

**Mémoire présenté le :  
pour l'obtention du diplôme  
de Statisticien Mention Actuariat  
et l'admission à l'Institut des Actuaire**

Par : Monsieur Andrew BERNET-ALINOT

*Titre du mémoire : Accélération du passage en invalidité*

Confidentialité :  NON  OUI (Durée :  1 an  2 ans)

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

*Membres présents du jury de  
l'Institut des Actuaire*

signature

*Entreprise :*

Nom :

*Signature :*

*Directeur de mémoire en  
entreprise :*

*Nom :*

*Signature :*

*Invité :*

*Nom :*

*Signature :*

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

Signature du responsable  
entreprise

Signature du candidat

*Membres présents du jury de la  
filiale*

# ***REMERCIEMENTS***

Avant de commencer ce rapport, je souhaite remercier toutes les personnes ayant contribué à sa rédaction, que ce soit en me donnant de précieux conseils ou simplement en me soutenant moralement à chaque instant.

Et je commence ainsi par remercier très chaleureusement Thomas PETEUL, mon tuteur durant mon alternance, qui, du fait de sa bienveillance, sa disponibilité, son expertise et son expérience, m'a sans cesse aidé et conseillé tout au long de l'écriture de ce mémoire.

Je tiens également à remercier Adeline GERARD pour m'avoir permis d'intégrer son équipe et sans qui ce mémoire n'aurait donc pas pu voir le jour. Elle m'a ensuite guidé, me mettant sur de nouvelles pistes et me challengeant pour pousser toujours plus loin ma réflexion et ainsi faire avancer et parfaire ce mémoire.

Je remercie encore l'ensemble de son équipe, David BOUGIER, Isabelle CHUPIN, Nadia EL MASRIOUI, Hanh LE HONG et Ingrid LEPETIT pour m'avoir si bien accueilli et soutenu en restant disponibles et en garantissant une bonne ambiance de travail.

Je remercie encore mon tuteur académique, Nicolas BOUSQUET, pour la relecture de mon mémoire, le temps qu'il m'a accordé et les conseils qu'il m'a prodigués.

Enfin, je remercie mes parents pour tout ce qu'ils m'ont apporté durant ma scolarité, leur soutien et leur confiance en moi. Mon frère et ma compagne ont également été des appuis de taille sur lesquels je me suis fréquemment reposé.

# *RESUME*

**Mots clés :** Invalidité, Incapacité, Projection, Intervalles de confiance, Tables d'expérience, Fonctions de survie, Estimateurs, Lissage, Interpolation et Extrapolation, Modèle multimodal.

L'objet de ce mémoire est de souligner et d'étudier une tendance s'exprimant depuis le début de la décennie et, si elle s'avère réelle, de tenter de l'expliquer et de projeter son impact dans le temps. En effet, il est remarqué par plusieurs organismes d'assurance que les délais de passage de l'état d'incapacité à celui d'invalidité tendent à diminuer, tout comme ceux de maintien en incapacité. Nous nous sommes donc attardés à vérifier cette impression en étudiant les données d'arrêt de travail d'un portefeuille d'assurés.

Cette étude fut construite en quatre étapes. Dans une première partie, nous avons rappelé les caractéristiques du risque « invalidité » dans son contexte sociodémographique actuel. Pour ce faire, nous avons tout d'abord décrit les différents régimes de prévoyance en France et leurs caractéristiques, avant de définir plus spécifiquement la garantie « Invalidité » : la façon dont est apprécié le niveau d'invalidité, et dont est provisionné et tarifé ce risque.

La deuxième partie de ce mémoire concerne le périmètre de notre étude, les différentes caractéristiques mises à notre disposition sur la durée des arrêts de travail. Il est précisé dans cette partie comment ont été fiabilisées les données et les statistiques que nous avons pu extraire.

Par la suite, dans une troisième partie, nous nous sommes davantage concentrés sur l'accélération du passage en invalidité. Il nous a fallu modéliser la durée des incapacités par suite d'un arrêt de travail ayant entraîné une invalidité. Après avoir testé plusieurs modèles unimodaux, nous nous sommes aperçu qu'une loi multimodale était la meilleure solution pour modéliser ces durées. Nous avons ainsi utilisé un mélange de lois gaussiennes s'ajustant sur les différents pics de passages. Cette modélisation par année de survenance des sinistres nous a permis de mettre en évidence l'accélération du passage en invalidité sur la dernière année étudiée. Si objectivement cela n'est pas suffisant pour nous prononcer de manière certaine sur le point de savoir si cette constatation traduit une réelle accélération durable dans le temps ou une simple particularité, il nous a paru intéressant de présupposer que cette tendance allait s'installer sur les années à venir, ce qui nous a ainsi conduits à étudier les impacts que cette accélération des délais de passage pourrait avoir économiquement sur les portefeuilles de l'entreprise si elle devait effectivement se poursuivre.

En vue de notre quatrième partie, nous avons donc construit des lois de maintien en incapacité et de passage en invalidité sur la base des délais « accélérés » car les lois actuelles du BCAC (Bureau Commun des Assurances Collectives) ne prennent pas en compte ce phénomène.

La difficulté des constructions reposait sur la faible volumétrie du portefeuille due au fait que seule l'année atypique était considérée. Afin d'obtenir tout de même les lois les plus robustes possibles, nous avons regroupé notre échantillon par tranches d'âge en prenant soin que chaque tranche vérifie le critère de Cochran. Puis nous avons modélisé les lois en utilisant l'estimateur de Kaplan-Meier, la technique de lissage de Whittaker-Henderson et en extrapolant et interpolant les résultats obtenus en s'appuyant sur les tables du BCAC.

Enfin, notre dernière partie fut consacrée à étudier les impacts dans le temps de l'accélération sur notre portefeuille. Nous avons donc projeté notre portefeuille d'assurés avec les tables construites précédemment pour obtenir les nombres d'arrêts de travail et d'invalidités qui en ont découlé.

Cela nous a permis de projeter ensuite les cotisations et les charges du portefeuille pour en déduire un indicateur de rentabilité destiné à déterminer si le portefeuille restera rentable malgré l'accélération remarquée.

# ***ABSTRACT***

**Keywords :** Disability, Incapacity, Projection, Confidence intervals, Experience tables, Survival functions, Estimators, Smoothing, Interpolation and Extrapolation, Multimodal model.

The purpose of this paper is to highlight and study a trend that has been expressed since the beginning of the decade and, if it is true, to try to explain it and project its impact over time. Indeed, it has been observed by several insurance agencies that the delays in moving from incapacity to disability tend to decrease, as do the delays in maintaining incapacity. We therefore focused on verifying this impression by studying the work stoppage data of an insured portfolio.

This study was built in four stages. In the first part, we recalled the characteristics of "disability" risk in its current socio-demographic context. To do this, we first described the various pension schemes in France and their characteristics, before defining more specifically the "Invalidity" guarantee: the way in which the level of disability is judged, and whose risk is provisioned and priced.

The second part of this thesis concerns the scope of our study, the different characteristics made available to us on the duration of work stoppages. This section explains how the data was made reliable and outlines the statistics we were able to derive from it.

Subsequently, in a third part, we focused more on accelerating the transition to disability. We had to model the duration of disabilities following a work stoppage that resulted in a disability. After testing several unimodal models, we realized that a multimodal law was the best solution to model these durations. We thus used a mixture of Gaussian laws adjusting to the different peaks of passages. This modelling by year of occurrence of claims has allowed us to highlight the acceleration of the transition to disability over the last year studied. If this did not allow us to comment on a real, lasting acceleration over time or a simple peculiarity, it seemed interesting to us to consider that this trend would take hold in the years to come. This allowed us to study the impact that this acceleration of transit times could have on the company's portfolios economically.

For our fourth part, we have therefore constructed incapacity retention and invalidity passage laws based on "accelerated" delays because current BCAC (Bureau Commun des Assurances Collectives) laws do not take this phenomenon into account.

The difficulty of the constructions was based on the small volume of the portfolio because only the atypical year was considered. To still obtain the most robust possible laws, we grouped our sample by age groups, taking care that each band checks the Cochran criterion. Then we modelled the laws using the Kaplan-Meier estimator, the Whittaker-Henderson smoothing technique and extrapolating and interpolating the results from the BCAC tables.

Finally, our last part was devoted to studying the impact in time of acceleration on our portfolio. We therefore projected our insured portfolio with previously constructed tables to obtain the number of work stoppages and disabilities that resulted.

This allowed us to then project the contributions and the costs of the portfolio to deduce an indicator of profitability which allowed us to know if the portfolio will remain profitable despite the acceleration noted.

# ***NOTE DE SYNTHÈSE***

## ***Contexte***

Les objectifs des régimes de prévoyance sont de prévenir et de couvrir les risques affectant le plus la vie d'un individu, à savoir le décès, la santé, le chômage, l'incapacité de travail, et l'invalidité.

Si la Sécurité Sociale garantit aujourd'hui des prestations financières en cas de réalisation d'un de ces risques, une personne peut choisir d'être couvert par un contrat individuel et/ou collectif afin de bénéficier d'un revenu de remplacement qui peut prendre différentes formes : capital, pension, ou indemnités journalières de base.

Ces contrats sont souscrits auprès d'assurances qui elles-mêmes, doivent pérenniser leur équilibre financier en tarifiant une prime, payée par l'assuré, devant au moins être égale au produit de la probabilité de sinistre par le coût prévisible de ce sinistre. De plus, l'assureur se doit de provisionner un certain montant, estimé à partir de statistiques et de probabilités, afin de toujours pouvoir faire face aux risques (principe comptable de prudence).

Or, en prévoyance (et plus précisément pour ce mémoire en incapacité/invalidité), le coût du sinistre et les provisions dépendent du nombre de jours passés par l'assuré en arrêt de travail et du temps qu'il met à changer d'état (d'incapacité à invalidité ou d'incapacité à validité...).

Afin de provisionner les prestations servies en cas d'incapacité et d'invalidité, les organismes assureurs ont le choix d'utiliser soit « des lois établies par leurs soins et certifiées par un actuair indépendant de cette entreprise et agréé par l'une des associations d'actuaires reconnues par l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) » (article 331-22 du code des assurances), soit les tables réglementaires du Bureau Commun des Assurances Collectives (BCAC). Ces dernières lois de provisionnement, sous forme de tables à double entrée (temps passé depuis l'entrée et âge), ont été entièrement reconstruites en 2013 mais il semblerait que depuis, les temps passés en arrêt de travail aient de nouveau évolué.

## ***Problématique***

Le comportement des délais de changements d'états semble en effet avoir été modifié, et il a été remarqué une diminution des délais de passages en invalidité et des variations de volumétries.

Les tables 2013 ne semblent donc plus « coller à la réalité » des portefeuilles de l'assureur. Ce mémoire cherche ainsi à statuer sur la véritable existence de cette évolution et ses causes potentielles. De plus, pour le cas où cette diminution des temps de passages en invalidité devait se poursuivre, ce mémoire développera de nouvelles lois de maintien en incapacité et de passage en invalidité afin de mieux anticiper les changements d'états futurs des assurés, et observera les impacts économiques que cette accélération pourrait avoir pour l'assureur.

## ***Validation de l'hypothèse d'accélération du passage en invalidité***

### **1. Visualisation du phénomène**

Après avoir analysé les délais moyens de passage en invalidité par année de survenance, nous observons qu'une tendance forte se dessine. Les délais s'amenuisent, passant de plus de 18 mois pour les assurés déclarés en incapacité en 2012 à seulement 12,6 mois pour ceux de 2016.

De plus, les parts d'invalidités déclarées le 36<sup>ème</sup> mois décroissent fortement en fonction des années (plus de 6% des passages en invalidité pour des assurés reconnus en état d'incapacité en 2012 se sont produits le 36<sup>ème</sup> mois contre seulement 1,5% pour les assurés entrés en 2015).

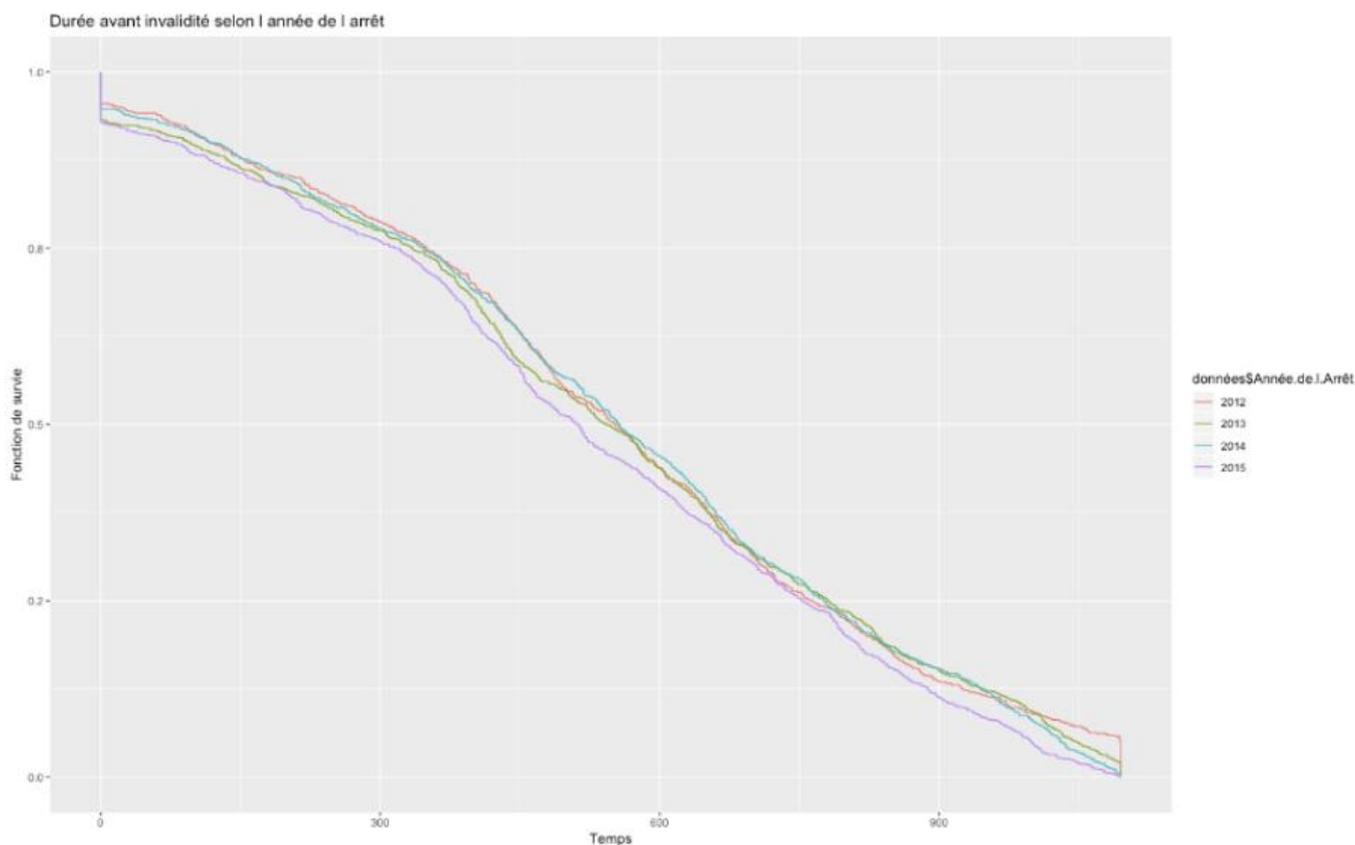
Ces résultats sur les portefeuilles de l'assureur viennent appuyer ceux entrevus lors de la présentation de Forsides sur les risques incapacité et invalidité du 18 avril 2019, lors de laquelle il a été constaté sur l'ensemble des portefeuilles d'Institution de Prévoyance un avancement du pic de passage en invalidité de sept mois entre les années 2013-2014 et les années 2015-2016.

## 2. Analyse des délais de passage en invalidité

Afin d'affirmer l'hypothèse d'un comportement différent des temps de passage en invalidité, nous cherchons à modéliser le temps restant avant le passage de l'état d'incapacité à celui d'invalidité. Nous devons pour ce faire, estimer une fonction de survie. L'analyse de la survie pour des êtres vivants étant une étude statistique qui cherche à modéliser le temps restant avant la mort (espérance de vie), ce qui est représenté graphiquement sous la forme d'une courbe de survie.

Pour estimer cette fonction de survie, il a été choisi d'utiliser l'estimateur de Kaplan-Meier. Cet estimateur est non-paramétrique et repose sur le fait que la période de survie est discrétisée en sous-périodes délimitées par des dates.

Nous avons tracé l'estimateur pour chaque année de survenance et il ressort très clairement que les assurés déclarés invalides des suites d'une incapacité en 2015 sont déclarés plus rapidement invalides que ceux qui ont connu une incapacité dans les années antérieures. Les probabilités « 2015 » (courbe violette ci-dessous) que le passage se produise à un instant « t » semblent toutes inférieures aux probabilités des années précédentes qui, elles, semblent décroître à la même vitesse.



## 3. Modélisation des densités

En traçant les densités des délais de passage en invalidité par année de survenance, nous remarquons là encore que les délais de passage « 2015 » semblent suivre une répartition bien différente des autres années.

Afin de savoir si cette répartition des délais de passage en invalidité des arrêts survenus en 2015 est bel et bien significativement différente des répartitions des délais des survenances 2012, 2013 et 2014, nous allons

modéliser, à l'aide de mélanges de gaussiennes, ces taux de passages pour ainsi comparer à l'aide de tests statistiques les distributions de chaque année.

Le modèle de mélange gaussien est une combinaison linéaire de plusieurs composantes gaussiennes. Ce mélange suscite un intérêt important en raison :

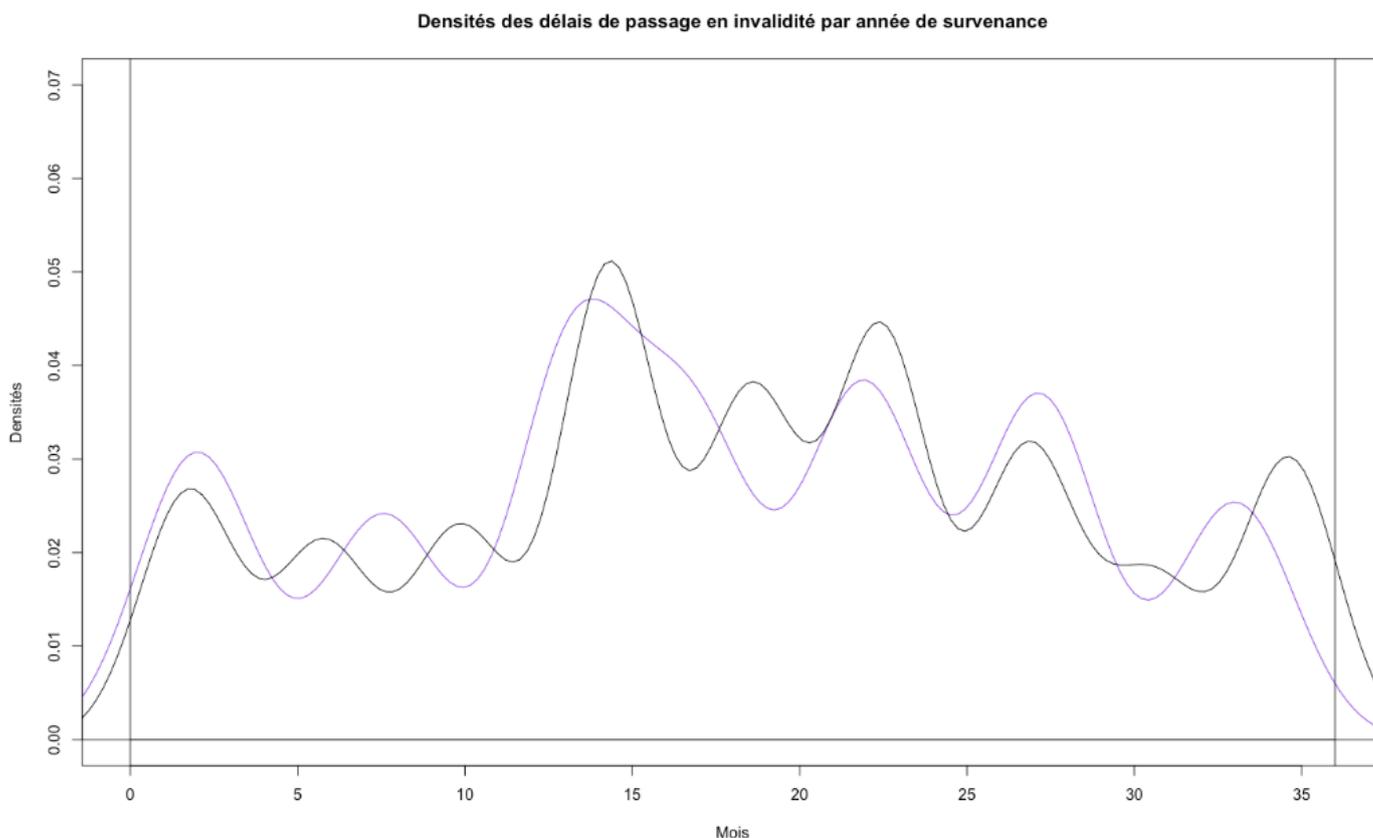
- De sa flexibilité ;
- De sa faculté à approcher une grande variété de densités : on peut toujours modéliser des données continues par un mélange gaussien ;
- De son usage mathématiquement simple et de la généralité de la loi normale qu'atteste le théorème central limite.

Ce mélange étant paramétrique, nous estimons les paramètres en utilisant une approche par maximisation de la vraisemblance.

Toutefois, maximiser la vraisemblance ne peut se faire analytiquement en raison de données cachées. Il faut donc trouver les solutions à l'aide d'algorithmes itératifs. Parmi ces algorithmes, figure l'algorithme EM ou « Expectation- Maximisation ».

L'algorithme garantit que la vraisemblance augmente à chaque itération, ce qui conduit donc à des estimateurs de plus en plus corrects.

Nous avons alors commencé par ajuster une loi de mélange de gaussiennes aux densités des délais de passage en invalidité de l'année de survenance 2015 (en violet) ainsi qu'aux densités des délais de passage en invalidité des assurés entrés en incapacité avant 2015 (en noir) à l'aide de l'algorithme EM.



Ces courbes montrent très explicitement l'accélération du passage en invalidité puisque l'on se rend bien compte du décalage vers la gauche de chaque cloche de la courbe violette ainsi que l'écart conséquent entre les proportions de passage en invalidité dans les derniers mois (la courbe noire étant plus haute et plus à droite que la violette).

Ces impressions « visuelles » sont confirmées par des tests de Kolmogorov-Smirnov et de Student qui concluent sur une accélération des délais de passage en invalidité, ceux des assurés entrés en incapacité en 2015 étant en moyenne significativement inférieurs aux délais des années antérieures. Il nous faut donc à présent créer de nouvelles tables de maintien en incapacité et de passage en invalidité afin de pouvoir mesurer l'impact économique de cette accélération pour l'assureur.

## *Création des tables de maintien en incapacité et de passage en invalidité*

### **1. Estimation des taux bruts**

Nous avons choisi d'utiliser l'estimateur de Kaplan-Meier pour la construction de nos lois. En effet, la loi de maintien se calcule à partir des taux de sortie conditionnels obtenus en intégrant les sorties de chaque période à la population initiale. Or le principe de cet estimateur repose sur le classement des observations par durées observées passées en incapacité en dénombrant les sorties.

Après avoir calculé les taux estimés, il nous est indispensable de vérifier leur fiabilité et leur robustesse. Nous utilisons pour cela la variance de Greenwood. Cette dernière n'est rien d'autre que la variance de l'estimateur de Kaplan-Meier.

Cette variance souligne le manque de profondeur de nos données avec de fortes valeurs pour les jeunes âges ainsi que pour les âges supérieurs à 60 ans. Les taux ne sont donc pas suffisamment fiables pour ces âges-là.

Nous décidons donc de former des tranches d'âges pour obtenir des estimateurs des fonctions de survie plus fiables. Or d'après Aurélie Gaumet (2 000), 1 000 observations sont nécessaires à la bonne réalisation d'une table de maintien en incapacité.

Nous décidons donc de former les tranches d'âge de telle sorte qu'il y ait toujours au moins 1 000 observations dans chacune d'entre elles. Nous recalculons alors des estimateurs de Kaplan-Meier beaucoup plus fiables.

L'étape suivante consiste à lisser ces fonctions de survie pour effacer les légères irrégularités existantes.

### **2. Lissage de Whittaker-Henderson**

La méthode de lissage choisie est celle de Whittaker-Henderson. Ce lissage non paramétrique est un outil simple pour régulariser une série de taux bruts ou estimés.

La méthode est fondée sur deux critères : un critère de fidélité  $F$  entre la courbe de référence et la courbe lissée, et un critère de régularité  $S$ , qui dépend d'un paramètre  $z$ . Elle permet d'obtenir les valeurs qui minimisent la combinaison linéaire  $M$  de ces deux critères :

$$M = F + h * S,$$

Où  $h$  est un paramètre du modèle contrôlant le poids du critère de régularité dans le lissage. Tout comme  $z$ , degré du polynôme auquel va s'ajuster la courbe lissée.

Le choix de  $h$  et de  $z$  a été fait empiriquement (nous les avons fait varier pour trouver leurs valeurs optimales).

### **3. Interpolation et extrapolation**

Pour interpoler les taux, nous commençons par calculer la moyenne d'âge des classes effectuées auparavant. Puis nous considérons que les fonctions de survie estimées déjà lissées sont celles de ces âges moyens.

Enfin, nous utilisons une méthode d'interpolation linéaire pour obtenir les taux entre nos deux âges moyens extrêmes.

Il reste encore à extrapoler les taux de maintien aux âges extrêmes. Nous choisissons alors, comme pour l'interpolation, de procéder à une extrapolation linéaire pour la table de maintien en incapacité, et en usant de coefficients de proportionnalité obtenus par rapport aux taux du BCAC pour la table de passage en invalidité (pour pallier aux manques d'information que nous avons, particulièrement aux jeunes âges).

## *Projection et visualisation des impacts de l'accélération sur un portefeuille*

Pour mesurer l'impact de l'accélération constatée, nous décidons de projeter un portefeuille d'assurés.

Pour ce faire, nous utilisons nos lois calculées ci-avant, une loi d'incidence d'entrée en incapacité interne à l'assureur, et les lois de mortalité de l'INSEE calculées en 2018 (TGH/TGF). Ces dernières nous permettent de calculer un unique taux pour les deux sexes en sachant que notre portefeuille est constitué à 99,5% de femmes.

Nous calculons aussi les taux d'entrée dans le portefeuille pour générer de nouveaux adhérents pour chaque âge entre 20 et 60 ans. Nous utilisons pour cela les données actuelles du portefeuille et la pyramide des âges 2019 de l'INSEE.

Enfin, la table de maintien en invalidité choisie est celle du BCAC.

Nous commençons par projeter la population active et celle présente en incapacité. Si, dans un monde idéal, nous projetterions nos assurés chaque mois pour suivre leur évolution et celles de nos charges et cotisations de plus près, ici nous simplifions le modèle et suivons l'évolution de nos variables année par année seulement par souci de clarté et de lisibilité.

Pour cela, nous avons décomposé les assurés en incapacité en trois groupes: ceux étant dans leur première année de maintien, ceux étant dans leur deuxième année et ceux étant dans leur troisième année.

Les résultats montrent que la population valide totale de notre portefeuille ne fera que diminuer durant les 20 prochaines années ce qui s'explique par le départ à la retraite progressif de la génération des baby-boomers depuis le début des années 2010. Le constat est naturellement le même pour le nombre d'incapacités.

La projection de ce portefeuille avec les tables « 2015 », puis avec les tables « avant 2015 » marque la nette influence de l'accélération des délais de sorties d'incapacité.

Nous projetons ensuite la population invalide en partant du principe qu'un assuré ne peut passer de l'état « valide » à celui « d'invalide » et ni de l'état « d'invalidité » à celui de « validité » et « d'incapacité ». Cela implique que la seule façon de sortir de l'état d'invalidité est le décès de l'assuré. Les taux de décès des tables TGH/TGF étant calculés sur un échantillon de population sans spécificité apparente, nous avons décidé qu'ils n'étaient pas assez représentatifs d'une population d'individus invalides qui a normalement plus de probabilités de décéder.

Nous avons donc majoré les taux de mortalité de ces tables de 20%.

Les résultats des projections nous montrent qu'à partir des tables créées avant l'accélération de passage en invalidité, le nombre d'invalides est bien plus important qu'à partir des tables d'expérience 2015.

En confrontant les résultats à ceux obtenus pour la volumétrie des assurés en incapacité, nous pouvons constater que l'accélération du passage en invalidité dont nous avons fait l'étude tout au long de ce mémoire peut aussi être assimilée à une accélération générale des sorties de l'état d'incapacité.

Cette dernière a pour effet de diminuer la volumétrie des incapacités longues et donc de diminuer l'accumulation des individus dans cet état.

Tandis que l'accélération du passage en invalidité augmente la fréquence de passage aux jeunes anciennetés et la diminue aux anciennetés les plus élevées (voir tables de passage en invalidité 2014 et 2015).

Ce double effet constaté implique donc un nombre moins important d'invalides (avec les tables 2015) puisque les assurés en incapacité sont moins nombreux, donc il y a moins d'individus soumis au risque de passage en invalidité.

Par la suite, la projection des cotisations, obtenues en multipliant la population active d'un âge  $x$  par le montant moyen des cotisations annuelles que paie la population de cet âge, expose un montant de cotisations total plus élevé à partir des tables « post accélération ».

Nous projetons ensuite les charges « incapacité ». Ces charges sont obtenues en sommant les prestations « incapacité » et les provisions « incapacité ».

De même que pour la projection de la population en état d'incapacité, nous divisons les prestations et les provisions en 3 périodes d'un an : les prestations et les provisions pour des assurés restant un an en incapacité, pour des assurés restant 2 ans et pour des assurés restant 3 ans dans cet état.

Nous constatons que les prestations payées pour l'incapacité projetée à partir des tables 2014 sont bien plus élevées que celles payées à partir des tables 2015.

L'accélération analysée dans ce mémoire implique ainsi des bonis non-négligeables pour le portefeuille de prestations de l'assureur, estimés de l'ordre de 19% sur ce portefeuille.

Concernant les provisions mises en place par l'assureur pour faire face au risque, elles augmentent jusqu'en 2031 avant de diminuer sur les années suivantes.

Ce phénomène est expliqué par la croissance du nombre d'invalidités pendant cette période. La stabilisation des invalidités et la décroissance des incapacités dans les années 2030 impliquent la baisse des provisions pour cette décennie.

Comme pour les prestations, le gain estimé de l'accélération du passage en invalidité est lui de 4% par an en moyenne pour les provisions.

Les résultats nous confortent dans l'idée que l'accélération du passage en invalidité permettra aux assureurs de payer moins de charges et d'engranger en plus davantage de cotisations (les tables 2015 étant plus avantageuses que celles de 2014).

L'accélération a donc des effets positifs sur le portefeuille de l'assureur et cela peut être une des raisons qui ont poussé à simplifier les démarches pour sortir de l'état d'incapacité.

De plus, si la génération des baby-boomers semble avoir eu une influence sur les nombres croissants d'arrêts et d'invalidités observés ces dernières années, son passage progressif au-delà de la barre des 60 ans dans un futur proche aura pour conséquence logique de diminuer le nombre d'assurés en incapacité.

Si toutefois le nombre d'incapacités continue d'augmenter comme c'est le cas depuis 2012 mais sans cause épidémiologique cette fois, il sera nécessaire de s'interroger sur la possible dégradation de l'environnement social des assurés et particulièrement de la tranche de travailleurs la plus âgée.

Enfin, pour donner plus de valeur à ce travail, il sera nécessaire de le reprendre d'ici trois à quatre ans afin d'obtenir un recul indispensable à l'affirmation définitive des résultats obtenus ici.

# ***SUMMARY***

## *Context*

The objectives of provident schemes are to prevent and cover the risks that most affect an individual's life, namely death, health, unemployment, incapacity for work and disability.

If Social Security today guarantees financial benefits in the event of one of these risks, a person can choose to be covered by an individual and / or group contract in order to benefit from a replacement income which can take different forms: capital, pension, or basic daily allowances.

These contracts are taken out with insurance companies, which must ensure their financial equilibrium by pricing a premium, paid by the insured, which must at least be equal to the product of the probability of loss by the estimated cost of this loss. In addition, the insurer must provision a certain amount, estimated from statistics and probabilities, in order to always be able to face risks (accounting principle of prudence).

However, in providence (and more specifically for this incapacity / invalidity report), the cost of the claim and the provisions depend on the number of days spent by the insured person on sick leave and the time it takes to change state (incapacity to invalidity or incapacity to validity, etc.).

In order to fund the benefits provided in the event of incapacity and invalidity, insurers have the choice of using either « laws established by them and certified by an actuary independent of this company and approved by one of the associations actuaries recognized by the Prudential Control and Resolution Authority (ACPR) » (article 331-22 of the Insurance Code), i.e. the regulatory tables of the Common Bureau of Collective Insurance (BCAC). These latest provisioning laws, in the form of double-entry tables (time spent since entry and age), were entirely reconstructed in 2013, but it seems that since then, times spent on sick leave have changed again.

## *Problematic*

The behavior of the delay for changes of state seems indeed to have been modified and it was noted a reduction in the delays of passages in invalidity and variations of volumetries.

The 2013 tables no longer seem to "stick to reality" in the insurer's portfolios. This thesis thus seeks to rule on the true existence of this evolution and its potential causes. In addition, in the event that this reduction in disability transfer times is confirmed, this paper will develop new disability maintenance and disability transfer laws in order to better anticipate changes in the future state of the insured and observe the impacts that this acceleration could have on the insurer.

## *Validation of the acceleration assumption of the transition to disability*

### **1. Visualization of the phenomenon**

We therefore begin by analyzing the average time taken to become disabled by year of occurrence, and we observe that a strong trend is emerging. Deadlines are getting shorter, from over 18 months for those declared incapacitated in 2012 to just 12.6 months for those in 2016.

In addition, the shares of invalidity declared in the 36th month decrease sharply depending on the year (more than 6% of invalidity passages for insured persons recognized as incapables in 2012 occurred in the 36th month against only 1.5% for insured persons who entered in 2015).

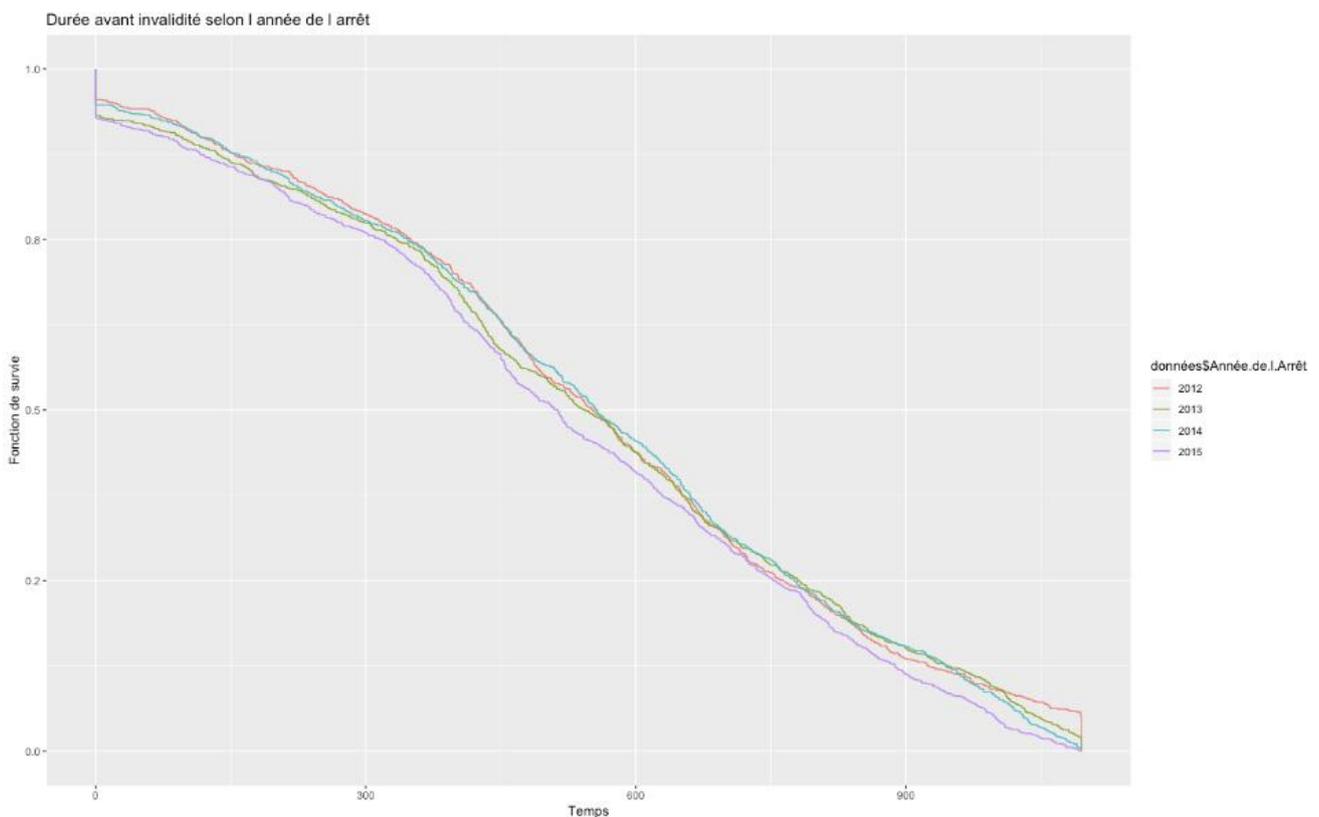
These results on the insurer's portfolios support those observed during Forsides' presentation on incapacity and invalidity risks on April 18, 2019, during which progress was noted on all of the Pension Fund portfolios the peak in disability for seven months between the years 2013-2014 and the years 2015-2016.

## 2. Analysis of delays in becoming disabled

In order to confirm the hypothesis of a behavior different from the invalidity transition times, we seek to model the time remaining before the transition from the state of incapacity to that of invalidity. To do this, we need to estimate a survival function. The survival analysis for living beings being a statistical study which seeks to model the time remaining before death (life expectancy), which is represented graphically in the form of a survival curve.

To estimate this survival function, it was chosen to use the Kaplan-Meier estimator. This estimator is non-parametric and is based on the fact that the survival period is discretized into sub-periods delimited by dates.

We have plotted the estimator for each year of occurrence and it is very clear that the insured declared disabled following a disability in 2015 are declared more quickly disabled than those who experienced a disability in previous years. The “2015” probabilities (purple curve below) that the changeover occurs at an instant  $t$  all seem lower than the probabilities of previous years which they seem to decrease at the same speed.



## 3. Modeling of densities

By plotting the densities of the disability transition times by year of occurrence, we note here again that the “2015” transition times seem to follow a distribution very different from other years.

In order to know if this distribution of the delays in the invalidity of the judgments in 2015 is indeed significantly different from the distributions of the delays in the 2012, 2013 and 2014 occurrences, we will model, using Gaussian mixtures, these rates of passages so as to compare, using statistical tests, the distributions of each year.

The Gaussian mixing model is a linear combination of several Gaussian components. This mixture evokes significant interest due to :

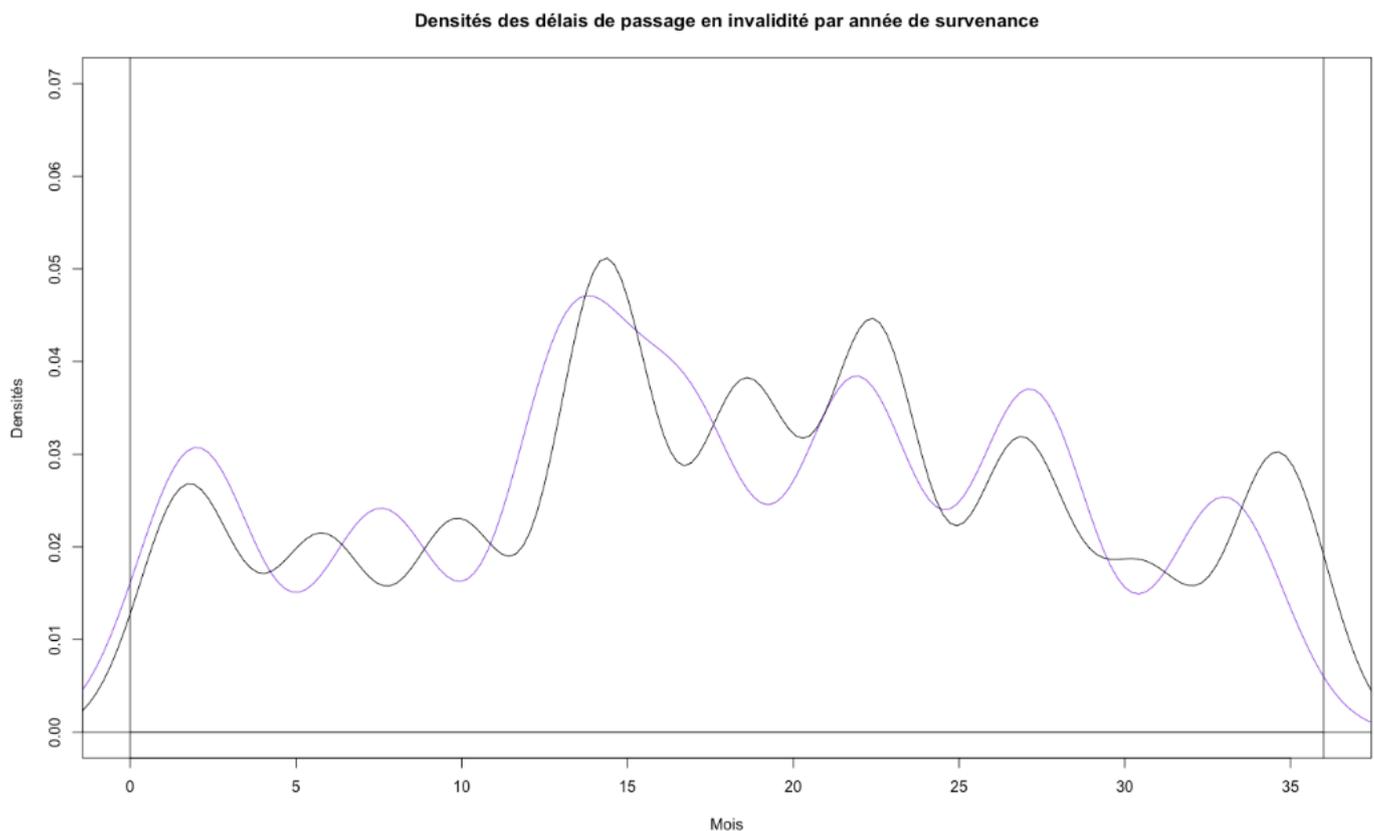
- Its flexibility
- Its faculty to approach a great variety of densities : one can always model continuous data by a Gaussian mixture.
- Its mathematically simple use and the generality of the normal law attested by the central limit theorem.

This mixture being parametric, we estimate the parameters by using a likelihood maximization approach.

However, maximizing likelihood cannot be done analytically due to hidden data. Solutions must therefore be found using iterative algorithms. Among these algorithms is the EM or “Expectation-Maximization” algorithm.

The algorithm guarantees that the likelihood increases with each iteration, which therefore leads to increasingly correct estimators.

We then started by adjusting a Gaussian mixing law to the densities of the disability transition times of the year of occurrence 2015 (in purple) as well as to the densities of the disability transition times of insured persons who became incapacitated before 2015 (in black) using the EM algorithm.



These curves show very explicitly the acceleration of the transition to disability since we are well aware of the shift to the left of each bell of the purple curve as well as the significant difference between the proportions of disability transition in the last months (the black curve being higher and more to the right than the violet).

These “visual” impressions are confirmed by Kolmogorov-Smirnov and Student tests which conclude that the delays in becoming disabled become faster, those of the insured who became incapacitated in 2015 being on average significantly lower than the delays in previous years. We therefore now need to create new incapacity maintenance and disability transfer tables in order to be able to measure the economic impact of this acceleration for the insurer.

## *Creation of disability maintenance and disability transition tables*

### **1. Estimate of gross rates**

We have chosen to use the Kaplan-Meier estimator for the construction of our laws. Indeed, the maintenance law is calculated from the conditional exit rates obtained by integrating the exits from each period to the initial population. However, the principle of this estimator is based on the classification of observations by observed durations spent in incapacity by counting outings.

After calculating the estimated rates, it is essential for us to check their reliability and robustness. We use the Greenwood variance for this. The latter is nothing other than the variance of the Kaplan-Meier estimator.

This variance highlights the lack of depth of our data with strong values for young ages as well as for ages over 60 years. The rates are therefore not sufficiently reliable for these ages.

We therefore decide to form age groups to obtain more reliable estimators of the survival functions. According to Aurélie Gaumet (2,000), 1,000 observations are necessary for the successful completion of a disability maintenance table.

So we decide to form the age groups so that there are always at least 1,000 observations in each of them. We then recalculate much more reliable Kaplan-Meier estimators.

The next step is to smooth out these survival functions to clear up any slight irregularities that exist.

### **2. Smoothing of Whittaker-Henderson**

The smoothing method chosen is that of Whittaker-Henderson. This non-parametric smoothing is a simple tool to regularize a series of gross or estimated rates.

The method is based on two criteria : a fidelity criterion  $F$  between the reference curve and the smoothed curve, and a regularity criterion  $S$ , which depends on a parameter  $z$ . It makes it possible to obtain the values which minimize the linear combination  $M$  of these two criteria:

$$M = F + h * S,$$

Where  $h$  is a parameter of the model controlling the weight of the smoothness regularity criterion. Just like  $z$ , degree of the polynomial to which the smoothed curve will adjust.

The choice of  $h$  and  $z$  was made empirically (we varied them to find their optimal values).

### **3. Interpolation and extrapolation**

To interpolate the rates, we start by calculating the average age of the classes completed before. Then we consider that the estimated survival functions already smoothed are those of these average ages.

Finally, we use a linear interpolation method to obtain the rates between our two extreme mean ages.

It remains to extrapolate retention rates at extreme ages. We then choose, as for the interpolation, to proceed to a linear extrapolation for the incapacity maintenance table, and by using coefficients of proportionality obtained compared to the BCAC rates for the invalidity transition table (to compensate for the lack information that we have, especially at young ages).

## *Projection and visualization of the impacts of acceleration on a portfolio*

To measure the impact of the acceleration observed, we decided to project a portfolio of policyholders.

To do this, we use our laws calculated above, a law of incidence of incapacity internal to the insurer, and the mortality laws of INSEE calculated in 2018 (TGH / TGF). These allow us to calculate a single rate for both sexes, knowing that our portfolio is made up of 99.5% women.

