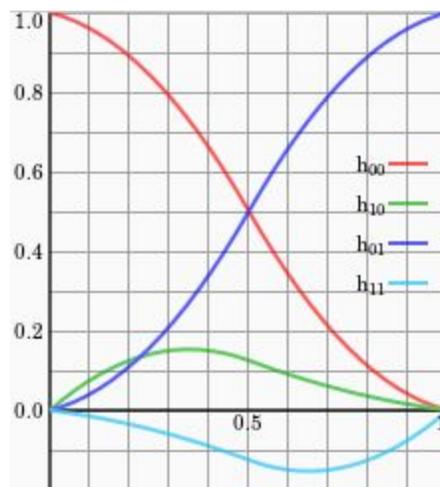
De même, nous utiliserons :

$$p_{n+1} = x_{n+1}$$

Afin de pouvoir utiliser l'interpolation, nous posons :

$$t = (x - x_k) / (x_{k+1} - x_k)$$

Cela nous permet d'être dans le domaine $[0, 1]$, qui est plus simple pour pouvoir utiliser les fonctions d'interpolation d'Hermite.



Les polynomes d'interpolation d'Hermite (source: Wikipedia.fr)

La formule est alors :

$$\mathbf{p}(x) = h_{00}(t)\mathbf{p}_k + h_{10}(t)(x_{k+1} - x_k)\mathbf{m}_k + h_{01}(t)\mathbf{p}_{k+1} + h_{11}(t)(x_{k+1} - x_k)\mathbf{m}_{k+1}.$$

Avec les valeurs h qui sont les fonctions de base des polynomes d'Hermite, définies ci-dessous. Il est intéressant de noter que les valeurs des tangentes ont été remises à l'échelle avec les valeurs $x_{k+1} - x_k$.

	Expression des polynomes de base ⁵
$h_{00}(t)$	$2t^3 - 3t^2 + 1$
$h_{10}(t)$	$t^3 - 2t^2 + t$
$h_{01}(t)$	$-2t^3 + 3t^2$
$h_{11}(t)$	$t^3 - t^2$

Nous pouvons alors utiliser les informations contenues ici afin de pouvoir déterminer les interpolations d'Hermite.

4.3 Cas des multiples vecteurs

⁵ Bartels, R. H.; Beatty, J. C.; and Barsky, B. A. "Hermite and Cubic Spline Interpolation." Ch. 3 in An Introduction to Splines for Use in Computer Graphics and Geometric Modelling. San Francisco, CA: Morgan Kaufmann, pp. 9-17, 1998.

Dans de nombreux cas, il est impossible d'isoler un vecteur unique, car nous ne disposons pas d'informations suffisantes pour pouvoir déduire la valeur d'une garantie unique, et donc colinéaire à un unique vecteur de base dans notre espace.

Il nous faut alors pouvoir composer la valeur de multiples garanties basiques et pouvoir en déduire un prix.

Pour cela, il faut pouvoir pondérer les différentes garanties en termes de prix.

Nous possédons une base de données nous donnant les frais réels moyens pour un bénéficiaire, à la fois par acte, et par fréquence.

En utilisant cette base de données, nous pouvons écrire un modèle de pondération en terme de prix pour les différentes garanties, où chaque poste a un poids ω_i associé à la garantie, afin de pouvoir comparer les différents postes.

Cette pondération vérifie évidemment l'équation suivante:

$$\sum_{i=1}^N \omega_i = 1$$

Posséder cette pondération nous permet de ne pas être bloqué par la résolution de la solution à multiples variables.

Bien qu'introduisant un certain bruit si les distributions de remboursements n'ont pas la même forme, cette solution nous permet d'arriver à un point où nous pouvons obtenir pour chaque garantie une pondération grossière, mais qui est le meilleur choix que l'on puisse faire sans données.

4.4 Application

Afin d'illustrer cette méthode, nous pouvons tout d'abord prendre une catégorie de garanties qui nous est disponible.

L'hospitalisation est un excellent exemple. Nous devons quantifier les différents poids dans les garanties d'hospitalisation, car la différence entre deux plans, comme les plans CNM 1111 et 1211 est basée uniquement sur les soins hospitaliers.

Grâce à notre source de données, nous pouvons utiliser les valeurs suivantes :

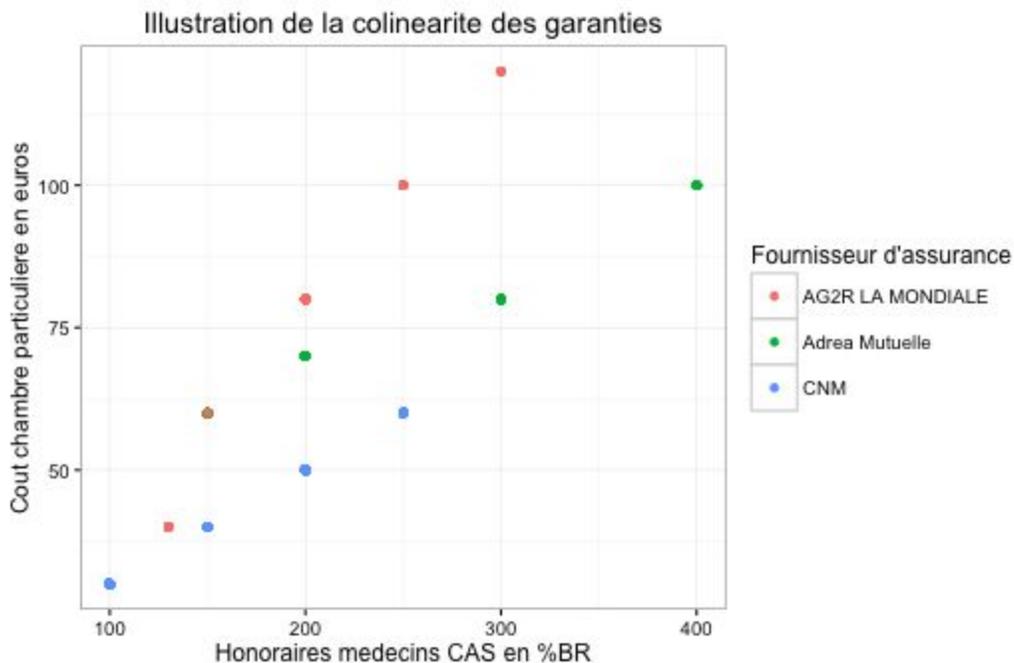
Type d'acte	Poids du prix
Forfait hospitalier	17%
Chambre particulière	20%
Frais de séjour	30%
Honoraires	33%

Ces poids sont calculés en termes de sinistralité par rapport au coût global du secteur de l'hospitalisation. On simplifie grandement le problème en estimant que les distributions sont équivalentes, et donc qu'on peut simplement appliquer à chaque garantie le facteur multiplicateur ci-dessus.

On peut remarquer que les valeurs ne s'ajustent pas forcément directement sur le tableau de garanties que l'on a choisi, ni sur les plans offerts par certains fournisseurs. Dans ce cas, nous utilisons l'expérience pour déterminer quelles valeurs sont importantes et pondérer entre les différents postes. Par exemple, il est possible que pour l'amélioration d'un contrat, un fournisseur ajoute fortement sur un poste, comme les chambres particulières, ce qui coûtera plus cher dans le détail que les autres garanties en Hospitalisation. Il serait alors faux d'extrapoler cet écart de prix comme

venant à 20% (voir tableau ci-dessus) du poste chambre particulière. Pour la simplicité du modèle, nous ignorerons ces cas spéciaux.

Le graphe ci-dessous illustre la colinéarité de l'augmentation des garanties, mais suggère toutefois que les facteurs de colinéarité seraient différents selon le fournisseur d'assurance. Comme nous ne possédons pas les données exhaustives de sinistralité de ces fournisseurs, nous accepterons de ne pas voir varier les pourcentages du tableau plus haut varier avec le fournisseur.



Il est important de réaliser que ces valeurs ne sont de toutes façons pas basées sur l'historique de notre démographie cible, ce qui n'est pas aujourd'hui possible de prédire, et donc que le résultat sera approximatif. Le fait que notre résultat soit approximatif, n'est pas un ennui en soi, car même dans le cas où l'on aurait déjà un historique de données, les hypothèses faites sur le type de sinistralité ne seront pas exactes. Des facteurs comme la distribution, le type de clauses contre l'anti-sélection, sont des facteurs qui peuvent couvrir l'approximation.

Dans le cas CNM, entre les prix 1111 et 1211, pour une entreprise de 10 salariés tous cadres, avec une moyenne d'âge de 20 ans, nous avons ainsi, pour une différence de prix de 0,81 euros:

Une différence du prix de la chambre particulière de 30 à 40 euros par jour. Grâce à notre méthodologie, nous inférons que le coût de ce changement est de 20% du prix du changement de garantie, ce qui veut dire que nous pouvons y mettre un prix à 0,162 euros.

Nous pouvons faire identiquement pour chacun des changements.

L'hypothèse critique ici est que l'évolution des plans se fait de manière globalement uniforme.

5. Complétion des nappes de prix par interpolation inter-assureur

5.1 Description du procédé

Les techniques d'interpolation présentées au-dessus nous permettent de résoudre des questions de sparsité, du caractère épars, de nos données.

Une autre solution que nous pouvons employer est une solution inter-assureurs. En effet, nous avons pu acquérir des données sur les différents plans de plusieurs assureurs à travers notre scraper. Utiliser ces données entre différents assureurs nous permet d'affiner nos modèles.

Lorsque l'on compare les prix des différents assureurs, nous réalisons rapidement qu'ils ne sont pas en adéquation pour le même niveau de garantie. Ainsi, un assureur proposera un produit plus cher que son concurrent, à paramètres intrinsèques et extrinsèques égaux.

Ce phénomène est dû au caractère non fluide du marché de la complémentaire santé, ainsi qu'à de multiples facteurs (structures de coûts différentes, rentabilité visée différente, stratégie d'acquisition de clients...). Le client peut par ailleurs valoriser d'autres éléments, comme la marque.

On ne peut donc pas se reposer sur la loi du prix unique.

La loi du prix unique stipule que dans un marché efficient, chaque bien identique doit avoir le même prix en tout point de ce marché. Cette loi a été énoncée par Gustav Cassel qui dicte le principe de l'unicité du pouvoir d'achat de chaque unité de monnaie.

Cette loi est une suite logique des mécanismes de marchés (d'ajustement du prix) : l'existence de prix de vente ou d'achat différents en plusieurs points du marché pousse les agents économiques à choisir les prix les plus avantageux pour eux, et par ajustement les prix d'un bien convergent vers une valeur unique.

Source: Wikipedia

Ce qu'on peut faire en revanche, c'est adopter un modèle pour la différence de prix entre différents assureurs. Pour cela, considérons pour chaque assureur le panier minimal réglementaire. On pourra le nommer X_0 pour l'assureur X.

Pour n'importe quel assureur Y, nous pourrons ainsi écrire son plan en fonction de l'assureur X :

$$P(Y_0) = P(X_0) + \Delta P_{0,XY}$$

Une fois que l'on a comparé les prix des différents paniers minimaux, il nous faut comparer les prix des paniers avec garanties augmentées.

Nous faisons l'hypothèse que les augmentations de prix sont proportionnelles au coût supplémentaire. Ceci est explicable par le fait qu'un organisme d'assurance raisonne en fonction de pourcentages de charge, en termes de distribution, gestion, administration et marge.

C'est un postulat, le plus simple modèle après la translation pure et simple (les plans d'assurance coûtent tant plus cher que la mutuelle Y). Il est donc nécessaire de prendre un modèle plus complexe, qui est justifié par les chargements.

Ainsi, nous pouvons écrire :

$$P(Y) = P(Y_0) + P(\Delta Y) = P(X_0) + \Delta P_{0,XY} + h_{XY}P(\Delta X)$$

h étant une valeur de transition entre les assurances X et Y. h est spécifiquement le multiplicateur de toute garantie d'une assurance à l'autre. En supposant des taux de chargements 10% plus élevés pour la société Y que pour la société X, on aura de manière plausible :

$$h = 1.1$$

En faisant cette supposition, nous pouvons alors transcrire le prix d'un panier de garanties pour une certaine assurance à notre assurance de référence.

A partir de là, nous pouvons déduire les valeurs h et $\Delta P_{0,XY}$ des données que nous avons.