We also calculate the entry rates into the portfolio to generate new members for each age between 20 and 60. To do this, we use current portfolio data and the INSEE 2019 age pyramid. Finally, the disability maintenance table chosen is that of the BCAC.

We start by projecting the active population and that present in incapacity. If, in an ideal world, we would project our policyholders each month to follow their evolution and that of our charges and contributions more closely, here we simplify the model and we will follow the evolution of our variables year by year only for the sake of clarity and readability.

For this, we have broken down the insured in disability into three: those in their first year of maintenance, those in their second year and those in their third year.

The results show that the total valid population of our portfolio will only decrease during the next 20 years which is explained by the gradual retirement of the baby boomer generation since the beginning of the 2010s. The observation is naturally the same for the number of disabilities.

The projection of this portfolio with the "2015" tables, then with the "before 2015" tables, marks the clear influence of the acceleration of disability exit times.

We then project the disabled population on the premise that an insured cannot go from the "valid" state to that of the "invalid" and neither from the disabled state to those of validity and incapacity. This implies that the only way out of the state of disability is the death of the insured. The death rates from the TGH / TGF tables being calculated on a sample of population with no apparent specificity, we decided that they were not sufficiently representative of a population of disabled individuals who are normally more likely to die.

We have therefore increased the mortality rates of these tables by 20%.

The results of the projections show that from the tables created before the acceleration into disability, the number of disabled is much greater than from the experience tables 2015.

By comparing the results with those obtained for the volume of disabled insured persons, we can see that the acceleration of the transition to invalidity, which we have studied throughout this thesis, can also be assimilated to a general acceleration of retirement from state of incapacity.

The latter has the effect of decreasing the volume of long-term disabilities and therefore decreasing the accumulation of individuals in this state.

While the acceleration of the transition to disability increases the frequency of transition to young seniority and decreases it to the highest seniority (see disability transition tables 2014 and 2015).

This double effect observed therefore implies a lower number of disabled people (with the 2015 tables) since the disabled are fewer, so there are fewer individuals subject to the risk of becoming disabled.

Afterwards, the projection of contributions, obtained by multiplying the working population of an age  $x$  by the average amount of annual contributions paid by the population of this age, shows a higher total contribution amount from the "post acceleration" tables.

We then project the "incapacity" charges. These charges are obtained by claiming "disability" benefits and "disability" provisions.

As with the projection of the incompetent population, we divide benefits and provisions into three one-year periods: benefits and provisions for insured staying one year incapacitated, for policyholders remaining for 2 years and for insured remaining 3 years in this state.

We find that the disability benefits projected from the 2014 tables are much higher than those paid from the 2015 tables.

The acceleration analyzed in this brief thus implies non-negligible bonuses for the insurer's benefit portfolio, estimated at around 19% on this portfolio.

With regard to the provisions put in place by the insurer to deal with the risk, they increase until 2031 before decreasing over the following years.

This is explained by the increase in the number of disabilities during this period. The stabilization of disabilities and the decrease in disability in the 2030s mean that provisions will fall for this decade.

As with benefits, the estimated gain from the acceleration of the disability transition is 4% per year on average for provisions.

The results support the idea that the acceleration of the disability transition will allow insurers to pay less expenses and earn more premiums (the 2015 tables being more advantageous than those of 2014).

The acceleration therefore has a positive effect on the insurer's portfolio and this may be one of the reasons why the process of exiting the disability state has been simplified.

Moreover, while the baby boomer generation appears to have had an influence on the increasing numbers of stoppages and disabilities observed in recent years, its gradual move beyond the 60-year mark in the near future will logically result in reduce the number of incapacitated.

If, however, the number of disabilities continues to increase as has been the case since 2012 but without an epidemiological cause this time, it will be necessary to consider the possible deterioration of the social environment of the insured and in particular of the oldest group of workers.

Finally, to give more value to this work it will be necessary to resume it in three to four years in order to obtain a step back essential to the definitive affirmation of the results obtained here.

# SOMMAIRE

<b>REMERCIEMENTS</b> .....	<b>2</b>
<b>RESUME</b> .....	<b>3</b>
<b>ABSTRACT</b> .....	<b>4</b>
<b>NOTE DE SYNTHÈSE</b> .....	<b>5</b>
<b>SUMMARY</b> .....	<b>11</b>
<b>SOMMAIRE</b> .....	<b>17</b>
<b>INTRODUCTION ET PROBLEMATIQUE</b> .....	<b>19</b>
<b>PARTIE I : LE RISQUE DE PASSAGE EN INVALIDITE</b> .....	<b>21</b>
<b>CHAPITRE 1. LA PREVOYANCE EN FRANCE</b> .....	<b>21</b>
<b>I.1.1. Définition</b> .....	<b>22</b>
<b>I.1.2. Les régimes de prévoyance</b> .....	<b>22</b>
I.1.2.1. Le régime de base de la Sécurité Sociale : la prévoyance étatique .....	22
I.1.2.2. Le régime complémentaire obligatoire qui comprend les obligations légales et conventionnelles de l'employeur .....	25
<b>I.1.3. La prévoyance « privée » en chiffres</b> .....	<b>28</b>
<b>CHAPITRE 2. LA GARANTIE « INVALIDITE »</b> .....	<b>30</b>
<b>I.2.1. Le contexte sociodémographique</b> .....	<b>30</b>
<b>I.2.2. Définition de l'invalidité</b> .....	<b>31</b>
<b>I.2.3. Les trois catégories d'invalidité</b> .....	<b>31</b>
I.2.3.1. Le montant de la pension du régime de base général de la Sécurité Sociale .....	31
I.2.3.2. Les conditions d'attribution de la pension d'invalidité .....	32
<b>I.2.4. La possibilité de couverture complémentaire</b> .....	<b>33</b>
<b>CHAPITRE 3. LE PROVISIONNEMENT DU RISQUE « INVALIDITE »</b> .....	<b>34</b>
<b>I.3.1. Définition</b> .....	<b>34</b>
<b>I.3.2. Les provisions</b> .....	<b>34</b>
<b>I.3.3. Le provisionnement pour les rentes d'invalidité</b> .....	<b>34</b>
<b>PARTIE II : LE PERIMETRE D'ETUDE</b> .....	<b>36</b>
<b>CHAPITRE 1. PRESENTATION DES DONNEES</b> .....	<b>36</b>
<b>II.1.1. Vision « stock »</b> .....	<b>37</b>
II.1.1.1. Les variables disponibles.....	37
II.1.1.2. Fiabilisation des données.....	37
<b>II.1.2. Vision « flux »</b> .....	<b>38</b>
II.1.2.1. Les variables disponibles.....	38
II.1.2.2. Fiabilisation des données.....	38
<b>CHAPITRE 2. STATISTIQUES DESCRIPTIVES</b> .....	<b>40</b>
<b>II.2.1. La vision « stock »</b> .....	<b>40</b>
II.2.1.1. Les assurés en incapacité .....	40
II.2.1.2. Les assurés en invalidité .....	42
II.2.1.3. Taux de passage en invalidité .....	44
II.2.1.4. Délai de passage en invalidité .....	45
II.2.2.1. Les assurés en incapacité.....	46
II.2.2.2. Les assurés en invalidité .....	48
II.2.2.3. Taux de passage en invalidité .....	49
II.2.2.4. Délai de passage en invalidité.....	50
<b>PARTIE III : ETUDES STATISTIQUES SUR LES ASSURES PASSANT EN INVALIDITE</b> .....	<b>52</b>
<b>CHAPITRE 1. ANALYSE DES DELAIS DE PASSAGE EN INVALIDITE</b> .....	<b>52</b>
<b>III.1.1. Fonction de survie et Estimateur</b> .....	<b>52</b>

III.1.1.1. Fonction de survie.....	52
III.1.1.2. Estimateur de Kaplan-Meier .....	52
<b>III.1.2. Histogrammes et densités .....</b>	<b>54</b>
<b>III.1.3. Modélisation des densités .....</b>	<b>56</b>
III.1.3.1. Loi de Weibull .....	56
III.1.3.2. Loi Normale .....	58
<b>CHAPITRE 2. LOI MULTIMODALE : MELANGE DE LOIS GAUSSIENNES .....</b>	<b>63</b>
<b>III.2.1. Théorie .....</b>	<b>63</b>
<b>CHAPITRE 3. CREATION DES TABLES DE MAINTIEN EN INCAPACITE ET DE PASSAGE EN INVALIDITE .....</b>	<b>70</b>
<b>III.3.1. Table de maintien en incapacité .....</b>	<b>70</b>
III.3.1.1. Vérification des résultats précédents sur ce jeu de données.....	70
III.3.1.2. Création de la table de maintien en incapacité .....	73
<b>III.3.2. Table de passage en invalidité.....</b>	<b>84</b>
III.3.2.1. Estimateurs de Kaplan-Meier des taux bruts et variances de Greenwood.....	84
III.3.2.2. Lissage de Whittaker-Henderson.....	88
III.3.2.3. Interpolation et extrapolation .....	89
<b>PARTIE IV : PROJECTION D'UN PORTEFEUILLE D'ASSURES A L'AIDE DES TABLES D'EXPERIENCE .....</b>	<b>94</b>
<b>CHAPITRE 1. LE PORTEFEUILLE UTILISE.....</b>	<b>94</b>
<b>IV.1.1. Le portefeuille, ses caractéristiques.....</b>	<b>94</b>
<b>IV.1.2. Quelques statistiques descriptives.....</b>	<b>95</b>
<b>IV.1.3. Les lois utilisées.....</b>	<b>96</b>
<b>CHAPITRE 2. PROJECTIONS DES POPULATIONS A PARTIR DES TABLES D'EXPERIENCE 2014 ET 2015.....</b>	<b>98</b>
<b>IV.2.1. Projection de la population active et en état d'incapacité.....</b>	<b>98</b>
<b>IV.2.2. Projection de la population invalide.....</b>	<b>103</b>
<b>CHAPITRE 3. PROJECTIONS DES COTISATIONS ET DES CHARGES A PARTIR DES TABLES D'EXPERIENCE 2014 ET 2015.....</b>	<b>106</b>
<b>IV.3.1. Projection des cotisations.....</b>	<b>106</b>
<b>IV.3.2. Projection des charges « incapacité » .....</b>	<b>107</b>
<b>CONCLUSION .....</b>	<b>111</b>
<b>BIBLIOGRAPHIE .....</b>	<b>113</b>

# ***INTRODUCTION ET PROBLEMATIQUE***

Obtenir une compensation à la perte de salaire due à un arrêt de travail en cas d'incapacité ou d'invalidité est l'une des garanties fondamentales en France. La Sécurité Sociale verse dans ce but des prestations financières sous forme d'indemnités journalières dans la limite de 48,08 € par jour.

Mais, additionnellement, une personne peut être couverte dans son entreprise par un contrat collectif de prévoyance et/ou peut choisir d'être couvert par un contrat individuel afin de permettre le maintien partiel ou total de sa rémunération. Ces contrats sont souscrits auprès d'assurances qui doivent assurer leur équilibre financier en tarifiant une prime, payée par l'assuré, devant au moins être égale au produit de la probabilité de sinistre par le coût estimé de ce sinistre. De plus, l'assureur se doit de provisionner un certain montant, estimé à partir de statistiques et de probabilités, afin de toujours pouvoir faire face aux risques (principe comptable de prudence).

Or, en prévoyance (ici incapacité/invalidité), le coût du sinistre et les provisions dépendent du nombre de jours passés par l'assuré en arrêt de travail et du temps qu'il met à changer d'état (d'incapacité à invalidité ou d'incapacité à validité...).

Des lois de provisionnement, sous forme de tables à double entrée (temps passé depuis l'entrée et âge), ont donc été construites pour provisionner les prestations servies en cas d'incapacité et d'invalidité. Ces lois ont été entièrement reconstruites en 2013 mais il semble que depuis, les temps passés en arrêt de travail aient évolué.

Le comportement des délais de changements d'états semble en effet, avoir été modifié, et il a été remarqué une diminution des délais de passages en invalidité et des variations de volumétries.

La présentation de Forsides sur les risques incapacité et invalidité du 18 avril 2019 à laquelle j'ai assisté, a renforcé cette impression d'accélération du passage en invalidité remarqué sur nos portefeuilles : il a été constaté sur l'ensemble des portefeuilles d'Institution de Prévoyance un décalage du pic de passage en invalidité de -7 mois entre les années 2013-2014 et les années 2015-2016.

*Nous avons donc été amenés à étudier plus en détail ce phénomène, afin de pouvoir nous prononcer sur la véritable existence de cette évolution, ses causes potentielles et les impacts économiques que cette accélération pourrait avoir sur l'assureur.*

Pour ce faire, nous commencerons par rappeler les caractéristiques du risque « invalidité » dans son contexte sociodémographique actuel.

Puis, nous expliciterons le périmètre de notre étude, nous fiabiliserons les données afin d'avoir une base solide pour la mener à bien, et nous calculerons de premières statistiques qui viendront décrire notre portefeuille et ses caractéristiques et nous donner des premiers éléments de réponse.

Par la suite, nous modéliserons la durée des incapacités ayant entraîné une invalidité à l'aide d'un mélange de lois gaussiennes.

Les résultats de cette troisième partie nous pousseront à faire un choix sur la création ou non de nouvelles tables d'expérience, afin que l'assureur puisse pallier les risques que cette accélération pourrait avoir sur ses sinistres, ses prestations et donc aussi sur ses provisions.

C'est pourquoi, en cas de construction de nouvelles tables d'expérience (à partir d'une estimation des taux bruts grâce à l'estimateur de Kaplan-Meier, un lissage de Whittaker-Henderson, une extrapolation et une

interpolation), nous tenterons, dans une dernière partie, de projeter un portefeuille d'assurés pour visualiser les variations que ces changements de tables auraient sur le provisionnement de l'assureur, et ainsi déterminer si les variations des charges à payer à court terme vont se prolonger ou s'atténuer à plus long terme.

Les résultats que nous fournirons tout au long de ce mémoire nous permettront également de réfléchir et de proposer des hypothèses sur les raisons de ces changements de comportements des délais en arrêt de travail, comme par exemple l'arrivée massive à l'âge de la retraite des baby-boomers (1946-1980) âgés en 2012 de 36 à 66 ans, qui disparaissent donc progressivement du paysage de la prévoyance en France.

Cependant, il paraît important de souligner certains points qui pourraient venir nuancer les résultats de ce mémoire.

Un premier point concerne la population sélectionnée. En effet, la population choisie pour vérifier le phénomène d'accélération du passage en invalidité et pour construire nos lois d'expérience est très ciblée (aides à domicile). Cela nous amène naturellement à nous interroger sur la représentativité de ce corps de métier par rapport à la population globale. Une population à plus forte majorité masculine ou plus « à risque » (d'accidents notamment) aurait pu donner une étude et des conclusions différentes.

Pour autant, le fait qu'une étude, menée par un cabinet de conseil (Forsides)<sup>i</sup> sur une population beaucoup plus globale, arrive aux mêmes conclusions vient renforcer notre étude et ses résultats.

Un deuxième point concerne la taille de notre échantillon et surtout le recul dont nous disposons. Ce faible passif a logiquement joué sur le nombre de passages en invalidité observés peu élevé. Nous avons donc trouvé des stratégies d'adaptation (grouper par classes d'âge, coefficients de proportionnalité pour extrapoler).

Enfin, un dernier point d'attention sur certaines hypothèses prises dans ce mémoire desquelles dépendent les résultats. Ainsi, des choix d'estimateurs (Kaplan-Meier, méthode de lissage de Whittaker-Henderson) ont été faits et ces estimateurs auraient pu être mis en concurrence mais nous n'avons pas eu le temps nécessaire pour finaliser totalement cette étude.

Bien entendu, des idées, des pistes qui permettraient avec plus de temps de challenger et ainsi d'améliorer la fiabilité des analyses, et des modélisations de cette étude seront retranscrites tout au long du mémoire.

# ***PARTIE I : LE RISQUE DE PASSAGE EN INVALIDITE***

## ***Chapitre 1. La Prévoyance en France***

« *La prévoyance ne doit pas stériliser l'avenir, elle doit préparer ses voies* » (Tristan BERNARD).

Ainsi, être prévoyant, c'est faire en sorte d'être garanti à un moment de la vie où tout va bien, contre des événements que l'on ne peut encore prévoir mais qui peuvent survenir à tout moment.

Il existe naturellement plusieurs types de garanties et plusieurs types d'événements à garantir.

Mais ceux qui sont les plus susceptibles de bouleverser la vie d'un individu, et qui par essence sont les plus imprévisibles et partant les plus importants à garantir, sont ceux affectant la santé, tels le décès, l'invalidité (qui fera l'objet de notre mémoire), l'incapacité de travail, ou encore les frais médicaux occasionnés.

Pour ces risques-là dont les conséquences peuvent parfois être dramatiques, dès lors qu'ils affectent de manière significative l'équilibre budgétaire d'une famille, le législateur a été conduit à intervenir en instituant certains régimes de prévoyance obligatoires.

Le plus connu d'entre eux est la sécurité sociale, fondée sur la solidarité nationale (article L111-1 du code de la sécurité sociale) et sur lequel nous reviendrons, s'agissant du régime de base.

Mais il existe également des régimes complémentaires qui ont été rendus pour certains obligatoires (notamment par les CCN), afin de pallier une éventuelle insuffisance du régime général.

Ainsi, en prenant l'exemple des huissiers de justice, en vue de compléter les IJSS plafonnées à 60% du salaire annuel dans une limite de 1/730<sup>ème</sup> du PASS, qui ne leur permet pas de maintenir leur niveau de vie, leur CCN rend obligatoire la souscription par l'employeur d'une prévoyance complémentaire auprès de la CARCO qui va venir compléter les IJSS pour garantir 95% du brut pendant 1095 jours.

Ces régimes tendent tous au même but : permettre aux personnes qui se retrouvent dans l'incapacité temporaire ou définitive de travailler, [ou à leurs ayant-droits (veuve, et/ ou enfants en cas de décès prématuré)], de pouvoir bénéficier d'un revenu de remplacement qui peut prendre différentes formes : capital, pension, ou indemnités journalières de base.

Nous verrons toutefois que ce revenu de remplacement tiré de la mise en œuvre des régimes obligatoires n'est pas toujours suffisant pour permettre de faire face aux obligations matérielles et financières de l'assuré ou de sa famille.

C'est dans ce contexte que des organismes privés, telles les compagnies d'assurance, les Institutions de Prévoyance, les Mutuelles, ou les banques, ont été amenés à développer et proposer des produits qui viennent renforcer la protection de la personne et la garantir suffisamment pour qu'en sus des problèmes de santé qu'elle peut être amenée à rencontrer, elle ne soit pas également contrainte de faire face à une perte de revenus, qui pourrait la placer, sa famille ou elle-même, en situation de précarité.

Pour que ces produits ne fassent pas double emploi avec les régimes légaux obligatoires de protection, il apparaît nécessaire pour la personne qui entend y recourir, de faire préalablement procéder à un bilan personnalisé.

Il pourra alors lui être proposé le produit complémentaire individuel le plus adapté.

Mais pour établir le profil, la nécessité et la rentabilité des produits à développer, les assurances/mutuelle/banques, principaux organismes privés concernés, doivent préalablement connaître les régimes de prévoyance déjà existants, et leur fonctionnement.

C'est l'objet plus précisément, du présent chapitre.

## I.1.1. Définition

La prévoyance est généralement définie par l'article 1 de la loi n°89-1009 de la loi du 31 décembre 1989, dite Loi Évin, de la manière suivante : il s'agit des « opérations ayant pour objet la prévention et la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque chômage ».

La loi Évin est intervenue spécifiquement dans le cadre des contrats collectifs de prévoyance. Elle impose ainsi aux assureurs, certaines obligations à l'égard de leurs assurés sur lesquelles nous reviendrons plus avant, et qui visent essentiellement à éviter aux salariés un défaut de couverture sur les risques liés à la santé.

De cette définition, et de manière générale, nous pouvons considérer que la prévoyance représente l'ensemble des contrats individuels ou collectifs, obligatoires ou facultatifs, qu'une personne a souscrit ou auxquels elle a adhéré auprès d'organismes publics ou privés, pour faire face aux événements auxquels elle pourrait se retrouver confrontée dans sa vie personnelle ou professionnelle et qui la priverait de manière temporaire ou définitive, de la faculté de travailler.

## I.1.2. Les régimes de prévoyance

Le système de protection sociale français repose sur 3 niveaux :

- Le régime de base,
- Le régime complémentaire obligatoire qui comprend les obligations légales et conventionnelles de l'employeur,
- La prévoyance individuelle.

Schématiquement, ces niveaux peuvent être représentés ainsi qu'il suit :



### I.1.2.1. Le régime de base de la Sécurité Sociale : la prévoyance étatique

Il existe deux types de régime de base (le régime général et les régimes spéciaux) dont le rôle est défini par l'article L111-1 du code de la sécurité sociale :

« Elle assure, pour toute personne travaillant ou résidant en France de façon stable et régulière, la couverture des charges de maladie, de maternité et de paternité ainsi que des charges de famille.

Elle garantit les travailleurs contre les risques de toute nature susceptible de réduire ou de supprimer leurs revenus. Cette garantie s'exerce par l'affiliation des intéressés à un ou plusieurs régimes obligatoires.

*Elle assure la prise en charge des frais de santé, le service des prestations d'assurance sociale, notamment des allocations vieillesse, le service des prestations d'accidents du travail et de maladies professionnelles ainsi que le service des prestations familiales dans le cadre du présent code, sous réserve des stipulations des conventions internationales et des dispositions des règlements européens. »*

Après avoir présenté ce régime, nous présenterons son fonctionnement.

### **I.1.2.1.A. Présentation du régime de base de la Sécurité Sociale**

a) **Le régime de base dit général** concerne en France plus de 80% de la population.

Il réunit les salariés du privé, leurs ayant droits, mais également les bénéficiaires de l'Allocation Adulte Handicapé, ou encore de la Couverture Maladie Universelle.

b) **Les régimes spéciaux, à savoir :**

- **Sécurité Sociale Indépendants** (ex : Régime Social des Indépendants – RSI) qui concerne tous les travailleurs indépendants à l'exception des travailleurs agricoles.

Ce régime spécial englobe :

- Les artisans et commerçants non-salariés ainsi que leurs conjoints collaborateurs déclarés à ce titre
- Les professions libérales
- Certains chefs d'entreprise.

- **La Mutualité Sociale Agricole (MSA)**

Ce régime englobe les exploitants et salariés agricoles.

- **Les salariés appartenant à certains corps de métiers** et qui, à ce titre ne relèvent pas du régime général.

Ceux-ci bénéficient d'ailleurs de prestations plus avantageuses.

Il s'agit notamment :

- Des agents des Industries Électriques et Gazières ou encore de la RATP
- Des militaires
- Des fonctionnaires
- Des ministres du culte.

- **Les Étudiants**

Il s'agit des jeunes majeurs qui poursuivent des études supérieures et dont le régime de protection a été confié aux Mutuelles Étudiantes, lesquelles sont des organismes privés.

### **I.1.2.1.B. Fonctionnement du régime de base général de la Sécurité Sociale**

Qu'il s'agisse du régime général ou des régimes spéciaux, ces organismes sont alimentés par le paiement de cotisations sociales obligatoires calculés en pourcentage de leurs revenus.

Concernant plus précisément les salariés, les cotisations sont financées pour partie par l'employeur, pour partie par le salarié (rappelons que le montant moyen des charges sociales retenues aux salariés est de 26%).

Le régime général couvre tous les risques de la personne : décès, invalidité, incapacité temporaire de travail, maladie, vieillesse.

### a) Le décès du salarié assuré

Il est ici versé aux ayants droits de l'assuré diverses prestations destinées à leur permettre de faire face aux conséquences financières que le décès de leur conjoint, père ou mère va avoir sur leur quotidien.

Ces prestations concernent :

- Le versement d'un **capital décès** non imposable.

Le montant de ce capital est soumis à différentes conditions tenant notamment à la personne même de l'assuré (activité salariée ou situation de maintien de droits dans les 3 mois précédant le décès, être bénéficiaire d'une pension d'invalidité ou d'une rente d'incapacité).

Son montant maximum au 1<sup>er</sup> janvier 2019 était de 10.131 € (avec un minimum de 405,24 €).

- **L'allocation veuvage** d'un montant de **607,54 €** mensuel.

Elle nécessite que le conjoint survivant ait moins de 55 ans. Ce montant peut être réduit en fonction des revenus.

- La **pension de réversion** qui représente 54% de la pension de retraite du défunt.

Pour pouvoir en bénéficier, le conjoint survivant doit avoir plus de 55 ans et des ressources inférieures à **20.300,80 €** (plafond 2017).

### b) L'incapacité temporaire de travail

L'incapacité temporaire de travail concerne la période pendant laquelle un salarié est inapte à exercer la moindre activité professionnelle des suites d'un accident ou d'une maladie.

Pendant cette période, celui-ci va donc percevoir des indemnités journalières qui lui seront versées par la sécurité sociale soit directement, soit, par le mécanisme de la subrogation en cas de maintien de salaire prévu par une prévoyance collective complémentaire, à son employeur, à charge pour celui-ci de les lui reverser :

- A compter du 4<sup>ème</sup> jour d'arrêt de travail
- Jusqu'au 1095<sup>ème</sup> jour en cas d'affection longue durée.

Le salaire journalier de référence pris en compte par la sécurité sociale pour le calcul des indemnités journalières est le salaire brut moyen des 3 derniers mois divisé par 30,42.

Le montant de l'indemnité journalière est de 60% du salaire journalier de base avec un montant maximum de 202,78 € pendant les 28 premiers jours.

A partir du 29<sup>ème</sup> jour, elle est portée à 80% avec un montant maximum plafonné à 270,37 €.

Elle cesse d'être versée au 1095<sup>ème</sup> jour (soit après 3 ans à date anniversaire de l'arrêt de travail initial).

Dès à présent, il ressort de ces premières observations d'ordre très général, que le régime de base obligatoire de la sécurité sociale ne permet pas au salarié qui n'est plus en mesure de travailler, de maintenir une équivalence de revenus.

Relevons à ce titre, que la différence de revenus de substitution est encore plus flagrante pour les indépendants à qui la SSI ne permet de percevoir qu'une indemnité journalière égale à 1/730<sup>ème</sup> des revenus moyens des trois dernières années civiles.

Ainsi, pour un revenu annuel moyen sur les 3 dernières années civiles de 42.000 €, l'indépendant ne percevra de la SSI qu'une indemnité journalière de  $42.000 \times 1/730 = 57,53$  €, ou encore ramenée au mois : 1.750 €, et à l'année civile : 21.002 €... soit une perte sèche de 50% de ses revenus.

Compte tenu de l'endettement français moyen, la prévoyance du régime de base s'est avérée insuffisante, ce qui a conduit à la mise en place de prévoyances complémentaires, par voie d'adhésion à un contrat collectif ou de souscription individuelle.

### c) L'invalidité permanente

Le passage en invalidité intervient soit lorsque la capacité de travail de l'assuré est réduite de 2/3 lors de la stabilisation ou de la consolidation de l'état de l'assuré après une maladie ou un accident d'origine non professionnelle, soit à l'expiration de la période de versement des indemnités journalières.

Il existe trois catégories d'invalidité. C'est au médecin conseil de la CPAM qu'il appartiendra de placer l'assuré en remplissant les conditions, dans l'une ou l'autre de ces catégories.

Catégorie d'invalidité en fonction de la situation du demandeur	
Catégorie	Situation
1-catégorie	Invalides capables d'exercer une activité rémunérée
2-catégorie	Invalides absolument incapables d'exercer une profession quelconque
3-catégorie	Invalides qui, étant absolument incapables d'exercer une profession, sont, en outre, dans l'obligation d'avoir recours à l'assistance d'une tierce personne pour effectuer les actes ordinaires de la vie

S'agissant ici d'une partie se rapportant plus précisément et directement à notre mémoire, nous la développerons plus avant au chapitre I.2.1 auquel il sera renvoyé.

### I.1.2.2. Le régime complémentaire obligatoire qui comprend les obligations légales et conventionnelles de l'employeur

Le régime de base de la sécurité sociale ne permettant pas de conserver un niveau de vie suffisant, comme il vient d'être démontré, il a été imposé aux employeurs, soit par le législateur (obligations légales), soit par les partenaires sociaux au travers d'accords de branches (obligations conventionnelles), de souscrire au profit de leurs salariés à des régimes de protections complémentaires qui vont permettre d'assurer à ces derniers des indemnités complémentaires à celles versées par le régime de base de la sécurité sociale.

On parle ici de prévoyance collective confiée par les employeurs à des organismes privés, laquelle prévoyance collective se trouve définie par l'article L141-1 du code des assurances de la manière suivante :

*« Est un contrat d'assurance de groupe le contrat souscrit par une personne morale ou un chef d'entreprise en vue de l'adhésion d'un ensemble de personnes répondant à des conditions définies au contrat, pour la couverture des risques dépendant de la durée de la vie humaine, des risques portant atteinte à l'intégrité*

*physique de la personne ou liés à la maternité, des risques d'incapacité de travail ou d'invalidité ou du risque de chômage. Les adhérents doivent avoir un lien de même nature avec le souscripteur. »*

### **I.1.2.2.A. Le régime complémentaire obligatoire légal**

#### **a) Mensualisation**

La loi de mensualisation du 19 janvier 1978 est venue faire obligation aux entreprises de garantir leurs salariés ayant plus de trois ans d'ancienneté en leur apportant une compensation supplémentaire aux indemnités journalières octroyées par le régime général de la sécurité sociale avec un délai de carence de 10 jours.

La loi du 25 juin 2008 a ramené la condition d'ancienneté à 1 an quelle que soit la catégorie du salarié (cadre ou non cadre) et réduit le délai de carence à 7 jours.

Si l'arrêt de travail est consécutif à une maladie ou un accident d'origine professionnelle, le délai de carence est supprimé.

En application de cette loi, les salariés concernés auront une garantie de 90% de leur rémunération brute durant les 30 premiers jours d'arrêt de travail, et 66,66% de celle-ci pendant les trente jours suivants.

Chacune de ces périodes d'indemnisation sera augmentée de 10 jours, tous les 5 ans d'ancienneté sans pouvoir excéder 90 jours, soit une garantie complémentaire obligatoire d'une durée totale maximale de 180 jours.

Elle peut se présenter sous le tableau suivant :

Ancienneté	Maintien de 90% de la rémunération brute pendant	Maintien de 66,66% de la rémunération brute pendant
Entre 1 et 5 ans	30 jours	30 jours
Au moins 6 ans	40 jours	40 jours
Au moins 11 ans	50 jours	50 jours
Au moins 16 ans	60 jours	60 jours
Au moins 21 ans	70 jours	70 jours
Au moins 26 ans	80 jours	80 jours
Au moins 31 ans	90 jours	90 jours

#### **b) L'obligation de cotiser à une retraite complémentaire**

L'employeur a l'obligation d'affilier ses salariés à une caisse de retraite complémentaire et d'y cotiser.

Il existe plusieurs tranches de cotisations en fonction du montant du salaire et en se référant au plafond de la sécurité sociale.

L'AGIRC est la caisse de retraite obligatoire des cadres depuis 1947.

Elle a fusionné avec l'ARCCO au 1<sup>er</sup> janvier 2019.

Les caisses de retraite complémentaires sont fondées sur un principe de solidarité à l'instar du régime général de la sécurité sociale, et fonctionnent selon un système de points.

Ainsi les cotisations versées pour le compte du salarié sont converties en points.

Au moment de la liquidation de ses droits à retraite, le salarié va bénéficier d'un complément de retraite égal au nombre de points cotisés multiplié par la valeur du point à ce jour.

### **c) L'obligation de proposer une mutuelle complémentaire**

La loi ANI du 14 juin 2013 impose à tous les employeurs de souscrire une complémentaire santé au profit de leurs salariés comprenant un socle de garanties minimales qui sont définies par l'alinéa 2 de l'article L911-7 du code de la sécurité sociale.

Cette obligation s'applique à toutes les entreprises et donc à tous les salariés sans aucune distinction depuis le 1<sup>er</sup> janvier 2016.

### **I.1.2.2.B. Le régime complémentaire obligatoire conventionnel**

Il s'agit d'un régime complémentaire obligatoire librement négocié avec les partenaires sociaux dans le cadre des conventions collectives nationales ou des accords de branche étendus.

Ce régime vient en complément du régime de base de la sécurité sociale et assure des garanties plus avantageuses au salarié.

#### **a) La mise en place du régime complémentaire conventionnellement obligatoire**

Le contrat d'assurance collective qui va être souscrit intervient entre les acteurs suivants :

- L'assureur – organisme privé d'assurance ou de réassurance
- Le souscripteur (l'employeur)
- Les assurés (les salariés)
- Les bénéficiaires (les assurés ou leurs ayant-droits ou même un tiers)

Il est mis en place simultanément au contrat de travail.

#### **b) Les garanties communément souscrites**

Comme le régime de base, le régime complémentaire va prendre en charge la couverture de certains risques liés à l'incapacité temporaire ou définitive de travail du salarié mais cette fois, dans la limite de ce qui aura été négocié en amont par les partenaires sociaux.

Ainsi, contrairement au régime général de la sécurité sociale qui octroie une protection identique pour tous les salariés, l'étendue de la protection des régimes complémentaires négociés sont fonction des accords intervenus.

Ils comprennent quoiqu'il en soit :

- Une garantie décès en capital ou en cas de Perte Totale et Irréversible d'Autonomie (PTIA)
- Une garantie quant au maintien du salaire en cas d'arrêt de travail pour maladie ou accident

Cette garantie est susceptible d'intervenir pendant toute la durée d'indemnisation par la sécurité sociale (soit jusqu'au 1095<sup>e</sup> jour) et même au-delà si le salarié se trouve ensuite placé en invalidité et bénéficie à ce titre d'une rente.

Le versement du maintien de salaire cesse au plus tard au jour de la liquidation par le salarié de ses droits à la retraite.

Il est noté, ce qui est particulièrement avantageux, que depuis la loi EVIN du 31 décembre 1989, les prestations sont dues à l'assuré ou au bénéficiaire, dès lors que le risque qu'elles couvrent s'est réalisé pendant la période de couverture, peu important une résiliation ultérieure du contrat.

De la même façon, le salarié qui quitte son emploi va pouvoir, sous certaines conditions, continuer de bénéficier des avantages liés au régime de prévoyance pendant une durée maximale de 1 an. C'est ce que l'on appelle la portabilité.

### I.1.2.2.C. Le régime complémentaire individuellement souscrit

Il ressort de ce qui précède que le système de prévoyance français permet de maintenir à chaque salarié, ou à ses ayants droits, un minimum de revenus pour faire face à une impossibilité de travailler liée à une maladie ou à un accident.

Il arrive toutefois que ce minimum, individuellement, ne soit pas suffisant.

Les assureurs proposent alors des produits destinés à venir compléter les garanties déjà souscrites à collectivement.

Il peut s'agir :

- d'une « convention obsèques » ;
- d'une « Garantie Accidents de la Vie » ;
- d'une prévoyance individuelle qui viendra soit compléter les indemnités Sécurité Sociale en cas d'incapacité de travail ou d'invalidité, soit prévoir une aide en cas de dépendance temporaire ou définitive.

Pour les professions indépendantes, dont le système de protection sociale est moins avantageux que celui des salariés, il est souvent proposé et souscrit :

- des produits « Madelin » pour couvrir les risques santé ;
- des produits plus spécifiques tels « Homme clés » pour garantir les revenus tirés d'une activité qui ne peut plus être exercée.

Ces différents produits sont généralement souscrits après une étude individuelle des besoins de la personne à assurer dès lors qu'ils visent à venir combler un déficit de couverture en tentant d'identifier le plus clairement possible ses véritables besoins, au regard des revenus, mais aussi des charges de la personne concernée ou de sa famille.

### I.1.3. La prévoyance « privée » en chiffres

En chiffres, les contrats de prévoyance complémentaires, individuels et collectifs réunis, représentaient en 2016, en termes de cotisations, non moins de 56,7 Md€.

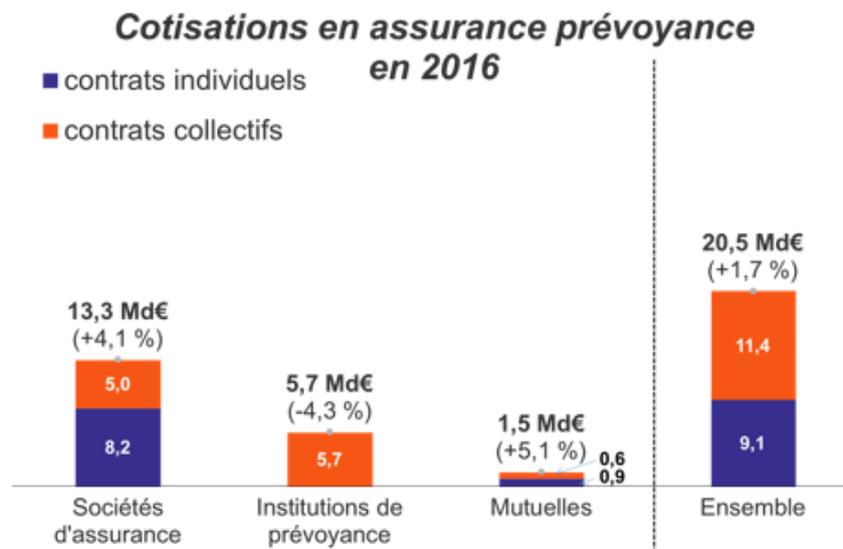
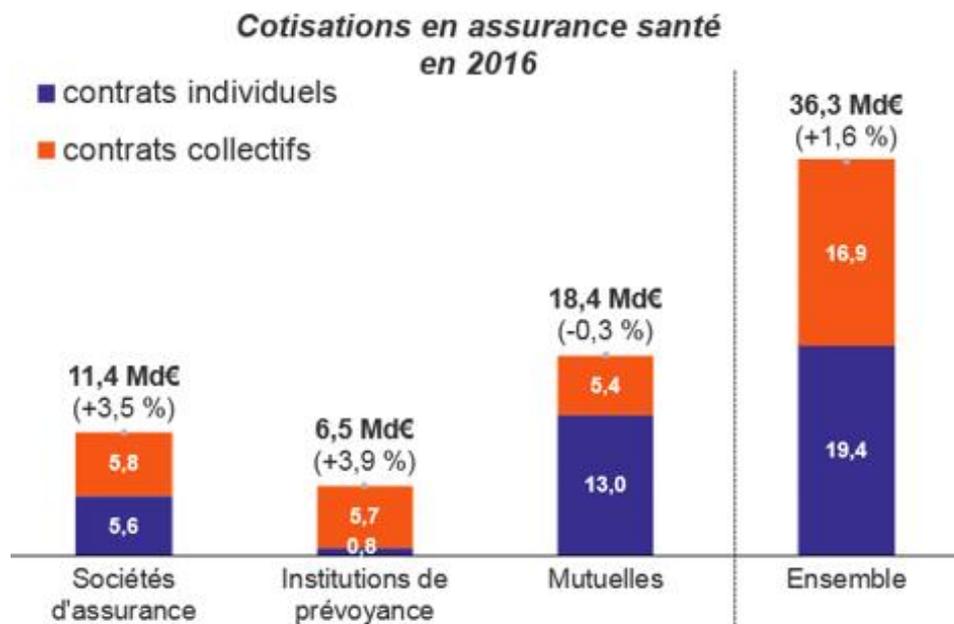
Cotisations 2016 en santé et en prévoyance

	Cotisations (en Md€)			Évolution (en %)		
	contrats individuels	contrats collectifs	Ensemble	contrats individuels	contrats collectifs	Ensemble
Institutions de Prévoyance	0,8	11,4	12,2	-8,5%	0,6%	0%
Sociétés d'assurance	13,8	10,8	24,6	1,8%	6,6%	3,8%
Mutuelles	13,9	6,1	19,9	0,0%	0,2%	0,1%
<b>Ensemble</b>	<b>28,5</b>	<b>28,3</b>	<b>56,7</b>	<b>0,6%</b>	<b>2,7%</b>	<b>1,6%</b>

Sources : CTIP, FFA, FNMF

Nota : en raison des arrondis, la somme des cotisations santé et prévoyance n'est pas toujours égale à l'ensemble.

Ces cotisations se répartissent en :



Les chiffres notent une augmentation de 1,6% par rapport à l'année précédente ce qui témoigne d'une demande constante.

Les organismes privés disposent ainsi d'un marché important, en constante évolution puisque lié aux évolutions législatives mais également sociologiques et démographiques.

L'invalidité, garantie spécifique et sujet de notre mémoire, est naturellement impactée et a nécessité une nouvelle analyse du risque chez les professionnels (principalement les sociétés d'assurance et les IP).

## Chapitre 2. La garantie « Invalidité »

Après avoir présenté le contexte général dans lequel s'inscrit la problématique du mémoire, nous allons donc désormais nous attacher plus spécifiquement à la garantie particulière qui en est l'objet, à savoir l'invalidité.

### I.2.1. Le contexte sociodémographique

Le vieillissement de la population est aujourd'hui une évidence.

Ainsi d'une espérance de 63 ans chez les hommes en 1950 et 69 ans chez les femmes à la même époque, nous sommes passés à une espérance de vie moyenne d'un homme en 2016 de 79 ans, contre 85 ans pour une femme.

Cette évolution s'explique certes par les importants progrès réalisés par la Médecine, mais également par ceux techniques et technologiques, qui sont venus alléger la pénibilité au travail.

De son côté, le progrès social incontestablement lié à la diminution des heures de travail, à l'augmentation des congés payés, au remboursement des soins, apparaît par ailleurs comme ayant joué un rôle non négligeable dans cette évolution.

Mais face à ce vieillissement de la population, et pour faire face à des retraites dont la durée s'avérait être de plus en plus longue, les pouvoirs publics ont dû prendre des mesures pour maintenir les caisses à l'équilibre. Car en effet il sera rappelé la spécificité du régime français qui fonctionne par répartition, les actifs cotisant ainsi pour les retraités.

Les taux de cotisations ont ainsi été revus à la hausse, mais également l'âge de la retraite.

A cela s'ajoute le travail féminin devenu de plus en plus fréquent et qui a donc accru la masse salariale.

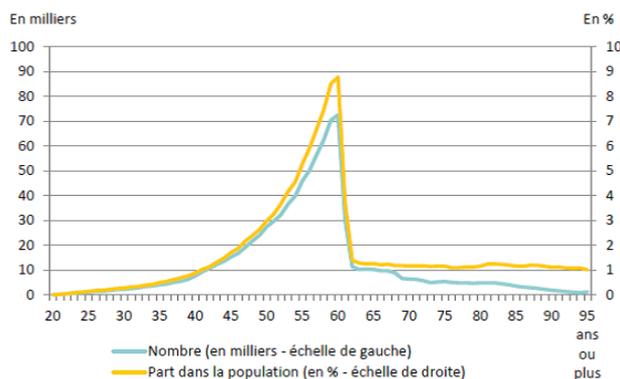
La situation à laquelle les assureurs font donc désormais face est la suivante :

- Plus d'actifs (plus de femmes et plus de seniors)
- Ce qui implique plus de risques d'arrêts de travail et/ou d'invalidités.

Une des conséquences directes de ces différents phénomènes a été l'augmentation du nombre de pensions d'invalidité servies.

Une analyse des chiffres publiés par le Ministère des Affaires Sociales (DRESS) montre sans nul doute possible que ce passage en invalidité concerne particulièrement les 50-59 ans.

#### Nombre et part des pensions d'invalidité en fonction de l'âge (en 2014)



De la même façon, on constate que le report de l'âge de départ à la retraite entraîne un report simultané de l'âge à partir duquel les pensions d'invalidité sont transformées en pensions de retraite.

De même la DRESS constate une évolution de la catégorie socio-professionnelle des bénéficiaires de ces pensions : Alors qu'antérieurement les bénéficiaires étaient essentiellement des ouvriers, les études les plus récentes démontrent une augmentation de ce phénomène chez les cadres.

Le contexte sociodémographique met en conséquence incontestablement l'accent sur une évolution du risque « invalidité » qui est devenu l'un des enjeux majeurs de notre société et donc naturellement des organismes d'assurances privés.

## **I.2.2. Définition de l'invalidité**

D'après l'article L341-1 du Code de la Sécurité sociale, l'invalidité est reconnue lorsque l'assuré « *présente une invalidité réduisant dans des proportions déterminées, sa capacité de travail ou de gain, c'est-à-dire le mettant hors d'état de se procurer, dans une profession quelconque, un salaire supérieur à une fraction de la rémunération normale perçue dans la même région par des travailleurs de la même catégorie, dans la profession qu'il exerçait avant la date de l'interruption de travail suivie d'invalidité ou de la date de constatation médicale de l'invalidité si celle-ci résulte de l'usure prématurée de l'organisme* ».

L'état d'invalidité est constaté lorsqu'il réduit au moins des 2/3 la capacité de travail ou de gain de l'assuré.

C'est au médecin conseil de la CPAM qu'il appartient de classer l'individu qui en remplit les conditions dans l'une ou l'autre des 3 catégories d'invalidité explicitées ci-dessous.

## **I.2.3. Les trois catégories d'invalidité**

Selon l'article L341-4 du Code de la Sécurité sociale, l'invalidité se distingue en trois catégories :

- 1<sup>ère</sup> catégorie : invalides capables d'exercer une activité rémunérée ;
- 2<sup>ème</sup> catégorie : invalides absolument incapables d'exercer une activité professionnelle quelconque ;
- 3<sup>ème</sup> catégorie : invalides qui, étant absolument incapables d'exercer une profession quelconque, sont en outre dans l'obligation d'avoir recours à l'assistance d'une tierce personne pour effectuer les actes ordinaires de la vie courante.

Nous nous sommes attachés à établir un tableau statistique tenant compte de la catégorie et de l'âge des personnes placées en invalidité dans le régime général.

Celui-ci s'établit comme suit :

Régime général	Catégorie 1	Catégorie 2	Catégorie 3
>50 ans	80%	88%	81%
<50 ans	20%	12%	19%

### **I.2.3.1. Le montant de la pension du régime de base général de la Sécurité Sociale**

La pension est calculée à partir d'un salaire annuel moyen de base en fonction d'un taux tenant compte du classement de l'invalidé dans l'une des 3 catégories énumérées ci-dessus.

Elle peut être complétée, sous certaines conditions, par une allocation supplémentaire ou par les prestations résultant d'un régime de prévoyance complémentaire souscrit par l'employeur.

Le taux de la pension est fixé ainsi qu'il suit :

Catégorie d'invalidité	Pourcentage du salaire annuel moyen	Montant mensuel minimum	Montant mensuel maximum
1 <sup>re</sup> catégorie	30 %	289,90 €	1 013,10 €
2 <sup>e</sup> catégorie	50 %	289,90 €	1 688,50 €
3 <sup>e</sup> catégorie	50 %, majoré de 40 % au titre de la <u>majoration pour tierce personne</u>	1 411,82 €	2 810,43 €

La pension d'invalidité est liquidée sur la base du salaire annuel moyen des 10 années civiles d'assurance dont la prise en considération est la plus avantageuse pour l'assuré et précédant soit l'interruption de travail suivie d'invalidité, soit la constatation médicale de l'invalidité lorsque celle-ci résulte de l'usure prématurée de l'organisme.