5.2 Application

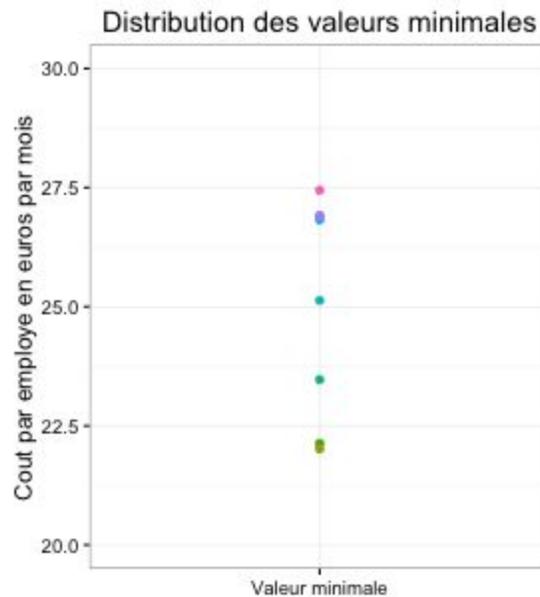
En pratique, nous voyons parfois des différences de prix parmi les données collectées, ce qui ne permet pas toujours d'avoir un résultat exact.

Une bonne métrique à utiliser pour le benchmarking est aujourd'hui le minimum dans le panier ANI, pour une convention collective comme Syntec.

On peut comparer le prix pour chacun des fournisseurs d'assurance dont on a collecté des données. Puisque nous avons le plus de plans collectés sur CNM et grâce à leur plus grande granularité dans les plans, il est intéressant de baser les prix sur ceux de CNM. On suppose alors une moyenne d'âge fixe.

On a alors $P(X_0) = P(CNM_{1111}) = 23,78EUR$.

Pour illustration, on a recueilli des données pour les différents fournisseurs, qui fournissent le panier minimal:



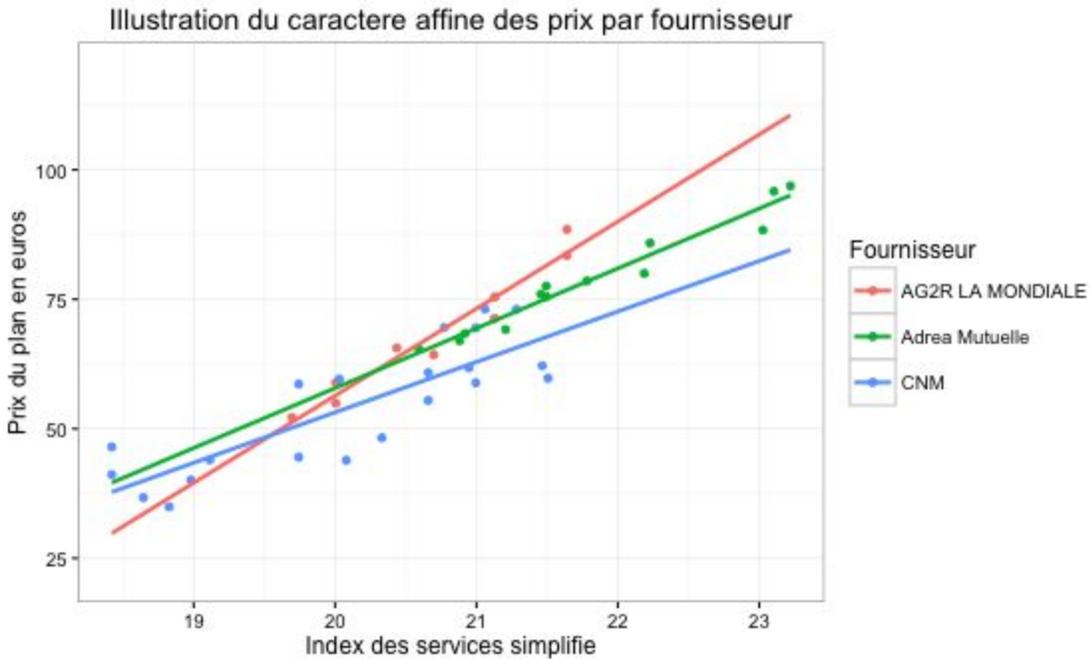
Pour prendre un exemple concret, on peut alors définir la différence grâce à AG2R - La Mondiale dont on a aussi récolté les données, afin d'avoir la valeur suivante :

$$P(Y_0) = P(AG2R_{base\ obligatoire}) = P(X_0) + \Delta P_{0,XY} = 23,75EUR$$

Cela nous permet simplement d'avoir le prix $\Delta P_{0,XY} = -0,03 EUR$.

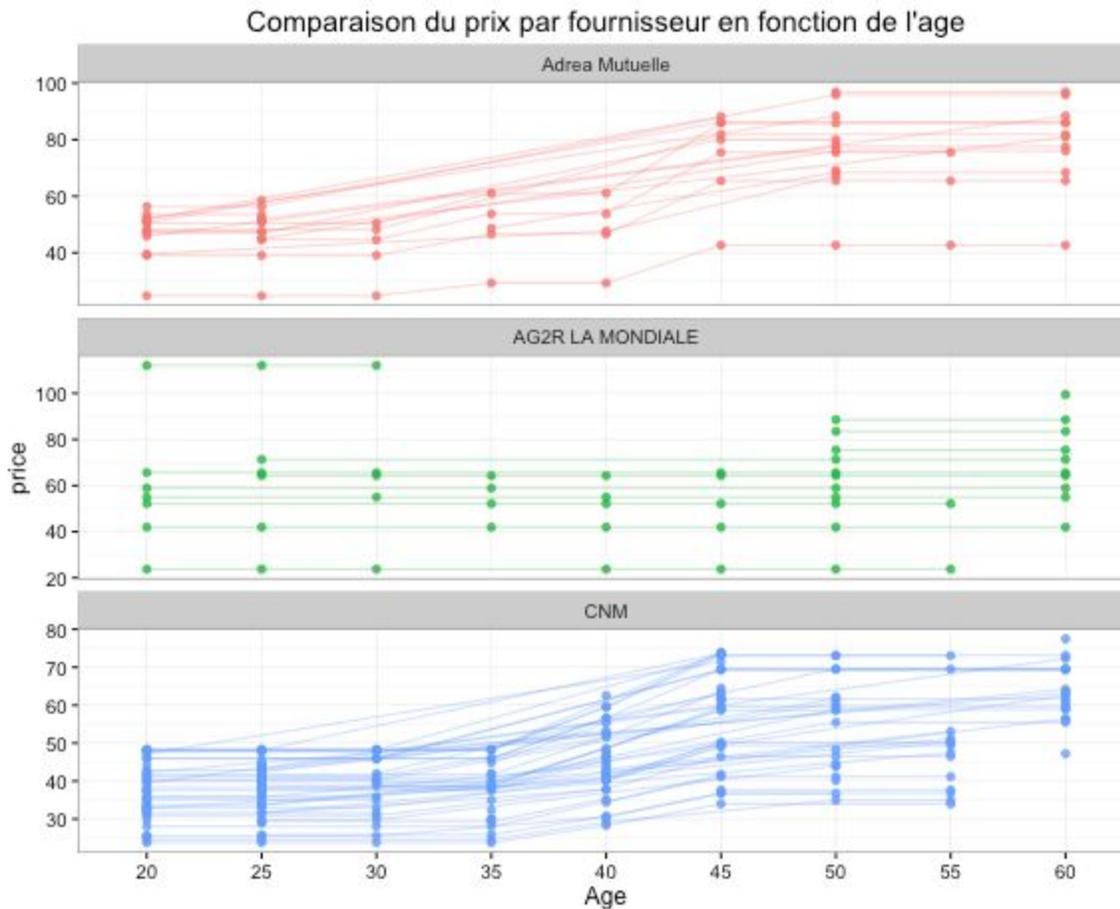
Nous pouvons alors constater que le panier minimum de CNM est légèrement plus cher que celui de AG2R. Au contraire, on pourra voir que le facteur homothétique h est supérieur à 1, ce qui veut dire que les prix et coûts augmentent plus fort du côté de AG2R.

Il existe aussi une difficulté différente : les homothéties ne sont pas forcément équivalentes lorsque les moyennes d'âge diffèrent. Il faut alors calibrer de manière similaire chaque bande de prix par âge.



Au contraire, comme illustré dans le graphe ci-dessus, on observe que les homothéties sont respectées. Pour l'illustration ci-dessus, nous avons créé un indice des services simplifié en une seule valeur, représentant chacune des catégories de garanties. Les prix, en revanche, varient en fonction des valeurs de l'âge, de manière indépendante de la variation d'autres facteurs du produit.

On observe alors la figure suivante, sur quelques exemples de mutuelles :



Alors que certaines mutuelles comme AG2R ne varient pas leurs plans en fonction de l'âge, d'autres, comme Adrea ou CNM, fonctionnent en plateaux, mais augmentent leurs tarifs en fonction de l'âge moyen dans l'entreprise.

Cette variation conduit AG2R à offrir des plans plus chers, et donc à justifier le facteur homothétique fort.

6. Construction de produits originaux

6.1 Description du procédé

Maintenant que nous avons calibré nos pricings, et que nous pouvons reconstruire chaque garantie indépendamment, il est possible de créer des produits originaux en additionnant le prix de chaque garantie.

Comme nous l'avons écrit plus haut, l'équation se vérifie :

$$P(X) = P\left(\sum_{i=1}^N x_i e_i\right) = \sum_{i=1}^N P(x_i e_i)$$

Il est alors simple d'ajouter les valeurs de chaque garantie pour créer un produit d'assurance.

Nous définissons cette méthode comme notre *tarificateur*.

6.2 Benchmarking: Source de Données

Afin de réaliser notre benchmarking, nous devons chercher des prix et des produits qui sont hors de notre échantillon, et lui donner un prix en utilisant notre méthode.

Trois étapes sont alors nécessaires :

1. Décomposer le produit en chacun de ses éléments
2. Donner un prix à ces éléments
3. Additionner les prix

Plusieurs solutions sont possibles pour obtenir les données de benchmark : garder une partie de données scrapées en dehors de notre échantillon d'entraînement, ou bien obtenir d'autres données utilisables.

Nous utiliserons la deuxième méthode, en appliquant notre solution à des produits utilisés par des startups réelles qui nous ont gracieusement partagé leurs plans d'assurance santé.

Nous avons collecté dans le cadre de notre activité plus d'une dizaine de plans proposés à différentes sociétés qui sont nos beta-testeurs. Nous avons passé leurs plans dans le procédé que nous avons construit.

Encadré : Liste des questions d'interview des clients bêta-testeurs

Société:

- *Interlocuteur*
- *Nombre d'employés*
- *Structuration du capital*
- *Convention collective*
- *Âge moyen*
- *Valeurs d'entreprise pour les employés*

Assurance complémentaire santé actuelle :

- *Quel est le fournisseur actuel ?*
- *Qui a fait le choix dans la société ?*
- *Comment a été fait le choix ? Par quel canal ?*
- *Comment avez-vous ressenti l'expérience ? Combien de documents à remplir à la main ?*
- *Quelle est la procédure pour demander un remboursement ?*
- *Quelle est la procédure pour ajouter un employé ?*
- *Quelle partie payez-vous en tant qu'employeur ?*
- *Quel est le prix mensuel ? Comment est-il exprimé dans le contrat ?*
- *De quelle flexibilité avez-vous besoin dans les plans ? Quelle part de vos employés a pris des options ?*
- *Savez-vous pour quels remboursements vous avez signé ? Décrivez-les.*
- *Pouvez-vous partager avec nous le contrat ?*
- *Avez-vous accès à des interfaces web ?*

- *Qu'est-ce qui vous ferez changer d'assureur ?*

Wellness:

- *Quelles données mesurez-vous sur le wellness de vos équipes ?*
- *Quels sont les outils pour mesurer les absences ?*
- *Qu'est-ce que vous aimeriez offrir à vos équipes ?*
- *Est-ce que vous utilisez des outils de fitness ?*

Développement produit :

- *Comment stockez-vous la liste des employés ?*
- *La liste des emails des employés ?*
- *Quel est votre outil de paie ?*
- *Voulez-vous devenir bêta-testeur du produit au cours de son développement ?*

Pour la plupart des plans, notre tarificateur est arrivé à un prix cohérent avec le prix offert, mais il existait parfois des déviations.

Les raisons pour ces déviations peuvent être expliquées pour beaucoup par le fait que nous n'avons malheureusement pas accès au facteur d'homothétie introduit précédemment.