Seules les années entières sont retenues à l'exclusion de l'année d'interruption du travail.

Les salaires des années civiles retenues sont revalorisés chaque année au 1<sup>er</sup> avril.

La pension d'invalidité est toujours concédée à titre temporaire et un contrôle des droits du titulaire est effectué chaque année.

La pension d'invalidité prend fin à l'âge légal d'ouverture d'une pension de retraite.

### **I.2.3.2. Les conditions d'attribution de la pension d'invalidité**

Les conditions d'attribution d'une pension d'invalidité sont les suivantes :

#### **a) Être en état d'invalidité**

L'état d'invalidité est apprécié globalement en tenant compte de la capacité de travail restante, de l'état général, de l'âge, des facultés physiques et mentales de l'assuré ainsi que de ses aptitudes et de sa formation professionnelle :

- Soit après consolidation de la blessure ;
- Soit à l'expiration de la période pendant laquelle l'assuré a bénéficié des prestations de l'assurance maladie ;
- Soit après stabilisation de son état intervenue avant l'expiration du délai permettant de bénéficier des prestations de la sécurité sociale ;
- Soit au moment de la constatation médicale de l'invalidité lorsque celle-ci résulte d'une usure prématurée de l'organisme.

#### **b) Affiliation préalable :**

Pour pouvoir invoquer le bénéfice de l'assurance invalidité, l'assuré doit être affilié depuis 12 mois au 1<sup>er</sup> jour du mois au cours duquel est survenue l'interruption de travail suivie d'invalidité.

**c) Durée de l'activité :**

L'assuré doit justifier :

- Soit d'avoir cotisé sur un salaire au moins égal à 2.030 fois la valeur du SMIC horaire au 1<sup>er</sup> janvier précédant la période de référence au cours des 12 mois civils précédant l'interruption du travail ;
- Soit avoir effectué au moins 600 heures de travail salarié ou assimilé au cours des 12 mois civils ou des 365 jours précédant l'interruption de travail.

#### **I.2.4. La possibilité de couverture complémentaire**

Un individu en état d'invalidité peut voir compléter la pension d'invalidité qui lui est servie par la sécurité sociale,

- Soit par un organisme de prévoyance collective auquel son employeur aura cotisé dans le cadre de ses obligations légales ou conventionnelles ;
- Soit par une assurance individuelle souscrite à titre particulier.

Cette indemnité complémentaire, versée sous forme de rente dépend des garanties souscrites et vient compléter la pension versée par la sécurité sociale.

Il est bien évidemment fondamental pour les assureurs d'étudier l'ensemble des données pouvant influencer sur le passage en invalidité, et la durée de celle-ci, afin de pouvoir établir un provisionnement suffisant.

## ***Chapitre 3. Le provisionnement du risque « Invalidité »***

Maintenant que la notion d'invalidité a été présentée, le provisionnement de cette garantie est décrit dans cette partie.

Le provisionnement est une partie primordiale du travail de l'assureur : afin de faire face à ses engagements futurs, il doit provisionner au mieux les risques auxquels il est confronté, c'est-à-dire évaluer de la façon la plus juste et précise possible, le montant de ses engagements envers les assurés.

### **I.3.1. Définition**

La notion de provisionnement nous fait revenir au principe même de l'assurance.

En effet, à la souscription du contrat, l'assureur s'engage, en contrepartie du versement d'une prime de la part de l'assuré, à verser une prestation au bénéficiaire en cas de survenance d'un risque.

Il s'agit donc d'un cycle de production inversé, caractéristique propre à l'assurance. Ce cycle entraîne la nécessité d'estimer les prestations à verser à l'avenir en cas de sinistres et de constituer une provision qui permettra de régler ces sinistres.

### **I.3.2. Les provisions**

Les provisions présentées ci-après sont des provisions techniques. Elles ne sont pas les mêmes selon qu'il s'agisse d'assurance vie ou d'assurance non-vie.

L'invalidité étant apparentée à une opération d'assurance non-vie, ce sont les provisions associées au second cas qui seront étudiées ci-dessous.

Elles sont regroupées en huit catégories : les provisions mathématiques de rente, les provisions pour primes non acquises, les provisions pour risques en cours, les réserves de capitalisation, les provisions pour sinistres à payer, les provisions pour risques croissants, les provisions pour égalisation, et les provisions pour risques d'exigibilité.

En cas d'arrêt de travail, les **provisions pour sinistres à payer** (qui nous intéressent le plus ici) se répartissent en trois catégories distinctes :

- La provision incapacité en cours : elle doit permettre de couvrir l'engagement de l'assureur lié au maintien de l'assuré en incapacité ;
- La provision invalidité en attente : elle est censée couvrir l'engagement de l'assureur lié au risque de passage de l'état d'incapacité à celui d'invalidité ;
- La provision invalidité en cours : son rôle est de couvrir l'engagement de l'assureur lié au maintien de l'assuré en invalidité.

### **I.3.3. Le provisionnement pour les rentes d'invalidité**

Le provisionnement pour le risque invalidité est encadré par différents textes de loi. Ces derniers ont évolué avec le temps.

Ainsi, avant le 31 décembre 1989, les assureurs n'étaient soumis à aucune obligation au niveau des provisions. Le principe de répartition imposait simplement que les primes reçues lors d'une année couvrent les sinistres

survenant la même année. L'on utilisait alors les indications fournies par le Bureau Commun d'Assurances Collectives (BCAC), fondées sur les travaux réalisés par M. WETZEL en 1960 pour les provisions techniques.

Cependant, ces indications ont montré leurs limites : la loi de maintien en invalidité n'anticipait ni entrée en invalidité avant 50 ans, ni passage en invalidité au cours de la première année d'incapacité.

La loi Évin a alors mis en vigueur de nouvelles règles.

Ainsi, l'article 7 précise que *« l'engagement doit être couvert à tout moment pour tous les contrats ou conventions souscrits, par des provisions représentées par des actifs équivalents »*.

Le décret n°90-768 du 30 août 1990 indique le montant minimum à provisionner :

- *« Pour les prestations dues au titre du risque incapacité, lorsque les versements peuvent s'échelonner sur plus de 365 jours et peuvent se poursuivre par des prestations d'invalidité : deux fois le montant annuel des prestations d'incapacité servies au cours de l'exercice ;*
- *« Pour les prestations dues au titre du risque invalidité : six fois le montant annuel des prestations d'invalidité servies au cours de l'exercice. »*

Plus tard, le 28 mars 1996, en découlèrent la mise au point par le BCAC et l'introduction des tables des provisions dans l'article A.331-22 du Code des Assurances. Ce qui a entraîné la révision des méthodes de provisionnement. Cet article précise en effet que les provisions techniques des prestations d'incapacité et d'invalidité sont désormais la somme :

- *« Des provisions correspondant aux prestations d'incapacité de travail à verser après le 31 décembre de l'exercice au titre des sinistres en cours à cette date, majorées des provisions dites pour rentes en attente relatives aux rentes d'invalidité susceptibles d'intervenir ultérieurement au titre des sinistres d'incapacité en cours au 31 décembre de l'exercice ;*
- *« Des provisions correspondant aux prestations d'invalidité à verser après le 31 décembre de l'exercice au titre des sinistres d'invalidité en cours à cette date. »*

Ce même article précise également que le calcul des provisions techniques de prestations d'incapacité de travail et d'invalidité est effectué à partir des éléments suivants :

- *« Les lois de maintien en incapacité de travail et en invalidité indiquées en annexe du Code. Toutefois, il est possible pour une entreprise d'assurances d'utiliser une loi de maintien établie par ses soins et certifiée par un actuair indépendant de cette entreprise, agréé à cet effet par l'une des associations d'actuaire reconnues par la commission de contrôle des assurances ;*
- *« Un taux d'actualisation qui ne peut excéder 75 % du taux moyen des emprunts de l'État français calculé sur la base semestrielle, sans pouvoir dépasser 4,5 %. »*

Ce dernier taux a depuis été révisé. En effet, l'article 1 de l'arrêté du 23 décembre 2010 a modifié l'article précédemment cité en indiquant notamment que le taux d'actualisation *« ne pouvait excéder 75 % du taux moyen au cours des vingt-quatre derniers mois des emprunts de l'État français, sans pouvoir dépasser 4,5 % »*.

L'établissement de ces tables du BCAC a largement amélioré notre connaissance du risque et permet donc à présent d'obtenir de meilleures estimations des provisions.

# ***PARTIE II : LE PERIMETRE D'ETUDE***

## ***Chapitre 1. Présentation des données***

Dans ce chapitre, nous présenterons la façon dont les données ont été préparées pour l'étude. Ces dernières proviennent de deux types de fichiers mis à disposition : les données extraites des fichiers des comptes utilisés pour calculer les provisions mathématiques (PM), et les données de prestations ligne à ligne. Les assurés travaillent dans la branche des aides à domicile.

**Précautions :** Il est important de signaler que cette population très spécifique nous amène naturellement à nous interroger sur la représentativité de ce corps de métier par rapport à la population globale. L'utilisation d'une population autre pourrait donner lieu à des conclusions différentes.

En effet, des biais existent : l'âge moyen de cette population, la répartition hommes-femmes (grande majorité de femmes dans ce portefeuille), la catégorie socio-professionnelle, le salaire moyen, ainsi que l'exposition des assurés à des risques d'arrêts de travail ou d'accidents de travail sont autant de facteurs qui peuvent mener à des résultats particuliers. Tous les résultats qui suivent sont donc à prendre avec précaution et à vérifier sur d'autres portefeuilles avec une population plus globale.

Par exemple, en faisant des tests de comparaison de deux moyennes observées<sup>ii</sup> (âge ou salaire moyen de la branche aides à domicile vs celui d'un portefeuille plus lambda). Ce test consiste à vérifier l'hypothèse selon laquelle la différence des moyennes est nulle.

Nous aurions également pu effectuer des tests d'appariement, comme celui de Kruskal-Wallis<sup>iii</sup> (d'après William Kruskal et Wilson Allen Wallis). C'est une méthode non paramétrique utilisée pour tester si des échantillons trouvent leur origine dans la même distribution. Ce test s'intéresse aux médianes de k populations et propose comme hypothèse nulle que les k échantillons sont confondus et proviennent d'un même échantillon (combiné) d'une population.

Le premier fichier (PM) est une photographie des assurés en arrêt de travail à une date fixée (31 décembre de l'année) qui présente tous les assurés avec leur date d'arrêt s'ils sont en incapacité, et leur date d'entrée en invalidité le cas échéant : il s'agit d'une vision « *stock* ». Le deuxième fichier représente la vision « *flux* » : il recense toutes les prestations en arrêt de travail (incapacité ou invalidité) pour tous les assurés pendant une période donnée. Typiquement, ce fichier est transmis trimestriellement.

L'utilisation de deux fichiers différents issus du même portefeuille s'explique du fait d'un souci d'efficacité de l'entreprise pour laquelle il est plus rapide d'obtenir et de travailler sur le fichier extrait des comptes que sur la vision « flux ».

Ce premier jeu de données sert donc à obtenir des indications sur une potentielle accélération du passage en invalidité, ce qui peut servir d'alerte pour l'entreprise. Mais ces données, déjà moins complètes et précises, sont aussi dépourvues d'informations sur les dates de sortie de l'état d'incapacité si celui-ci n'a pas engendré d'invalidité. Or en cas d'accélération de la reconnaissance de l'invalidité, il nous faudra créer une nouvelle table de passage en invalidité mais aussi une nouvelle table de maintien en incapacité. C'est la raison principale pour laquelle les données de prestations ligne à ligne ont été utilisées puisqu'on peut plus facilement y suivre l'évolution d'un assuré.

Travailler sur les fichiers « comptes » est ainsi plus utile et réaliste dans la pratique où nous ne disposons que de peu de temps à consacrer à une étude. Mais dans un cadre de recherche comme c'est ici le cas de ce mémoire, les fichiers « flux » semblent plus corrects et rigoureux pour approfondir les résultats aperçus avec les premiers fichiers.

Dans la suite, nous détaillerons les variables que nous avons retenues pour notre étude et les traitements faits au préalable sur nos données.

Nous précisons que dû à la franchise de 60 jours en arrêt de travail, le volume de déclarations des premiers mois est altéré. En effet, les assurés, connaissant cette franchise, ne déclarent pas forcément tous leur arrêt de travail à l'assureur. Cela est particulièrement le cas pour les jeunes assurés qui ne souhaitent pas forcément perdre de temps à cause de démarches administratives sachant qu'ils retourneront en général rapidement au travail (ce sont eux qui ont les périodes d'incapacité les plus courtes en moyenne comme nous le verrons plus tard).

Ces cas constituent des pertes d'informations pour l'assureur qui ne peut pas recenser la durée de tous les arrêts de façon exhaustive, et est par conséquent moins juste dans la création de ses tables de maintien en incapacité.

### **II.1.1. Vision « stock »**

Dans ce premier jeu de données, les références contractuelles sont celles des fichiers 2012 à 2018. **L'extraction brute avant fiabilisation des données comportait 57 815 personnes.**

#### **II.1.1.1. Les variables disponibles**

Pour chaque référence contractuelle extraite, les variables suivantes ont été retenues :

- Les informations sur le souscripteur
  - Nom ;
  - Prénom ;
  - Date de Naissance.
- Les informations sur la référence contractuelle :
  - Code risque RGP : cette information nous permet de savoir dans quel état se trouve notre assuré en arrêt de travail (32 pour l'incapacité et 22 pour l'invalidité) ;
  - Date d'entrée en Incapacité ;
  - Date d'entrée en Invalidité ;
  - Année de compte.

#### **II.1.1.2. Fiabilisation des données**

Pour fiabiliser le périmètre extrait, toutes les variables ont été analysées : quelles valeurs ces variables contenaient-elles ? Ne fallait-il pas exclure certaines références contractuelles du périmètre retenu compte tenu de leurs caractéristiques ?

Les traitements apportés pour rendre les données exploitables ont consisté en :

- **L'harmonisation des noms et prénoms**

En effet certaines écritures pouvaient varier d'une année sur l'autre. Il nous fallait donc corriger ces erreurs de saisie pour pouvoir suivre au mieux les assurés à travers le temps.

- **La suppression des doublons**

Nous avons retiré les lignes de données se répétant (mêmes nom, prénom, date de naissance et date d'arrêt). Ces doublons correspondaient aux assurés se retrouvant dans les photographies des arrêts de travail plusieurs années consécutivement.

Nous avons donc gardé l'information la plus récente disponible pour chaque arrêt. Ainsi il nous restait 38 370 lignes et donc autant d'arrêts.

- **La création des variables « Année de l'Arrêt », « Âge à l'Arrêt », « Âge à l'Entrée en Invalidité » et « Délais de Passage en Invalidité »**

Nous avons alors enlevé les 34 cas où la date de l'arrêt n'appartenait pas à notre périmètre d'étude (01/01/2012-31/12/2018).

Puis nous avons supprimé toutes les lignes dans lesquelles l'assuré avait plus de 67 ans à l'entrée en incapacité (âge limite pour entrer en incapacité). Soit 63 références contractuelles.

Enfin, nous avons pu retirer tous les délais supérieurs aux trois années (1 095 jours) maximales de maintien en incapacité. Soit 498 cas.

Remarque : les âges ont été calculé de la manière suivante :

$$\text{Troncature} \left( \frac{\text{Date d'Arrêt ou d'Invalidité} - \text{Date de Naissance}}{365,25} \right).$$

**L'extraction après fiabilisation contenait donc 11 variables pour 37 775 personnes en incapacité.**

**De plus, sur ces arrêts, nous avons recensé 5 344 passages en invalidité.**

## **II.1.2. Vision « flux »**

Dans ce jeu de données, les références contractuelles sont celles des fichiers 2014 à 2018. Si le commencement de notre période d'étude sur cette vision est retardé de 2 ans par rapport à celle de la vision « stock », c'est en raison du fait de l'indisponibilité des informations en vision « flux » avant 2014. Ce type de fichier ne fut intégré aux bases de données de l'assureur qu'en 2014.

**L'extraction brute avant fiabilisation des données comportait 133 360 références contractuelles.**

### **II.1.2.1. Les variables disponibles**

Pour chaque référence contractuelle extraite, les variables suivantes ont été retenues :

- Les informations sur le souscripteur :
  - Numéro de Sécurité Sociale
- Les informations sur la référence contractuelle :
  - Code risque RGP ;
  - Date d'entrée en Incapacité ;
  - Date de sortie d'Incapacité ;
  - Date d'entrée en Invalidité ;
  - Année de compte.

### **II.1.2.2. Fiabilisation des données**

De même que pour les fichiers « comptes » les données ont été retravaillées pour les rendre exploitables. Nous avons donc procédé à :

- **La suppression des doublons :**

Nous avons retiré les lignes de données se répétant (même numéro de sécurité sociale et même date d'arrêt). Nous avons donc gardé l'information la plus récente disponible pour chaque arrêt.

Pour plus de prudence, si les dates de sorties différaient pour un même arrêt, nous n'avons gardé que celle qui donnait la plus longue période d'incapacité.

Ainsi il nous restait 77 453 lignes et donc autant d'arrêts.

- **Une vérification de la cohérence des informations :**

En effet, chaque assuré pouvait avoir plusieurs entrées en incapacité. Il nous fallait donc vérifier qu'une entrée ne se faisait pas alors que l'assuré était encore en incapacité dû à son arrêt précédent ou alors qu'il venait d'en sortir. Dans ce cas, par prudence, nous avons fusionné les deux arrêts en gardant la date d'entrée la plus ancienne et la date de sortie la plus récente.

Cela nous a fait réduire notre base de 2 404 lignes.

- **La création des variables « Date de Naissance », « Année de l'Arrêt », « Âge à l'Arrêt » et « Délais de Passage en Invalidité »**

Nous avons alors enlevé les cas où la date de l'arrêt n'appartenait pas à notre périmètre d'étude (01/01/2014-31/12/2018).

Puis nous avons supprimé toutes les lignes dans lesquelles l'assuré avait plus de 67 ans (âge limite pour entrer en incapacité).

Enfin, nous avons pu retirer tous les délais supérieurs aux trois années (1 095 jours) maximales de maintien en incapacité (0,28% des cas).

Remarque : les dates de naissance ont été reconstituées à partir du numéro de Sécurité sociale en supposant que le jour de naissance était le milieu du mois (le 15).

**L'extraction après fiabilisation contenait donc 74 672 contrats contenant 10 variables.**

**De plus, nous avons recensé 4 082 passages en invalidité (contre 3 322 sur la même période dans le jeu de données précédent).**

Dans le chapitre suivant nous analyserons les jeux de données avec des statistiques descriptives pour mieux se rendre compte du contenu et des informations de notre portefeuille.

## Chapitre 2. Statistiques descriptives

### II.2.1. La vision « stock »

#### II.2.1.1. Les assurés en incapacité

L'objectif de cette section est de définir le profil type d'un assuré à sa date d'arrêt de travail et d'étudier son évolution en fonction des années de survenance.

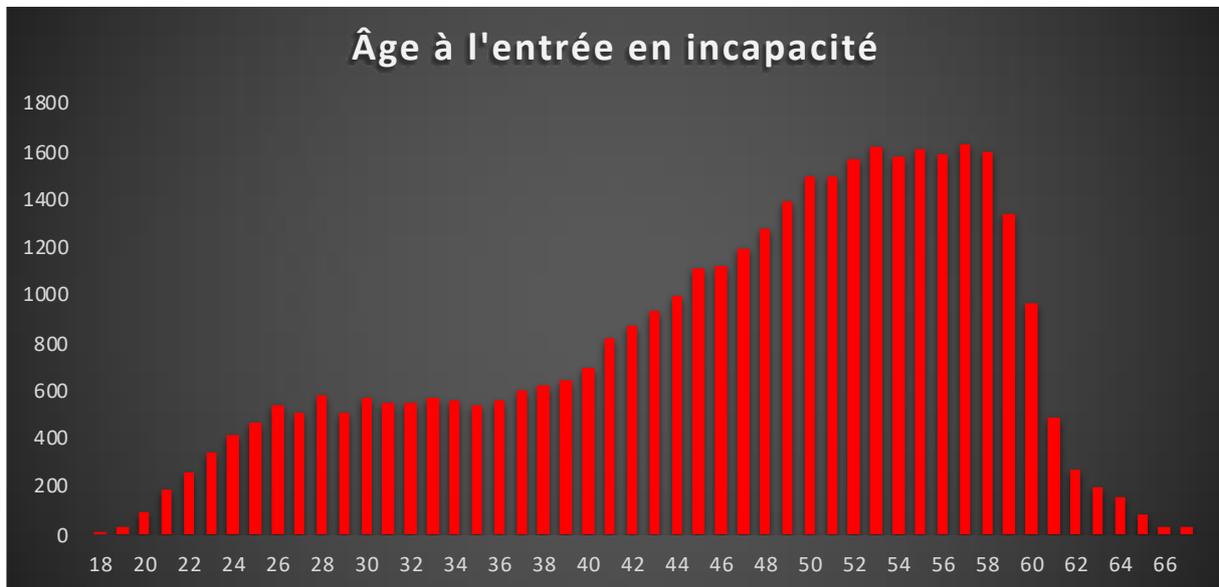
Il est important de préciser qu'il ne sera pas possible de tirer de conclusions sur l'entrée en arrêt de travail à partir des statistiques qui seront faites dans cette section car seuls les assurés en arrêt de travail sont connus, la base des assurés n'est pas connue.

Les années de survenance du portefeuille sélectionné varient **entre 2012 et 2018**.

**Par année de survenance, les dossiers sont répartis ainsi :**

		Nombre d'arrêts
Année de survenance	2012	4303
	2013	4932
	2014	5029
	2015	6125
	2016	6395
	2017	5855
	2018	5136
	Total	37775
	Total 2012 à 2015	20389

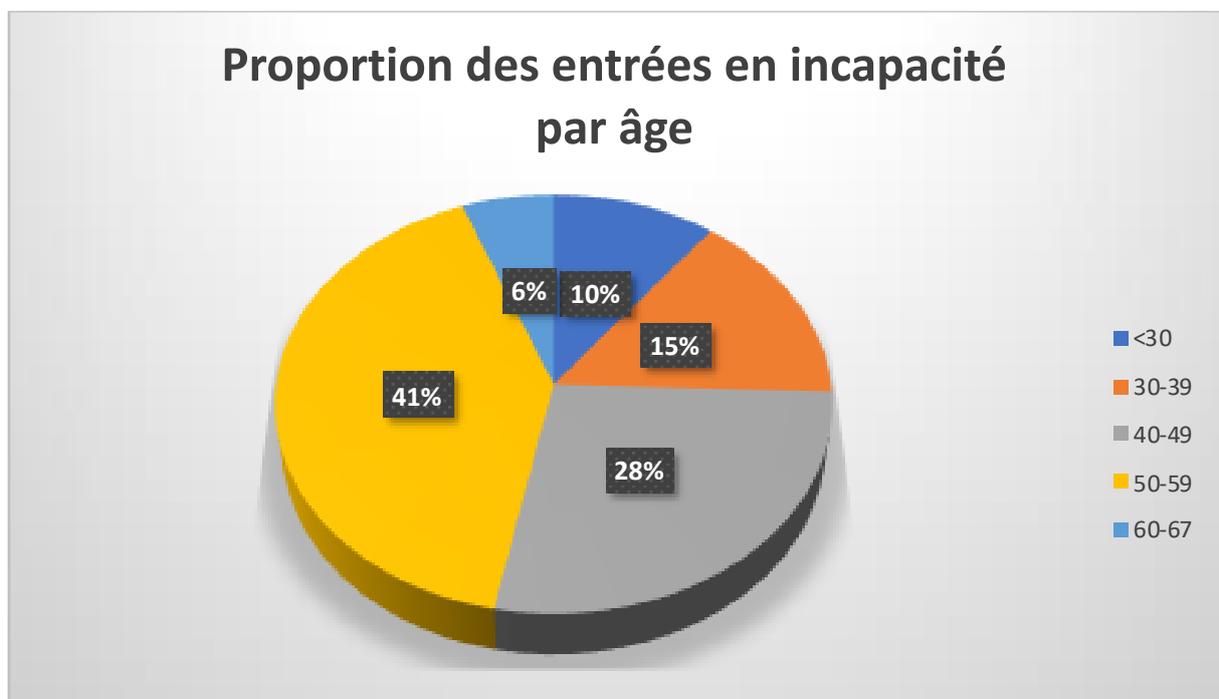
Nous remarquons un **nombre croissant d'arrêts** jusqu'en 2016. Si ce nombre décroît pour les deux dernières années, il se peut que ce soit dû aux sinistres survenus et non encore déclarés et donc non encore connus de l'assureur. Ces derniers, une fois déclarés, augmenteront le nombre d'arrêts et permettront de valider ou pas la tendance observée sur les années antérieures. Cette tendance peut trouver une explication générationnelle puisque les baby-boomers commencent à arriver à l'âge de la retraite depuis 2012 mais aussi car le départ à la retraite a été repoussé progressivement à 62 ans entre 2013 et 2017. Ainsi plus d'assurés implique plus d'arrêts.



Sur la pyramide des âges ci-dessus, nous pouvons observer un âge maximal de **67 ans** et un âge minimal de **18 ans**. En ne s'intéressant qu'aux invalides, l'âge minimum à la survenance est de **22 ans** et le maximum de **61 ans**.

Nous pouvons souligner également que ce graphique laisse apparaître clairement la **bosse des accidents** qui survient aux alentours de la trentaine. Le nombre de survenances **croît** ensuite régulièrement avec l'âge, se **stabilise** aux environs de la **cinquantaine (avec un pic à 53 ans)**, avant de **diminuer** fortement à partir de **59 ans**.

Enfin, nous remarquons que l'âge auquel il y a le plus d'entrées en incapacité est **57 ans**.



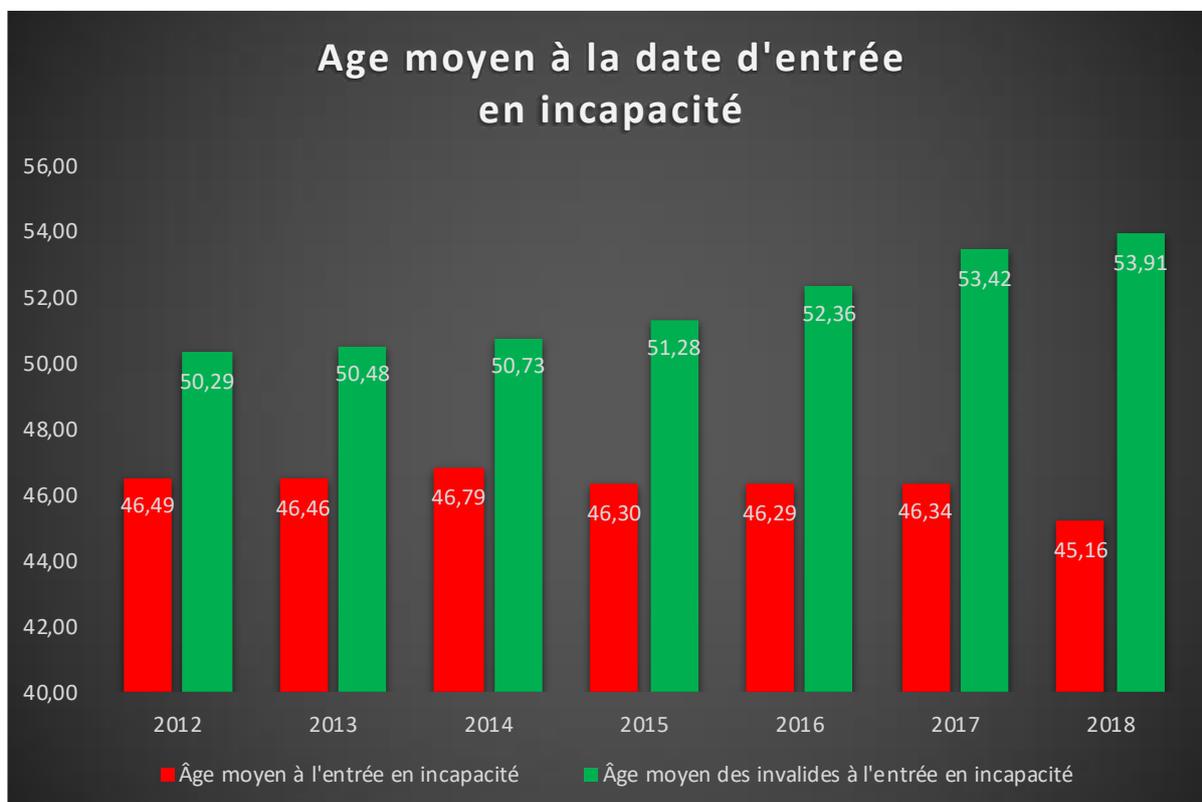
		Nombre d'arrêts	Nombre d'arrêts des 50-60 ans	Part d'arrêts des 50-60 ans
Année de survenance	2012	4 303	1 890	44%
	2013	4 932	2 165	44%
	2014	5 029	2 311	46%
	2015	6 125	2 743	45%
	2016	6 395	2 808	44%
	2017	5 855	2 565	44%
	2018	5 136	1 981	39%
	Total	<b>37 775</b>	<b>16 463</b>	<b>44%</b>

On peut constater que :

- Les arrêts entraînant une incapacité **surviennent principalement entre 40 et 59 ans** (à moindre mesure entre 30 et 39 ans) puisque presque 69% des incapacités sont déclarés dans cette tranche d'âge ;
- La tranche 50-60 ans représente près de 50% des arrêts recensés depuis 2012 ;
- Il y a **peu d'assurés** entrant en incapacité pour **les âges inférieurs à 30 ans**, ce qui s'explique en partie par le fait que les jeunes salariés sont en meilleure santé que les assurés plus âgés mais aussi car ils ont souvent des emplois plus précaires qui les poussent à retourner rapidement au travail. Sachant que la franchise en place ne leur permettra pas de toucher d'indemnités, les jeunes assurés se passent alors des démarches administratives servant à déclarer leur arrêt auprès de l'assureur, ce qui explique le faible nombre d'incapacités à ces âges dans la base de données de l'assureur ;
- **Les arrêts sont encore moins nombreux au-delà de 60 ans** ce qui s'explique sûrement par les nombreux départs à la retraite (62 ans) dans cette tranche d'âge et par la précarité des emplois lorsque l'on est proche de la retraite.

### II.2.1.2. Les assurés en invalidité

Les graphiques ci-dessous nous donnent des renseignements sur les invalides entrés en incapacité entre 2012 et 2018. Nous remarquons que les assurés déclarés invalides sont **en moyenne plus vieux** que le reste du portefeuille à leur date d'arrêt de travail. En effet, l'âge moyen global à la date de l'arrêt de travail est de **46 ans et 3 mois**. Or pour les assurés étant passés de l'état d'incapacité à celui d'invalidité, cet âge moyen monte à **51 ans et 2 mois**.



L'âge moyen lors de l'entrée en incapacité semble **stable chaque année** (environ 46 ans et 6 mois). Nous remarquons cependant une petite anomalie en 2018, avec une baisse significative (plus d'un an) de cet âge moyen.

Cela est dû à une part moins importante de la tranche d'âge des 50-60 ans dans notre portefeuille d'arrêts global (comme vu dans le tableau précédent).

Cette part moins grande d'individus de cet âge peut peut-être s'expliquer par le départ progressif des baby-boomers de cette tranche d'âge (et leur entrée dans la tranche d'âge 60-70 ans, ce qui expliquerait que la génération suivante prenne « moins de place » dans notre portefeuille.

Cela peut aussi simplement être dû à un manque de réactivité de ces individus plus âgés en suite de leur arrêt, ce qui impliquerait que l'âge moyen des individus ayant connu un sinistre mais ne l'ayant pas encore déclaré est bien plus élevé.

Au contraire, l'âge moyen des invalides, lui, **croît en fonction du temps**. Cet âge moyen a en effet augmenté de plus de 3 ans durant la période d'étude. Cela peut être le résultat de la réforme des retraites augmentant le risque de passage en invalidité pour les âges les plus élevés.



Nous notons que **très peu d'arrêts avant l'âge de 40 ans** entraînent une invalidité, tout comme les **arrêts au-delà de 60 ans**.

D'autre part, concernant les arrêts pour incapacité intervenant entre 40 et 60 ans, le nombre d'assurés reconnus invalides **croît assez régulièrement avec l'âge** pour atteindre un pic pour un âge à l'entrée en incapacité **de 57 ans**.

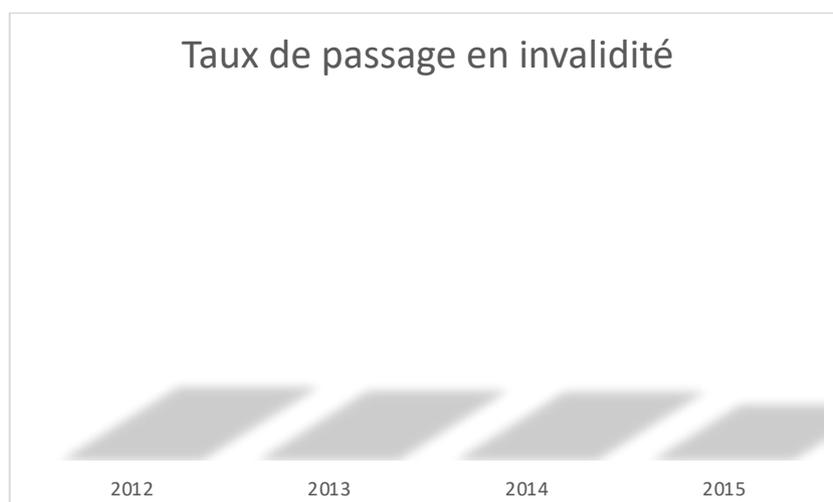
### II.2.1.3. Taux de passage en invalidité

En observant que le nombre d'arrêts augmentait au fil des ans, nous pourrions penser que le nombre d'invalidités augmenterait de manière proportionnelle.

Pour vérifier cette hypothèse, nous allons mettre en place un indicateur pour analyser au mieux l'augmentation du volume d'invalides. Nous observons donc ici le taux de passage en invalidité selon les années d'entrée en incapacité et l'âge.

Les années observées sont celles de **2012 à 2015** puisqu'un taux de passage en invalidité sur les années suivantes serait faussé par le simple fait que la durée maximale de maintien en incapacité est de 3 ans et que donc tous les assurés dans cet état durant ces années ne sont pas encore sortis de cet état.

En effet, si l'on prend pour exemple une personne entrée en incapacité le 22 mars 2016 qui sera déclarée invalide au bout de 3 ans (le 16 mars 2019), cette personne ne sera pas prise en compte dans notre calcul du taux de passage en invalidité puisque nos données ne vont que jusqu'au 31 décembre 2018.



Il ressort tout d'abord que **le taux de passage décroît** entre 2012 et 2015. Il est passé de plus de 22% à un peu moins de 17%. Il n'avait pourtant perdu qu'un point en 3 ans avant d'en perdre quatre entre 2014 et 2015. Le volume d'incapacités ne croît donc pas au même rythme que celui d'invalidités. Cela peut s'expliquer par un manque d'informations concernant les derniers passages en invalidité des assurés entrés en incapacité en 2015 car les différentes démarches administratives peuvent parfois retarder l'officialisation du passage en invalidité.



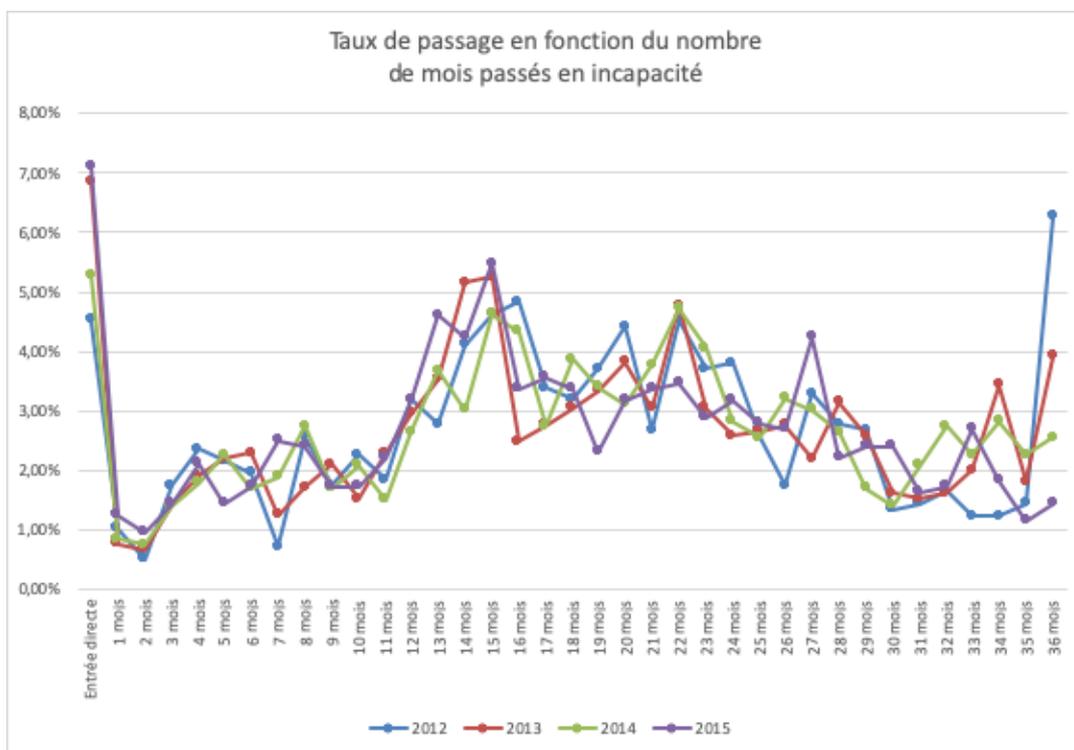
Nous remarquons ici assez logiquement par rapport aux pyramides des âges que nous avons exposées, que plus l'âge d'entrée en incapacité est élevé, plus la proportion d'entrée en invalidité est grande, avec un pic à **57 ans**. À cet âge-là, près de 2 assurés sur 5 passent en invalidité pendant leur période d'incapacité.

#### II.2.1.4. Délai de passage en invalidité

Concernant les délais de passages en invalidité, il ressort que, pour les années inférieures ou égales à 2015, **la moyenne est de 540 jours, alors qu'en y ajoutant les années 2016, 2017 et 2018, cette moyenne chute de 50 jours**. Cela s'explique par le fait qu'à partir de 2016 les arrêts ont été observés sur une période de moins de 3 ans. Ainsi, certains assurés ont pu être déclarés invalides après la fin de la période d'étude et augmenteraient donc la moyenne totale des délais de passage en invalidité.

		Délai moyen de passage
Année de survenance	2012	552
	2013	543
	2014	551
	2015	517
	2016	384
	2017	204
	2018	92
	Total	490
	Total 2012 à 2015	540

Nous remarquons ici une **accélération du passage en invalidité** : de 18,1 mois en moyenne en 2012, nous ne sommes plus qu'à 17,0 mois en moyenne en 2015 et 12,6 mois en 2016.



En analysant ce graphique, nous confirmons ce que nous montre le tableau précédent puisque **les pics de passage en invalidité de l'année 2015 interviennent avant ceux des années antérieures.**

De plus, les parts d'invalidités déclarées le 36<sup>ème</sup> mois décroissent fortement en fonction des années (plus de **6%** des passages en invalidité pour des assurés reconnus en état d'incapacité en 2012 se sont produits le 36<sup>ème</sup> mois contre seulement **1,5%** pour les assurés entrés en 2015).

## II.2.2. La vision « flux »

### II.2.2.1. Les assurés en incapacité

Dans cette section, les années de survenance du portefeuille sélectionné varient **entre 2014 et 2018.**

**Par année de survenance, les dossiers sont répartis ainsi :**

		Nombre d'arrêts
<b>Année de Survenance</b>	<b>2014</b>	12506
	<b>2015</b>	20968
	<b>2016</b>	18006
	<b>2017</b>	13478
	<b>2018</b>	9714
	<b>Total</b>	<b>74672</b>
	<b>Total 2014-2015</b>	<b>33474</b>

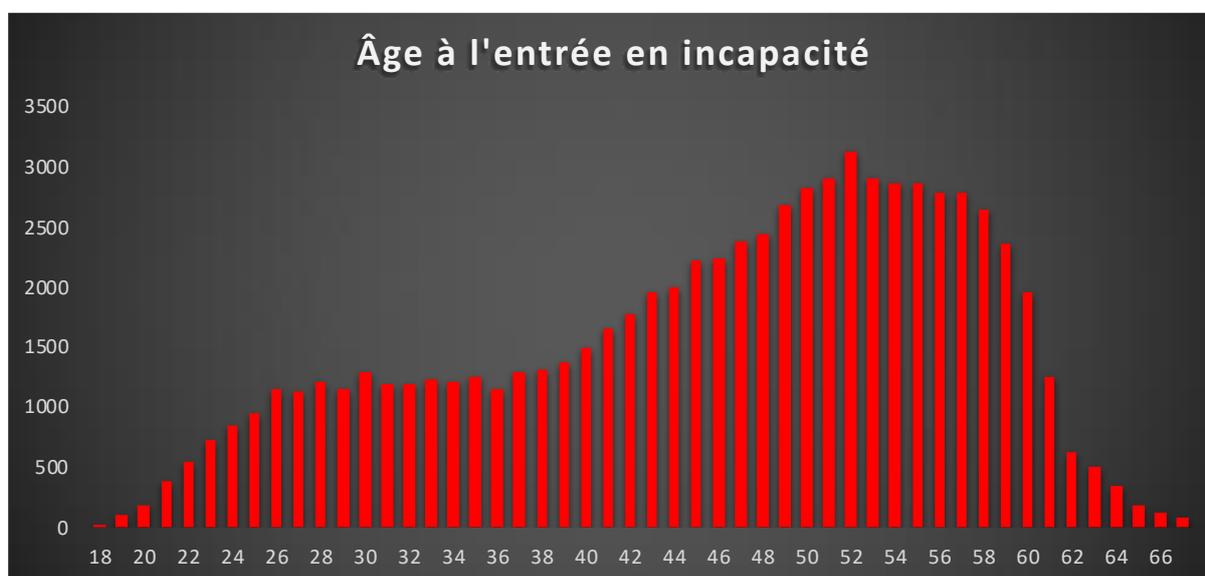
Nous remarquons **nettement plus d'arrêts en 2015 qu'en 2014**.

Cela est en partie dû à un contexte épidémiologique favorable en 2014 avec moins d'épisodes de grippe que les deux années suivantes, ainsi que du fait que les arrêts non indemnisés par l'assureur sont intégrés à la base de données à partir de 2015 et non de 2014.

Le nombre décroissant d'entrées en incapacité à partir de 2017 s'explique quant à lui, par le fait de la longueur de notre période d'observation qui ne nous permet pas de prendre en compte les sinistres survenus mais non déclarés que l'on ne connaîtra que plus tard.

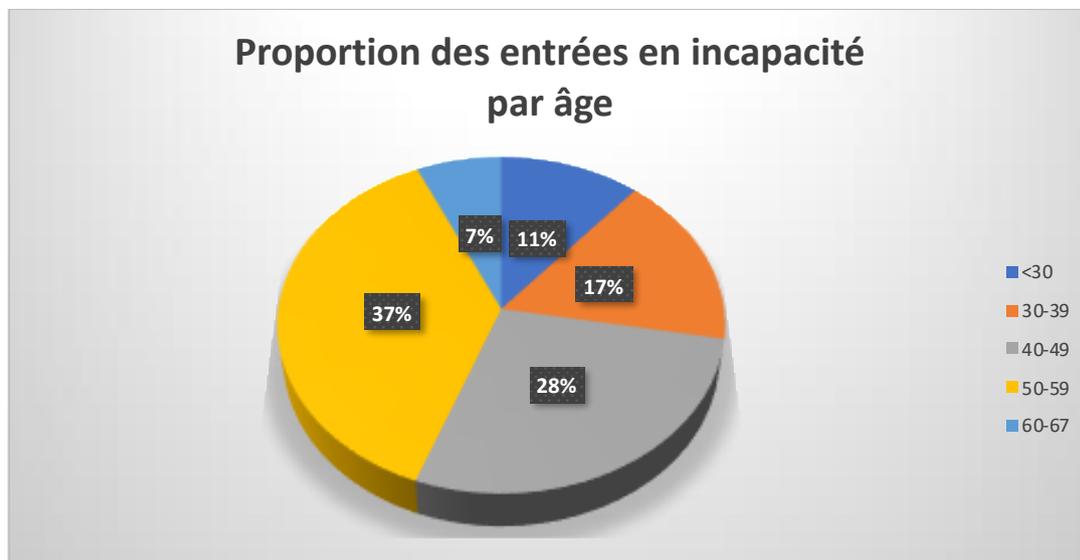
Les nombres d'arrêts beaucoup plus élevés que dans la vision « stock » s'expliquent par la définition même des deux visions. En effet, la plupart des périodes d'incapacités courtes présentes dans cette vision qui nous permet de suivre pas à pas un assuré ne sont pas visibles sur une vision « stock » qui n'est qu'une « photographie » des assurés en incapacité au 31 décembre seulement.

Un assuré sorti de l'état avant n'y est donc pas.



Nous pouvons souligner la présence de la **bosse des accidents** autour de 30 ans sur cette pyramide des âges. Le nombre d'entrées en incapacité **croît** ensuite régulièrement avec l'âge, connaît **un pic à l'âge de 52 ans**, avant de **diminuer** fortement à partir de **59 ans**.

La tendance globale de la courbe est sensiblement identique à celle de la vision précédente.