D'autres facteurs d'explication peuvent être les suivants :

- Les prix sont parfois obtenus par l'intermédiaire d'un courtier, qui touche une commission supplémentaire sur le prix de l'assurance. Cette commission n'existe pas lorsque les plans sont souscrits par internet. Du moins, c'est ce que les sites proclament ("Les vrais prix, sans frais cachés", "Un service gratuit & sans engagement"). S'il existe des commissions, elles seront inférieures à celles du courtage en direct, dû au mode de distribution.

- Les prix donnés sur internet sont des prix à l'entrée. Par opposition, le prix qui existe pour les entreprises qui nous ont communiqué leurs plans a pu dériver à cause d'une sinistralité inattendue. Ainsi, la valeur différente est potentiellement un indice de sinistralité exceptionnelle de la société

Ce benchmarking est important, car il permet de nous rendre compte que bien que notre méthode soit approximative sur les points détaillés plus haut, elle aboutit à un resultat correct.

Le tableau suivant détaille les prix payés par employé dans certaines entreprises en beta-test pour Alan.

Entreprise	Prix par employé	Nombre d'employés	Part couverte	Budget mensuel
A	55	9	50%	247.5
B	90	30	100%	2700
C	56	15	50%	420
D	70	6	100%	420
E	55	700	80%	30800
F	47	11	100%	517
G	45	12	50%	270
H	45	80	100%	3600

Nous avons changé les noms des entreprises afin de conserver leur anonymat.

6.3 Second benchmarking

Afin de pouvoir être certain que l'utilisation de cette méthode était saine pour notre entreprise, il était indispensable de comparer notre méthode avec une méthode différente, que l'on pourrait qualifier de *bottom-up*. C'est la méthode qui serait privilégiée si nous possédions de l'historique de sinistralité.

Nous avons donc fait appel à un acteur extérieur, un cabinet de conseil en actuariat, Optimind Winters, pour utiliser un tarificateur basé sur cette approche.

Optimind Winter se présente comme le leader français indépendant du conseil en actuariat et en gestion des risques.

Optimind travaille avec les assureurs, mutuelles, administrations, banques et grandes entreprises qui souhaitent un partenaire métier les accompagnant dans leurs projets.

Parmi ses 180 collaborateurs, il y a 70 actuaires diplômés membres de l'Institut des actuaires.

En utilisant la sinistralité collectée à travers les différents partenaires avec qui ils ont pu travailler, le cabinet de conseil Optimind a pu construire au fil des années un tarificateur qui utilise la sinistralité historique pour projeter une sinistralité attendue sur les populations qui nous intéressent.

L'utilisation de ces chiffres de sinistralité complétés par un ajustement de taux de chargements, ainsi que les taux d'imposition, permettent à notre partenaire de nous offrir une évaluation du coût et du prix final de notre produit assurantiel.

La comparaison de ce produit créé et son prix avec la méthode que l'on a expliquée ci-dessus nous offre la réalisation que les deux prix sont très similaires, et ainsi que la méthode développée dans ce mémoire est solide.

Un avantage supplémentaire de cette méthode complémentaire est qu'elle nous permet de voir clair dans les taux de chargements que l'on peut appliquer pour nos produits.

En effet, un des inconvénients de l'approche du mémoire est qu'elle ne permet de raisonner que sur des tarifs TTC, ce qui conditionne sur des chargements standards pour le marché. Alan ayant une approche radicalement différente, avec beaucoup d'investissement sur les outils informatiques, et une distribution directe, les taux de chargements sont vraiment différents.

Optimind Winter accompagne Alan dans la mise en place d'offres santé collective et individuelle. Dans ce cadre, l'intervention d'Optimind Winter a porté sur :

- La définition des niveaux de garanties
- La tarification des offres de santé collective et individuelle

Les tarifications ont été réalisées à l'aide du logiciel de tarification développé par Optimind Winter, éprouvé avec succès sur de nombreux travaux de tarification santé.

L'outil de tarification santé développé par Optimind Winter applique une méthode usuelle, pour chaque acte de la grille de prestations: "Coût unitaire x Fréquence".

Optimind Winter base son évaluation tarifaire du risque santé sur les données de consommation médicale nationale par acte ajustées régulièrement en fonction de l'évolution de la réglementation, des comportements de soins des assurés et des dérives de coût de santé :

- Nature des actes (dentaire, optique, soins courants) avec la prise en compte des spécificités de chacun (Base de Remboursement, Ticket Modérateur...) et des dépassements d'honoraires
- Fréquence des actes

- Dispersion des coûts par acte (minimum - moyen - maximum)

Par ailleurs, pour chaque coût ainsi calculé pour les actes, des correctifs sont appliqués pour tenir compte :

- De l'âge de la démographie
- Du lieu géographique
- Du type d'acte (dentaire, optique,...)
- Des effets d'anti-selection

Le tarif global correspond ainsi à l'addition des différents coûts obtenus pour chaque acte de la grille de garanties.

Les tarifs ainsi obtenus sont, par la suite, challengés grâce à une analyse technique du marché selon le niveau (bas, milieu ou haut de gamme) et le type d'offres.

6.4 Application à Alan

La méthode développée nous permet donc de produire un plan avec différentes garanties, de lui assigner un prix.

Le produit proposé par Alan a été défini en choisissant des garanties en fonction d'un prix fixé initialement.

Le prix est fixé de manière stratégique afin de pouvoir s'adresser à un public particulier tout en conservant des marges saines.

Le prix en a été fixé à 55 euros pour une entreprise sous la convention collective Syntec, basée à Paris, et dont la moyenne d'âge des employés est de 30 ans.

La raison pour laquelle ce prix a été fixé est simple : c'est la médiane du prix payé par des entreprises qui ont été sélectionnées pour le test beta.

A partir de là, nous pouvons utiliser les items qui ont été pricés précédemment pour pouvoir composer un produit assurantiel qui soit solide et intéressant pour les assurés.

Le produit offert par Alan a ainsi la forme qui suit :

SOINS MEDICAUX (SS incluse)	
Consultation Généraliste (médecins adhérents au CAS)	300% BR
Consultation Généraliste (médecins non adhérents au CAS)	200% BR
Consultation Spécialiste (médecins adhérents au CAS)	300% BR
Consultation Spécialiste (médecins non adhérents au CAS)	200% BR
Consultation Psychiatrie et assimilé (médecins adhérents au CAS)	300% BR
Consultation Psychiatrie et assimilé (médecins non adhérents au CAS)	200% BR
Infirmiers / Auxiliaires médicaux	200% BR
Analyses	200% BR
Radiologie (médecins adhérents au CAS)	300% BR
Radiologie (médecins non adhérents au CAS)	200% BR
Pharmacie	100% BR

HOSPITALISATION (SS incluse)	
Frais de séjour / Etablissements conventionnés ou non conventionnés	300% BR
Honoraires médecins adhérents au CAS	300% BR
Honoraires médecins non adhérents au CAS	200% BR
Forfait journalier	FR
Transport	100% BR
Chambre particulière	80€/jour limitée à 30 jours

Lit accompagnant	50€/jour limitée à 30 jours
Maternité / Honoraires d'accouchement (médecins adhérents au CAS)	400% BR
Maternité / Honoraires d'accouchement (médecins non adhérents au CAS)	200% BR
Maternité / Chambre particulière	100€/jour limitée à 30 jours

OPTIQUE (SS incluse)

Monture + 2 verres simples 1 paire tous les deux ans pour les adultes sans prescription médicale	150€ Dont monture maximum 150€
Monture + 1 verre simple + 1 verre complexe 1 paire tous les deux ans pour les adultes sans prescription médicale	200€ Dont monture maximum 150€
Monture + 2 verres complexes 1 paire tous les deux ans pour les adultes sans prescription médicale	300€ Dont monture maximum 150€
Monture + 1 verre simple + 1 verre très complexe 1 paire tous les deux ans pour les adultes sans prescription médicale	200€ Dont monture maximum 150€
Monture + 1 verre complexe + 1 verre très complexe 1 paire tous les deux ans pour les adultes sans prescription médicale	300€ Dont monture maximum 150€
Monture + 2 verres très complexes 1 paire tous les deux ans pour les adultes sans prescription médicale	400€ Dont monture maximum 150€
Lentilles acceptées	100% BR+ 150€/an
Lentilles refusées	150€/an
Chirurgie Myopie	500€/œil/an

FRAIS DENTAIRES (SS incluse)

Consultation dentiste	100% BR
Orthodontie acceptée	300% BR

Prothèses dentaires prises en charge ou non par la Sécurité Sociale (y compris Inlay/Onlay)	300% BR
Implant dentaire	200€/an
Détartrage bi-annuel	100% BR

AUTRES (SS incluse)	
Audioprothèses	100% BR + 100€/an
Cures	100% BR
Prime de naissance	300€/an
Contraception	100€/an
Traitement anti-Tabac	100€/an
Visite annuelle du sport	30€/an
Vaccins internationaux	50€/an
Prothèses	200% BR
Appareillage	200% BR
Actes de prévention pris en charge par la Sécurité Sociale	100% BR

Comme évoqué en début de mémoire, nous avons combiné cette approche à de la donnée historique pour obtenir nos prix pour les bénéficiaires supplémentaires.

Notre approche est mono-produit avec la tarification évoluera selon l'âge moyen et le nombre de bénéficiaires par utilisateur.

L'évolution des primes par utilisateur est calculée dans le tableau suivant :

en €	Contrat unique
Prime Alan	55
Prime pour bénéficiaire supplémentaire	55
Prime pour enfant supplémentaire	40
Taxes	13.27%

7. Limites

La méthode présentée dans ce mémoire présente plusieurs limites que nous avons identifiées au cours de la construction du Produit d'Assurance Complémentaire Santé. Nous tâchons de les décrire ci-dessous :

- Tarifs initiaux

Notre solution est biaisée pour un vrai pricing de complémentaires santé, car il porte uniquement sur des nouveaux contrats.

On peut légitimement supposer que pour des fins de distribution, les organismes fournisseurs d'assurance cassent les prix pour des nouveaux clients, qui ne possèdent pas encore d'historique de sinistralité. Les années suivantes, les organismes ont tendance à augmenter les primes au plus haut niveau historique de sinistralité, basé sur le ligne à ligne.

Ainsi, notre solution aura tendance à sous-estimer le prix de ces contrats.

- Entre les différentes mutuelles

Notre modèle estime que les différentes mutuelles ont une relation affine pour la fixation de leurs primes. En réalité il se peut que leurs cibles, leurs sinistralités, leurs structure de coûts et leurs objectifs de rentabilité soient différents et mettent l'accent sur de différentes garanties. Par exemple, une mutuelle qui sera particulièrement orientée

famille pourra avoir un coût plus grand pour certaines prestations enfant, ce que nous ignorons dans le modèle présenté dans ce mémoire.

- Indépendance des coûts de postes de garanties

Dans ce mémoire, nous faisons la supposition que les garanties sont entièrement indépendantes entre elles, c'est à dire que chaque poste a un prix unique, peu importe quelles sont les autres garanties du plan.

En réalité, il est possible que certaines garanties élevées poussent à la consommation d'autres garanties, ou bien la réduisent. Ainsi, on pourrait observer qu'un meilleur remboursement des opérations de chirurgie laser réduise les prestations en optique pour les ports de lunettes.

A l'opposé, il est possible aussi que certains remboursements puissent pousser à la consommation de services remboursés.

- Nécessité de ligne-à-ligne

L'approche que l'on a adoptée dans ce mémoire est une approche *top-down*, de haut en bas : nous partons de prix déjà constitués, et nous en déduisons des prix pour chacun des postes, ce qui permet de donner un prix à notre plan. Une autre approche, qui serait plus exacte et plus proche de la solution utilisée par les assureurs aujourd'hui, serait d'obtenir de la donnée ligne-à-ligne, qui nous permettrait de faire un réel calcul de sinistralité prédictive et historique. Il est alors possible d'y ajouter un premium pour les frais de gestion, d'administration, et de distribution, afin d'obtenir un prix commercial.

Malheureusement, pour un acteur nouveau, cette donnée n'existe pas encore, ce qui nous a poussé à créer le cheminement ci-dessus. En revanche, certains cabinets d'actuariat possèdent des échantillons de données collectées à travers leurs partenaires. La complétion du prix passe par combinaison de ces méthodes. C'est ce que nous avons décrit dans la partie précédente.

Conclusion

Ce mémoire consacré à la création d'un nouveau produit d'assurance complémentaire santé dans un contexte de start-up en assurance, s'est attaché dans un premier temps à décrire le cadre contextuel, marché et réglementaire pertinent pour l'Assurance Santé.