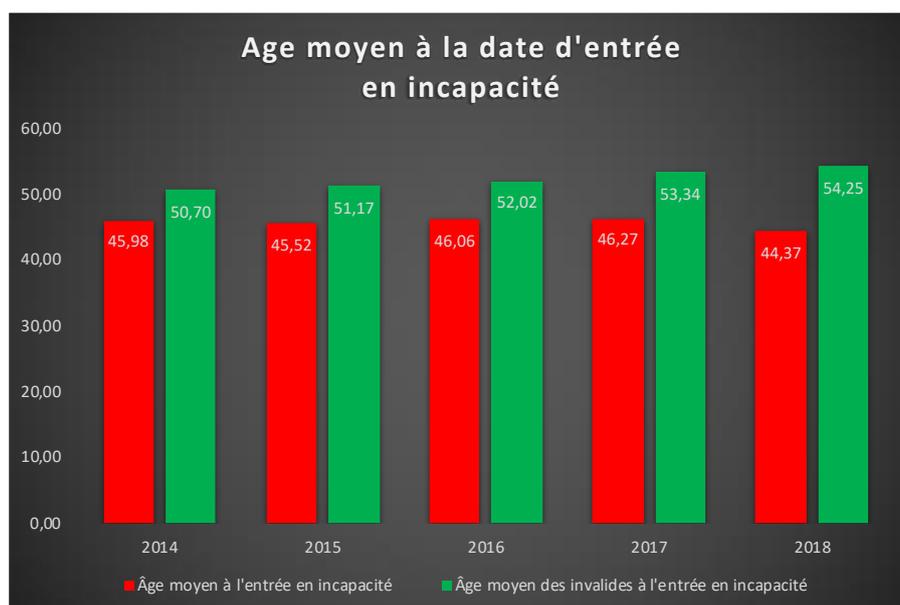


On peut constater, comme pour les résultats de la vision « stock », que :

- Les arrêts entraînant une incapacité **surviennent principalement entre 40 et 59 ans** (à moindre mesure entre 30 et 39 ans) puisque 65% des incapacités sont déclarées dans cette tranche d'âge ;
- Il y a **peu d'assurés** entrant en incapacité pour **les âges inférieurs à 30 ans**, ce qui s'explique en partie par le fait que les jeunes salariés ont souvent des emplois plus précaires ;
- **Les arrêts sont encore moins nombreux au-delà de 60 ans** ce qui s'explique sûrement par les nombreux départs à la retraite (62 ans) dans cette tranche d'âge et par la précarité des emplois lorsque l'on est proche de la retraite.

## II.2.2.2. Les assurés en invalidité

Les graphiques ci-dessous nous donnent des renseignements sur les invalides entrés en incapacité entre 2014 et 2018. Nous remarquons que les assurés déclarés invalides sont **en moyenne plus vieux** que le reste du portefeuille à leur date d'arrêt de travail. Ce qui valide la conclusion faite pour la vision « stock ».



L'âge moyen lors de l'entrée en incapacité semble **stable** chaque année (environ 46 ans et 6 mois), contrairement à celui des invalides qui **croît** en fonction du temps. Nous remarquons tout de même et comme précédemment dans la vision « stock », que l'âge moyen d'entrée en incapacité en 2018 a chuté de près de deux ans. L'explication de ce phénomène a été donnée dans la vision précédente.

L'âge moyen en invalidité a quant à lui, augmenté de plus de 3 ans durant la période d'étude, ce qui confirme une nouvelle fois les conclusions tirées de la vision précédente.

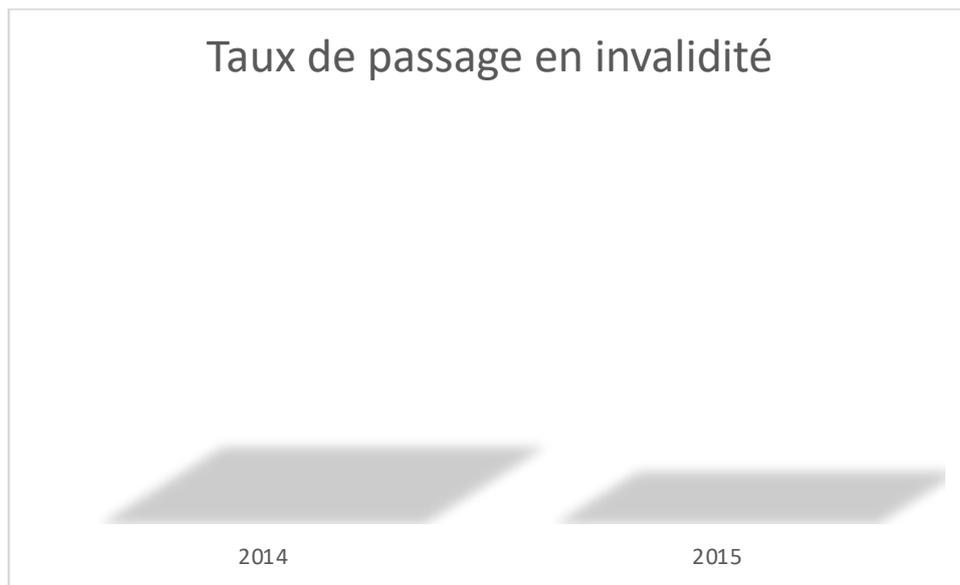


Nous notons que **très peu d'arrêts avant l'âge de 40 ans** entraînent une invalidité, tout comme les **arrêts au-delà de 60 ans**.

D'autre part, concernant les arrêts pour incapacité intervenant entre 40 et 60 ans, le nombre d'assurés reconnus invalides **croît assez régulièrement avec l'âge** pour atteindre un pic pour un âge à l'entrée en incapacité de **54 ans**. Ce qui reste proche des conclusions précédentes.

### II.2.2.3. Taux de passage en invalidité

Nous avons ici observé le taux de passage en invalidité selon les années d'entrée en incapacité et l'âge. Les années observées sont celles de 2014 et de 2015 pour les mêmes raisons que dans la partie précédente.



Il ressort tout d'abord que le taux de passage a **fortement chuté**. Il est passé de presque **10% à un peu plus de 6,5%**. Si cette chute (3,2%) est assez proche de celle observé avec l'autre jeu de données (4%), les chiffres sont eux plus de deux fois moins importants. Cela peut s'expliquer avec les mêmes arguments que pour la volumétrie totale du nombre d'assurés en incapacité puisque les arrêts plus courts non présents au 31 décembre augmentent le nombre d'assurés en incapacité mais en général pas le nombre d'invalides.

En effet peu d'assurés sont déclarés invalides lors de courts arrêts surtout si ceux-ci découlent effectivement d'un épisode de grippe ou si ceux-ci ne sont pas indemnisés par l'assureur.

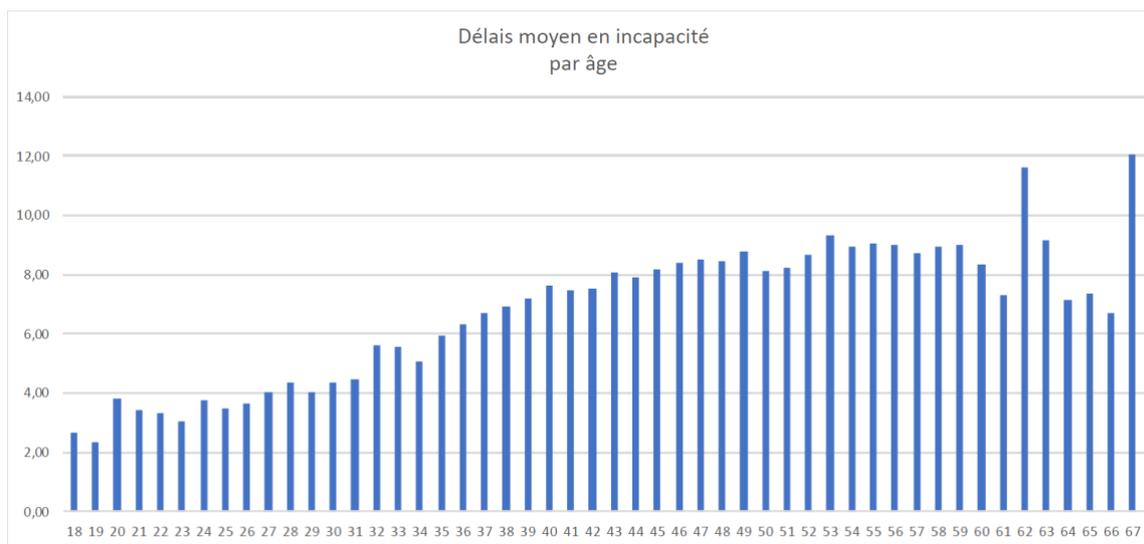


Nous remarquons que plus l'âge d'entrée en incapacité est élevé, plus la proportion d'entrée en invalidité est grande, avec un pic à **57 ans**. À cet âge-là, près de 1 assuré sur 6 passent en invalidité pendant leur période d'incapacité.

En nous appuyant sur le graphique ci-contre, l'augmentation des taux de passage avec l'âge paraît logique.

En effet, comme nous le constatons, plus l'assuré (entre 18 et 60 ans) est entré tardivement en incapacité, plus il reste en incapacité longtemps : un assuré de 18 ans reste environ en moyenne 3 mois en incapacité contre 9 mois pour un assuré de 55 ans.

On peut donc en déduire que plus on est en incapacité longtemps, plus le retour à l'emploi est difficile et plus la probabilité de passage en invalidité va être importante.



#### II.2.2.4. Délai de passage en invalidité

Concernant les délais de passages en invalidité, il ressort que, pour les années 2014 et 2015, **la moyenne est de 680 jours, alors qu'en y ajoutant les années 2016, 2017 et 2018, cette moyenne chute de 57 jours**. Cela s'explique par le fait qu'à partir de 2016 les arrêts ont été observés sur une période de moins de 3 ans. Ainsi, certains assurés ont pu être déclarés invalides après la fin de la période d'étude et augmenteraient donc la moyenne totale des délais de passage en invalidité.

		Délai moyen de passage	Délai moyen en incapacité
Année de Survenance	2014	717	302
	2015	647	211
	2016	598	214
	2017	354	171
	2018	171	95
	Total	623	205
	Total 2014-2015	680	245

Nous remarquons ici une **accélération du passage en invalidité** : de 24,2 mois en 2014, nous ne sommes plus qu'à 21,8 mois en 2015. Mais également un raccourcissement du maintien en incapacité avec en moyenne 3 mois de moins passés en incapacité en 2015 par rapport à 2014.

L'étude descriptive de ces deux fichiers nous laisse constater plusieurs phénomènes :

- Les données du premier fichier (« stock ») sont moins complètes et donc sûrement plus déficientes que celle du second jeu de données ;
- Les deux fichiers n'ont donc pas les mêmes résultats ;
- Cependant, les graphiques montrent que les données des deux jeux suivent les mêmes tendances :
  - Un nombre d'entrées en incapacité de plus en plus élevé avec le temps, expliqué par les épidémies, les modifications des modes d'enrichissement des bases de données de l'entreprise en 2015, et le vieillissement de la population ;
  - Une concentration des entrées en incapacité plus forte entre 40 et 60 ans, tranche d'âge dans laquelle la qualité de l'emploi est la meilleure (moins précaire que pour les jeunes) ;
  - Les assurés entrant en invalidité sont en grande majorité entrés en incapacité après 45 ans, et cette entrée en incapacité s'est faite pour eux en moyenne de plus en plus tard (51 ans en 2014 contre 54 en 2018) ;
  - Plus l'âge d'entrée en incapacité est tardif, plus les taux de passage en invalidité sont grands ;
  - Les délais de passage en invalidité et de maintien en incapacité ont diminué entre 2012 et 2015.

Ce dernier point concentrera la suite de ce mémoire puisque nous étudierons si cette accélération du passage en invalidité est assez significative pour, le cas échéant, revoir les tables de passage et de maintien actuelles avant de regarder l'impact de cette accélération sur les provisions de l'assureur.

# ***PARTIE III : ETUDES STATISTIQUES SUR LES ASSURES PASSANT EN INVALIDITE***

## ***Chapitre 1. Analyse des délais de passage en invalidité***

Nous travaillerons dans ce chapitre avec les données issues du fichier « stock » étant donné que, pour l'assureur, il sera plus utile et efficace d'arriver à détecter une accélération de ces délais de passage avec des données auxquelles il a accès plus facilement.

### **III.1.1. Fonction de survie et Estimateur**

#### **III.1.1.1. Fonction de survie**

L'analyse de la survie pour des êtres vivants est une étude statistique qui cherche à modéliser le temps restant avant la mort (espérance de vie), ce que l'on représente graphiquement sous la forme d'une courbe de survie. Dans cette section, nous chercherons à modéliser le temps restant avant le passage de l'état d'incapacité à celui d'invalidité.

Cette analyse de la survie va s'appuyer sur des séries temporelles de données résultant du suivi d'une population dans le temps en fonction d'un évènement de départ. Dans le cas où les évènements d'intérêt ne se sont pas produits avant la fin de la période d'observation (ici l'assuré n'est pas sorti de l'état d'incapacité pour entrer en invalidité), on parle de censure de la série de données.

La fonction de survie, notée  $S$  par convention, est définie par :  $S(t) = P(T > t)$ , où  $t$  est la variable temps et  $T$  une variable aléatoire symbolisant le moment, dans notre cas, du passage en invalidité.

C'est donc la probabilité que le passage en invalidité se produisent après un instant  $t$ .

$F(t) = 1 - S(t)$  est la fonction de distribution de longévité. La dérivée de cette fonction, notée  $f$ , est la densité de la distribution de longévité. Dans notre cas, cette dérivée est le taux de passage en invalidité par unité de temps.

On a donc  $S(t) = \int_t^{\infty} f(u)du$ .

Cette fonction est continue monotone non croissante et telle que  $\lim_{t \rightarrow \infty} S(t) = 0$  et  $S(0)=1$ .

#### **III.1.1.2. Estimateur de Kaplan-Meier**

##### **III.1.1.2.A. Principe**

L'estimateur de la fonction de survie le plus utilisé est celui de Kaplan-Meier<sup>iv</sup>. Il a été introduit en 1958 par E. KAPLAN et P. MEIER et est non-paramétrique, c'est-à-dire que l'on ne fait pas d'hypothèses a priori sur la forme de la loi de survie qu'on estime.

Il repose sur le principe suivant : la période de survie est discrétisée en sous-périodes délimitées par des dates.

La probabilité d'un assuré de ne pas changer d'état (d'incapacité à invalidité) lors d'une sous-période est équivalent au fait de ne pas changer d'état juste avant cette sous-période et de ne pas changer d'état lors de cette sous-période. La probabilité de ne pas changer d'état à une sous-période donnée peut alors se calculer conditionnellement au fait de ne pas changer d'état juste avant cette sous-période.

La courbe d'estimation de Kaplan-Meier pour la fonction de survie est une série de marches horizontales qui, quand un échantillon suffisamment grand est utilisé, permet d'approcher la fonction de survie réelle dans cette population.

### III.1.1.2.B. Définition

Soit  $S(t)$  la probabilité qu'un membre d'une population donnée ait une durée de vie (durée en incapacité) supérieure à  $t$ . Pour un échantillon de taille  $N$  dans cette population, les durées observées jusqu'à chaque décès (pour nous chaque sortie en invalidité) des membres de l'échantillon  $N$  sont :

$$t_1 \leq t_2 \leq t_3 \leq \dots \leq t_N.$$

À chaque  $n_i$  correspond un  $t_i$ ,  $n_i$  étant le nombre de personnes « à risque » juste avant le temps  $t_i$  (donc pour nous le nombre de personnes encore présentes en incapacité après  $i$  jours), et  $d_i$  le nombre de passages en invalidité au temps  $t_i$ .

On note que les intervalles entre chaque évènement ne sont pas uniformes.

L'estimateur de Kaplan-Meier est l'estimation du maximum de vraisemblance non-paramétrique de  $S(t)$ . C'est un produit de la forme :

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

Lorsqu'il n'y a aucune censure,  $n_i$  est le nombre de survivants (d'assurés encore en incapacité) juste avant le temps  $t_i$ .

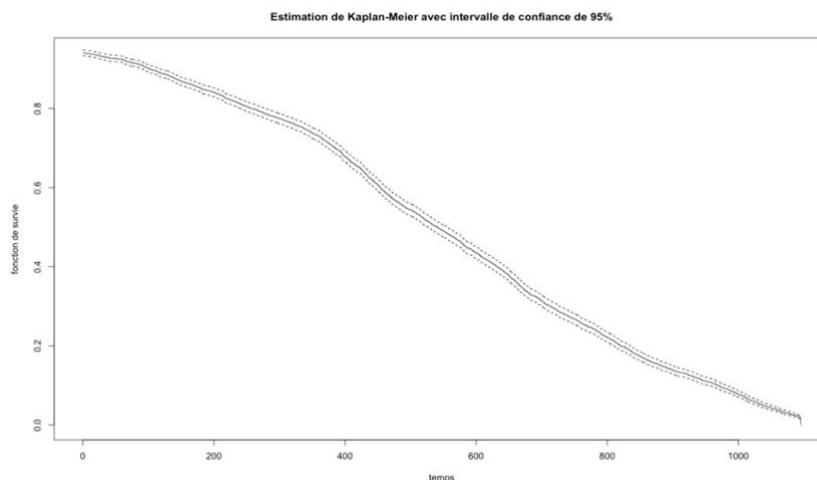
Lorsqu'il y a censure,  $n_i$  est le nombre de survivants moins le nombre de pertes (cas censurés).

Ce sont seulement ces cas survivants qui continuent à être observés (qui n'ont pas encore été censurés) qui sont « à risque » de passage en invalidité observé.

Nous travaillons ici avec les données « stock » sur la période 2012 à 2015 car il serait absurde de travailler sur des délais de passage en invalidité avec des assurés dont on ne sait même pas encore s'ils vont passer en invalidité un jour.

Nous évitons donc les censures puisque chaque assuré a pu être suivi sur l'intégralité (3 ans maximum) de son arrêt de travail.

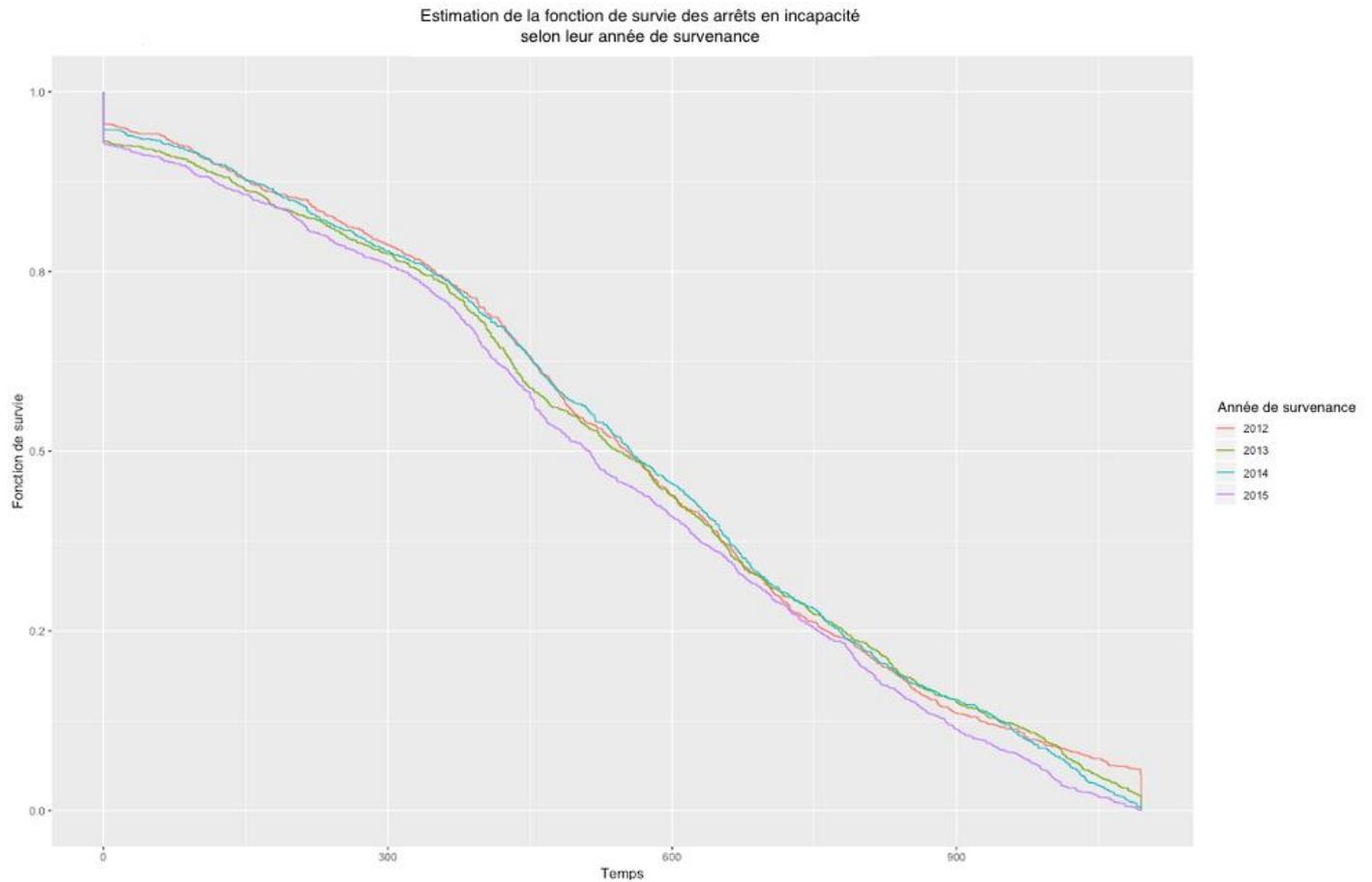
Sur l'ensemble des arrêts de travail survenus de 2012 à 2015, l'estimation de la fonction de survie est la suivante :



Graphique 1 : Fonction de survie estimée à l'aide de Kaplan-Meier

Ici (graphique 1), les sous-périodes sont assez petites pour obtenir la forme d'une droite bien que cela reste une fonction en escalier.

Nous remarquons aussi que la différence entre les deux bornes de l'intervalle de confiance de 95% est assez faible ce qui valide une bonne qualité de l'estimateur.



En nous concentrant maintenant sur l'estimateur de chaque année de survenance (graphique 2), nous observons clairement que les assurés déclarés invalides des suites d'une incapacité en 2015 sont déclarés plus rapidement invalides que ceux qui ont connu une incapacité dans les années antérieures.

Les probabilités « 2015 » (courbe violette du graphique 2) que le passage se produise à un instant  $t$  semblent toutes inférieures aux probabilités des années précédentes qui elles, semblent décroître à la même vitesse.

Nous allons donc essayer de vérifier cette impression d'accélération en analysant plus profondément ces délais de passage d'un état à l'autre.

### III.1.2. Histogrammes et densités

L'histogramme suivant donne une vue globale et détaillée de la distribution des délais dans notre population.

Nous traçons de plus les densités empiriques pour chaque année de survenance en utilisant la méthode non paramétrique des noyaux (kernel en anglais). Elle se fonde sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support.

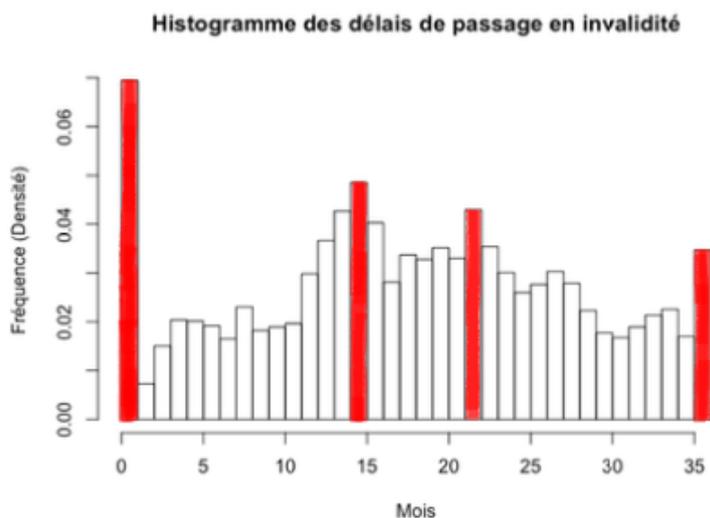
En ce sens, cette méthode généralise astucieusement la méthode d'estimation par un histogramme.

Si  $t_1, t_2, \dots, t_N \sim f$  est un échantillon i.i.d. d'une variable aléatoire (ici ce sont les N délais observés), alors l'estimateur non-paramétrique par la méthode du noyau de la densité est :

$$\hat{f}(t) = \frac{1}{Nh} \sum_{i=1}^N K \times \frac{(t - t_i)}{h}$$

Où K est un noyau et h un paramètre nommé fenêtre, qui régit le degré de lissage de l'estimation (ici nous le mettrons égal à 1).

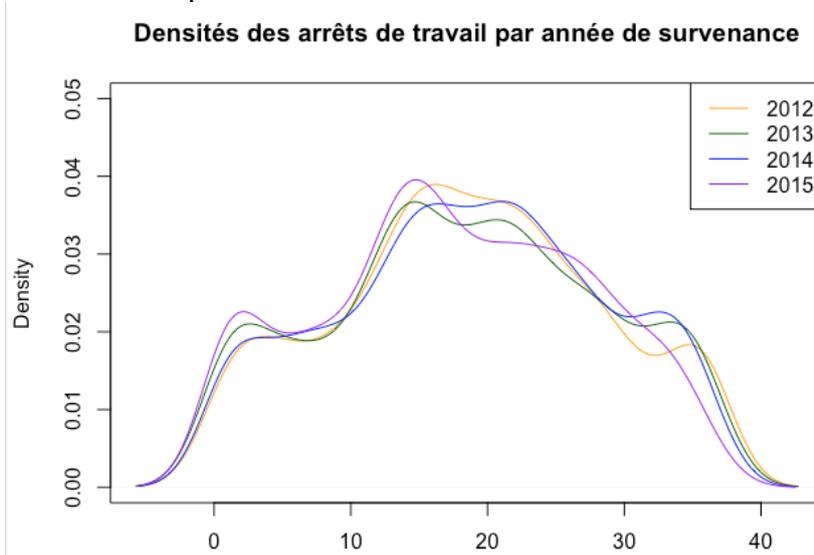
Généralement, K est choisi comme la densité d'une fonction gaussienne standard (espérance nulle et variance unitaire) :  $K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$ . C'est ce que nous choisissons ici.



Graphique 3 : Histogramme des délais de passage en invalidité

Nous visualisons ainsi très bien dans le graphique 3 la tendance globale des délais de passage pour notre portefeuille d'assurés en incapacité :

- Près de 7% (une majorité) des invalides sont passés dans cet état au cours du premier mois. Cela est dû aux assurés déclarés invalides directement après leur arrêt de travail. Ils ne sont pas réellement passés de l'état d'incapacité à celui d'invalidité mais plutôt de celui de validité à celui d'invalidité. Ils restent très importants à observer car leur augmentation au cours des années fait partie intégrante du phénomène d'accélération du passage en invalidité ;
- Des pics interviennent aux 15<sup>ème</sup> et 22<sup>ème</sup> mois ;
- Le dernier mois connaît lui aussi un fort taux de passage puisque les assurés encore présents en incapacité passent automatiquement en invalidité à l'issue de ce mois-ci.



Graphique 4 : Distribution (en fréquence) des délais de passage en invalidité

Nous remarquons là encore dans le graphique 4 que les délais de passages « 2015 » (en violet) semblent suivre une répartition bien différente des autres années, bien qu'il puisse y avoir un peu de ressemblance avec l'année 2013 (en vert).

Nous nous apercevons tout de même que la densité 2015 est plus élevée avant le 18<sup>ème</sup> mois et donc qu'une proportion plus grande d'assurés en état d'incapacité est passée avant ce jour en invalidité par rapport aux années précédentes.

En effet le pic observé chaque année aux environs de ce mois est décalé vers la gauche pour l'année de survenance 2015 et est plus élevé, ce qui signifie qu'il a eu lieu plus tôt que les autres années et que la proportion d'invalides déclarés à cette période est plus grande que pour les années de survenance 2012 (en jaune), 2013, et 2014 (en bleu).

Il y a également beaucoup moins de passages en fin d'incapacité (au bout des 3 ans) pour cette année 2015.

Nous allons à présent essayer de modéliser ces taux de passages pour ainsi comparer à l'aide de tests statistiques les distributions de chaque année.

Cela nous permettra de savoir si la répartition des passages en invalidité de l'année 2015 est significativement différente de celle des autres années et s'il est donc nécessaire de créer une nouvelle loi de passage en invalidité (et donc de maintien en incapacité).

Nous allons commencer par voir si ces densités peuvent s'approcher de lois unimodales.

### III.1.3. Modélisation des densités

#### III.1.3.1. Loi de Weibull

Nous allons tenter à présent d'approcher ces distributions empiriques par une loi existante pour en tirer des conclusions sur une réelle différence de distribution entre les années de survenance.

Nous essayons d'abord d'ajuster une loi de Weibull<sup>v</sup>.

Avec deux paramètres (pour un temps  $t > 0$ ) la densité de probabilité de cette loi est :

$$f(t, k, \lambda) = \frac{k}{\lambda} \left(\frac{t}{\lambda}\right)^{k-1} e^{-\left(\frac{t}{\lambda}\right)^k},$$

Où

- $k > 0$  est le paramètre de forme ;
- $\lambda > 0$  le paramètre d'échelle de la distribution.

Sa fonction de survie est définie par :

$$S(t, k, \lambda) = e^{-\left(\frac{t}{\lambda}\right)^k},$$

et sa fonction de répartition complémentaire de défaillance est définie par :

$$F(t, k, \lambda) = 1 - e^{-\left(\frac{t}{\lambda}\right)^k}.$$

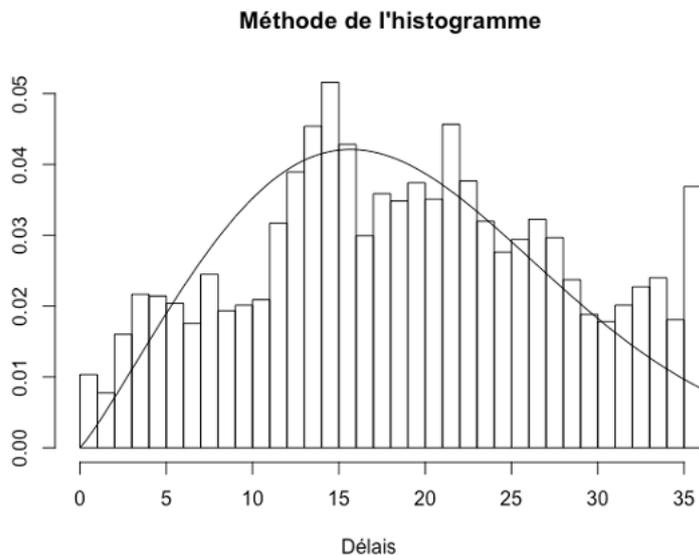
Bien que cette loi ait peu de chances de s'adapter correctement à nos données à raison du fait que nous avons une grande quantité de passage au temps 0 et que le temps doit être supérieur à 0 pour une loi de Weibull,

nous tentons quand même d'ajuster cette distribution car elle est souvent utilisée dans le domaine de l'analyse de la durée de vie, grâce à sa flexibilité : elle permet en effet de représenter au moins approximativement une infinité de lois de probabilité.

Nous commençons par approximer la densité en déterminant les paramètres d'échelle et de forme.

Puis nous vérifions l'ajustement visuellement à l'aide de la méthode de l'histogramme.

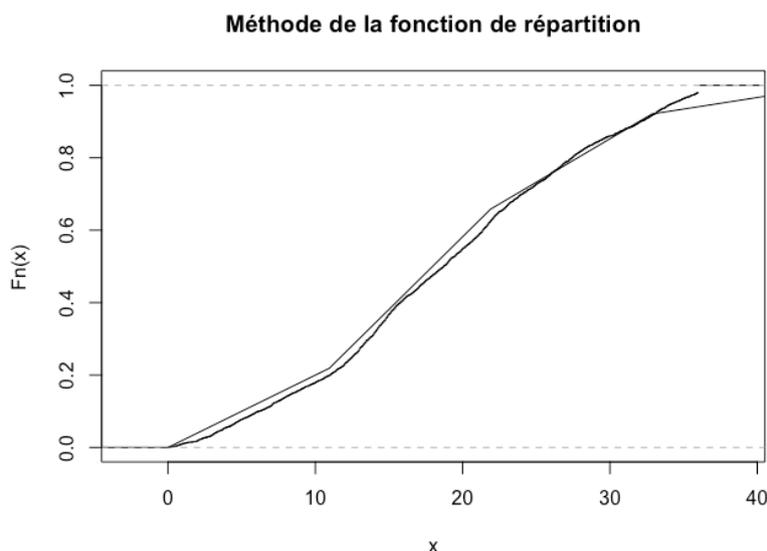
Par cette méthode nous observons dans le graphique 5 ci-dessous que la loi choisie ne semble pas correspondre à notre distribution.



Graphique 5 : Histogramme des délais de passage en invalidité et comparaison avec modèle de Weibull

Nous traçons tout de même dans le graphique 6 la courbe de la fonction de répartition empirique pour la comparer avec celle obtenue par la loi de Weibull. Cette fonction de répartition empirique est définie par :

$$F_N(t) = \frac{\text{Nombre de délais} \leq t}{N} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{t_i \leq t}$$



Graphique 6 : Fonctions de répartition empirique et de Weibull

Une fois encore les fonctions de répartition semblent significativement différentes et nous concluons donc que cette loi ne nous permet pas de modéliser les délais de passage en invalidité de nos assurés.

Il nous faut donc essayer d'ajuster une autre loi à nos données.

### III.1.3.2. Loi Normale

Pour rappel, la densité de probabilité d'une loi normale<sup>vi</sup> est :

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}.$$

Sa fonction de répartition, elle, vaut :

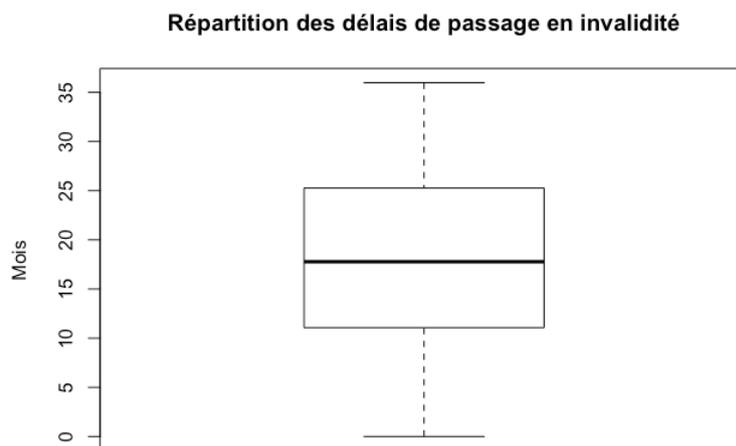
$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}x^2} dx.$$

Cette loi, au vu de notre histogramme, semble potentiellement convenir à notre modèle.

En effet, en faisant abstraction des mois 1 et 36, la distribution pourrait faire penser à une courbe en cloche, dont le pic serait atteint au milieu de la période maximale de maintien en incapacité.

De plus cette loi peut comprendre un temps de départ nul et est l'une des lois de probabilité les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs évènements aléatoires.

En traçant la boîte à moustache représentant les quartiles de nos délais de passage en invalidité, il ressort également que la loi normale pourrait bel et bien s'ajuster correctement.



Graphique 7 : Boîte à moustache des délais de passage en invalidité

En effet, aucune valeur extrême ne se signale et la répartition semble bien symétrique autour de la médiane. Nous pouvons donc être dans le cas d'une répartition normale.

Nous calculons à présent les coefficients d'asymétrie et d'aplatissement.

Le premier vaut  $\frac{N}{(N-1)\times(N-2)} \times \sum_{i=1}^N \left(\frac{t_i - E(t)}{\text{Var}(t)}\right)^3$ ,  $E(t)$  étant la moyenne des  $N$  délais  $t_i$  et  $\text{Var}(t)$  la variance des  $t_i$ .

Dans notre cas, ce coefficient est négatif ce qui implique que notre distribution est asymétrique à droite.

Le second coefficient vaut  $\frac{N \times (N+1)}{(N-1)\times(N-2)\times(N-3)} \times \sum_{i=1}^N \left(\frac{t_i - E(t)}{\text{Var}(t)}\right)^4 - \frac{3 \times (N-1)^2}{(N-2)\times(N-3)}$ .

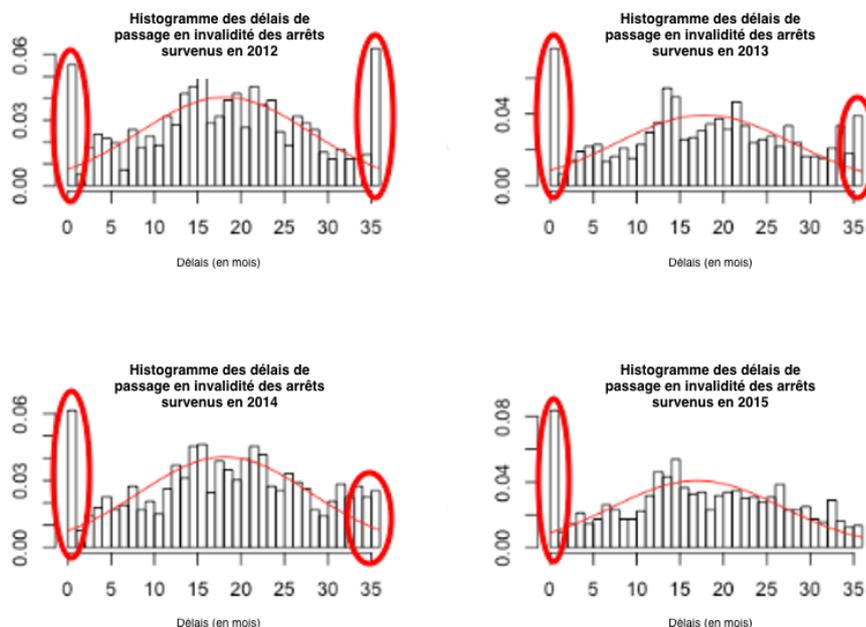
Il est ici négatif ce qui montre que la distribution est platikurtique (plus aplatie que la distribution normale).

Ces coefficients ayant par ailleurs des valeurs relativement faibles, nous pouvons continuer à considérer correcte l'idée de modéliser nos délais par une loi gaussienne.

Toujours dans le but de prouver l'existence d'une accélération du passage en invalidité à partir de l'année 2015, nous décidons d'étudier la possible normalité de la distribution de nos données pour chaque année de survénance. Ainsi en cas de bon ajustement, nous pourrions comparer les écarts des modèles et ainsi savoir si ces derniers sont assez significatifs pour conclure à une accélération.

Comme nous pouvons le voir ci-dessous dans le graphique 8, les passages « directs » du premier mois semblent augmenter avec le temps. Ainsi, moins de 6% des invalides étaient déclarés directement après leur arrêt de travail en 2012 contre plus de 8% en 2015.

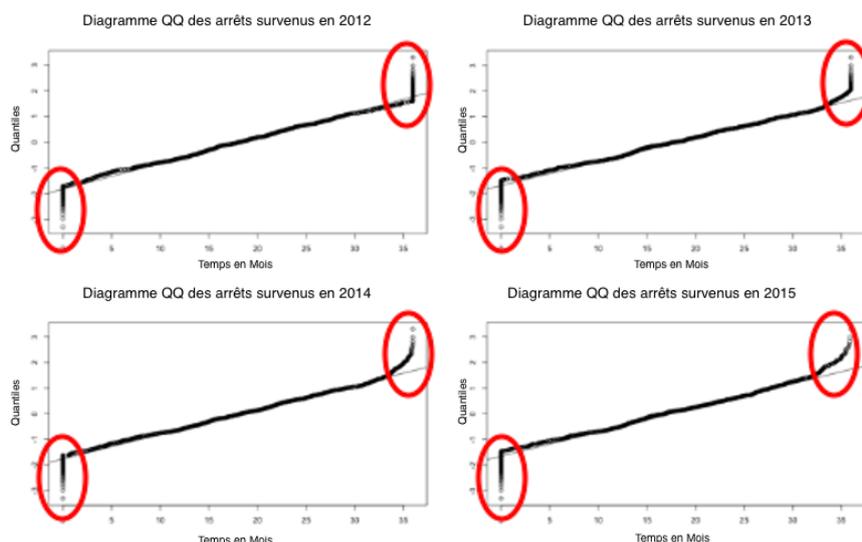
De plus, les fonctions de répartition théorique des lois normales d'espérance égale à la moyenne empirique et de variance égale à la variance empirique semblent s'ajuster plus ou moins correctement aux histogrammes des divers échantillons.



Graphique 8 : Histogramme des délais de passage en invalidité et comparaison avec modèle normale selon l'année de survénance

Les plus grosses erreurs d'ajustement semblent se produire au niveau des mois extrêmes (1 et 36). Mais également sur le fait que nos délais semblent connaître au moins deux pics (15<sup>ème</sup> et 22<sup>ème</sup> mois) quand une loi normale ne comporte qu'une « cloche ».

Nous traçons les nuages de points des diagrammes QQ :



Graphique 9 : Diagrammes Quantile-Quantile selon l'année de survénance des arrêts

Comme nous pouvons le constater, et hormis en début et fin de queue de distribution nos nuages de points s'alignent très correctement à la QQ line (graphique 9). Ce qui indique que notre loi s'ajuste assez bien à notre jeu de données.

Nous réalisons donc un test de Kolmogorov-Smirnov, pour conclure définitivement sur cette modélisation.

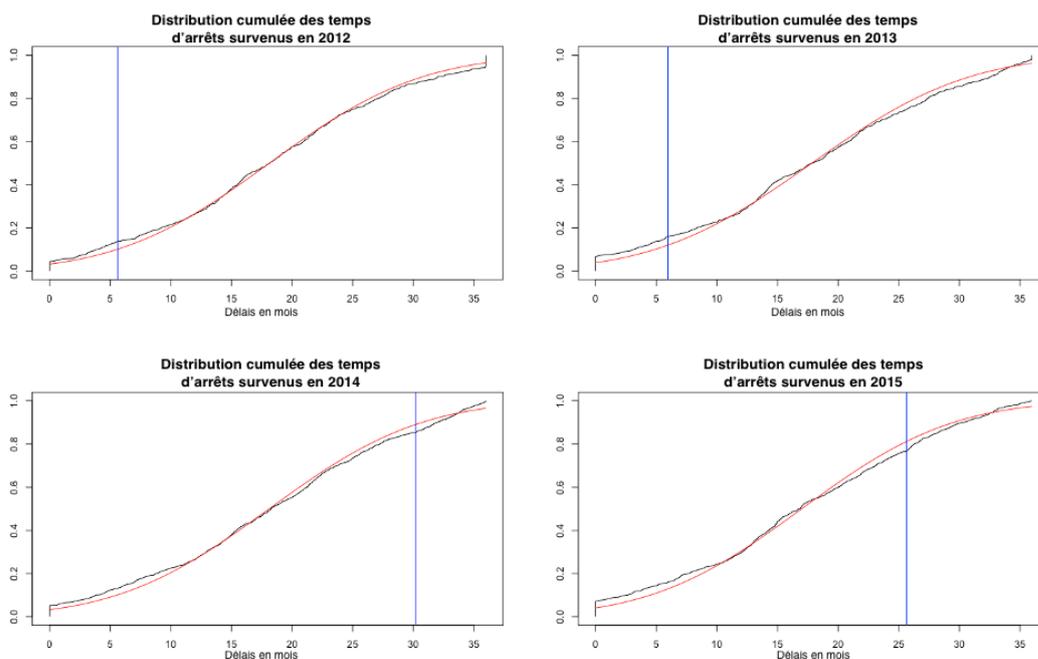
Ce test est un test d'hypothèse non paramétrique utilisé pour déterminer si un échantillon suit bien une loi donnée connue (ici Gaussienne) par sa fonction de répartition continue. Le principe est simple : on mesure l'écart maximum qui existe entre notre fonction de répartition empirique et la fonction de répartition théorique (ici Gaussienne).

Si  $F_N$  est la fonction de répartition empirique et  $F$  la fonction de répartition de la loi de probabilité théorique, alors on a  $d = \max(|F_N(t) - F(t)|)$ . Soit  $D$  la variable aléatoire qui prend la valeur de  $d$ .

Notons  $H_0$  l'hypothèse : « *La distribution des délais de passages par mois suit une loi gaussienne* ».

Sous l'hypothèse  $H_0$ ,  $d$  tend vers 0. La distribution de  $D$  fait l'objet des tables de Kolmogorov, qui prennent en compte l'effectif de l'échantillon et le seuil de risque accepté : il suffit alors de comparer  $d$  à la valeur idoine de  $D$  dans la table.

Ce test est représenté dans les graphiques ci-après :



Graphique 10 : Tests de Kolmogorov-Smirnov selon l'année de survenance de l'arrêt

La courbe noire représente la distribution cumulée observée tandis que la courbe rouge est la distribution cumulée théorique.

La ligne bleue indique le  $D_{max}$ , c'est-à-dire l'endroit où l'écart entre les données observées et théoriques est le plus important.

Mais hormis pour l'année 2015 lors de laquelle cet écart maximal est de 0,05, ce qui correspond à une probabilité de dépassement de 0,02 (ce qui est inférieur à 0,05 (notre seuil de risque) et nous conduit donc à rejeter l'hypothèse selon laquelle les délais des arrêts survenus en 2015 et ayant entraîné une invalidité suivent une loi normale), nous ne pouvons rejeter la normalité pour les autres années car nos p-valeur (probabilité de dépassement) sont supérieures à 0,05.

En revanche, vu que les diagrammes QQ de toutes les années de survenance semblaient plutôt conclure à une normalité de chaque distribution des délais de passage en invalidité et que le test de KS peut parfois s'avérer inexact, nous allons tout de même soumettre nos quatre échantillons au test de Student (t-test) pour savoir si l'un d'eux suit une distribution significativement différente des autres.

Pour réaliser ce test il était nécessaire :

- De satisfaire à l'hypothèse de la normalité,
- Que les échantillons soient indépendants,
- Qu'il n'y ait ni d'asymétrie marquée ni valeurs extrêmes.

Ces hypothèses sont satisfaites pour chacune des années étudiées (cf ci-avant).

Mais ce test est également un test paramétrique dans lequel la statistique de test calculée suit une loi de Student lorsque l'hypothèse nulle est vraie. L'hypothèse nulle postule l'égalité entre les variances de deux échantillons.

Il nous faut donc encore prouver l'homoscédasticité et donc procéder à un test du F (de Fisher-Snedecor).

Ce test repose sur le fait qu'une somme de carrés d'écart à une moyenne observée (variables centrées au carré) divisée par une vraie variance, suit une loi du  $\chi^2$ .

Or sous l'hypothèse de l'égalité des variances, les deux variances sont identiques. Le test est donc un simple rapport de deux variances estimées de deux échantillons, qui suit une loi de Fisher-Snedecor de statistique  $F(n_1 - 1 ; n_2 - 1)$  (car cette loi est celle du rapport entre deux v.a. indépendantes qui suivent chacune une loi du  $\chi^2$ .)

Ici dans chacun des tests, l'hypothèse de l'égalité des variances est vérifiée et nous pouvons donc réaliser le t-test.

Le t-test calcule dans un premier temps l'erreur-type de la différence des moyennes à partir de l'écart-type et de la taille de chaque groupe.

L'erreur-type :

$$S_{X_1-X_2} = \sqrt{\frac{(n_1 - 1)S^2 + (n_2 - 1)S^2}{(n_1 - 1) + (n_2 - 1)}}$$

Avec  $n_i$  le nombres d'observations dans l'échantillon  $i$  et  $S$  la variance supposée égale des deux échantillons.

Puis il calcule la valeur de  $t$  :

$$t = \frac{X_1 - X_2}{S_{X_1-X_2}}$$

Où  $X_i$  est le vecteur des observations de l'échantillon  $i$ .