Après un rappel du Marché de l'Assurance Complémentaire Santé et des acteurs qui le composent, nous nous sommes intéressés à l'évolution de la réglementation d'abord dans le cadre de la Solvabilité avec la mise en application de Solvabilité II en tant qu'Assurance et ensuite sur le marché spécifique de la Santé avec l'implémentation de l'ANI et du contrat responsable et solidaire. Il s'agit d'un marché avec une grande quantités d'acteurs (plus de 500) dont la compétition et l'évolution a conduit à une consolidation sur les 10 dernières années avec des niveaux de profitabilités relativement bas. Ce système est dépendant des éléments existants que sont l'interconnexion aux systèmes de la Sécurité Sociale ainsi que de la protection des données des utilisateurs pour laquelle un cadre réglementaire existe pour les Assurances. Tous ces éléments sont des opportunités pour une nouvelle Société d'Assurance d'offrir un service innovant, et digital qui réduit la structure de coût tout en améliorant la valeur pour l'utilisateur final.

Dans un second temps, nous nous sommes concentrés sur les briques élémentaires qui permettent la construction de ce dit produit, allant de la collecte de données qui nécessite la création de robots, dits *scrapers*.

La collection de ces données nous a permis de pouvoir comparer les offres présentes sur le marché. Grâce à une normalisation des données, nous pouvons alors les utiliser pour en déduire le prix, pour certaines caractéristiques intrinsèques et extrinsèques

fixées, de garanties itémisées présentes dans les plans offerts sur les comparateurs internet.

Ces prix nous ont permis de pouvoir déduire le coût estimé du produit d'assurance complémentaire santé que Alan souhaite proposer à des souscripteurs prospectifs.

Un benchmark extérieur nous permet d'augmenter la confiance que l'on a dans notre calcul.

Il sera maintenant important de collecter notre propre historique de sinistralité pour pouvoir avoir de la donnée ayant valeur de vérité sur notre base d'utilisateurs.

Bibliographie

[1] DREES (2016), *Assurance complémentaire santé : les contrats collectifs gagnent du terrain*, Etudes & Résultats, Numéro 0952.

[2] Montaut A. (2015), *Rapport du gouvernement au Parlement sur la situation financière des organismes complémentaires assurant une couverture santé*, DREES.

[3] Le Palud V. (2013), *Comment les organismes complémentaires fixent leurs tarifs*, Etudes et Résultats, DREES, n°850

[4] Données de la DREES: www.data.drees.sante.gouv.fr

[5] Tosetti A., Behar T., Fromenteau M., Menart S., *Assurance, Comptabilité, réglementation, actuariat*, Economica

[6] Denuit M., Charpentier A., *Mathématiques de l'assurance non-vie : Tome 1, Principes fondamentaux de théorie du risque*, Economica

[7] Articles 41 à 48 de la directive Solvabilité II : <https://eiopa.europa.eu/>

[8] Haut conseil pour l'avenir de l'assurance maladie, *La généralisation de la couverture complémentaire en santé*

[9] Bartels, R. H.; Beatty, J. C.; and Barsky, B. A., *Hermite and Cubic Spline Interpolation*. Ch. 3 in *An Introduction to Splines for Use in Computer Graphics and Geometric Modelling*. San Francisco, CA: Morgan Kaufmann, pp. 9-17, 1998.

[10] ACPR, *Préparation à Solvabilité II. Traduction partielle du document de l'EIOPA sur les hypothèses sous-jacentes à la formule standard pour le calcul du SCR*. 6 août 2014
– Version 1.0

[11] CNIL, *Pack de conformité Assurance*, Edition Juillet 2014

[12] Lambert-Faivre Y., Leveneur L., *Droit des assurances*, 13ème édition, Dalloz

[13] Légifrance, *Décret n° 2014-1374 du 18 novembre 2014 relatif au contenu des contrats d'assurance maladie complémentaire bénéficiant d'aides fiscales et sociales*.

[14] Wikipedia, *fr.wikipedia.org*.

[15] Optimind-Winter, *Tarification de l'offre santé Alan*. 16 mai 2016.

[16] StackOverflow, *Selenium with scrapy for dynamic page*.

<http://stackoverflow.com/questions/17975471/selenium-with-scrapy-for-dynamic-page>

[17] MOULIN P.J., *Construction de bases de tarification en frais de santé*, Mémoire d'actuariat ISFA, 2003.

Annexes

Extraits du code du *scraper*

```
# -*- coding: utf-8 -*-
```

```
import scrapy
```

```
import time
```

```
from selenium import webdriver
```

```
from selenium.webdriver.support.ui import Select
```

```
from selenium.webdriver.common.keys import Keys
```

```
import logging
```

```
from unidecode import unidecode
```

```
import json
```

```
from random import randint
```

```
from selenium.webdriver import ActionChains
```

```
class PlansPricesSpider(scrapy.Spider):
```

```
    name = "product_spider"
```

```
    allowed_domains = ["www.assurancepro.com"]
```

```
    params_sel = {'channel': '0', # Sur internet  
                 'couverture': '1'} # Salaries
```

```
    params_input = {'raisonSociale': 'monEntreprise',  
                   'nom': 'monNom',  
                   'prenom': 'monPrenom',  
                   'fonction': 'PDG',  
                   'adresse': '123 rue Bidon',  
                   'adresseMail': 'monmail@monmail.com',  
                   'telephone': '0610204050',  
                   'conventionCollectiveLibelle': 'syntec',  
                   'codePostal': '75010',  
                   'nbSalarie': '20',  
                   'nbCadres': '20',
```

```
    'ageMoyenCadre': '35'}
```

```
params = {}
```

```
provider = ""
```

```
def __init__(self, params={}, provider=""):
```

```
    self.driver = webdriver.Firefox()
```

```
    # Make the scraper wait 5sec
```

```
    self.driver.implicitly_wait(5)
```

```
    # Make it full size
```

```
    self.driver.set_window_size(2000, 1000)
```

```
    # Set the parameters
```

```
    self.params_input.update(params)
```

```
    self.params = params
```

```
    self.provider = provider
```

```
    # To make scrapy work with selenium
```

```
    self.start_urls = ["http://www.lassurancepro.com/105848-mutuelle-tpe-pme"]
```

```
    logging.info("Init finished")
```

```
def parse(self, response):
```

```
    logging.info("parse started")
```

```
    self.driver.get(response.url)
```

```
    self.fill_first_form(self.params_input, self.params_sel)
```

```
    self.fill_second_form(self.params_input)
```

```
    self.fill_comparison_page()
```

```
    items = self.scrape_final_page()
```

```
    return items
```

```
def fill_first_form(self, params_input, params_sel):
```

```
    # While on the first page, try to fill up the form
```

```
    while self.driver.title.split(' ')[0] == 'Comparez':
```

try:

```
# Click on the civillite radio button  
self.driver.find_elements_by_class_name("iradio_square-blue")[0].send_keys(Keys.SPACE)
```

```
selects = self.driver.find_elements_by_tag_name("select")
```

```
for sel in selects:
```

```
    if params_sel.__contains__(sel.get_attribute("id")):  
        Select(sel).select_by_value(params_sel[sel.get_attribute("id")])
```

```
inputs = self.driver.find_elements_by_tag_name("input")
```

```
for input_el in inputs:
```

```
    if not params_input.__contains__(input_el.get_attribute("id")):  
        continue
```

```
    input_id = input_el.get_attribute("id")
```

```
    inpu = self.driver.find_element_by_id(input_id)
```

```
# Clear the field
```

```
inpu.clear()
```

```
# Fill the field
```

```
inpu.send_keys(params_input[input_id])
```

```
# Validate by going to the next
```

```
inpu.send_keys(Keys.TAB)
```

```
self.driver.execute_script("window.scrollTo(0, 50);")
```

```
self.driver.find_element_by_id("btnValider").click()
```

```
time.sleep(5)
```

```
except:
```

```
    time.sleep(5)
```

```
    continue
```

```
print "Finished page 1"
```

```
def fill_second_form(self, params_input):
```

```
# wait 3 seconds for the page to load
```

```
if self.driver.title.split(' ')[0] != 'Votre':
```

```
    time.sleep(3)
```

```

# While on the first page, try to fill up the form
while self.driver.title.split(' ')[0] == 'Votre':
    try:
        inputs = self.driver.find_elements_by_tag_name("input")
        for input_el in inputs:
            if not params_input.__contains__(input_el.get_attribute("id")):
                continue

            input_id = input_el.get_attribute("id")
            inpu = self.driver.find_element_by_id(input_id)

            # Clear the field
            inpu.clear()
            time.sleep(1)

            # Fill the field
            inpu.send_keys(params_input[input_id])
            time.sleep(1)

            # Validate by going to the next
            inpu.send_keys(Keys.TAB)

            # Hack to make the browser believe it's focused
            if input_el.get_attribute("id") == 'nbCadres':
                current_handle = self.driver.current_window_handle
                self.driver.execute_script("alert('Test')")
                self.driver.switch_to.alert.accept()
                self.driver.switch_to.window(current_handle)
                inpu.send_keys(Keys.TAB)

    except:
        continue

# Get the size of the slides
slide0 = self.driver.find_element_by_id("slider0")

```

```

width = slide0.size["width"]

# Select levels at random
for elem in self.driver.find_elements_by_class_name("noUi-handle"):
    percent = randint(0,100)
    ActionChains(self.driver)\
        .drag_and_drop_by_offset(elem,percent*width/100,0)\
        .perform()

self.driver.find_element_by_id('btnValider').send_keys(Keys.ENTER)
time.sleep(5)

while self.driver.title.split(' ')[0] != 'Vos':
    time.sleep(3)
print "Finished page 2"

def fill_comparison_page(self):
    # if provider is specified, only select it
    if self.provider != "":
        for elem in self.driver.find_elements_by_xpath("//table[@id='tabValListAssureur']/tbody/tr/td"):
            if elem.text.split(' ')[0] == self.provider:
                continue
            elem.find_element_by_name('chkAssureur').send_keys(Keys.SPACE)

# Click on the more offers button
for i in xrange(5):
    self.driver.find_element_by_id("plusdOffres").send_keys(Keys.ENTER)
    time.sleep(1)

# Select plans available
elems = self.driver.find_elements_by_xpath("//*[contains(@class,'infoOffers') and not(contains(@class,
'notDisplay')) and not(contains(@class, 'hide'))]/td[1]/div/input")

# Select 4 plans randomly
plansSelected = []
while len(plansSelected) < 4:
    ind = randint(0, len(elems))

```

```

if ind in plansSelected:
    continue

# Select offers available randomly
try:
    elems[ind].send_keys(Keys.SPACE)
    plansSelected.append(ind)
except:
    continue

# Validate form
self.driver.find_element_by_class_name("compareDetailsBtn").send_keys(Keys.ENTER)

# Wait for the page to load
while self.driver.title.split(' ')[0] != 'LassurancePro.com':
    time.sleep(3)
print "Finished page 3"

def scrape_final_page(self):
    # Scrape the final page (there are up to 4 plans on the page)
    items = []
    for itemInd in xrange(0, 4):
        item = PricescraperItem()

        # Find name of the plan
        try:
            sel = self.driver.find_elements_by_xpath("//table[@class="compareTable"]/thead/tr/th/p")[itemInd]
        except:
            break

        item["provider"] = unicode(sel.text.split(' - ')[0])
        item["name"] = unicode(sel.text.split(' - ')[1])

        # Find price of the plan
        item["price"] = unicode(self.driver
            .find_element_by_xpath("//table[@class="compareTable"]/tbody/tr[1]/td["+str(itemInd+2)+"']").text)

```

```

# Find all the data
item['data'] = {}
for sel in self.driver.find_elements_by_xpath('//table[@class="compareTable"]/tbody/tr):
    # Find the cells
    elems = sel.find_elements_by_xpath('.//td')
    name = unicode(elems[0].text)

    # It not in the format of 3 values, skip
    if len(elems) < 3 or name.strip(' ') == "":
        continue

    # assign in data
    val = unicode(elems[itemInd+1].text)
    (item['data'])[name] = val

    # Set params in the item
    item['params'] = self.params

    # Set a specific id for each plan scraped (combination of data)
    item['id'] = json.dumps([item['name'], item['params'].values()])

items.append(item)

self.driver.close()
return items

```

```

class PricescraperItem(scrapy.Item):
    name = scrapy.Field()
    provider = scrapy.Field()
    price = scrapy.Field()
    data = scrapy.Field()
    id = scrapy.Field()
    params = scrapy.Field()

```