Ce calcul nous indique à combien d'unités d'erreurs-types se situe la différence observée de la moyenne populationnelle de 0.

Après calcul, nous obtenons des p-valeurs supérieures à 0,05 lorsque l'on compare les années 2012, 2013 et 2014 entre elles, ce qui signifie qu'elles ne sont pas significativement différentes.

Mais lorsqu'on les compare à l'année 2015, cette p-valeur est assez faible pour conclure à une différence significative des échantillons.

Nous mettons ci-après les résultats obtenus entre les échantillons réunis des 3 premières années et celui de l'année 2015.

```
> var.test(DélaisSANS15,Délais15)
```

```
F test to compare two variances
```

```
data: DélaisSANS15 and Délais15
```

```
F = 1, num df = 3000, denom df = 1000, p-value = 0.5
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
```

```
0.9 1.1
```

```
sample estimates:
```

```
ratio of variances
```

```
1
```

```
> #L'hypothèse de l'égalité des variances n'est pas rejetée, on peut donc utiliser le test de student (t). Si ce n'avait pas été le cas, on aurait fait le test de Welch.
```

```
> t.test(DélaisSANS15,Délais15)
```

```
Welch Two Sample t-test
```

```
data: DélaisSANS15 and Délais15
```

```
t = 3, df = 2000, p-value = 0.004
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
0.3 1.7
```

```
sample estimates:
```

```
mean of x mean of y
```

```
18 17
```

Nous pouvons donc conclure dans ce cas sur une accélération des délais de passage en invalidité entre 2014 et 2015 puisque nous avons constaté que les délais de passage des assurés entrés en incapacité en 2015 étaient en moyenne significativement inférieurs aux délais des années antérieures.

## Chapitre 2. Loi multimodale : Mélange de lois Gaussiennes

Si la section précédente nous a permis de constater statistiquement une accélération du passage en invalidité, nous allons à présent tenter de mieux approcher ces densités en utilisant une loi multimodale<sup>vii</sup>. En effet, les ajustements effectués avec les lois normales et les différents histogrammes vus ci-avant (où nous constatons plusieurs « pics ») nous ont fait penser que ce type de loi, en vue d'une modélisation, pourrait être beaucoup plus efficace et précis.

Le modèle de mélange gaussien<sup>viii</sup> est une combinaison linéaire de plusieurs composantes gaussiennes. Il est particulièrement utilisé dans le cas où les données en études ne peuvent pas être modélisées par une simple gaussienne. En d'autres termes, si la structure de données est composée naturellement de plusieurs groupes, il faut les représenter par un modèle de mélange gaussien plutôt qu'une simple distribution gaussienne, ce qui semblent être le cas dans notre échantillon.

Ces mélanges qui supposent les populations conditionnelles distribuées selon une loi normale, suscitent un intérêt important en raison :

- De leur flexibilité ;
- De leur faculté à approcher une grande variété de densités : on peut toujours modéliser des données continues par un mélange gaussien ;
- De leur usage mathématiquement simple et de la généralité de la loi normale qu'atteste le théorème central limite.

### III.2.1. Théorie

Les modèles de mélange de lois de probabilité consistent à supposer que les données proviennent d'une source contenant plusieurs sous-populations homogènes appelées composants.

La population totale est un mélange de ces sous-populations. Le modèle résultant est un modèle de mélange.

Soit  $X = (X_1, \dots, X_n)$  un échantillon de variables aléatoires indépendantes identiquement distribuées (i.i.d.) de loi mélange à  $K$  composants, de densité  $f$  dont la forme générale est :

$$f(x) = \sum_{k=1}^K \Pi_k f_k(x) \quad (1).$$

Avec :

- $\Pi_k$  les proportions respectives des sous populations telles que

$$0 < \Pi_k \leq 1 \text{ et } \sum_{k=1}^K \Pi_k = 1$$

- $f_k$  la densité du  $k^{\text{ième}}$  composant (la paramétrisation des densités des composants dépend de la nature continue ou discrète des données observées).

Les modèles de mélanges sont des modèles à données manquantes. En effet, si l'on échantillonnait dans une population formée de  $K$  sous-populations, on devrait avoir les couples  $(X_i, Z_i)$  où  $X_i = x_i$  représente la mesure faite sur le  $i^{\text{ème}}$  individu et  $Z_i = k$  indique le numéro de la sous-population à laquelle appartient cet individu.

En échantillonnant uniquement dans la sous-population  $k$  et en supposant  $X$  discrète, on obtiendrait le modèle  $P(X = x | Z = k) = f_k(x, \alpha_k)$  mais le paramètre  $\alpha_k$  est en général inconnu et propre à la  $k^{\text{ième}}$  sous population.

De même, dans les modèles de mélanges, les données manquantes sont ainsi  $Z = (Z_1, \dots, Z_n)$  avec  $Z_i = k$  si  $i$  provient du groupe  $k$ .

L'on observe donc seulement l'échantillon  $(X_1, \dots, X_n)$ . Le mélange (1) peut être vu comme la loi marginale de la variable  $X$  pour le couple  $(X, Z)$ . C'est donc un modèle à données manquantes.

Il faut noter que pour pouvoir estimer des paramètres, il faut que les paramètres soient identifiables, c'est à dire si  $\Theta = \Theta'$  alors  $P_\Theta = P_{\Theta'}$ .

En ce qui concerne les mélanges, les paramètres ne sont pas identifiables au sens classique. Il faut donc des conditions d'identifiabilités spécifiques aux mélanges.

On pourra dire que  $H = \{H(\cdot) \mid H(x) = \sum_{k=1}^K \Pi_k F(x, \alpha_k), \Pi_k \geq 0, \sum_{k=1}^K \Pi_k = 1\}$  est identifiable si :

$$H = \sum_{k=1}^K \Pi_k F(\cdot, \alpha_k) \equiv H' = \sum_{k=1}^K \Pi'_k F(\cdot, \alpha'_k) \Rightarrow \sum_{k=1}^K \Pi_k \delta_{\alpha_k} = \sum_{k=1}^K \Pi'_k \delta_{\alpha'_k}.$$

Autrement dit, il existe une permutation  $\sigma$  de  $\{1, \dots, K\}$  telle que pour tout  $k$ , on a  $\Pi_k = \Pi'_{\sigma(k)}$  et  $\alpha_k = \alpha'_{\sigma(k)}$  avec  $F(\cdot, \alpha_k)$  la fonction de répartition du  $k^{\text{ième}}$  composant du mélange. Il s'agit d'une notion d'identifiabilité à permutation des classes près.

Proposition :

Les modèles de mélanges gaussiens uni variés :

$\{H(\cdot) \mid H(x) = \sum_{k=1}^K \Pi_k F(x, \alpha_k), \Pi_k \geq 0, \sum_{k=1}^K \Pi_k = 1\}$ , avec  $F(\cdot, \alpha_k)$  la fonction de répartition d'une gaussienne de densité  $f(x, \alpha_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$  et  $\alpha_k = (\mu_k, \sigma_k)$ , sont identifiables.

Théorème (admis) :

La famille  $H = \{H(\cdot) \mid H(x) = \sum_{k=1}^K \Pi_k F(x, \alpha_k), \Pi_k \geq 0, \sum_{k=1}^K \Pi_k = 1\}$  est identifiable si et seulement si l'image de  $F$  par tout isomorphisme défini sur le sous espace vectoriel engendré par l'ensemble des fonctions de répartition  $F$  est une famille libre dans l'espace d'arrivée.

*Preuve de la proposition :*

Dans le cadre des mélanges gaussiens, l'isomorphisme qui nous intéresse est celui qui aux fonctions de répartition associent les transformées de Laplace correspondantes.

Rappelons la formule de la transformée de Laplace :  $L_X(z) = E[e^{zX}]$ .

La transformée de Laplace d'une gaussienne de densité  $f(x, \alpha_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$  est :

$$e^{\mu_k z + \frac{1}{2}\sigma_k^2 z^2}.$$

On veut montrer que la famille des transformées de Laplace d'une gaussienne de densité  $f(x, \alpha_k) = \frac{1}{(2\pi\sigma_k^2)^{\frac{1}{2}}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$  est libre.

Supposons donc que l'on ait :  $\sum_{k=1}^K \lambda_k e^{\mu_k z + \frac{1}{2} \sigma_k^2 z^2} = 0$  avec  $(\lambda_k)_{k=1, \dots, K}$  des réels non tous nuls. Sans perte de généralité, on suppose que  $\sigma_1^2 < \dots < \sigma_K^2$ .

$$\text{On a aussi : } \frac{1}{e^{\mu_K z + \frac{1}{2} \sigma_K^2 z^2}} \sum_{k=1}^K \lambda_k e^{\mu_k z + \frac{1}{2} \sigma_k^2 z^2} = 0 \Leftrightarrow \sum_{k=1}^K \lambda_k e^{(\mu_k - \mu_K)z + \frac{1}{2}(\sigma_k^2 - \sigma_K^2)z^2} = 0$$

$$\Leftrightarrow \lambda_k + \sum_{k=1}^{K-1} \lambda_k e^{(\mu_k - \mu_K)z + \frac{1}{2}(\sigma_k^2 - \sigma_K^2)z^2} = 0.$$

Or  $(\mu_k - \mu_K)z + \frac{1}{2}(\sigma_k^2 - \sigma_K^2)z^2$  est un polynôme de degré 2 à coefficient dominant négatif qui tend vers 0 quand  $z$  tend vers  $+\infty$ .

Ainsi quand  $z$  tend vers  $+\infty$ , on obtient  $\lambda_k = 0$  et par récurrence descendante on a finalement :

$$\lambda_1 = \dots = \lambda_K = 0.$$

*Fin de la preuve de la proposition.*

Ici nous créerons un modèle de mélange de gaussiennes. Ce mélange est un mélange paramétrique. De façon générale, les mélanges paramétriques se caractérisent par l'existence d'hypothèses sur la distribution de probabilités induisant une classification. La distribution de probabilités appartient à une famille paramétrique c'est à dire que l'espace des paramètres est de dimension finie.

Pour estimer les paramètres<sup>ix</sup> nous utiliserons une approche par maximisation de la vraisemblance et pour ce faire, nous utiliserons l'algorithme EM.

Comme son nom l'indique, l'approche « Maximum de vraisemblance » consiste à maximiser la vraisemblance c'est à dire maximiser  $L(\Theta, X) = \prod_{i=1}^n \sum_{k=1}^K \Pi_k f_k(X_i, \alpha_k)$  ou de façon équivalente à maximiser la log vraisemblance  $l(\Theta, X) = \sum_{i=1}^n \log (\sum_{k=1}^K \Pi_k f_k(X_i, \alpha_k))$  afin d'estimer le paramètre inconnu.

Avec :

- $\Theta = (\Pi_1, \dots, \Pi_K, \alpha_1, \dots, \alpha_K)$  le paramètre inconnu du modèle paramétrique ;
- $X = (X_1, \dots, X_n)$  un échantillon de  $n$  variables aléatoires i.i.d ;
- $f_k(\cdot, \alpha_k)$  la densité du  $k^{\text{ième}}$  composant du modèle paramétrique : le paramètre  $\alpha_k \in R^d$ .

Toutefois, ce problème de maximisation ne peut être résolu analytiquement en raison de données cachées. Il faut donc trouver les solutions à l'aide d'algorithmes itératifs. Parmi ces algorithmes, figure l'algorithme EM.

Cet algorithme est dû à Dempster, Laird et Rubin (1977). Il vise à fournir un estimateur lorsqu'il est impossible de calculer la solution en raison de la présence de données cachées ou manquantes ou plutôt, lorsque la connaissance de ces données rendrait possible l'estimation des paramètres.

L'algorithme EM tire son nom du fait qu'à chaque itération il opère deux étapes distinctes :

- La phase « Expectation », souvent désignée comme « l'étape E », procède à l'estimation des données inconnues, sachant les données observées et la valeur des paramètres déterminée à l'itération précédente ;
- La phase « Maximisation », ou « étape M », procède donc à la maximisation de la vraisemblance, rendue désormais possible en utilisant l'estimation des données inconnues effectuée à l'étape précédente, et met à jour la valeur du ou des paramètre(s) pour la prochaine itération.

L'algorithme garantit que la vraisemblance augmente à chaque itération, ce qui conduit donc à des estimateurs de plus en plus corrects.

De façon plus formelle et plus généralement dans le cas discret :

- Si l'on dispose d'observations i.i.d.  $X = (X_1, \dots, X_n)$  de vraisemblance notée  $P(X | \Theta)$  dont la maximisation est impossible,
- L'on considère des données cachées  $Z = (Z_1, \dots, Z_n)$  dont la connaissance rendrait possible la maximisation de la « vraisemblance des données complètes »,  $P(X, Z | \Theta)$ ,
- Étant donné que ces données  $Z$  sont inconnues, l'on estime la vraisemblance des données complètes en prenant en compte toutes les informations connues : pour cela on choisit comme estimateur :  $E_{Z|X, \Theta_m}[\log(P(X, Z | \Theta))]$  (« étape E » de l'algorithme),
- L'on maximise enfin cette vraisemblance estimée pour déterminer la nouvelle valeur du paramètre (« étape M » de l'algorithme).

Ainsi, le passage de l'itération  $m$  à l'itération  $m + 1$  de l'algorithme consiste à déterminer :

$$\Theta_{m+1} = \operatorname{argmax}(E_{Z|X, \Theta_m}[\log(P(X, Z | \Theta))]).$$

Prenons par exemple un modèle de mélange de deux lois gaussiennes.

Soit  $X = (X_1, \dots, X_n)$  un échantillon i.i.d. d'observations issues d'un mélange de deux gaussiennes bidimensionnelles, et soit  $Z = (Z_1, \dots, Z_n)$  la donnée cachée où  $Z_i$  détermine la distribution dont est issue  $X_i$  :

$$L(X_i | Z_i = 1) = N(\mu_1, \Sigma_1)$$

$$L(X_i | Z_i = 2) = N(\mu_2, \Sigma_2).$$

Avec  $P(Z_i = 1) = \Pi_1$  et  $P(Z_i = 2) = \Pi_2 = 1 - \Pi_1$ .

Ne connaissant que  $X$ , on cherche à estimer les 5 paramètres inconnus :

$$\Theta = (\Pi_1, \mu_1, \Sigma_1, \mu_2, \Sigma_2)$$

La vraisemblance des données complètes est :

$$L(X, Z | \Theta) = \prod_{i=1}^n \sum_{j=1}^2 1_{Z_i=j} \Pi_j f_j(X_i).$$

Où  $f_j: R \rightarrow R$  est telle que  $f_j(x) = \frac{1}{2\pi \det(\Sigma_j)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)}$  est une densité gaussienne bidimensionnelle de paramètres  $\mu_j$  et  $\Sigma_j$ . La log-vraisemblance des données complètes obtenue est :

$$\log(L(X, Z | \Theta)) = \sum_{i=1}^n \left[ \sum_{j=1}^2 1_{Z_i=j} \left( \log(\Pi_j) - \log(2\pi) - \frac{1}{2} \log(\det(\Sigma_j)) - \frac{1}{2} (X_i - \mu_j)^T \Sigma_j^{-1} (X_i - \mu_j) \right) \right].$$

À chaque itération, l'étape E nécessite de définir la distribution de  $\Theta_m$ .

On définit  $\tilde{p}_{i,j} = P(Z_i = j | X_i = x_i, \Theta_m) = \frac{\Pi_j f_j(x_i)}{\Pi_1 f_1(x_i) + \Pi_2 f_2(x_i)}$  la probabilité pour que le point  $X_i$  soit issu de la distribution  $f_j \equiv N(\mu_j, \Sigma_j)$  connaissant  $\Theta_m$ .

Alors, on a :

$$E_{Z|X,\Theta}[\log(L(X, Z | \Theta))] = \sum_{i=1}^n \left[ \sum_{j=1}^2 \tilde{p}_{i,j} \left( \log(\Pi_j) - \log(2\pi) - \frac{1}{2} \log(\det(\Sigma_j)) - \frac{1}{2} (X_i - \mu_j)^T \Sigma_j^{-1} (X_i - \mu_j) \right) \right]$$

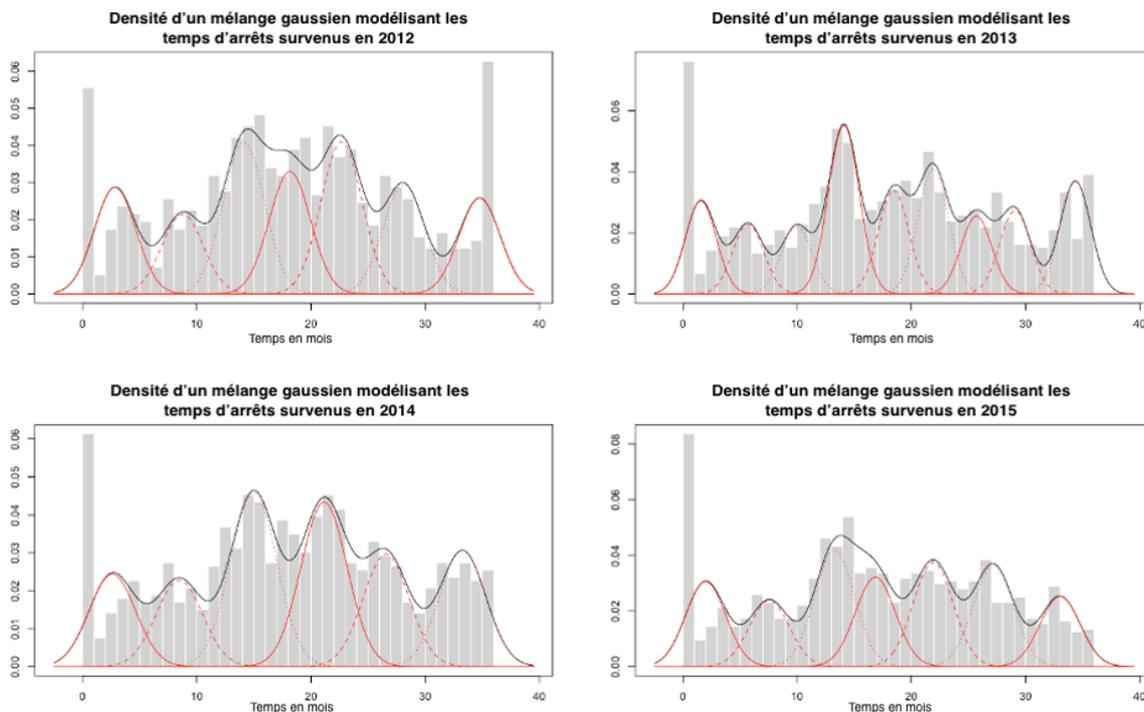
La maximisation en  $\Theta$  de cette expression conduit aux estimateurs suivants :

- $\Pi_j^{(m+1)} = \frac{\sum_{i=1}^n \tilde{p}_{i,j}}{n}$ ,
- $\mu_j^{(m+1)} = \frac{\sum_{i=1}^n \tilde{p}_{i,j} x_i}{\sum_{i=1}^n \tilde{p}_{i,j}}$ ,
- $\Sigma_j^{(m+1)} = \frac{\sum_{i=1}^n (\tilde{p}_{i,j} (x_i - \mu_j^{(m+1)})(x_i - \mu_j^{(m+1)})^T)}{\sum_{i=1}^n \tilde{p}_{i,j}}$ .

À la convergence de l'algorithme, on peut déduire une partition en rangeant chaque individu dans la classe maximisant la probabilité a posteriori  $\tilde{p}_{i,j}$  : il s'agit de la méthode du MAP ou maximum a posteriori.

### III.2.2. Pratique : Ajustement aux délais de passage en invalidité

Nous avons donc commencé par ajuster une loi de mélange de gaussiennes à chacune des quatre années qui constituent notre période d'étude à l'aide de l'algorithme EM.



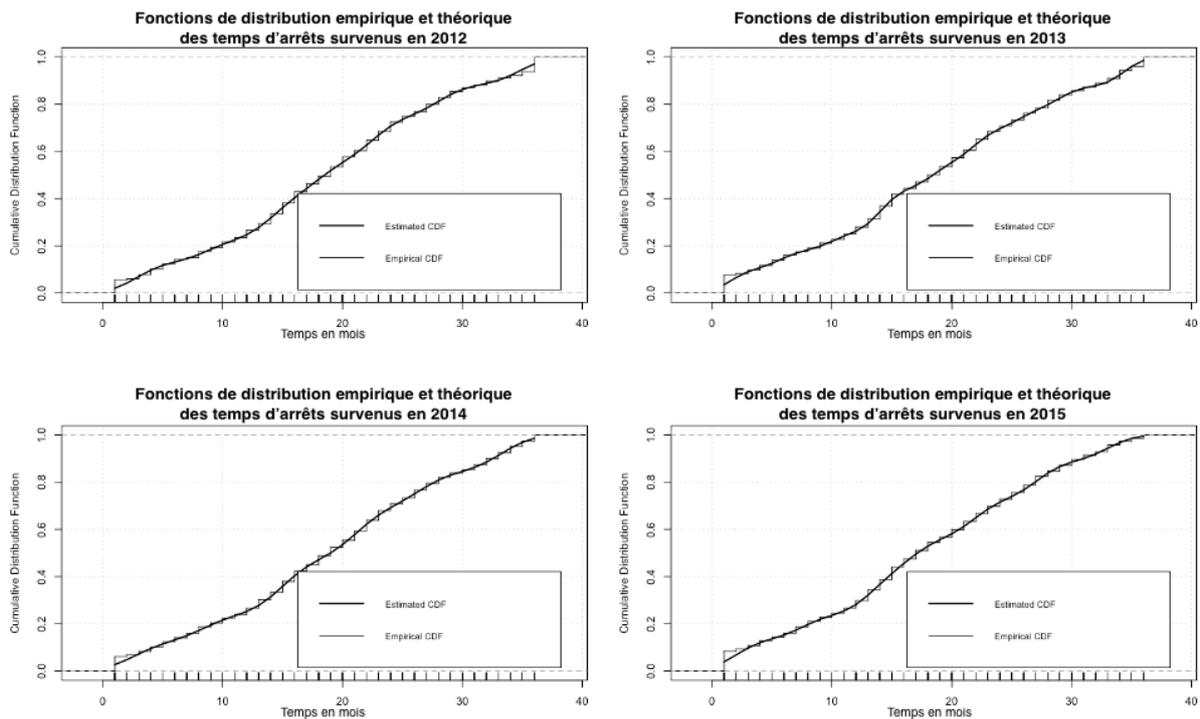
Graphique 11 : Modélisation de la densité des temps d'arrêts par un mélange de gaussiennes selon l'année de survenance

Les lois utilisent entre 6 (pour les délais dont l'entrée en incapacité des assurés s'est faite en 2014) et 9 (pour des entrées en 2013).

Les délais dont les entrées se sont faites en 2012 et 2015 sont eux, modélisés par des mélanges à 7 gaussiennes.

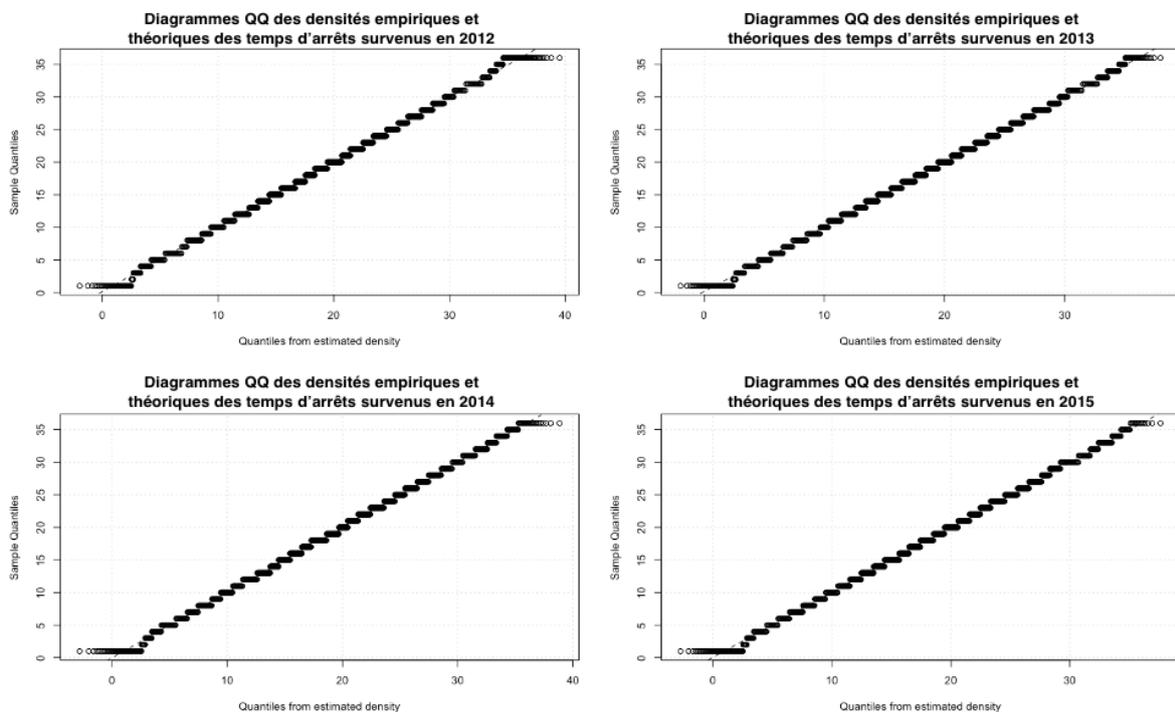
Les résultats (graphique 11) montrent que les modèles semblent très efficaces bien que la réalité du premier mois reste difficile à atteindre pour un modèle.

Nous représentons graphiquement désormais les fonctions de répartition empirique (en escalier) et théorique (en gras) pour chaque année observée.



Graphique 12 : Fonctions de distribution empiriques et de mélanges de gaussiennes selon l'année de survenance

Le résultat, visuellement, est une nouvelle fois très bon. Les fonctions de répartition des mélanges de gaussiennes correspondent parfaitement aux fonctions empiriques. Tout comme les graphiques quantiles ci-dessous, dont les nuages de points empiriques se confondent avec la première bissectrice, ce qui démontre de très bons ajustements de nos modèles.



Graphique 13 : Diagrammes QQ vérifiant l'ajustement des modèles multi-gaussiens selon l'année de survenance

Ces graphiques 12 et 13 nous montrent visuellement que le modèle de mélange de gaussiennes est celui s'ajustant au mieux à nos données.

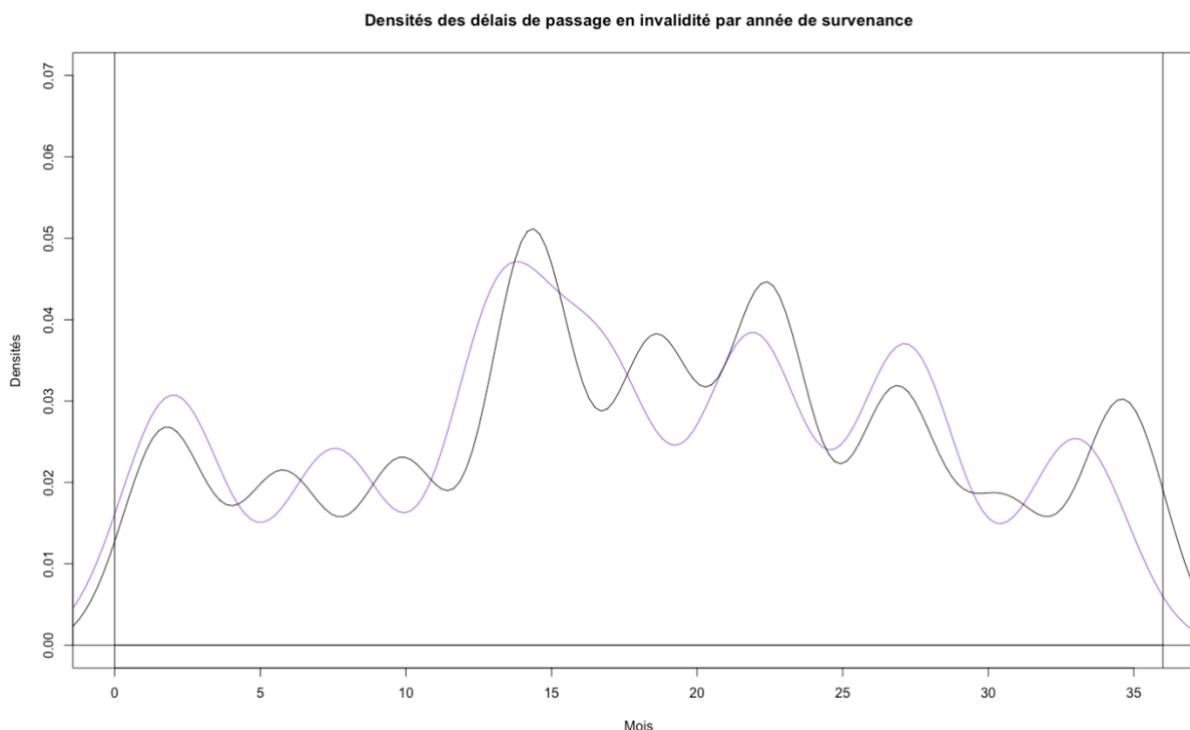
Pour le vérifier statistiquement, nous calculons le critère d'information d'Akaike<sup>x</sup> (AIC) qui est une mesure de la qualité d'un modèle statistique proposée par Hirotugu Akaike en 1973. Nous avons utilisé ce critère plutôt que le critère d'information bayésien (BIC) car l'AIC tente de sélectionner le modèle qui décrit le mieux

une réalité alors que le BIC essaie de trouver le véritable modèle parmi l'ensemble des candidats. Or, il paraît difficile que la réalité se retrouve exactement dans un des modèles considérés.

Le critère d'information d'Akaike s'écrit comme suit :  $AIC = 2k - 2\ln(L)$ , où  $k$  est le nombre de paramètres à estimer du modèle et  $L$  est le maximum de la fonction de vraisemblance du modèle.

Le modèle de mélange gaussien ayant le critère AIC le plus faible, c'est bien celui qui s'ajuste le mieux aux données. Le travail ayant été réalisé en deux temps, nous n'avons pas pu conserver les valeurs de l'AIC obtenues.

Enfin et pour valider les résultats précédents, nous traçons les densités calculées à partir des délais de passage en invalidité des assurés entrés en incapacité avant 2015 (en noir) et à partir des délais des assurés entrés en 2015 (en violet).



Graphique 14 : Densités modélisées des délais de passage en invalidité selon l'année de survenance

Ces courbes montrent très explicitement l'accélération du passage en invalidité puisque l'on se rend bien compte du décalage vers la gauche de chaque cloche de la courbe violette du graphique 14 ainsi que l'écart conséquent entre les proportions de passage en invalidité dans les derniers mois (la courbe noire étant plus haute et plus à droite que la violette).

Ces deux premiers chapitres ont clairement exposé une accélération significative des délais de passage en invalidité entre 2014 et 2015.

Cependant il nous est impossible de conclure sur une accélération durable dans le temps ou sur une simple année particulière qui restera un cas isolé dans le temps. Pour valider la première hypothèse, il nous faudrait réaliser à nouveau cette étude dans quelques années pour ainsi avoir l'ensemble des passages en invalidité des assurés entrés en incapacité après 2016 et donc bénéficier d'assez de recul pour confirmer ou infirmer cette hypothèse.

Dans la suite, nous supposerons cette hypothèse d'accélération vérifiée. Cela aura pour conséquence une modification des tables de maintien en incapacité, en invalidité et de passage en invalidité puisqu'il sera alors nécessaire à l'assureur de corriger ses provisions mathématiques faites pour le maintien en incapacité (il devra payer moins longtemps ses assurés en incapacité) et pour le maintien en invalidité (si ses assurés restent plus longtemps en invalidité, l'assureur aura des provisions de maintien dans cet état plus importantes).

## ***Chapitre 3. Création des tables de maintien en incapacité et de passage en invalidité***

L'accélération du passage en invalidité montrée ci-avant, nous pousse à revoir les tables BCAC actuelles. Cependant, le manque de recul sur ce phénomène ne nous permet pas à l'heure actuelle de créer une loi de maintien en invalidité étant donné que les assurés entrés en invalidité après ce phénomène ne sont pour l'immense majorité d'entre eux pas encore sortis de cet état.

Ce chapitre ne présentera donc que les révisions des tables de maintien en incapacité et de passage en invalidité.

Les données utilisées dans ce chapitre seront les données issues du fichiers « flux » qui nous permettent de mieux suivre les assurés en incapacité (puisque nous avons alors leur date de sortie même s'ils ne sont pas passés en invalidité contrairement aux données de la vision « stock ») et avoir un portefeuille plus fourni en nombre d'incapacités.

Cependant le faible historique de ce fichier (aucunes données antérieures à 2014) risquera de remettre en question la robustesse de l'estimation des taux de maintien en incapacité et de passage en invalidité quant au reflet de la réalité du portefeuille.

La période d'observation retenue est du 1<sup>er</sup> janvier 2014 au 31 décembre 2018 avec des données extraites à la date du 31 mars 2019. Cette période nous permet d'observer l'intégralité des sorties d'incapacité des assurés entrés entre le 1<sup>er</sup> janvier 2014 et le 31 décembre 2015.

Afin d'obtenir des lois robustes statistiquement en termes de volumes, nous regrouperons notre échantillon par tranches d'âge.

### **III.3.1. Table de maintien en incapacité**

Dans cette partie nous créerons donc une table de maintien en incapacité pour les années de survenance 2014 et 2015.

#### **III.3.1.1. Vérification des résultats précédents sur ce jeu de données**

Nous présentons tout d'abord nos données.

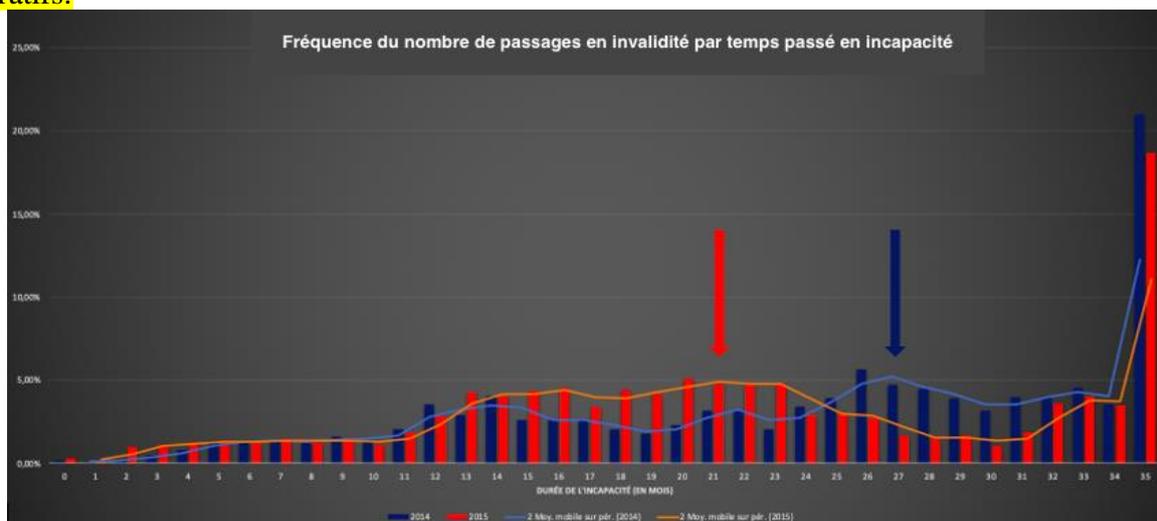
Pour rappel, le nombre plus conséquent d'entrées en incapacité en 2015 par rapport à celui de l'année 2014 s'explique en raison d'un contexte épidémiologique moins favorable en 2015.

Les entrées en incapacité sont plus nombreuses entre 45 ans et 60 ans (voir graphique 15), les tranches d'âge inférieures et supérieures reflétant la précarisation des emplois à ces âges-là (particulièrement pour les jeunes de moins de 30 ans).

Enfin la tranche d'âge (de 5 ans) la plus fournie est celle des 50-55 ans.



Les assurés mettant plus de temps à passer en invalidité sont seulement des exceptions liées à des retards administratifs.



Graphique 17 : Histogramme des fréquences de sorties d'incapacité vers l'invalidité par durée selon l'année de survenance

Enfin et comme montré dans le chapitre précédent, nous pouvons repérer l'accélération du passage en invalidité en remarquant les pics (hors 36<sup>ème</sup> mois) des moyennes mobiles des deux années, illustrés dans le graphique 17 : il est de 27 mois en 2014 (courbe bleue) contre 21 mois seulement en 2015 (courbe rouge).

Nous notons également que le nombre d'invalides entrés en incapacité en 2015 est plus important que celui des invalides entrés en incapacité en 2014 (1 555 contre 1 370).

Cette différence peut s'expliquer de plusieurs façons : soit par le fait que l'année 2015 est un cas isolé pendant lequel les assurés en incapacité ont été plus rapidement placés en invalidité ; soit par le fait de l'arrivée des baby-boomers à l'âge de la retraite, comme nous tenterons de le voir dans la partie IV, soit par une volonté pour la Sécurité Sociale de passer plus vite les assurés en invalidité car ils lui reviendront moins cher.

En moyenne les arrêts durent un tout petit peu moins de 10 mois s'ils sont survenus en 2014 contre un peu moins de 7 mois s'ils sont survenus en 2015.

Année de l'arrêt	Durée moyenne en incapacité
2014	302,49
2015	210,61

Avant de créer deux lois différentes, il nous est nécessaire de vérifier que, comme pour les données « stock », les deux échantillons ont des lois significativement différentes.

Pour commencer nous allons donc utiliser un test de Chi-2 pour déterminer si les échantillons de 2014 et de 2015 suivent la même loi. Pour cela, on construit un indice ( $\chi^2$ ) mesurant l'écart constaté entre les effectifs observés des deux années de survenance 2014 et 2015 et l'effectif théorique égale à l'effectif total des deux années de survenance (comme si les deux suivaient la même loi).

Le Chi-2<sup>xi</sup> est défini comme suit :

$$\chi^2 = \sum_{i=1}^l \frac{(O_i - T_i)^2}{T_i}$$

Où  $l$  est le nombre de classes (31 ici), les  $O_i$  sont les effectifs observés et les  $T_i$  sont les effectifs « théoriques ».

Pour pouvoir appliquer ce test il faudra tout d'abord vérifier que nos classes de durée (en mois) vérifient le critère de Cochran<sup>xii</sup>.

Ce dernier souligne qu'un test est valide si pour 80% des classes de durée  $i$ , la fréquence est supérieure à 5 et si toutes les classes ont une valeur non nulle.

De plus, Aurélie Gaumet<sup>xiii</sup> (2000) estimant que 1 000 observations sont nécessaires à la bonne réalisation d'une table de maintien en incapacité, nous décidons de former des tranches d'âge de telle sorte qu'il y ait toujours au moins 1 000 observations dans chacune d'entre elles.

Les tranches sont les suivantes : [18 ; 27], [28 ; 32], [33 ; 37], [38 ; 42], [43 ; 45], [46 ; 48], [49 ; 51], [52 ; 54], [55 ; 57], [58 ; 67].

Ce critère est validé en créant 21 classes de durée pour chacune des 10 tranches d'âge créées. Nous présentons ci-dessous les résultats pour la première tranche d'âge.

18-27 ans							
Classe de durée	2014-2015 assemblées	2014 (n1 = 991 observations)			2015 (n2 = 1 810 observations)		
	Proportion p	Fréquence	Proportion p1	Critère de Cochran n1*p	Fréquence	Proportion p2	Critère de Cochran n2*p
0 mois	15,99%	118	11,91%	159	330	18,23%	289
1 mois	15,14%	121	12,21%	150	303	16,74%	274
2 mois	19,56%	119	12,01%	194	429	23,70%	354
3 mois	16,60%	190	19,17%	165	275	15,19%	300
4 mois	11,50%	137	13,82%	114	185	10,22%	208
5 mois	6,78%	98	9,89%	67	92	5,08%	123
6 mois	4,18%	58	5,85%	41	59	3,26%	76
7 mois	1,64%	24	2,42%	16	22	1,22%	30
8 mois	0,89%	15	1,51%	9	10	0,55%	16
9 mois	0,86%	12	1,21%	8	12	0,66%	16
10 mois	0,64%	8	0,81%	6	10	0,55%	12
11 mois	0,54%	6	0,61%	5	9	0,50%	10
12 mois	0,54%	8	0,81%	5	7	0,39%	10
13 mois	0,68%	11	1,11%	7	8	0,44%	12
14-16 mois	1,46%	22	2,22%	15	19	1,05%	26
17-19 mois	0,93%	10	1,01%	9	16	0,88%	17
20-22 mois	0,61%	10	1,01%	6	7	0,39%	11
23-25 mois	0,43%	8	0,81%	4	4	0,22%	8
26-28 mois	0,29%	5	0,50%	3	3	0,17%	5
29-31 mois	0,29%	4	0,40%	3	4	0,22%	5
32-35 mois	0,46%	7	0,71%	5	6	0,33%	8
Valeur du $\chi^2$	157,16	> 31 donc on rejette l'homogénéité, les tables ne suivent pas la même loi					

Les valeurs du  $\chi^2$  sont pour chaque tranche d'âge, supérieures aux seuils de 5% (157,16 > 31 dans l'exemple ci-dessus) et nous pouvons donc rejeter l'homogénéité des tables, elles ne suivent donc pas la même loi.

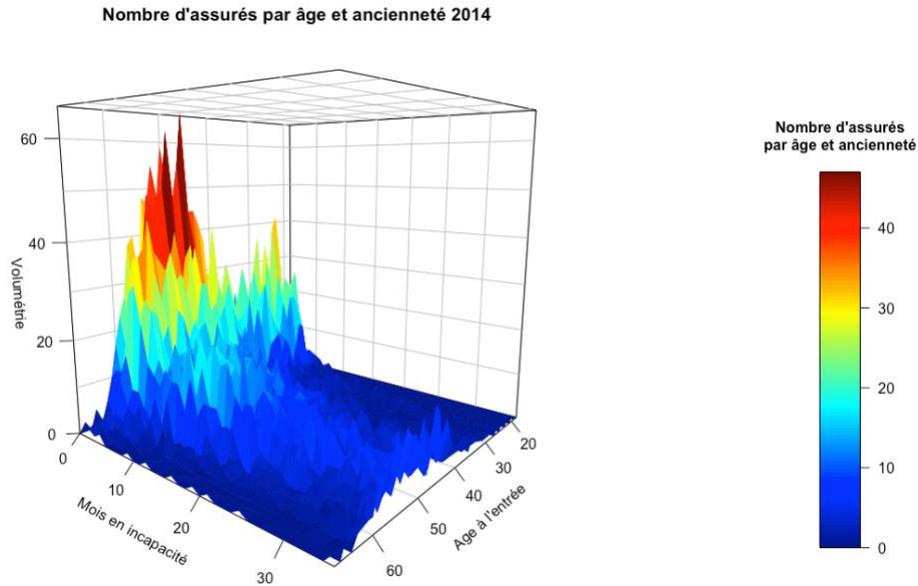
Nous aurions pu rendre ce test plus robuste en effectuant d'autres tests mais pour une question de temps, ces tests n'ont pu être réalisés. Nous pourrions ainsi citer des tests non-paramétriques plus puissants que le test paramétrique du Chi-2 comme celui de Kruskal-Wallis<sup>iii</sup> déjà cité plus haut, ou encore celui de Wilcoxon-Mann-Whitney<sup>xiv</sup>. Ce dernier, proposé par Frank Wilcoxon en 1945 et par Henry Mann et Donald Ransom Whitney en 1947 permet de tester l'hypothèse selon laquelle les médianes de chacun de deux groupes de données sont proches.

**Conclusion :** Sans surprise par rapport aux chapitres précédents, il sera pertinent de créer 2 lois différentes afin de les comparer et juger de l'impact du changement de comportement en arrêt de travail des assurés.

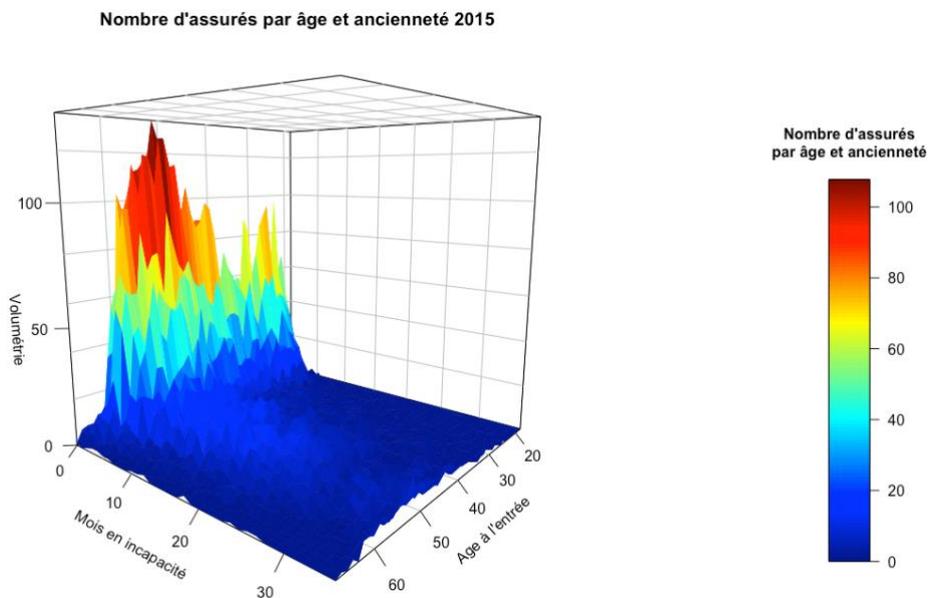
### III.3.1.2. Création de la table de maintien en incapacité

#### III.3.1.2.A. Estimateurs de Kaplan-Meier des taux bruts

Ci-dessous, nous représentons en 3D, la répartition du nombre d'assurés par âge à l'entrée en incapacité et par ancienneté dans l'état, sans classe d'âges car nous voulions d'abord représenter notre jeu de données dans son intégralité pour ainsi mieux comprendre les impératifs à utiliser des classes d'âges pour créer des tables plus robustes.



Graphique 18 : Graphe du nombre de sorties d'incapacité vers l'invalidité par mois (année de survénance 2014)



Graphique 19 : Graphe du nombre de sorties d'incapacité vers l'invalidité par mois (année de survénance 2015)

Comme exposé plus haut et visible dans les graphes 18 et 19 ci-dessus, les assurés sont plus nombreux entre 45 et 60 ans et les sorties s'effectuent le plus souvent les premiers mois.

Nous avons ensuite choisi d'utiliser l'estimateur de Kaplan-Meier pour la construction de nos lois. En effet, la loi de maintien se calcule à partir des taux de sortie conditionnels obtenus en intégrant les sorties de chaque période à la population initiale. De plus, nous pouvons utiliser cet estimateur car les observations sont indépendantes (voir la partie III.1.1.2. Estimateur de Kaplan-Meier).

Pour tester la robustesse de cet estimateur, nous tracerons dans la partie suivante la variance de Greenwood. Le principe de cet estimateur repose sur le classement des observations par durées observées passées en incapacité en dénombrant les sorties.

Soit  $x$  l'âge à l'entrée en incapacité. Estimer la fonction de survie revient à estimer la probabilité de sortie de l'arrêt. Il existe 3 façons de sortir de l'état d'incapacité : le retour au travail, le passage en invalidité et le décès. La probabilité de sortie à un instant  $t$  (entre 1 et 36 mois) sachant que l'assuré est encore en incapacité à  $t-1$  est :

$$q_{x,t} = P(T = t \mid T \geq t - 1),$$

Avec  $T$  la durée de maintien en incapacité.

La fonction de survie s'écrit donc :

$$S_x(t) = \prod_{i < t} (1 - q_{x,i}).$$

Soient :

- $d_{x,t}$  le nombre de sorties de l'état d'incapacité à la date t
- $n_{x,t}$  la population sous risque de maintien en incapacité avant la date t.

Étant donné que le nombre de sorties d'incapacité suit une loi binomiale  $B(n_{x,t}, q_{x,t})$ , la fonction de vraisemblance que l'on doit maximiser pour estimer les  $q_{x,t}$  est :

$$\ln(L(q_{x,1}, \dots, q_{x,36})) = \sum_{i=1}^{36 \text{ (mois)}} \ln \left( C_{n_{x,i}}^{d_{x,i}} \times q_{x,i}^{d_{x,i}} \times (1 - q_{x,i})^{(n_{x,i} - d_{x,i})} \right).$$

L'on obtient un estimateur de  $q_{x,i} \forall i \in \{1, \dots, 36\}$  égal à :

$$\hat{q}_{x,i} = \frac{d_{x,i}}{n_{x,i}}.$$

L'estimateur de Kaplan-Meier de la fonction de survie s'écrit donc pour un âge d'entrée en incapacité x fixé :

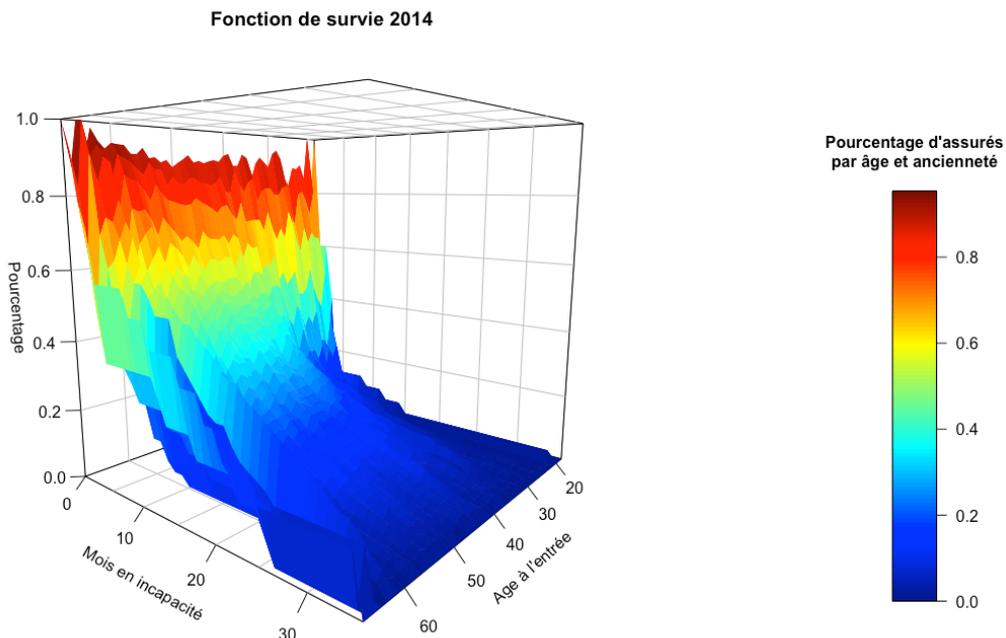
$$\hat{S}_x(t) = \prod_{i < t} (1 - \hat{q}_{x,i}) = \prod_{i < t} \left( 1 - \frac{d_{x,i}}{n_{x,i}} \right).$$

Dans notre cas, en supposant que nous ne disposons ni de censures ni de troncatures puisque chaque incapacité a été observée durant sa totalité,  $n_{x,i} = n_{x,i-1} - d_{x,i-1}$ , avec  $n_{x,0} = n$  et  $d_{x,0} = 0$ .

Ainsi nous obtenons définitivement :

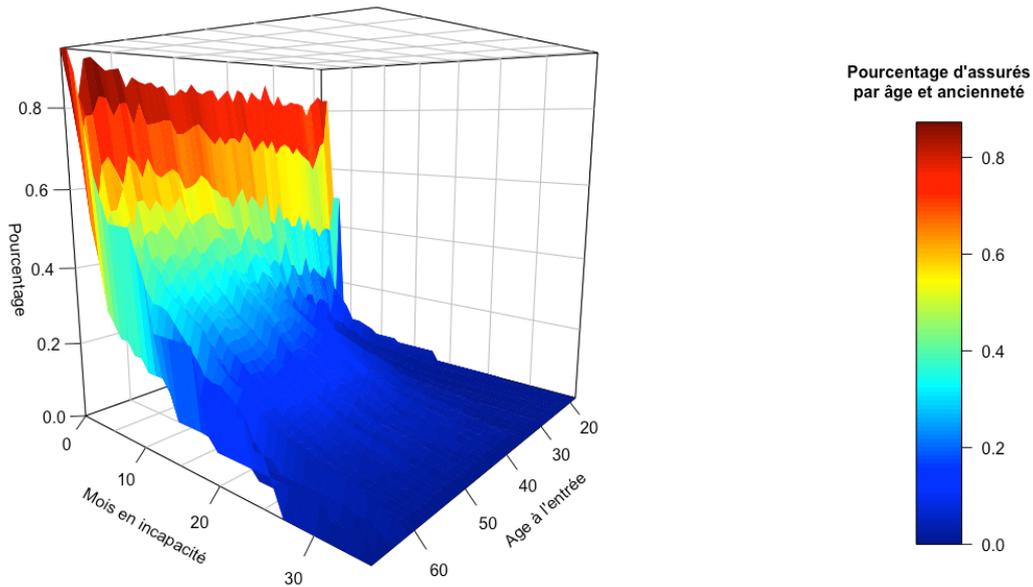
$$\hat{S}_x^{KM}(t) = \prod_{i < t} \left( 1 - \frac{d_{x,i}}{n_{x,i-1} - d_{x,i-1}} \right).$$

Les résultats sont exposés ci-dessous :



Graphique 20 : Graphe de la fonction de survie estimée par Kaplan-Meier (année de survenance 2014)

Fonction de survie 2015



Graphique 21 : Graphe de la fonction de survie estimée par Kaplan-Meier (année de survenance 2015)

Nous remarquons que les assurés entrés en incapacité en 2015 sortent en plus grande proportion au cours des premiers mois. Nous voyons en effet, que la fonction de survie 2015 tend plus vite vers 0, puisqu’au bout de 10 mois, seul 20% des assurés en incapacité de départ sont maintenus dans l’état contre environ 30% pour ceux entrés en incapacité en 2014.

Mais pour savoir si ces taux sont fiables et robustes, nous devons à présent calculer la variance de Greenwood.

**III.3.1.2.B. Variances de Greenwood et intervalles de confiance**

La variance de Greenwood<sup>xv</sup> n’est rien d’autre que la variance de l’estimateur de Kaplan-Meier.

L’estimateur de la variance de Greenwood est obtenu en utilisant l’approximation suivante :

$$\widehat{Var} \left( \log \left( \hat{S}_x^{KM}(t) \right) \right) \approx \sum_{i < t} \frac{d_{x,i}}{n_{x,i}(n_{x,i} - d_{x,i})}$$

et en appliquant la delta-méthode  $\left( Var(f(Z)) \approx [f'(E(Z))]^2 \times Var(Z) \right)$  pour montrer que :

$$\widehat{Var} \left( \log \left( \hat{S}_x^{KM}(t) \right) \right) \approx \frac{1}{\hat{S}_x^{KM}(t)^2} \times Var \left( \hat{S}_x^{KM}(t) \right).$$

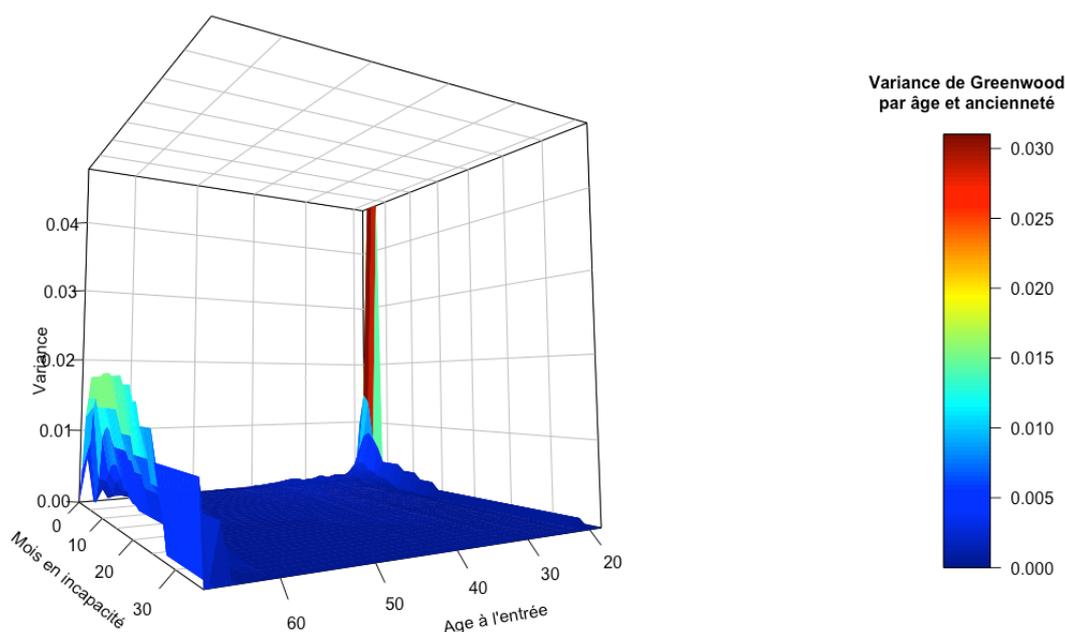
On obtient alors :

$$\widehat{Var} \left( \hat{S}_x^{KM}(t) \right) = \hat{S}_x^{KM}(t)^2 \times \sum_{i < t} \frac{d_{x,i}}{n_{x,i}(n_{x,i} - d_{x,i})}.$$

De manière théorique, ce résultat s’obtient de la propriété de normalité asymptotique de l’estimateur de Kaplan-Meier.

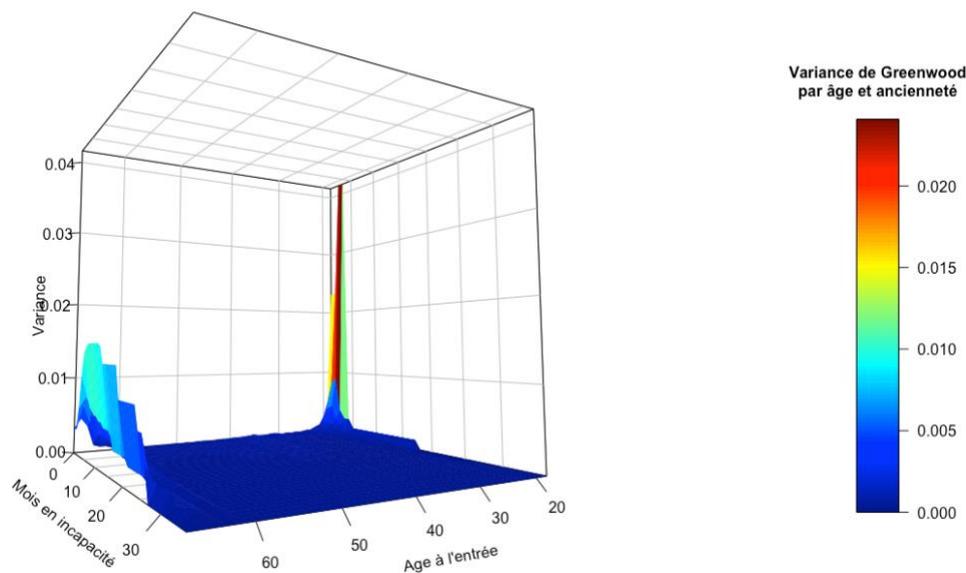
Le graphique ci-dessous représente les estimations de la variance de Greenwood obtenues pour chaque âge :

Variance de Greenwood 2014



Graphique 22 : Graphe de la Variance de Greenwood (année de survenance 2014)

Variance de Greenwood 2015

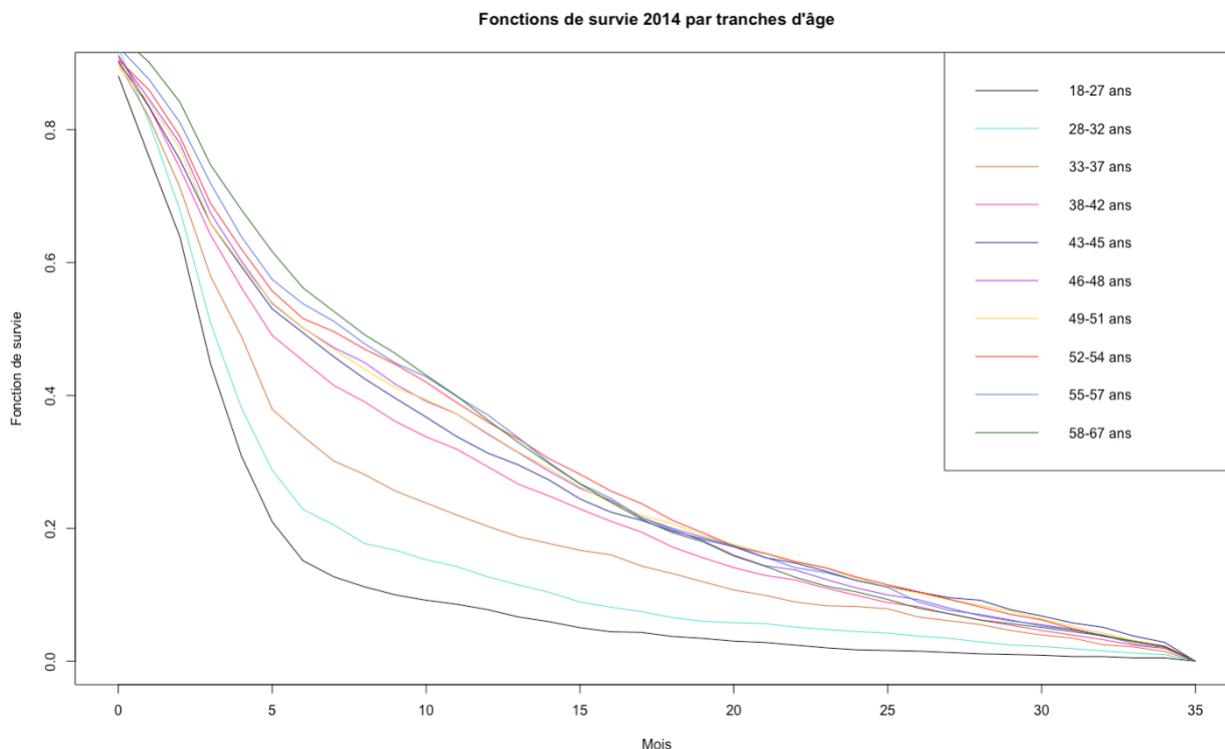


Graphique 23 : Graphe de la Variance de Greenwood (année de survenance 2015)

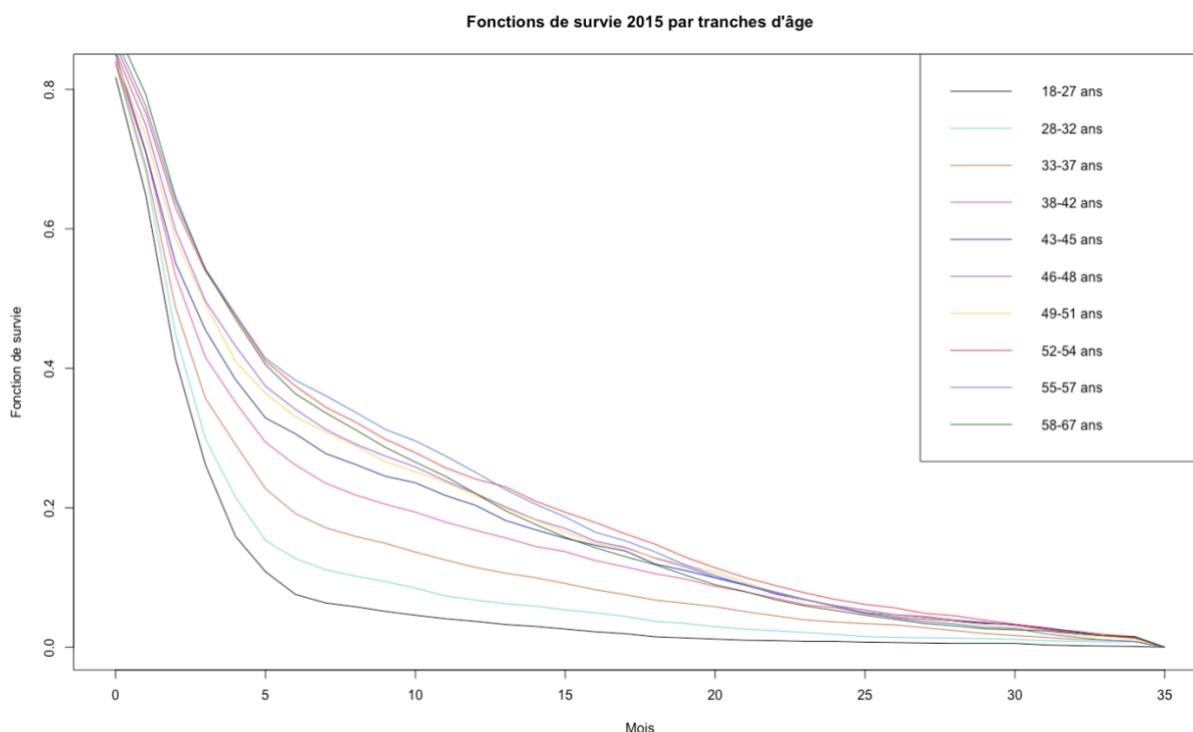
Nous remarquons dans ces deux graphes 22 et 23, en raison du manque de profondeur de nos données, de fortes valeurs pour les jeunes âges ainsi que pour les âges supérieurs à 60 ans. Les taux ne sont donc pas suffisamment fiables pour ces âges-là.

Nous décidons donc de former des tranches d'âges pour obtenir des estimateurs des fonctions de survie plus fiables. Nous prendrons les mêmes que celles créées ci-avant pour déterminer si les échantillons 2014 et 2015 suivaient la même loi.

Nous avons dès lors retracé les estimateurs de Kaplan-Meier mais par tranches d'âge (graphiques 24 et 25).



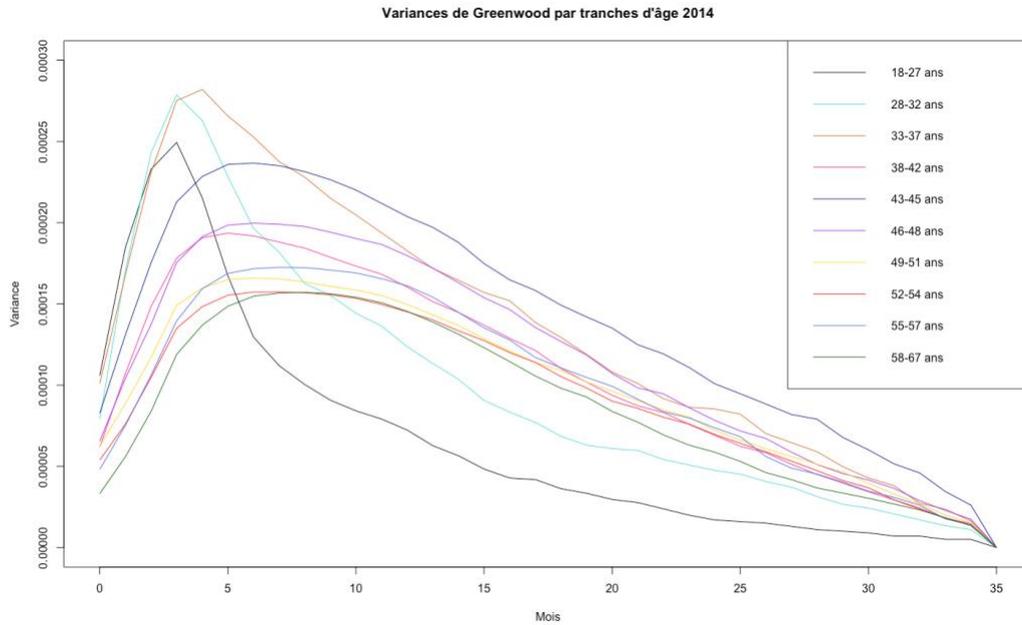
Graphique 24 : Graphe de la fonction de survie estimée par Kaplan-Meier par tranche d'âge (année de survenance 2014)



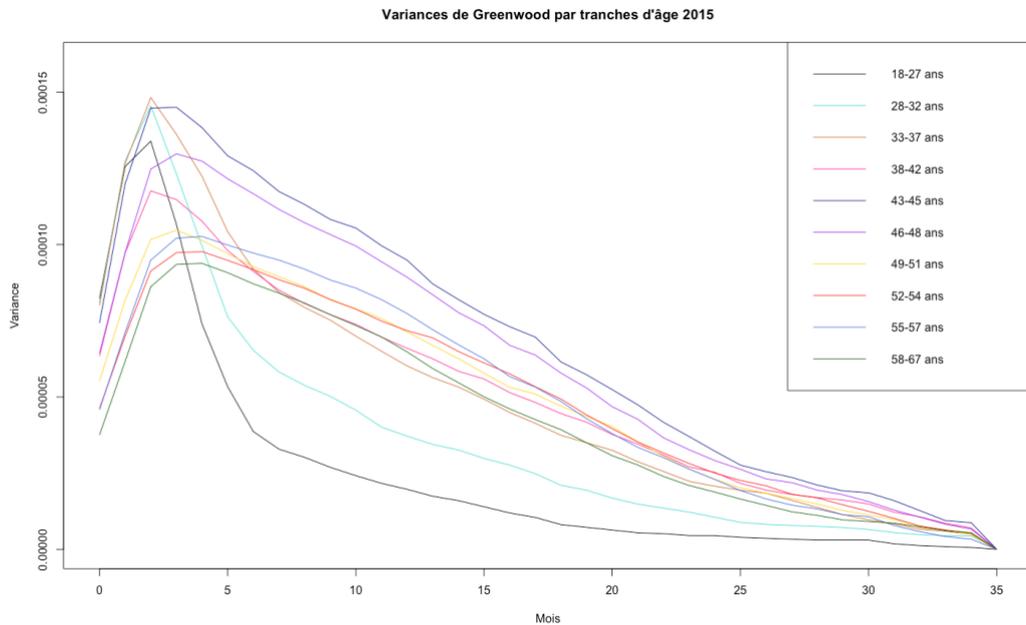
Graphique 25 : Graphe de la fonction de survie estimée par Kaplan-Meier par tranche d'âge (année de survenance 2015)

Les courbes ci-dessous nous montre clairement que les assurés plus jeunes passent moins de temps en incapacité que les assurés plus âgés. En effet seuls moins de 10% des assurés de la tranche la plus jeune sont maintenus plus de 10 mois contre plus de 30% pour la tranche la plus âgée.

Pour confirmer la meilleure fiabilité de nos taux estimés, nous retraçons les variances de Greenwood par tranche d'âges. Et bien entendu, nos estimateurs sont nettement plus fiables et robustes.



Graphique 26 : Graphe de la Variance de Greenwood par tranche d'âge (année de survenance 2014)



Graphique 27 : Graphe de la Variance de Greenwood par tranche d'âge (année de survenance 2015)

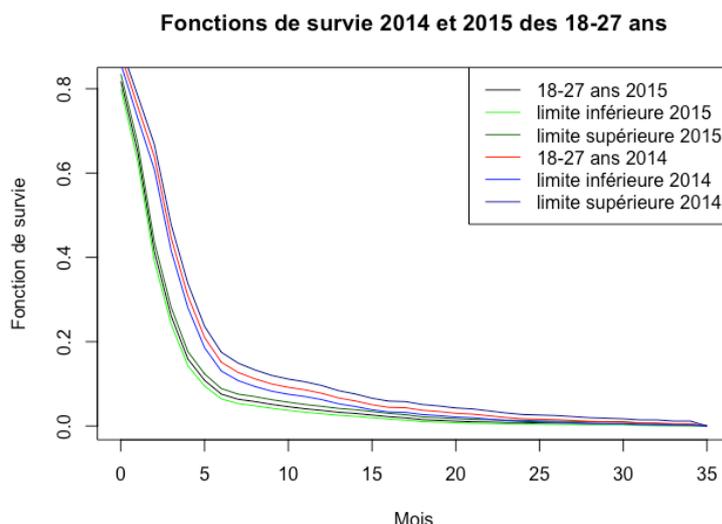
À l'aide de la variance de Greenwood, nous pouvons tracer des intervalles de confiance à 95%. Pour ce faire nous n'utiliserons pas un intervalle de confiance classique  $\left( \hat{S}_x^{KM}(t) \pm 1,96 \times \sqrt{\widehat{Var}(\hat{S}_x^{KM}(t))} \right)$  car lorsque  $\hat{S}_x^{KM}(t)$  sera proche de 0 ou de 1, les bornes pourront potentiellement dépasser ces valeurs. Or une probabilité est nécessairement contenue dans l'intervalle  $[0 ; 1]$ .

Nous utiliserons donc l'intervalle de confiance de Rothman qui contourne cet obstacle :

$$IC(\alpha) = \frac{K}{K + \left(\frac{z_\alpha}{2}\right)^2} \times \left[ \hat{S}_x^{KM}(t) + \frac{\left(\frac{z_\alpha}{2}\right)^2}{2K} \pm \frac{z_\alpha}{2} \times \sqrt{\widehat{Var}(\hat{S}_x^{KM}(t)) + \frac{\left(\frac{z_\alpha}{2}\right)^2}{4K^2}} \right]$$

Où  $K = \frac{\hat{S}_x^{KM}(t) \times (1 - \hat{S}_x^{KM}(t))}{\widehat{Var}(\hat{S}_x^{KM}(t))}$  et  $\frac{z_\alpha}{2}$  est le fractile de de rang  $100 \times \frac{\alpha}{2}$  de la distribution normale standardisée.

Nous montrerons ci-dessous les intervalles de confiance appliqués à l'estimateur de Kaplan-Meier pour la tranche d'âge des assurés de 18 à 27 ans entrés en incapacité en 2014 et 2015. Cela nous permet de voir clairement deux comportements distincts entre les deux années de survéance étudiées. Les assurés entrés en 2015 sont plus susceptibles d'avoir un arrêt plus court que ceux entrés en incapacité en 2014.



Graphique 28 : Graphe des fonctions de survie pour la tranche d'âge 18-27 ans avec intervalles de confiance (années de survéance 2014 et 2015)

En effet, comme nous pouvons le voir sur le graphique 28, la probabilité de rester en incapacité après un certain nombre de mois pour un assuré entré en 2015 (courbe noire) est toujours plus faible que celle d'un assuré entré en 2014 dans l'état. Ce qui signifie que les arrêts sont plus courts pour l'année de survéance 2015.

L'étape suivante consiste donc à lisser ces fonctions de survie pour effacer les légères irrégularités existantes particulièrement entre le 5<sup>ème</sup> et du 10<sup>ème</sup> mois.

### III.3.1.2.C. Méthode de lissage de Whittaker-Henderson

Nous réalisons le lissage des taux en utilisant la méthode de lissage de Whittaker-Henderson<sup>xvi</sup>.

Nous aurions également pu opter pour d'autres méthodes de lissages (comme le lissage par splines<sup>xvii</sup> ou encore par la méthode des noyaux<sup>xviii</sup>) afin de les comparer et choisir celle qui convient le mieux. Cependant, par choix, seule cette méthode a été utilisée, le but principal étant avant tout d'observer des impacts du changement de comportement sur le portefeuille de l'assureur. L'assureur pourra prendre davantage de temps pour construire des tables encore plus robustes si les résultats de cette étude sont concluants.

Ce lissage non paramétrique est un outil simple pour régulariser une série de taux bruts ou estimés.

En dimension 1, comme ce sera le cas ici, la méthode est fondée sur deux critères : un critère de fidélité entre la courbe de référence et la courbe lissée, et un critère de régularité. Cette méthode permet effectivement d'obtenir les valeurs qui minimisent la combinaison linéaire M de ces deux critères :

$$M = F + h * S$$

Où h est un paramètre du modèle, et :

- F est le critère de Fidélité :

Ce critère minimise la distance quadratique entre le taux de maintien lissé, noté  $i_t$ , et le taux de maintien estimé  $\hat{i}_t$ , chaque distance étant pondérée par un poids  $w_t$  :

$$F = \sum_{t=1}^{36} w_t (i_t - \hat{i}_t)^2 = (i - \hat{i})'W(i - \hat{i}),$$

Où  $W$  est la matrice diagonale des poids.

La distance représente l'écart aux valeurs estimées. Les poids sont eux, ajustables.

-  $S$  est le critère de Régularité :

Il permet d'évaluer la régularité de la courbe lissée :

$$S = \sum_{t=1}^{36-z} \left( \sum_{k=0}^z C_z^k (-1)^{z-k} i_{t+k} \right)^2 = i' K_z' K_z i,$$

Où :

$$\Delta^z i_t = \sum_{k=0}^z C_z^k (-1)^{z-k} i_{t+k},$$

$K_z$  est la matrice de taille  $(p-z, p)$  telle que  $\Delta^z i_t = K_z i$ , et  $z$  est un second paramètre du modèle entier positif.

Cette méthode implique le choix de ses deux paramètres  $z$  et  $h$  :

- $z$  est le degré du polynôme auquel va s'ajuster la courbe lissée. Ici nous le prendrons égal à 2 ce qui implique :  $\Delta^2 i_t = i_{t+2} - 2i_{t+1} + i_t$ .
- $h$  est un réel positif contrôlant le poids du critère de régularité dans le lissage. Ainsi, si  $h$  est grand, la régularité sera privilégiée (la courbe sera plus lisse), et si  $h$  est petit ce sera la fidélité qui aura l'avantage (plus proche des taux estimés).

On a donc :

$$M = F + h \times S = (i - \hat{i})' W (i - \hat{i}) + h i' K_z' K_z i = i' W i - 2i' W \hat{i} + \hat{i}' W \hat{i} + h i' K_z' K_z i.$$

Les  $i_t$  lissés sont alors trouvés en annulant la dérivée partielle de  $M$  par rapport à chacun des  $i_t$ .

On dérive  $M$  vectoriellement en égalisant à 0 :

$$\begin{aligned} \frac{\Delta M}{\Delta i} &= 2W i - 2W \hat{i} + 2h K_z' K_z i = 0 \\ \Leftrightarrow W i + h K_z' K_z i &= W \hat{i} \end{aligned}$$

On pose  $C = W + h K_z' K_z$  et on a alors :

$$C i = W \hat{i}.$$

$C$  étant définie positive, elle est donc inversible.

On obtient donc finalement l'expression suivante des taux lissés :

$$i = C^{-1} W \hat{i}.$$

Concernant le choix des poids, plusieurs valeurs sont possibles :

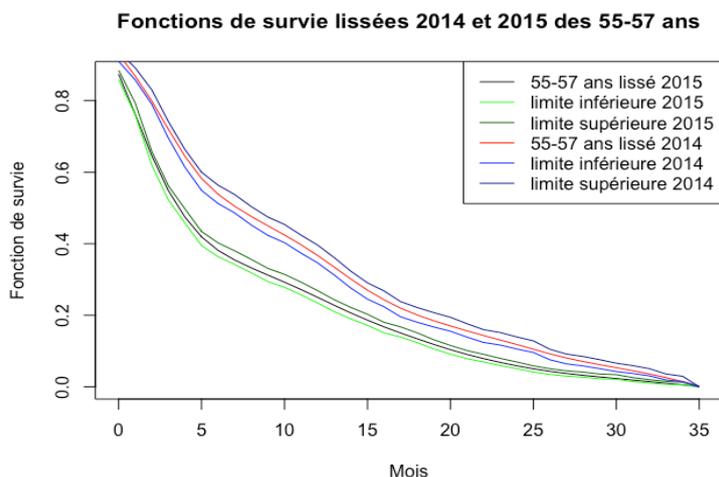
- A priori si l'on remarque des incohérences pour certaines estimations ou des absences de données. Dans ce cas, on accorde à chacun des vecteurs le même poids :  $w_t = 1$  si  $\hat{i}_t \neq 0$  et  $w_t = 0$  sinon.
- A posteriori si l'on remarque que le lissage tient trop compte de certaines valeurs estimées au regard de l'effectif présent à cet âge. Dans ce cas, on pondère par l'exposition au mois  $t$  rapporté à l'exposition moyenne pour tous les mois :  $w_t = \frac{n_t}{\bar{n}}$ , où  $\bar{n} = \frac{\sum_{t=1}^{36} n_t}{36}$ .

Nous choisissons d'utiliser la solution a posteriori car cela permet de maîtriser au mieux la fidélité des taux en limitant le poids donné aux points aberrants et en obtenant le même temps moyen de sortie avec les taux bruts et les taux lissés.

Enfin, pour le choix de  $h$ , nous le faisons varier pour trouver la valeur pour laquelle le lissage est optimal. Chaque valeur choisie dépend donc de la tranche d'âge considérée mais nous privilégierons toujours un peu plus la fidélité à la régularité.

Nous avons testé plusieurs  $h$  possibles entre 1 et 20 et il ressort que  $h=2$  est le paramètre le mieux adapté pour lisser chacune des courbes de survie estimée.

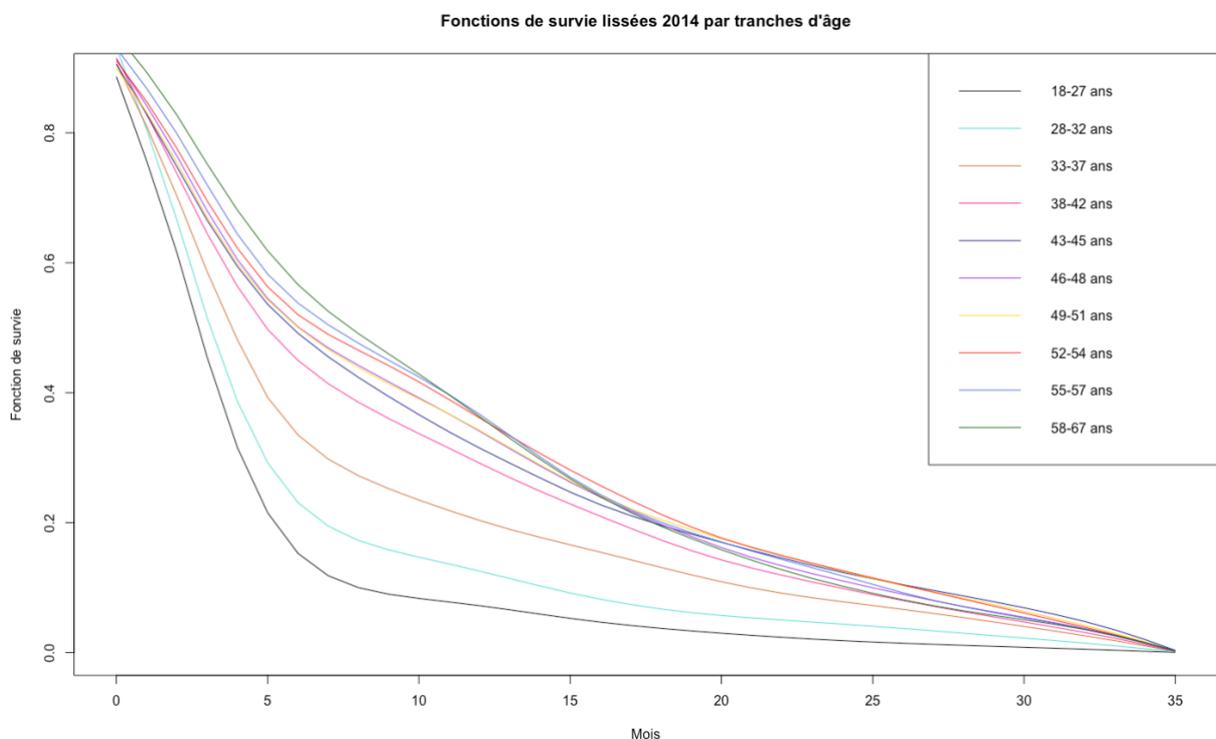
Nous présenterons ci-dessous seulement la courbe lissée des taux de survie estimés pour la tranche d'âge 55-57 ans avec intervalle de confiance pour les années de survénance 2014 (rouge) et 2015 (noire) :



Graphique 29 : Graphe des fonctions de survie lissées pour la tranche d'âge 55-57 ans avec intervalles de confiance (années de survénance 2014 et 2015)

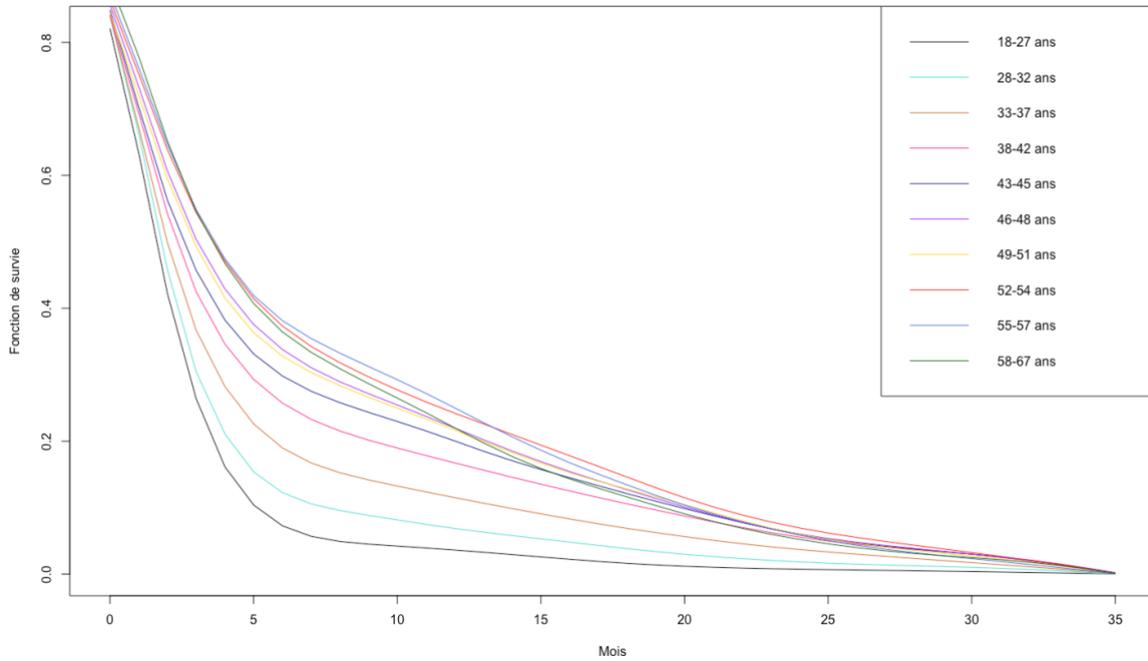
Comme nous pouvons le voir la courbe des taux lissés ne contient aucun point en dehors de l'intervalle de confiance asymptotique des taux bruts estimés. Le lissage est donc assez fidèle.

Et voici le graphique de chaque fonction de survie estimée lissée :



Graphique 30 : Graphe des fonctions de survie lissées par tranche d'âge (année de survénance 2014)

Fonctions de survie lissées 2015 par tranches d'âge



Graphique 31 : Graphe des fonctions de survie lissées par tranche d'âge (année de survenance 2015)

### III.3.1.2.D. Interpolation et extrapolation

Pour cette section, nous commençons par calculer la moyenne d'âge des classes effectuées auparavant. Puis nous considérons que les fonctions de survie estimées déjà lissées sont celles de ces âges moyens, c'est-à-dire : 24, 30, 35, 40, 44, 47, 50, 53, 56, et 60.

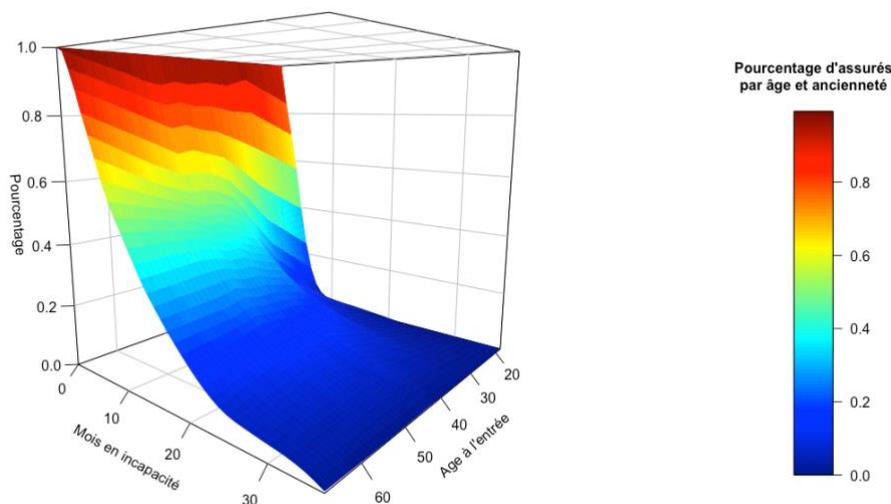
Enfin pour déterminer les fonctions de survie des âges compris entre 24 et 60 ans, nous ne nous sentons pas dans l'obligation de faire d'hypothèse sur la croissance de la fonction de survie à ancienneté fixée entre deux âges, car cette croissance est déjà suffisamment correctement prise en compte dans les classes d'âges constituées, qui ne sont distantes que de 3 ou 5 ans.

Nous pouvons donc utiliser une méthode d'interpolation linéaire.

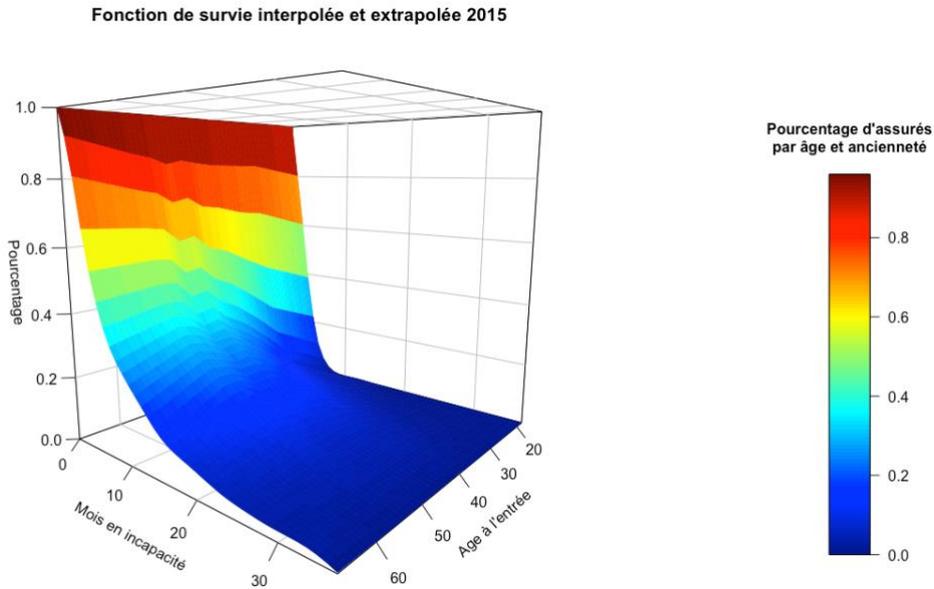
Cependant il reste tout de même à extrapoler les taux de maintien aux âges extrêmes (18-23 ans et 61-67 ans). Nous choisissons alors, comme pour l'interpolation, de procéder à une extrapolation linéaire<sup>xix</sup>.

Les résultats sont exposés ci-dessous :

Fonction de survie interpolée et extrapolée 2014



Graphique 32 : Graphe de la fonction de survie lissées à tout âge (année de survenance 2014)



Graphique 33 : Graphe de la fonction de survie lissées à tout âge (année de survenance 2015)

Nous remarquons dans les premiers mois que les pourcentages de la courbe « 2015 » tendent plus rapidement vers 0 que ceux « 2014 ». Ce qui prouve que la part de arrêts courts est plus importante pour l'année de survenance 2015. Et ainsi il apparaît alors normal d'avoir constaté une durée moyenne en incapacité plus courte à partir de cette année (211 contre 302 jours en 2014).

Il faut néanmoins maintenant vérifier la qualité de l'ajustement de nos lois d'expérience.

### III.3.2. Table de passage en invalidité

Dans cette partie nous créerons une table de passage en invalidité pour les années 2014 et 2015. Il s'agit, comme pour la table de maintien, d'estimer la probabilité de sortie de l'état d'incapacité pour chaque âge et chaque ancienneté mais en ne s'intéressant qu'aux sorties pour cause d'invalidité. On parle alors de passage en invalidité.

Il est important de préciser que nos données ont été modifiées pour le dernier mois. En effet, le faible taux de passage de ce mois alors que le passage est automatique pour les assurés étant encore en incapacité à ce moment nous laisse penser que le changement d'état de certains assurés n'a pas encore été rendu officiel (sûrement du fait de démarches administratives) et n'apparaît donc dans notre jeu de données.

**Nous avons donc pris l'hypothèse forte que tous les assurés sortis de l'état d'incapacité au cours du dernier mois l'avaient été pour invalidité.**

#### III.3.2.1. Estimateurs de Kaplan-Meier des taux bruts et variances de Greenwood

Le raisonnement que l'on suivra durant cette partie est sensiblement le même que celui utilisé dans la partie précédente.

En effet, nous nous fondons sur la population présente en incapacité pour calculer le taux de passage. Nous obtenons ces probabilités de passage conditionnellement au maintien en incapacité en divisant simplement terme à terme les effectifs de passage par ceux des assurés maintenus en incapacité.

L'effectif sous risque de passage en invalidité au mois  $t$  est le même que celui sous risque de maintien en incapacité. Cet effectif est donc  $n_{x,t} = n_{x,t-1} - d_{x,t-1}$  pour un âge  $x$  fixé et  $d_{x,t-1}$  étant le nombre de sortis pour toute cause de l'état d'incapacité au mois  $t-1$ .

L'estimateur du maximum de vraisemblance de la fonction de survie est lui donnée par :

$$\hat{S}_x(t) = \prod_{i < t} \left( 1 - \frac{d_{x,i}^{inv}}{n_{x,i}} \right),$$

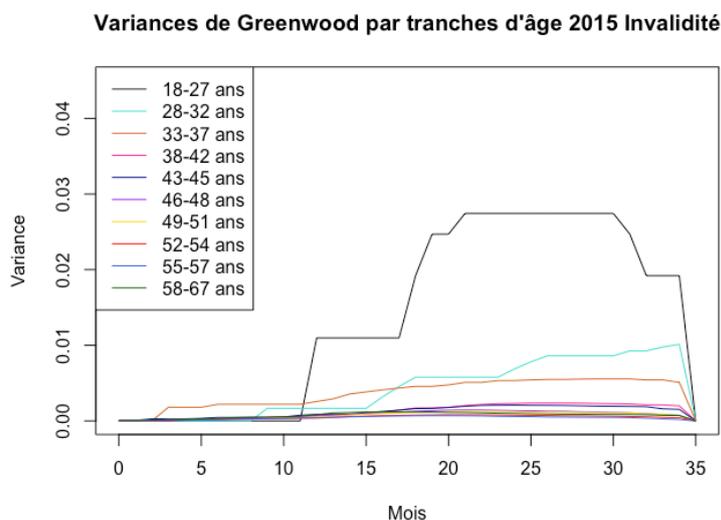
Avec  $d_{x,i}^{inv}$  le nombre de passages en invalidité au mois  $i$ .

Les données étant les mêmes que celles utilisées pour la construction de la loi de maintien en incapacité, nous devons être cohérent et calculer les estimateurs seulement par tranche d'âge.

Cependant, étant donnée la faible volumétrie de notre portefeuille invalidité, nous avons commencé par tracer la variance de Greenwood.

En effet, les invalidités ayant découlé d'une incapacité en 2014 et 2015 ne représentent que 8,78% des incapacités totales. Nous avons donc que très peu de volumétrie en invalidité.

Nous représentons ici que le graphique de l'année 2015, celui de 2014 amenant au même résultat :

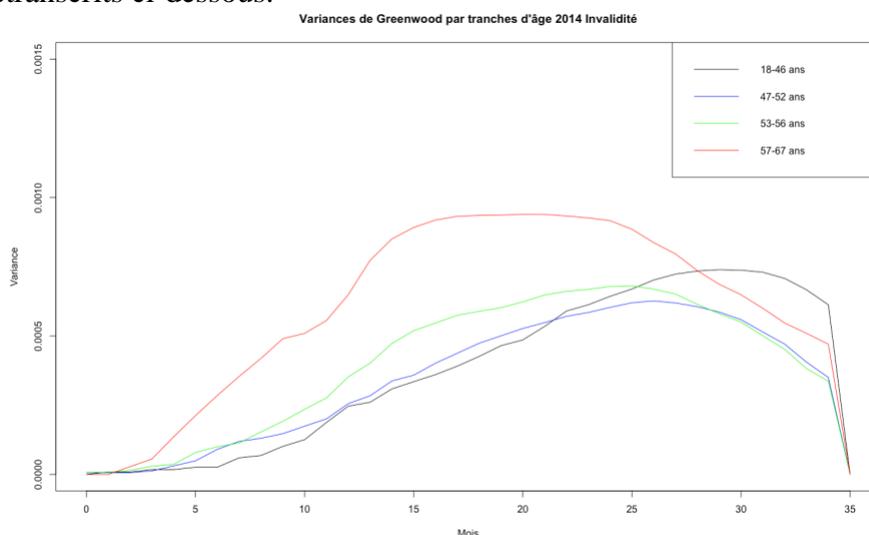


Graphique 34 : Graphe de la Variance de Greenwood (passage en invalidité) pour les mêmes tranches d'âge que pour les sorties d'incapacité (année de survenance 2015)

Celle-ci n'étant pas assez bonne, nous décidons de reformer de nouvelles tranches équivalentes en termes de volume et en gardant le maximum d'information par tranche.

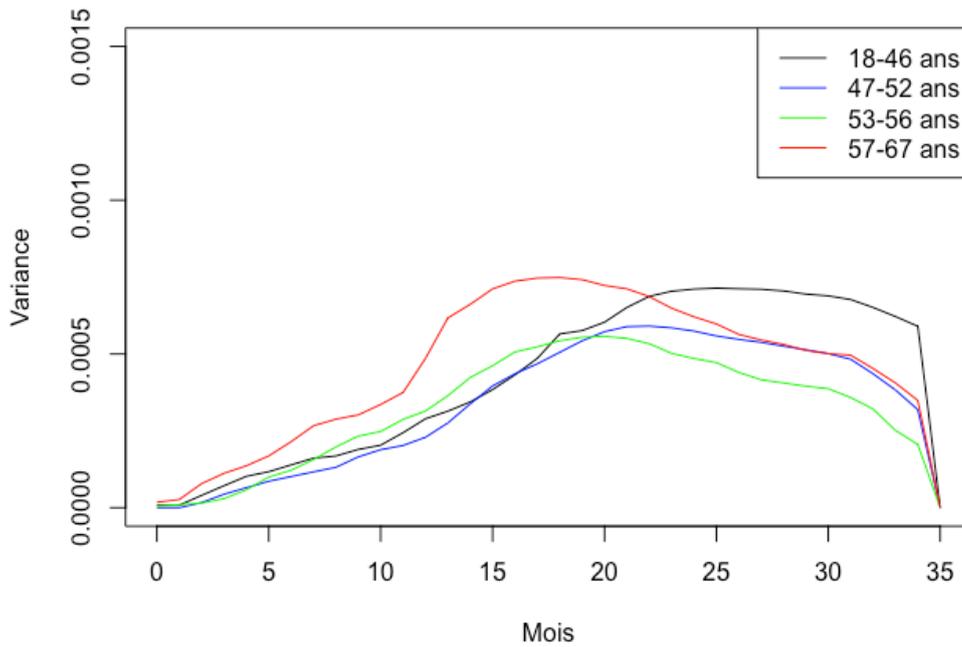
Nos nouvelles tranches sont donc les 18-46 ans, les 47-52 ans, les 53-56 ans et les 57-64 ans.

Les résultats sont retranscrits ci-dessous.



Graphique 35 : Graphe de la Variance de Greenwood (passage en invalidité) par tranche d'âge (année de survenance 2014)

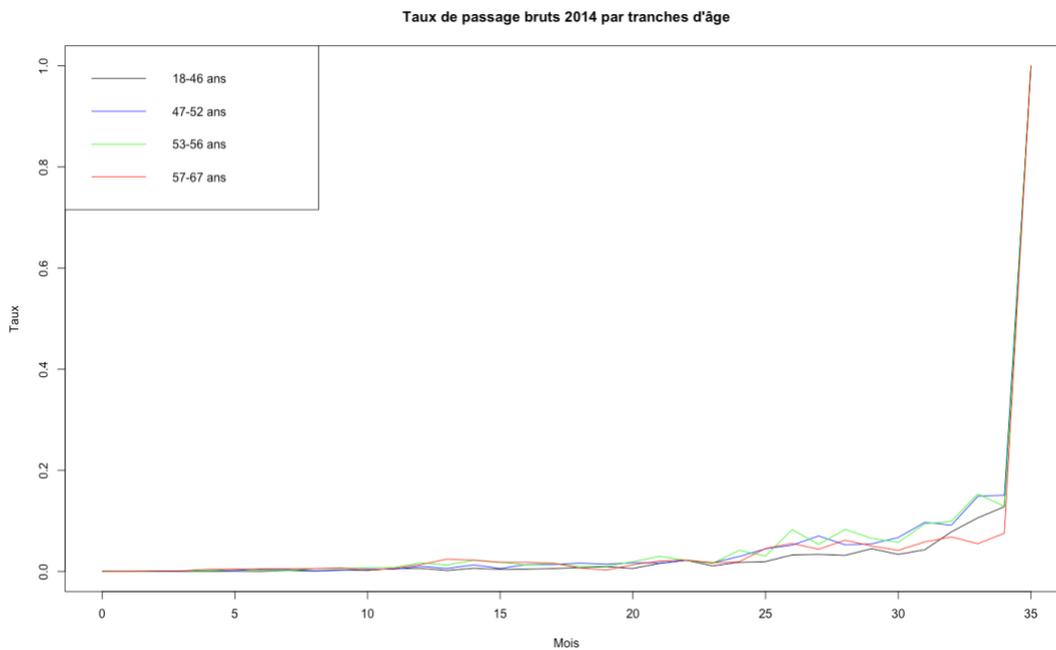
### Variances de Greenwood par tranches d'âge 2015 Invalidité



Graphique 36 : Graphe de la Variance de Greenwood (passage en invalidité) par tranche d'âge (année de survie 2015)

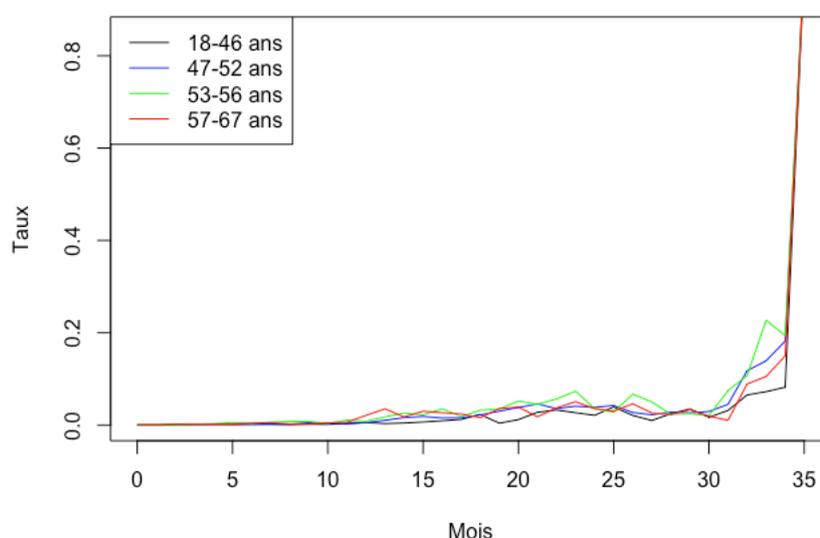
Nous obtenons dans ces graphes 35 et 36 de bien meilleurs résultats, nos taux seront donc plutôt robustes en dépit de notre faible effectif d'invalides.

Pour ces nouvelles tranches le taux de passages en invalidité estimés pour chaque tranche d'âge et chaque ancienneté sont :



Graphique 37 : Graphe des taux de passage en invalidité estimés par Kaplan-Meier par tranche d'âge (année de survie 2014)

### Taux de passage bruts 2015 par tranches d'âge



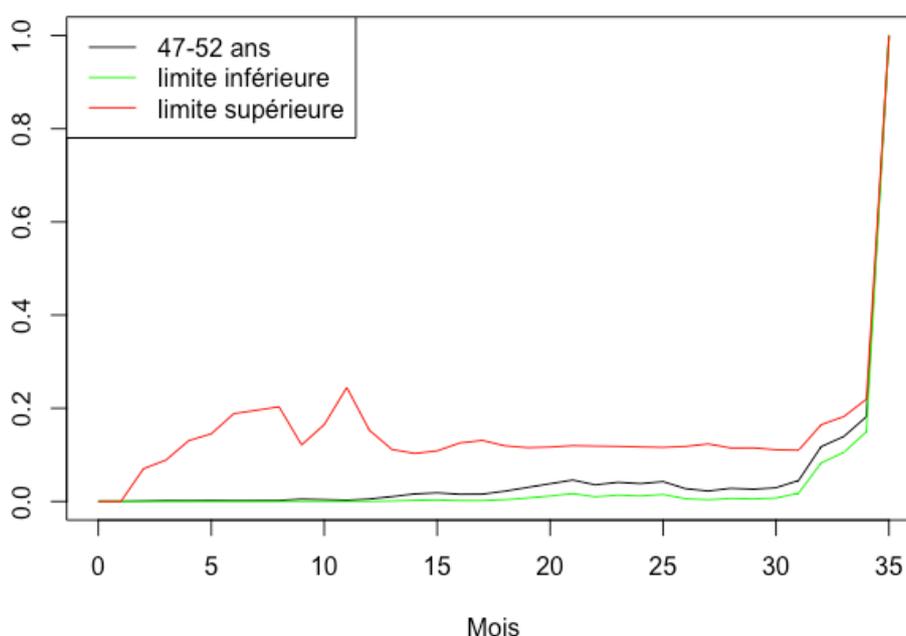
Graphique 38 : Graphe des taux de passage en invalidité estimés par Kaplan-Meier par tranche d'âge (année de survénance 2015)

On remarque que l'on constate que les taux de passage sont assez faibles si l'on fait abstraction de ceux du dernier mois, où tout assuré encore présent en incapacité passe automatiquement en invalidité. En effet, le mois où l'on constate le plus haut taux de passage est le 35<sup>ème</sup> et particulièrement pour les 53-56 ans passés en incapacité en 2015 (20% de passage ce mois-là).

Puis nous traçons les bornes de confiances de Rothman (graphique 39), calculées de la même manière que pour les tables de maintien en incapacité :

Nous les représentons seulement pour la tranche d'âge 47-52 ans :

### Taux de passage 2015 des 47-52 ans



Graphique 39 : Graphe des taux de passage en invalidité pour la tranche d'âge 47-52 ans avec intervalles de confiance (année de survénance 2015)

L'intervalle de confiance est assez large, ce qui est tout à fait logique au vu de la faible quantité de données à notre disposition.

Nous remarquons que nos taux sont nettement plus proches de la borne inférieure de notre intervalle, il sera alors intéressant dans une dernière partie de faire une étude de sensibilité en majorant nos taux de passage lors de la création de la table de passage en invalidité et d'observer ensuite les changements que cela implique dans notre projection.

### III.3.2.2. Lissage de Whittaker-Henderson

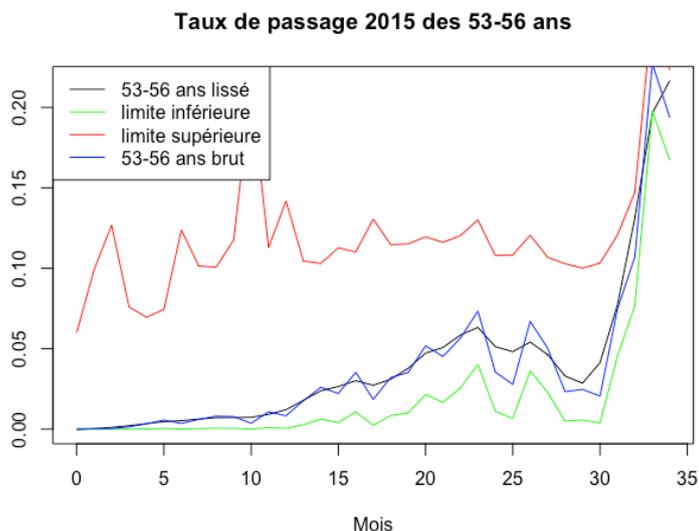
Comme pour les tables de maintien calculés dans les pages précédentes, nous utiliserons la méthode de lissage de Whittaker-Henderson pour gommer quelques-unes des irrégularités rencontrées dans les courbes des taux bruts estimés.

Nous ne liserons que les 35 premiers mois car le dernier angle est beaucoup trop droit et forcerait le modèle à surévaluer les derniers mois de passage pour adoucir l'angle. De plus, le dernier mois de passage sera de toute façon « forcé » à 100% en accord avec l'hypothèse pris au début de cette partie.

Nous appliquerons encore une fois un paramètre  $z = 2$  (pour rappel ce paramètre représente le degré du polynôme auquel va s'ajuster la courbe lissée).

Le paramètre  $h$  choisi sera lui de 5 pour les deux tranches les plus jeunes et de 3 pour les deux restantes.

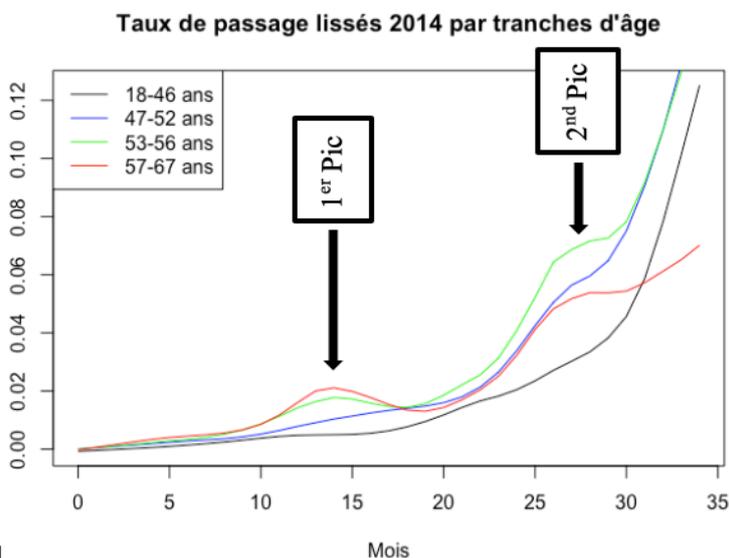
Les résultats avec intervalles de confiance sont exposés ci-dessous pour la tranche 53-56 ans des assurés entrés en incapacité en 2015 :



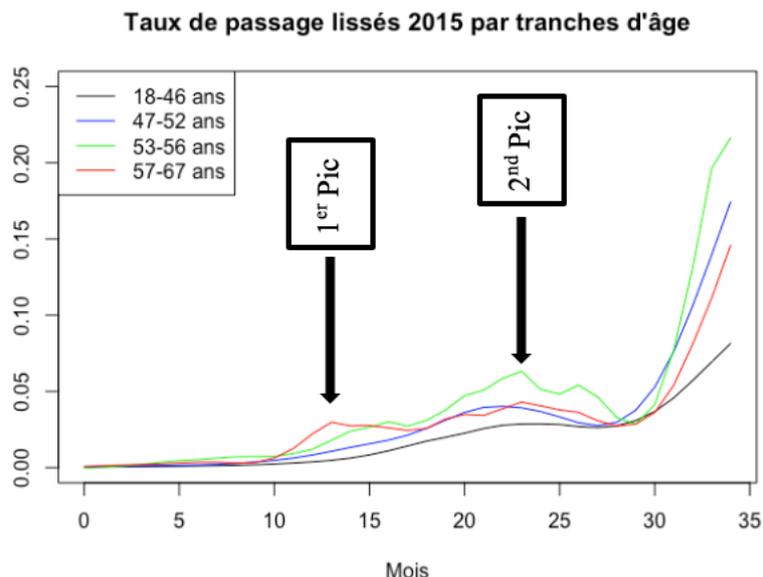
Graphique 40 : Graphe des taux de passages en invalidité lissés pour la tranche d'âge 53-56 ans avec intervalles de confiance (année de survénance 2015)

Comme nous pouvons le voir dans le graphique 40, le lissage adoucit la courbe brute. La courbe lissée obtenue reste dans l'intervalle de confiance en tout point, nous pouvons donc penser que le lissage est satisfaisant.

Les lissages sont effectués pour chaque tranche d'âge et sont retranscrits ci-dessous pour les années de survénance 2014 et 2015 entre les mois 1 et 35 :



Graphique 41 : Graphe des taux de passages en invalidité lissés par tranche d'âge (année de survénance 2014)



Graphique 42 : Graphe des taux de passages en invalidité lissés par tranche d'âge (année de survenance 2015)

Comme nous l'observons les assurés entrés en incapacité en 2015 connaissent un premier pic de passage en invalidité au 13<sup>ème</sup> mois et un second entre les 22<sup>ème</sup> et 23<sup>ème</sup> mois.

Ces pics interviennent plus tard pour les assurés déclarés en état d'incapacité en 2014 puisque nous ne les constatons qu'au 14<sup>ème</sup> mois et qu'entre le 26<sup>ème</sup> et le 28<sup>ème</sup> mois.

Des gains de temps considérables qui exposent une nouvelle fois les effets de l'accélération du passage en invalidité.

De plus, les taux observés dans le premier graphe sont moins importants durant les 25 premiers mois, ce qui signifie que la part d'assurés en incapacité passant en invalidité est moins importante durant cette période, mais pas nécessairement que le nombre d'invalidités est moins élevé.

En effet, les tables de maintien créées dans la section précédente montrent également que le taux de sortie est plus élevé à partir des données « 2015 » durant cette période. Il y a donc moins d'assurés déclarés en état d'incapacité en 2015 sous risque de passage en invalidité durant les 25 premiers mois.

La logique, au vu des taux constatés (de maintien et de passage), serait même une diminution du nombre de passages en invalidité à la suite d'une incapacité en 2015 par rapport au nombre de passages en suite d'une incapacité en 2014.

### III.3.2.3. Interpolation et extrapolation

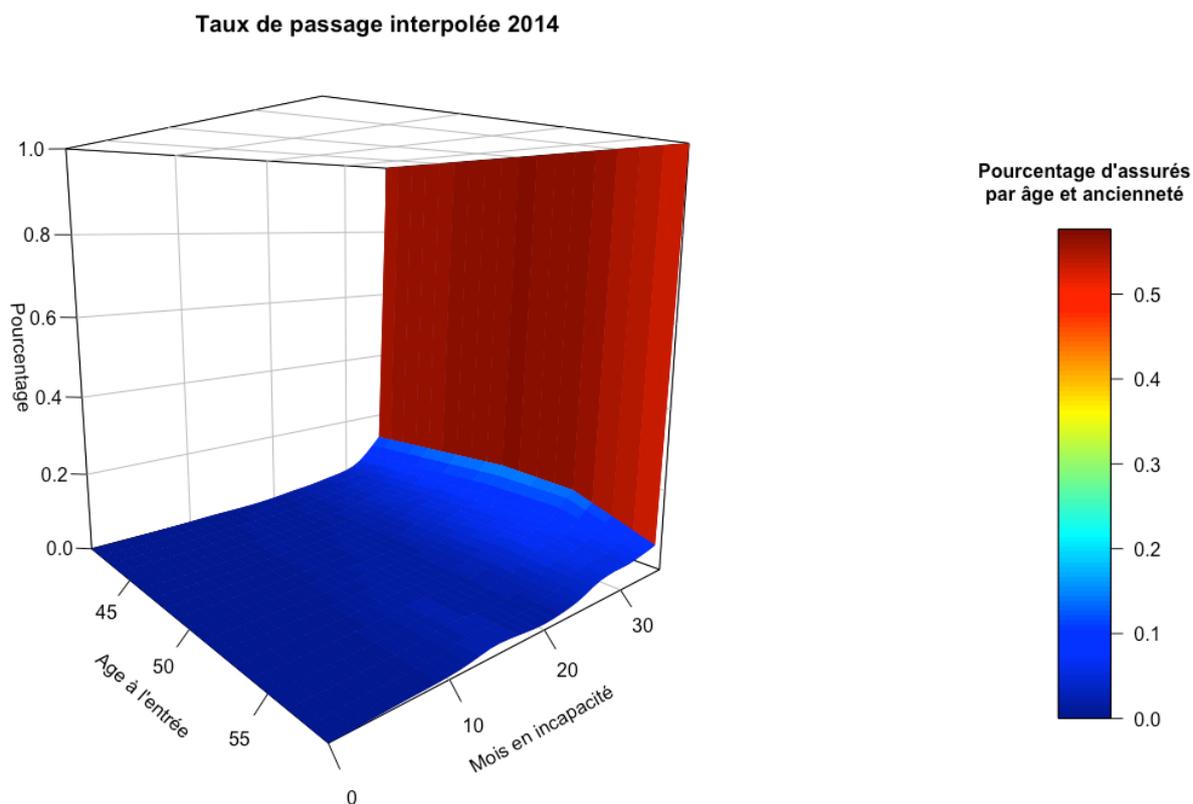
Étant donné le faible nombre de passage en invalidité avant 40 ans nous avons considéré qu'il serait d'abord plus intéressant d'exposer les résultats seulement pour les âges avec le plus de volumétrie et donc les résultats les plus robustes et fidèles à nos données initiales. Nous ne procéderons donc pas, en premier lieu, à d'extrapolations mais seulement aux interpolations.

Dans un second temps, nous extrapolerons les taux de passage pour les âges les plus extrêmes en utilisant les taux de passage déjà calculés dans les tables du BCAC.

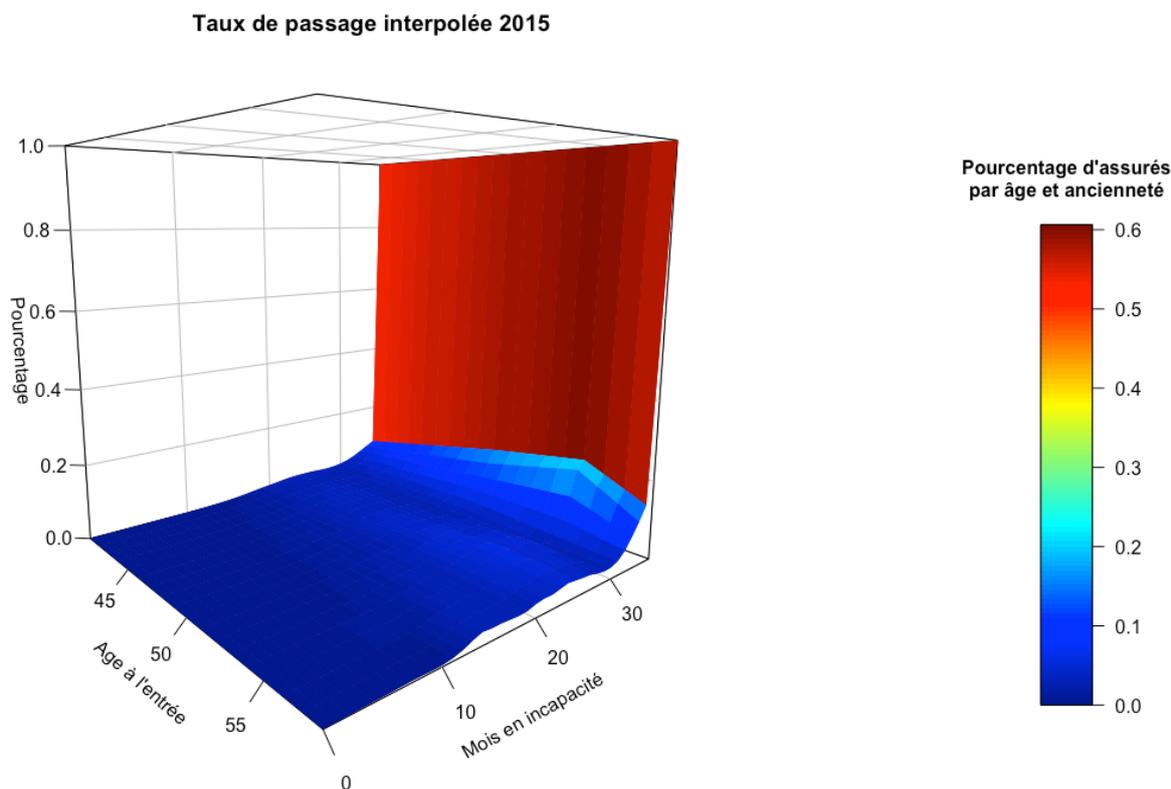
#### III.3.2.3.A. Sans extrapolation

Cette sous-section utilise la même méthode d'interpolation que celle utilisée dans le chapitre précédent.

Les résultats obtenus (nous avons en plus forcé le passage automatique le dernier mois) sont exposés ci-après pour les âges compris entre 41 et 58 ans, âges moyens de la tranche d'âges la plus jeune et de la plus vieille :



Graphique 43 : Graphe des taux de passages en invalidité lissés à tout âge sans extrapolation (année de survenance 2014)



Graphique 44 : Graphe des taux de passages en invalidité lissés à tout âge sans extrapolation (année de survenance 2015)

Ce type d'interpolation (entre 41 ans et 58 ans) reste efficace car les tranches d'âge sont assez petites entre ces âges-là.

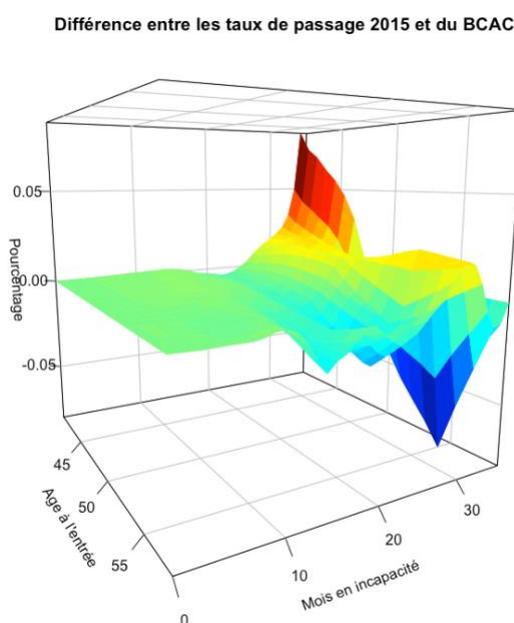
Mais une extrapolation de type linéaire (comme utilisée pour la table de maintien) peut sembler obsolète par l'hypothèse sur la croissance des taux entre 18 ans et 41 ans ainsi qu'entre 58 et 64 ans qui s'impose à nous pour réaliser l'extrapolation. Ces périodes étant bien trop étendues, nous devons trouver une autre méthode pour arriver à nos fins.

### III.3.2.3.B. Extrapolation en utilisant des coefficients de proportionnalité

Pour pallier ce manque d'information pour les jeunes âges, nous décidons de comparer les taux des âges de 41 à 58 ans à ceux du BCAC. Nous pourrions alors déduire les coefficients de proportionnalité permettant de modifier les taux de la table du BCAC<sup>xx</sup> pour obtenir à chaque âge de la tranche extrême, les taux de la table estimée et lissée.

Nous décidons de travailler sur les probabilités de sorties et nous commençons par étudier la différence entre les taux de sortie de notre table d'expérience et ceux de la table BCAC entre 41 et 58 ans et à toute ancienneté.

Nous montrons les résultats pour l'année 2015 ici :



Graphique 45 : Différence des taux de passages en invalidité calculés (année de survivance 2015) et du BCAC

Nous pouvons observer que si à âge fixé les différences varient très fortement selon les mois, à ancienneté fixée les différences selon les âges sont plutôt de la même grandeur entre 41 et 49 ans et entre 50 et 58 ans.

Nous décidons donc de calculer des coefficients pour chaque ancienneté.

Nous travaillons en considérant une relation linéaire entre les taux de passage d'expérience calculés en amont et les taux de la table du BCAC en séparant les âges inférieurs ou égaux à 49 et ceux supérieurs à cet âge. Le but est donc de déterminer les uniques coefficients  $\alpha_t$  et  $\beta_t$ , pour tout  $t$  appartenant à  $\{1, \dots, 36\}$ , permettant de minimiser le critère de moindre carré suivant :

$$\sum_{x=41}^{49} (\alpha_t \times q_{t,x}^{BCAC} + \beta_t - q_{t,x})^2,$$

pour pouvoir extrapoler les taux aux âges inférieurs à 41 ans ;

$$\text{et } \sum_{x=50}^{58} (\alpha_t \times q_{t,x}^{BCAC} + \beta_t - q_{t,x})^2,$$

pour pouvoir extrapoler les taux aux âges supérieurs à 58 ans.

En résolvant ce problème, nous obtenons des formules explicites pour les coefficients recherchés (ici pour extrapoler les jeunes âges) :

$$\Rightarrow \alpha_t = \frac{(49 - 41 + 1) \times \sum_{x=41}^{49} q_{t,x}^{BCAC} \times q_{t,x} - \sum_{x=41}^{49} q_{t,x}^{BCAC} \times \sum_{x=41}^{49} q_{t,x}}{(49 - 41 + 1) \times \sum_{x=41}^{49} (q_{t,x}^{BCAC})^2 - (\sum_{x=41}^{49} q_{t,x}^{BCAC})^2}$$

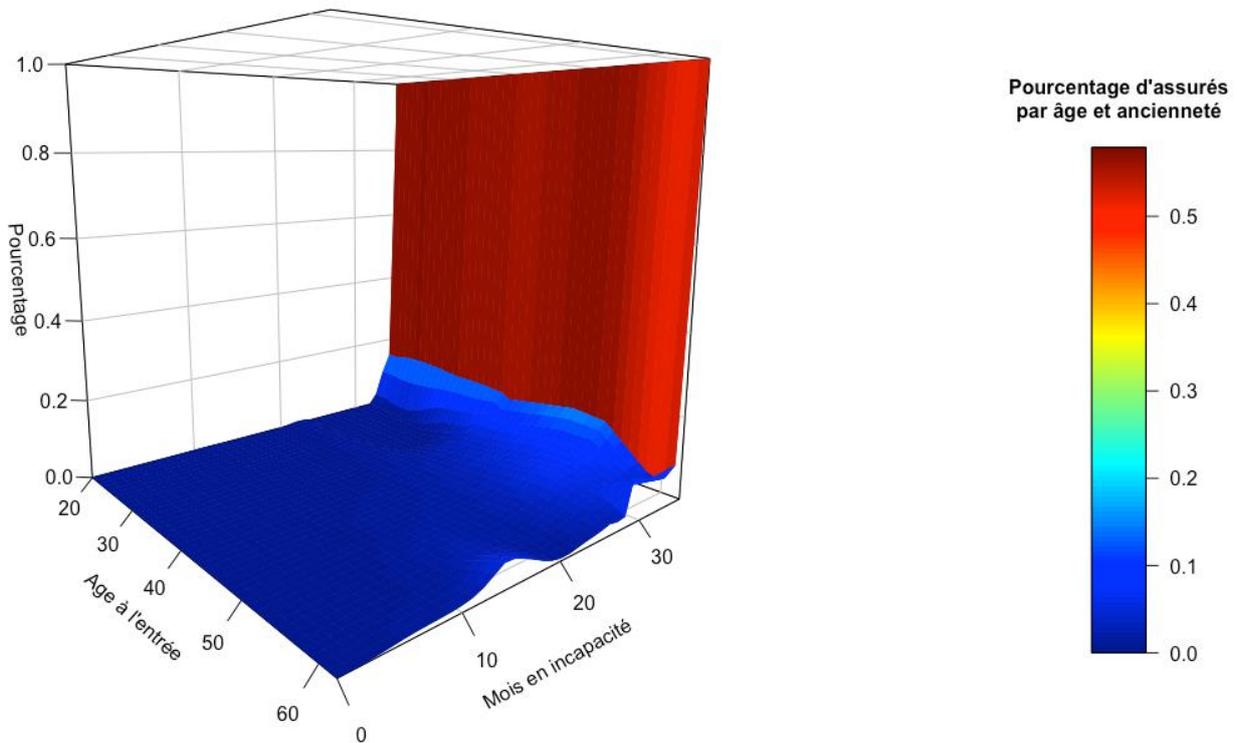
$$\text{et } \beta_t = \bar{q}_t - \alpha_t \times \overline{q_{t,x}^{BCAC}}.$$

Nous reconstituons alors la table d'expérience des probabilités  $q_{t,x}^*$  de passage en invalidité pour un mois  $t$  et un âge  $x$  inférieur à 41 par la formule suivante :

$$q_{t,x}^* = \alpha_t \times q_{t,x}^{BCAC} + \beta_t.$$

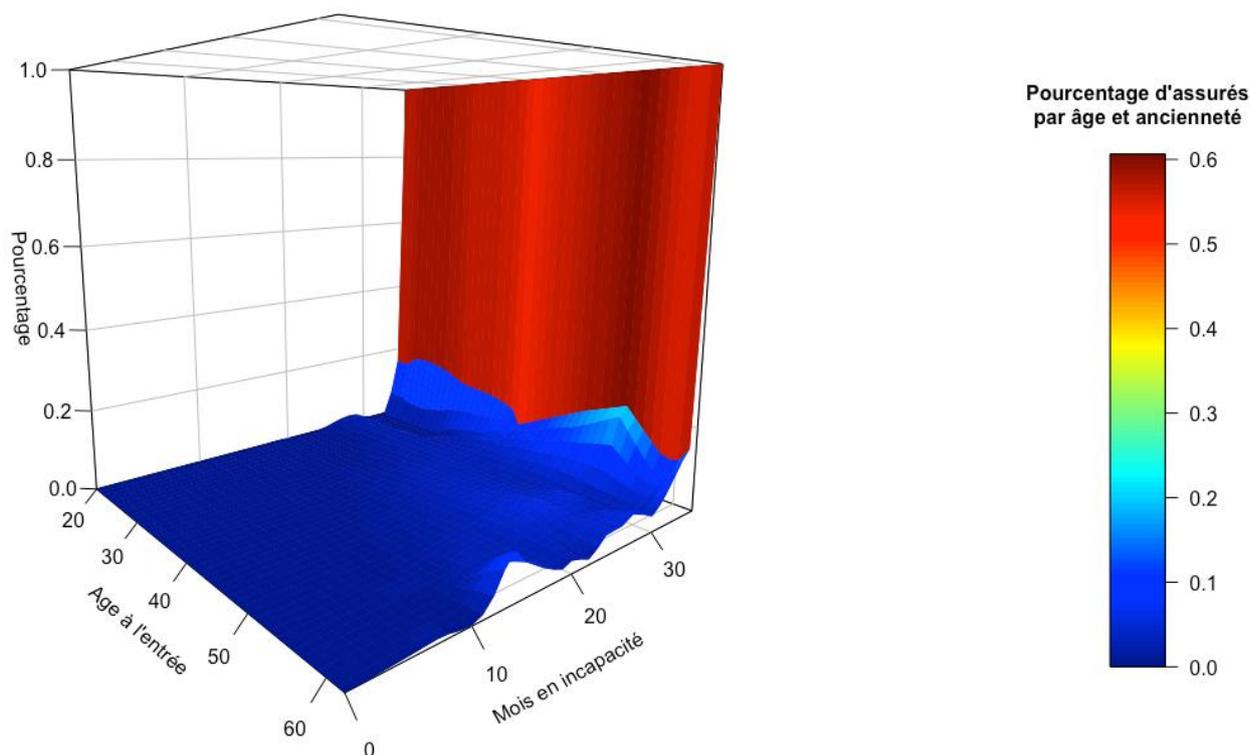
Nous affichons ici les résultats obtenus après extrapolation pour les tranches d'âges extrêmes :

**Table de passage 2014**



Graphique 46 : Graphe des taux de passages en invalidité lissés à tout âge (année de survenance 2014)

## Table de passage 2015



Graphique 47 : Graphe des taux de passages en invalidité lissés à tout âge (année de survenance 2015)

Ces tables maintenant créées nous serviront pour la suite à comparer les impacts que provoquerait l'accélération des délais de passage en invalidité si toutefois ces derniers étaient confirmés par la suite lorsque nous disposerons de plus de recul sur nos données.

Ces impacts se ressentiraient sur les portefeuilles d'assurés où la répartition des individus présents dans chaque état (valides, en incapacité, invalides) serait évidemment différente, mais aussi, et tout aussi logiquement, dans les charges de l'assureur et les cotisations des assurés.

Il serait également intéressant de se questionner sur les raisons de cette accélération du passage en invalidité.

La projection, qui sera faite dans la partie suivante pour évaluer les impacts évoqués ci-dessus, nous permettra donc aussi de conforter, sans le prouver pour autant, deux hypothèses que nous avons trouvées pour expliquer l'accélération et le nombre croissant d'arrêts de travail (2015 ayant connu plus d'arrêts que 2014 et même plus d'invalidités que 2014) :

- L'arrivée de la génération des baby-boomers à l'âge de la retraite ;
- Le fait qu'un assuré auquel on paie des indemnités journalières revient moins cher à l'assureur qu'un assuré présent dans l'état d'invalidité.

# ***PARTIE IV : PROJECTION D'UN PORTEFEUILLE D'ASSURES A L'AIDE DES TABLES D'EXPERIENCE***

## ***Chapitre 1. Le portefeuille utilisé***

### **IV.1.1. Le portefeuille, ses caractéristiques**

Le portefeuille utilisé est constitué principalement d'assistants maternels. Nous n'avons pas pu utiliser le même portefeuille car nous ne disposions que des données de sinistres. En effet, nous n'avons pas le portefeuille d'assurés de la BAD, et nous avons choisi celui-ci car il était « le plus proche » (même catégorie sociale, celle des aides à domicile) pour lequel nous disposions d'un tel jeu de données.

Le fait de travailler avec une population différente peut introduire des biais – notamment et principalement sur la répartition hommes/femmes -, le comportement en arrêt de travail n'étant pas forcément le même entre deux populations différentes.

Ce portefeuille est ainsi à grande majorité féminine (99%) quand le portefeuille précédent, lui, était constitué d'une population plus homogène, bien qu'à forte majorité féminine également. Dans le cas qui nous occupe, les tables construites dans la partie précédente sont indépendantes de cette variable de sexe. Mais les biais liés à la variables sexe se retrouvent dans d'autres variables comme la CSP, ou le salaire moyen.

Il sera donc important de prendre les résultats qui suivront avec toutes les précautions nécessaires liées à ce changement de population.

Pour le risque d'incapacité, le taux de couverture est de 76% du salaire brut de référence limité au net et il existe une franchise de 10 jours. Le contrat est valide jusqu'à la retraite ou au plus tard à 60 ans.

Concernant le risque d'invalidité, le taux de couverture est de 90% seulement pour les invalides de deuxième et troisième catégories (selon la Sécurité Sociale). L'individu est couvert jusqu'à la retraite ou jusqu'à une limite de 60 ans.

Pour pouvoir projeter dans un second temps les provisions et prestations de ce portefeuille, il nous est nécessaire de calculer les salaires moyens mensuels et annuels, ainsi que le montant des cotisations par tranches d'âge des assurés.

Pour cela, nous disposons des informations suivantes sur le portefeuille étudié :

- Le nombre de femmes et d'hommes pour chaque tranche d'âges. Ces tranches sont distantes de 10 ans : 20-29 ans ; 30-39 ans ; 40-49 ans ; et 50-59 ans ;
- La masse salariale que représente chaque tranche d'âges ;
- Les taux de cotisations par risques (incapacité et invalidité).

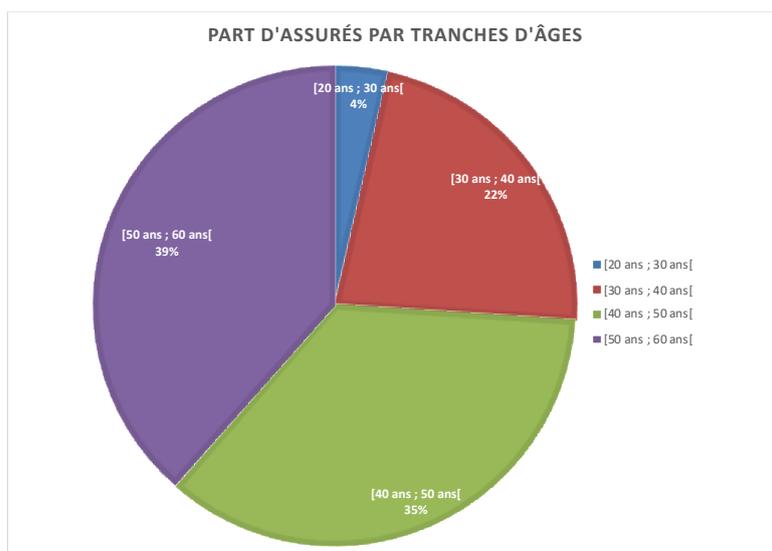
Ainsi pour une tranche d'âges  $i$  :

$$Salaire_i^{annuel} = \frac{Masse\_salariale_i}{Effectif_i} = 12 \times Salaire_i^{mensuel},$$

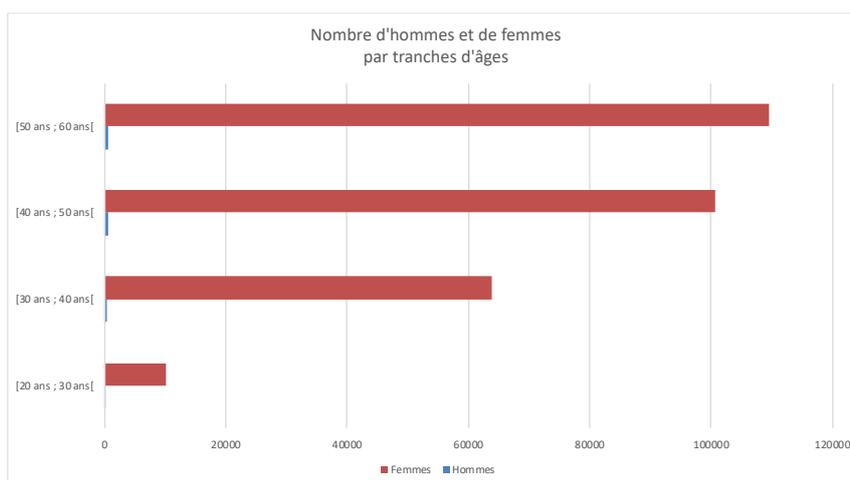
$$Cotisation_i = Salaire_i^{annuel} \times (tx^{incap} + tx^{inval}).$$

## IV.1.2. Quelques statistiques descriptives

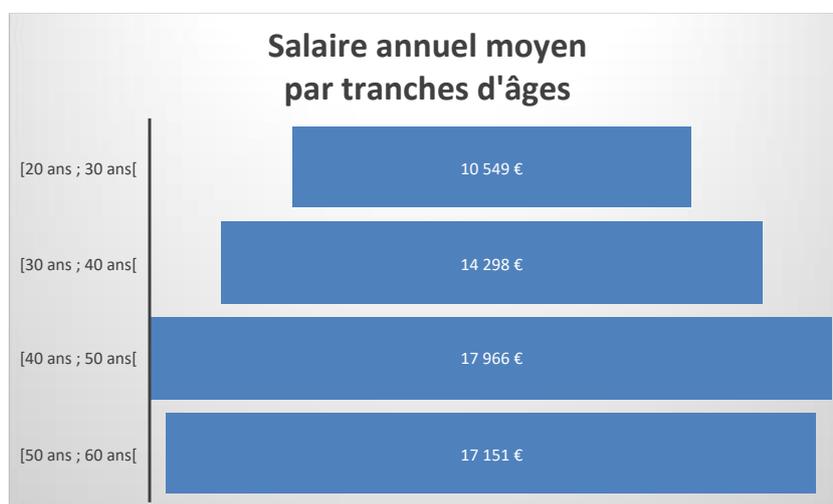
Nous commençons par donner la répartition des assurés par tranches d'âges :



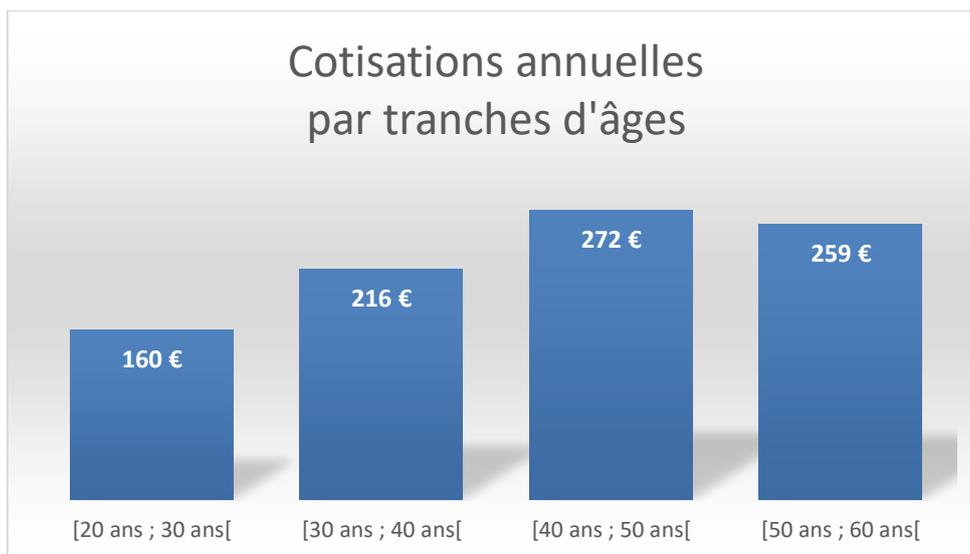
Les tranches les plus âgées sont également les plus nombreuses : Près de trois quarts des assistants maternels ayant souscrit ce contrat ont plus de 40 ans et seulement 4% ont moins de 30 ans.



Nous pouvons voir ici l'écrasante majorité de femmes présentes dans ce portefeuille puisque 99,50% des individus sont des femmes.



Nous remarquons que le salaire moyen est croissant avec l'âge jusqu'à 50 ans mais que les individus âgés de 50 à 59 ans sont moins payés que ceux âgés de 40 à 49 ans.



Logiquement, le constat est le même concernant les cotisations moyennes payées par les assurés puisque les cotisations dépendent de l'assiette des cotisations qui est une partie de la rémunération. Plus l'on prend de l'âge, plus les cotisations sont élevées à part pour les assurés les plus vieux qui paient moins que ceux de la tranche d'âges qui les précèdent.

### IV.1.3. Les lois utilisées

Nous utiliserons, pour ces projections, une loi d'incidence d'entrée en incapacité interne à l'assureur à laquelle nous avons appliqué un coefficient multiplicatif calculé en ajustant les prestations réelles de ce portefeuille à celle obtenues dans la projection.

En effet, la loi d'incidence ayant été calculée à la base pour un portefeuille emprunteur, elle ne correspondait pas forcément à la population assez spécifique du portefeuille utilisé dans cette partie.

Nous précisons que le coefficient multiplicatif calculé sera le même pour les deux projections (à partir de la table « 2014 » et de la table « 2015 »).

Les lois de mortalité utilisées seront celles de l'INSEE calculées en 2018 (mises en place chaque année à partir de l'enregistrement des décès à l'état civil). Nous calculons à partir de celles-ci un unique taux pour les deux sexes en sachant que notre portefeuille est constitué à 99,5% de femmes :

$$q_x^{general} = 99,5\% \times q_x^{femme} + 0,5\% \times q_x^{homme}.$$

Nous calculerons des taux d'entrée dans le portefeuille pour générer de nouveaux adhérents pour chaque âge entre 20 et 60 ans. Nous utiliserons pour cela les données actuelles du portefeuille et la pyramide des âges 2019 de l'INSEE.

Les tables de maintien en incapacité et de passage en invalidité seront, elles, celles que nous avons modélisées dans la partie précédente.

Enfin, la table de maintien en invalidité choisie qui servira à calculer nos provisions mathématiques sera celle du BCAC.

Il est à noter que toutes ces hypothèses (renouvellement des nouveaux adhérents) et lois (d'incidence, de mortalité, tables de maintien et de passage) pourraient être rendues mises en concurrence en en appliquant d'autres et en analysant les impacts de ces choix (test de sensibilité des hypothèses). Par exemple, nous pourrions comparer les résultats obtenus avec les lois de mortalité de l'INSEE et la table TF 00-02 établie suite à l'observation faite sur les femmes entre 2000 et 2002 ou encore avec la table générationnelle TGF05

qui fournissent le nombre de survivants âge par âge pour les femmes et pour toutes les générations entre 1900 et 2005. Nous pourrions ainsi observer sur quels aspects de nos résultats cette hypothèse de mortalité influe et juger de la table la plus en adéquation avec le sujet de l'étude et le niveau de prudence de l'assureur.

Cependant, à l'heure à laquelle ces lignes sont écrites, nous ne disposons plus des données de portefeuille nécessaires à la réalisation de ces tests.

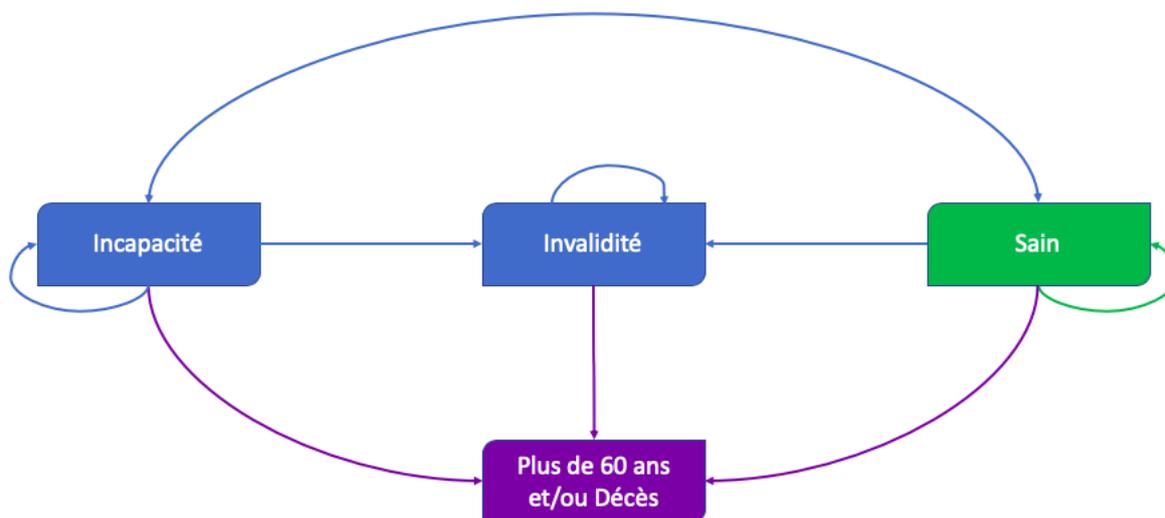
## Chapitre 2. Projections des populations à partir des tables d'expérience 2014 et 2015

### IV.2.1. Projection de la population active et en état d'incapacité

Il est nécessaire de projeter la population active et celle présente en incapacité car les deux dépendent l'une de l'autre.

En effet, la population active de notre portefeuille à l'année  $n$  est égale à la population qui n'est ni en incapacité, ni en invalidité, ni décédée.

Or, la population d'invalides dépend de celle présente dans l'état d'incapacité et ne peut revenir à l'état actif. Donc la population active ne dépend en réalité que de celle présente dans l'état d'incapacité (la seule population qui peut revenir à l'état actif) et des taux de mortalité.



La difficulté réside dans le fait que la population présente en incapacité n'a pas la même probabilité de passer dans l'état actif selon le temps qu'elle est restée maintenue dans cet état.

Dans un monde idéal, nous projetterions nos assurés chaque mois pour suivre leur évolution et celles de nos charges et cotisations de plus près.

Cependant, par souci de clarté et de lisibilité, nous simplifierons le modèle et suivrons donc l'évolution de nos variables année par année seulement.

Pour cela, nous décomposerons donc les assurés en incapacité en trois : ceux étant dans leur première année de maintien, ceux étant dans leur deuxième année et ceux étant dans leur troisième année.

Nous utiliserons dans les formules les notations suivantes :

- $i_x$  = le taux d'incidence à l'âge  $x$  (pour passer en incapacité)
- $r_x^{t\ an}$  = le taux de retour à la population active au bout de  $t$  années,  $t \in \{1,2,3\}$ , pour un assuré entré en incapacité à l'âge  $x$

- $e_x$  = le taux de nouveaux entrants (taux de chute) d'âge  $x$ . Il peut être négatif car il peut y avoir plus de résiliations de contrats que de nouveaux adhérents
- $m_x^{t\ an}$  = le taux de maintien en incapacité pour un âge  $x$  et une ancienneté de  $t \times 12$  mois
- $pc_x^{t\ an}$  = le taux de passage en invalidité cumulé pour un âge  $x$  et une ancienneté de  $t \times 12$  mois.
- $qx_x^{mois}$  = le taux de mortalité par mois pour un âge  $x$
- $c^{incap}$  = le taux de couverture pour le risque incapacité
- $N_{x,A_n}^{actifs}$  = le nombre d'actifs d'âge  $x$  au début de l'année  $n$ ,  $n \in \llbracket 0 ; 19 \rrbracket$
- $N_{x,A_n}^{incap,t\ an}$  = le nombre d'assurés en incapacité d'âge  $x$  et d'ancienneté comprise entre  $t-1$  et  $t$ , pour une entrée en incapacité à l'année  $n$ .

Nous commençons par calculer le taux d'entrée des nouveaux adhérents pour chaque âge.

Pour ce faire, nous avons consulté les données de l'INSEE sur le nombre de naissances en France chaque année. Cela nous a permis de calculer le nombre de Français par âge en 2019 en respectant la proportion d'hommes et de femmes de notre portefeuille (0,5% - 99,5%).

Puis nous avons récupéré le nombre d'adhérents présents dans le portefeuille par âge en 2019.

Nous avons ensuite calculé que le taux d'adhésions par rapport à la population française globale :

$$Tx_x^{adhésions} = \frac{Nb\ adhérents\ d'âge\ x}{Nb\ Français\ d'âge\ x}.$$

Par la suite, nous avons pris l'hypothèse que ce taux d'adhésions par âge n'évoluait pas avec le temps.

Il ne nous restait donc plus qu'à calculer le nombre d'adhérents 2020 par âge de la façon suivante :

$$Nb\ adhérents\ d'âge\ x\ en\ 2020 = (Nb\ Français\ d'âge\ x - 1\ en\ 2019) \times Tx_x^{adhésions}.$$

Pour les assurés de 20 ans, nous ferons ainsi entrer chaque année le nombre de nouveaux adhérents de cet âge en les calculant comme ci-dessus pour chaque année future.

Enfin, pour trouver le taux d'entrée des nouveaux adhérents pour un âge  $x$  entre 20 et 58 ans, nous avons réalisé le quotient du nombre d'assurés d'âge  $x+1$  en 2020 par le nombre d'assurés d'âge  $x$  en 2019 :

$$e_x = \frac{Nb\ adhérents\ d'âge\ x + 1\ en\ 2020}{Nb\ adhérents\ d'âge\ x\ en\ 2019}.$$

### Première année de projection :

Pour la seconde phase de cette projection, nous calculons le nombre d'incapacités par âge survenues à l'année actuelle ( $A_0$ ) en multipliant simplement le nombre d'assurés du portefeuille initial à chaque âge par le taux d'incidence de l'âge correspondant :

$$N_{x,A_0}^{incap,1\ an} = N_{x,A_0}^{actifs} \times i_x.$$

Cela implique que le nombre d'assurés actifs de l'année 1 est égale à la population active de l'année 0, à laquelle nous avons ôté le nombre d'assurés en incapacité étant entrés dans cet état dans l'année 0, et à laquelle nous multiplions le taux de chute plus 1. Il ne faut pas également oublier d'ajouter à ce nombre le nombre d'assurés de retour à la population active après avoir été maintenus en incapacité moins d'un an :

$$N_{x,A_1}^{actifs} = N_{x-1,A_0}^{actifs} \times (1 - i_{x-1}) \times (1 + e_{x-1}) + N_{x-1,A_0}^{incap,1 an} \times r_{x-1}^{1 an},$$

Sachant que  $r_x^{1 an}$  se calcule à partir des tables d'expérience construites en amont :

$$r_x^{1 an} = (1 - m_x^{1 an} - pc_x^{1 an}) \times qx_x^{mois}.$$

Pour calculer le nombre d'assurés en incapacité de cette année  $A_1$ , il faut tenir compte de la différence entre les assurés déjà présents et les assurés entrant cette année.

Ainsi, le nombre d'assurés entrant en incapacité en  $A_1$  s'obtient de la même façon que ceux entrés en  $A_0$  :

$$N_{x,A_1}^{incap,1 an} = N_{x,A_1}^{actifs} \times i_x.$$

Et le nombre d'assurés présents depuis au moins 12 mois se retrouve par simple produit du nombre d'assurés entrés l'année 0 par le taux de maintien en incapacité au bout de 12 mois :

$$N_{x,A_1}^{incap,2 an} = N_{x-1,A_0}^{incap,1 an} \times m_{x-1}^{1 an}.$$

### Deuxième année de projection :

Pour la deuxième année de projection ( $A_2$ ), la population valide est obtenue de la même manière que celle de l'année antérieure, mais en prenant en plus en compte les assurés de retour d'incapacité après avoir passé entre un et deux ans dans cet état que l'on ajoute à la somme précédente :

$$N_{x,A_2}^{actifs} = N_{x-1,A_1}^{actifs} \times (1 - i_{x-1}) \times (1 + e_{x-1}) + N_{x-1,A_1}^{incap,1 an} \times r_{x-1}^{1 an} + N_{x-1,A_1}^{incap,2 an} \times r_{x-2}^{2 an},$$

Avec :

$$r_x^{2 an} = (1 - m_x^{2 an} - pc_x^{2 an}) \times (qx_x^{mois,10\%}),$$

où  $qx_x^{mois,10\%}$  est le taux de mortalité par mois pour un âge  $x$  majoré de 10%.

Nous avons effectivement pris l'hypothèse considérant qu'un assuré présent en incapacité depuis plus d'un an avait une probabilité plus importante de mourir qu'un individu de la population générale. Nous avons donc majoré le taux de mortalité initial de 10% dans ce cas. Nous aurions pu choisir une table de mortalité aggravée (présente dans les tables du BCAC), mais étant donnée la difficulté des assureurs pour rattacher les sinistres décès à un arrêt de travail antérieur, et dans un souci de simplification du modèle dans l'outil, nous avons plutôt choisi d'appliquer une surmortalité uniforme sur la population en incapacité. Ici encore, un biais dû aux hypothèses est introduit et l'hypothèse pourrait être rediscutée avec une table plus adaptée (comme citée juste avant), mais nous ne disposons pas des ressources pour le faire.

Concernant les assurés en incapacité de cette année  $A_2$ , nous calculons désormais la troisième « classe » d'assurés en incapacité (celle des assurés présents en incapacité depuis plus de 24 mois) en plus des deux autres « classes » qui se calculent avec les mêmes formules que pour l'année précédente.

Cette nouvelle « catégorie » d'assurés en incapacité se calcule en multipliant le nombre d'assurés entrés en  $A_0$  par le taux de maintien en incapacité au bout de 24 mois :

$$N_{x,A_2}^{incap,3 an} = N_{x-2,A_0}^{incap,1 an} \times m_{x-2}^{2 an}.$$

Troisième année de projection et suivante :

Enfin, en utilisant le même raisonnement que pour les deux premières années de projection, la population active des années suivantes s'obtient en ajoutant à la formule utilisée pour les actifs au bout de deux ans, les actifs revenus valides après avoir été maintenus en incapacité entre 24 et 36 mois :

$$N_{x,A_{n \geq 3}}^{actifs} = N_{x-1,A_{n-1}}^{actifs} \times (1 - i_{x-1}) \times (1 + e_{x-1}) + N_{x-1,A_{n-1}}^{incap,1 an} \times r_{x-1}^{1 an} + N_{x-1,A_{n-1}}^{incap,2 an} \times r_{x-2}^{2 an} + N_{x-1,A_{n-1}}^{incap,3 an} \times r_{x-3}^{3 an},$$

Avec :

$$r_x^{3 an} = (1 - m_x^{3 an} - pc_x^{3 an}) \times (qx_x^{mois,20\%}),$$

où  $qx_x^{mois,20\%}$  est le taux de mortalité par mois pour un âge  $x$  majoré de 20%.

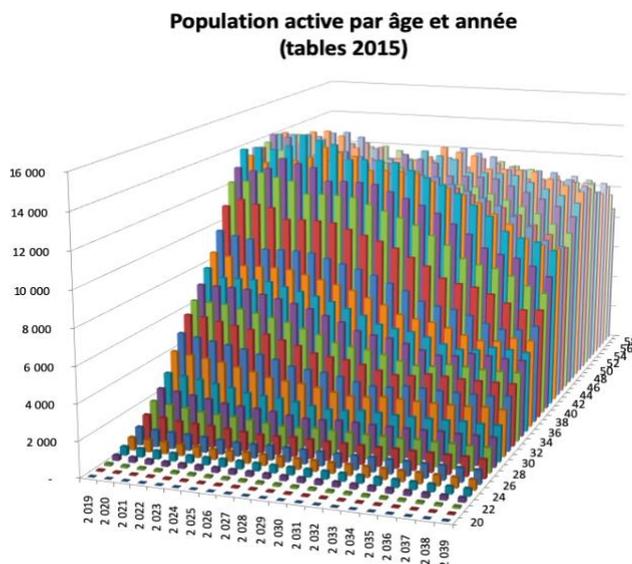
Nous avons effectivement considéré cette fois qu'un assuré présent depuis plus de 24 mois en incapacité avait une probabilité plus importante de décéder qu'un assuré de la population générale et même qu'un assuré présent depuis plus d'un an. Nous avons donc majoré le taux de mortalité « normal » de 20%. Comme pour l'hypothèse d'une surmortalité de 10% faite pour les assurés présents depuis un an en incapacité dans le paragraphe précédent, nous aurions pu choisir une table de mortalité aggravée. Pour les mêmes raisons que précédemment, nous avons opté pour une nouvelle fois pour un taux appliqué uniformément à tous les assurés revenant d'une incapacité de plus de deux ans.

Encore une fois, cette hypothèse devrait être rediscutée avec une table plus adaptée pour rendre cette étude plus fiable.

Le nombre des assurés présents en incapacité pour une année  $n > 1$  se retrouve encore en ajoutant les assurés en incapacité des 3 « classes » formées précédemment car un assuré ne peut de toute façon pas rester dans cet état plus de 3 ans.

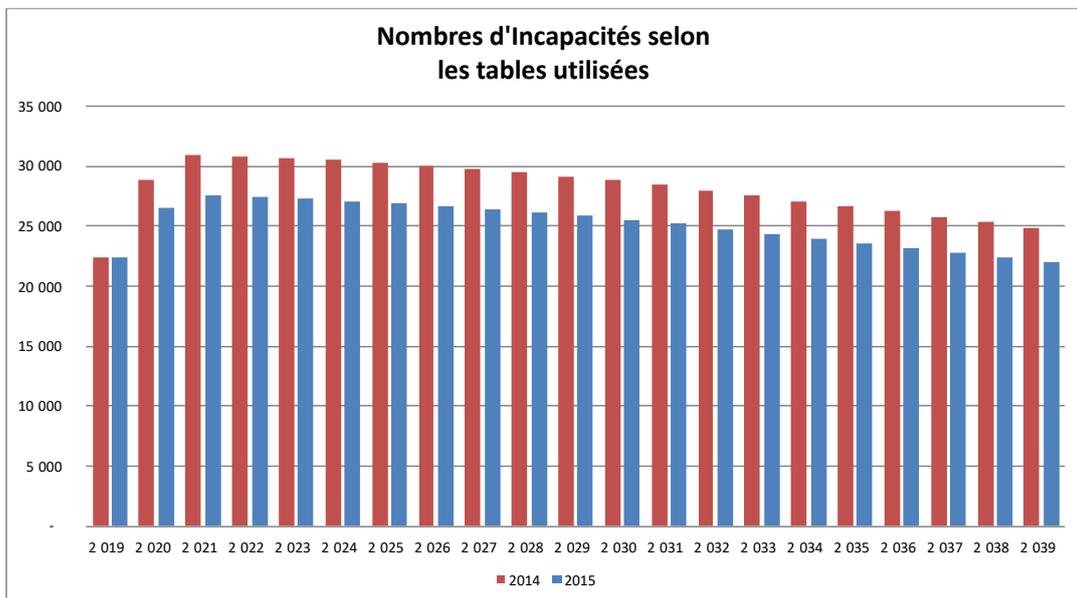
Nous traçons donc maintenant le graphique représentant l'évolution de la population active dans les 20 prochaines années.

Les résultats obtenus à partir de la table construite avec les données des assurés entrés en incapacité en 2015 sont exposés ci-après :



Les résultats montrent que la population valide totale de notre portefeuille ne fera que diminuer durant les 20 prochaines années ce qui s'explique par le départ à la retraite progressif de la génération des baby-boomers depuis le début des années 2010.

Le constat est naturellement le même pour le nombre d'incapacités. Nous remarquons tout de même l'augmentation du nombre d'incapacités visibles les 3 premières années du fait qu'une année correspond à l'empilement de 3 générations d'assurés en état d'incapacité et qu'ainsi les deux premières années ne disposent pas du recul nécessaire pour bénéficier de 3 générations.

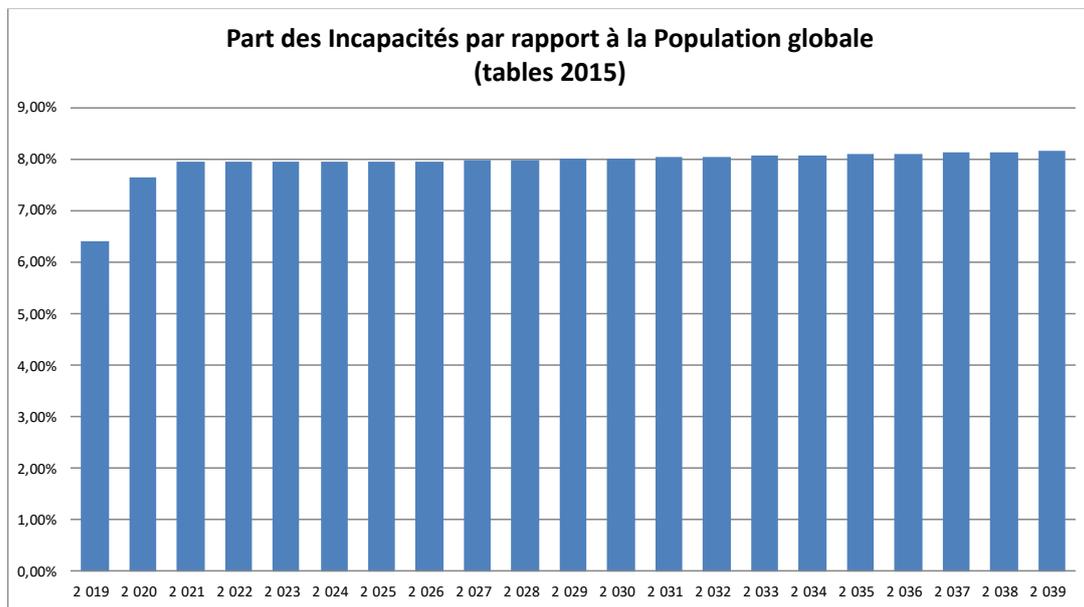


Nous remarquons en revanche la nette influence de l'accélération des délais de sorties d'incapacité constatée lors des parties précédentes et plus précisément lors de la création des tables de maintien en incapacité.

En effet, nous constatons que lors de l'empilement de la première génération, un premier écart conséquent (8% d'assurés en état d'incapacité en moins) se crée découlant du nombre plus important de sorties lors des 12 premiers mois constatées dans les données des assurés entrés en incapacité en 2015, et donc une accumulation moins nombreuse d'assurés en incapacité pour les années suivantes.

Cet écart grandira encore lors de la deuxième année pour la même raison (passant de 8 à 11%). L'écart restera ensuite le même pour les années suivantes.

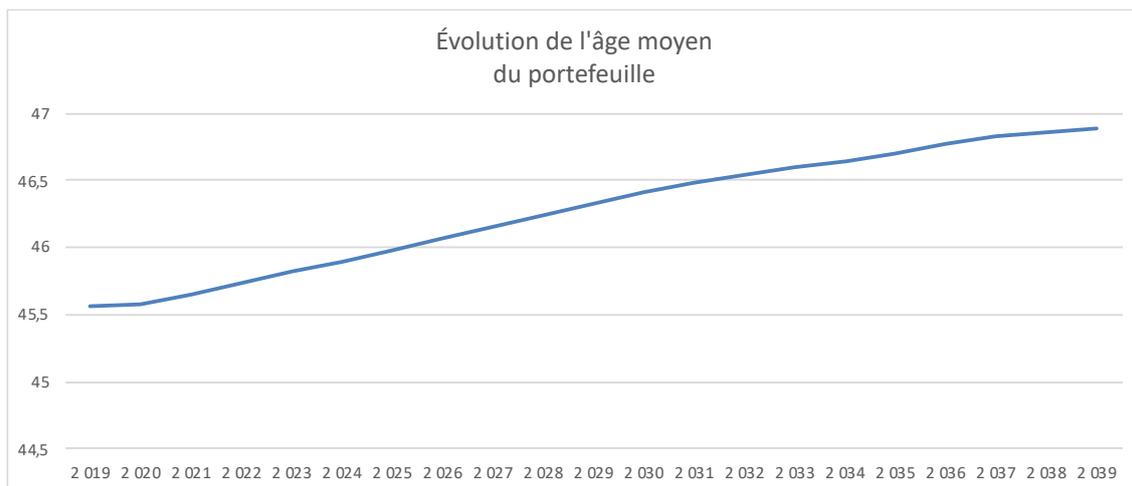
Nous traçons aussi le taux d'assurés en incapacité par rapport à la population globale (active, en incapacité et invalide).



En observant les proportions obtenues à partir des tables 2015, nous observons une légère tendance générale croissante (de 7,95 en 2021 à 8,15 en 2039). Le constat est identique avec les tables 2014 (bien que les taux soient plus élevés).

Cela signifie que si la population générale décroît, la part d'assurés en incapacité, elle, est stable et même légèrement croissante.

Cela s'explique par le vieillissement de la population dans notre modèle (comme illustré dans le graphique suivant) qui ne simule pas assez de nouvelles entrées de jeunes assurés pour compenser la part des individus plus âgés et vieillissants. Cet âge moyen a en effet augmenté de 1 an et presque 4 mois sur les 20 ans de notre simulation.



Nous allons à présent projeter la population présente dans l'état d'invalidité pour vérifier si l'accélération du passage en invalidité joue aussi un rôle sur la volumétrie des invalides.

## IV.2.2. Projection de la population invalide

Pour la projection des invalides, nous partons du principe qu'un assuré ne peut passer de l'état « valide » à celui d'« invalide » et ni de l'état d'invalidité à ceux de validité et d'incapacité. Cela implique que la seule façon de sortir de l'état d'invalidité est le décès de l'assuré.

Nous avons donc choisi d'utiliser les tables de maintien en invalidité du BCAC en considérant que la sortie de l'état était obligatoirement due au décès de l'assuré. Ces taux de décès étant calculés sur un échantillon de population en invalidité, nous avons décidé qu'ils étaient plus représentatifs que ceux d'une table de mortalité classique modélisée à partir d'une population sans spécificité apparente qui a normalement moins de probabilités de décéder. Nous aurions aussi pu nous servir des recommandations de l'Institut des Actuaire préconisant de majorer la table TF 00-02 de 50% pour les invalides entre 60 et 62 ans. Toutefois, il nous paraissait plus pertinent d'utiliser une table construite exclusivement sur l'observation des personnes invalides, sachant qu'il y a très peu de probabilités de sortir de l'état d'invalidité en vie.

Dans le but de rendre l'étude la plus complète possible, nous aurions dû mettre cette hypothèse en concurrence en analysant l'impact de l'application d'une surmortalité ou d'une autre table sur les résultats. Il ne nous est plus possible à l'heure actuelle de reprendre cette hypothèse.

Venons-en au calcul de la volumétrie des invalides ( $N_{x,A_{n \geq 1}}^{inval}$ ) pour chacune des années futures, sachant qu'à l'année initiale  $A_0$ , aucun invalide n'est présent, et ce, pour chaque âge.

Pour ce faire, nous n'avons qu'à faire la somme de quatre produits. Les trois premiers produits étant ceux des nombres d'individus en incapacité dans chacune des trois « classes » de maintien (1 an, 2 ans, 3 ans) par les probabilités de passage au bout de respectivement un an, deux ans et trois ans. Le quatrième étant celui du nombre d'invalides déjà présents par le taux de mortalité de l'âge concerné.

Ainsi nous obtenons la formule suivante pour tout âge  $x \in \llbracket 20; 59 \rrbracket$  et toute année  $n \geq 1$  :

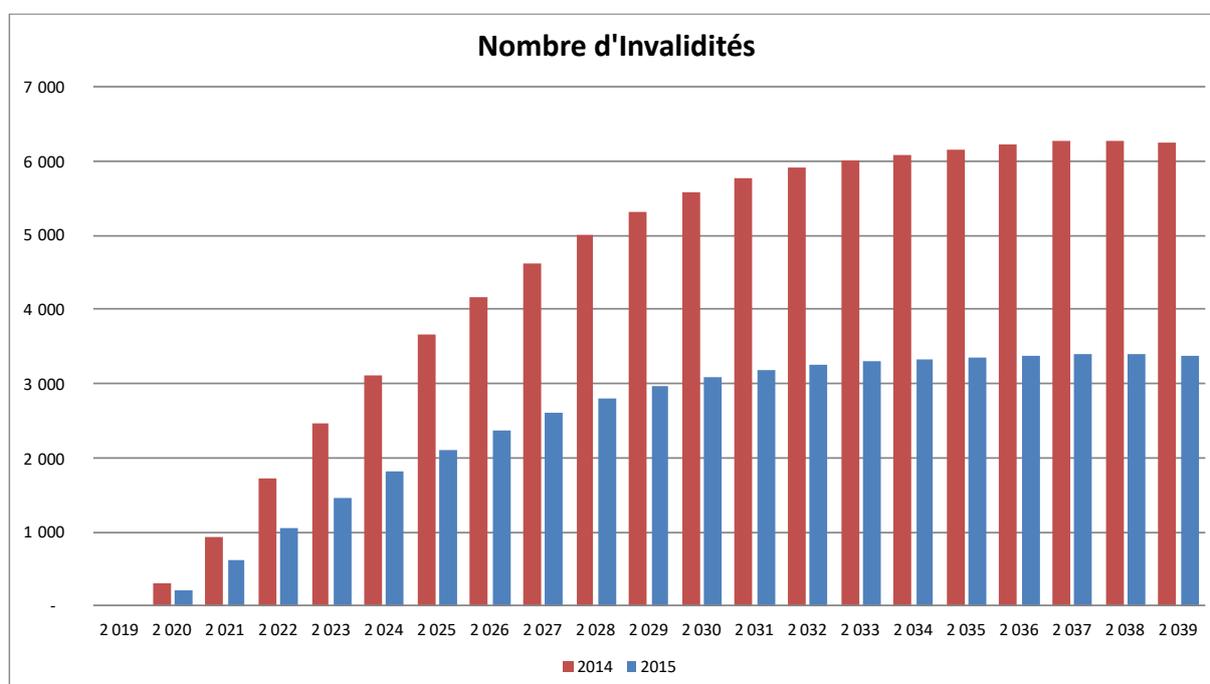
$$N_{x,A_{n \geq 1}}^{inval} = N_{x-1,A_{n-1}}^{inval} \times qx_{x-1}^{année,20\%} + N_{x-1,A_{n-1}}^{incap,1 an} \times pc_{x-1}^{1 an} + N_{x-1,A_{n-1}}^{incap,2 an} \times pc_{x-2}^{2 an} + N_{x-1,A_{n-1}}^{incap,3 an} \times pc_{x-3}^{3 an},$$

Où  $qx_x^{année,20\%}$  est le taux de mortalité annuel pour un âge  $x$  accru de 20%.

Cette formule s'applique bien pour les deux premières années de la même façon que pour le calcul de la population valide puisque :

- Pour la première année,  $N_{x-1,A_{n-1}}^{incap,3 an}$  et  $N_{x-1,A_{n-1}}^{incap,2 an}$  sont nuls étant donné qu'aucun assuré n'est déjà en incapacité depuis plus d'un an ;
- Pour la deuxième année, seul  $N_{x-1,A_{n-1}}^{incap,3 an}$  est nul étant donné qu'aucun assuré n'est dans l'état d'incapacité depuis plus de deux ans.

Voici les graphiques représentant la part des invalidités par rapport à la population valide obtenu selon les années à partir desquelles ont été construites les tables d'expérience :



Nous pouvons d'abord constater que le nombre d'invalides va croître fortement les 15 premières années avant de se stabiliser les 5 suivantes et ce dans les deux modèles.

Cette tendance signifie qu'il faudra 15 ans pour que les sorties de l'état d'invalidité compensent les entrées et ainsi régulent l'effet d'accumulation des assurés dans cet état.

Nous remarquons aussi instantanément qu'à partir des tables 2014 (avant l'accélération de passage en invalidité) la projection donne un nombre bien plus important d'invalides qu'à partir des tables d'expérience 2015.

En confrontant les résultats à ceux obtenus pour la volumétrie des assurés en incapacité, nous pouvons constater que l'accélération du passage en invalidité dont nous avons fait l'étude tout au long de ce mémoire peut aussi être assimilée à une accélération générale des sorties de l'état d'incapacité.

Cette dernière a pour effet de diminuer la volumétrie des incapacités longues et donc de diminuer l'accumulation des individus dans cet état.

Tandis que l'accélération du passage en invalidité augmente la fréquence de passage aux jeunes anciennetés et la diminue aux anciennetés les plus élevées (voir tables de passage en invalidité 2014 et 2015).

Ce double effet constaté implique dès lors un nombre moins important d'invalides (avec les tables 2015) puisque les assurés en incapacité sont moins nombreux, donc il y a moins d'individus soumis au risque de passage en invalidité.

Et le constat est d'autant plus fort que la dernière année d'incapacité, les assurés présents dans cet état sont moins nombreux et les taux de passages moins forts avec les tables de 2015, ce qui augmente significativement l'écart entre les deux modèles.

Nous pourrions pousser la réflexion en nous demandant si un surplus d'individus impliquant un plus grand nombre d'arrêts, n'impliquerait pas surtout une plus grande part d'arrêts « moins graves » et donc de plus petites périodes d'incapacité, ce qui réduiraient la répartition générale des délais de sortie de l'état d'incapacité.

Quant à celle des délais d'entrées en invalidité, elle se verrait également diminuer car un nombre croissant d'invalidités avec les années impliquerait peut-être une simplification du système décisionnel de passage d'un état à l'autre pour éviter des surcharges administratives, et ainsi passer les assurés en incapacité plus facilement et rapidement en invalidité.

La proportion, elle, serait moins importante que sans l'accélération du passage en invalidité car il y aurait, comme écrit juste au-dessus, une proportion d'arrêts « moins graves » (et donc ne passant pas en invalidité) plus importante.

Enfin, nous allons nous demander dans la partie suivante si cette accélération des sorties de l'état d'incapacité et donc de la part moins importante d'assurés en incapacité a des impacts positifs sur les charges et les cotisations que gèrent l'assureur.

# Chapitre 3. Projections des cotisations et des charges à partir des tables d'expérience 2014 et 2015

## IV.3.1. Projection des cotisations

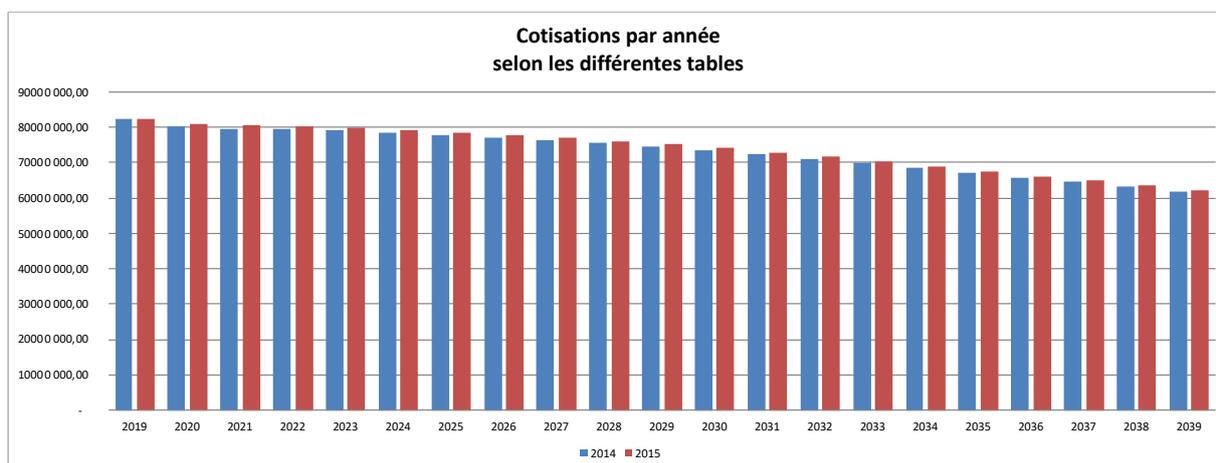
Les cotisations (également appelées primes) en assurance sont les sommes à payer par l'assuré en contrepartie des garanties accordées par l'assureur. Cela permet à l'assureur de payer, en mutualisant les risques, les charges dues en cas de réalisation des risques pour lesquels était couvert l'assuré.

Elles sont donc très importantes et permettent également de calculer des indicateurs permettant de savoir si les primes ont suffisamment permis de couvrir les sinistres.

Les cotisations ( $Cot_x^{totale}$ ) d'un âge  $x$  sont ici obtenues en multipliant simplement la population active d'un âge  $x$  par le montant moyen des cotisations annuelles ( $Cot_x^{moyenne}$ ) que paie la population de cet âge :

$$Cot_x^{totale} = N_{x,A_n}^{actifs} \times Cot_x^{moyenne}.$$

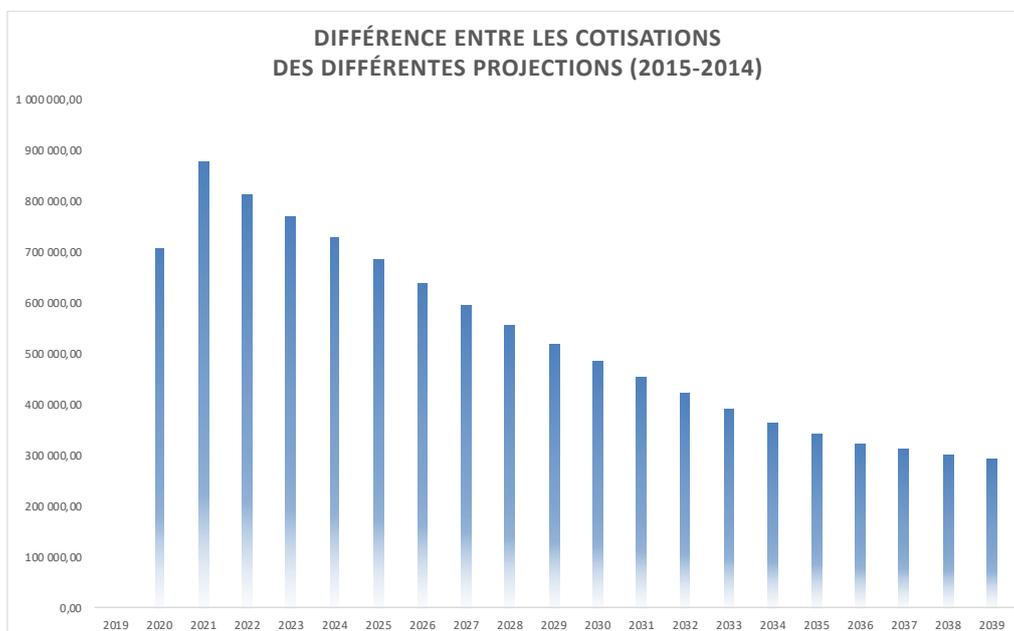
Voici les résultats que nous obtenons selon les différentes tables d'expérience utilisées :



Logiquement au vu de l'évolution de la population valide observée ci-avant, nous pouvons constater que les cotisations totales diminuent tout au long de la projection.

Ce qui correspond une nouvelle fois au départ à la retraite de la génération des baby-boomers, la génération suivante moins nombreuse paie donc moins au total.

La différence entre les deux projections est assez faible, c'est pourquoi nous traçons également les différences observées entre les deux projections :



Nous remarquons qu'à partir des tables 2015 nous obtenons des montants de cotisations globaux plus élevés pendant chacune des 20 prochaines années.

Cela paraît cohérent puisque l'accélération des délais de sorties de l'état d'incapacité implique un plus grand nombre d'assurés « valides » et donc plus de cotisations.

### IV.3.2. Projection des charges « incapacité »

Nous nous intéressons ici seulement aux charges « incapacité » étant donné qu'il nous est difficile d'évaluer la durée de maintien en invalidité de chacun des assurés du fait du manque d'expérience que nous avons sur le nouveau comportement des assurés invalides depuis l'émergence du phénomène d'accélération des délais de passages en invalidité.

Il ne nous semble donc pas cohérent de calculer le montant des prestations « invalidité » à partir de tables d'expérience de passage différentes mais de tables de maintien identiques.

Les charges « incapacité » cependant peuvent être calculées et comparées puisque nous disposons de tout ce qui nous est nécessaire pour faire les projections.

Ces charges sont obtenues en sommant les prestations « incapacité » et les provisions « incapacité ».

De même que pour la projection de la population en état d'incapacité, nous divisons les prestations et les provisions en 3 périodes d'un an.

Ainsi la première année  $A_0$ , les prestations ( $S_{x,A_0}^{1an}$ ) sont calculées à partir de la durée passée, entre 10 jours soit  $2/3$  de mois (franchise) et 11 mois, par un individu d'âge  $x$  dans l'état d'incapacité. Pour cet individu, nous multiplions ensuite son nombre de mois passés en incapacité par son salaire mensuel moyen (qui dépend de son âge) et par le taux de couverture pris en charge par l'assureur moins celui pris en charge par la sécurité sociale (76%-50% car tous les salaires sont dans la tranche A d'après les tranches de la Sécurité Sociale). Nous répétons l'opération pour chacun des assurés et nous sommons le tout pour finir :

$$S_{x,A_0}^{1an} = \left( \frac{2}{3} \times (m_x^{0\text{ mois}} - m_x^{1\text{ mois}}) + \sum_{i=1}^{11} i \times (m_x^{i\text{ mois}} - m_x^{i+1\text{ mois}}) \right) \times N_{x,A_0}^{incap,1an} \times \text{Salaire}_x^{\text{mensuel}} \times 26\%.$$

Les provisions sont, elles, obtenues la première année en observant le nombre d'assurés présents en incapacité depuis plus d'un an et moins de deux ans dans la deuxième année. Nous faisons ensuite le produit de ce nombre d'individus par leur salaire annuel moyen, puis par 26%, et enfin par la somme de deux coefficients : un « d'incapacité temporaire de travail » et un de « rente en attente de passage ». Ces coefficients sont spécifiques à l'assureur et sont calculés respectivement à partir des tables de maintien en incapacité et de passage en invalidité.

Nous notons :

- $pmitt_{t\ an}$  le coefficient d'incapacité temporaire de travail ;
- $pmrea_{t\ an}$  le coefficient de rente en attente ;
- $PM_{x,A_n}^{1\ an}$  le montant de la provision mathématique au bout d'un an pour un âge  $x$  l'année  $n$ .

$$PM_{x,A_0}^{1\ an} = N_{x,A_1}^{incap,2\ an} \times Salaire_{x-1}^{annuel} \times 26\% \times (pmitt_{1\ an} + pmrea_{1\ an}).$$

Pour la seconde année, nous calculons  $S_{x,A_1}^{1\ an}$ , mais aussi les prestations payées pour les assurés présents depuis plus de 11 mois de la même façon que pour ceux de moins de 11 mois :

$$S_{x,A_1}^{2\ an} = \left( \sum_{i=1}^{12} i \times (m_x^{i+11\ mois} - m_x^{i+12\ mois}) \right) \times N_{x-1,A_0}^{incap,1\ an} \times Salaire_{x-1}^{mensuel} \times 26\%.$$

Il ne reste plus qu'à sommer les deux résultats obtenus pour connaître les prestations payées la deuxième année.

Les provisions de cette année pour les assurés présents depuis plus de 11 mois sont calculées avec la même formule que ceux présents depuis moins de 1 an mais en utilisant le nombre d'assurés présents en incapacité depuis plus de 2 ans l'année suivante. On obtient :

$$PM_{x,A_1}^{2\ an} = N_{x,A_2}^{incap,3\ an} \times Salaire_{x-2}^{annuel} \times 26\% \times (pmitt_{2\ an} + pmrea_{2\ an}).$$

Enfin pour les prestations et provisions « 3 ans », nous utilisons les mêmes raisonnements que pour les prestations et provisions « 1 an » et « 2 an ».

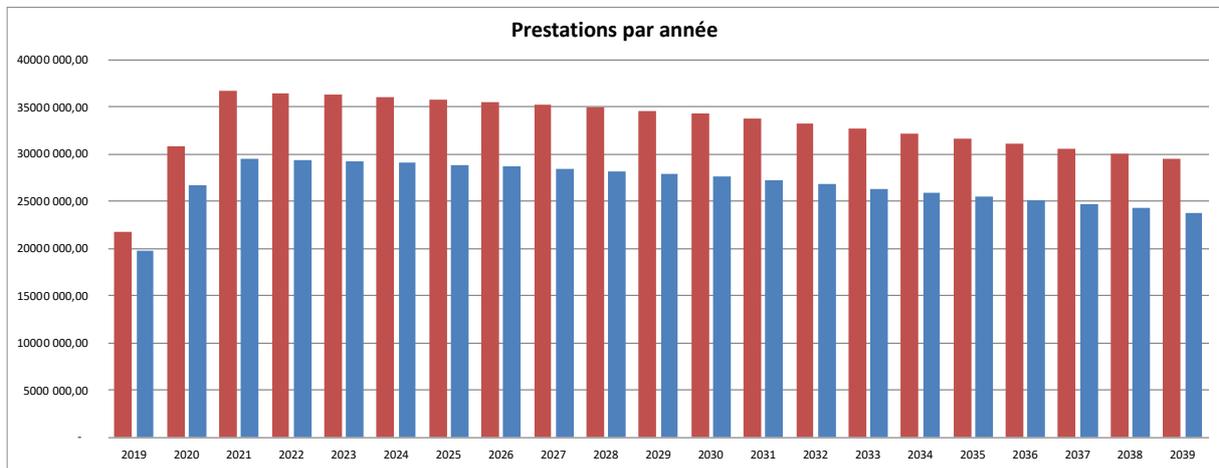
Les formules obtenues sont les suivantes en introduisant un nouveau coefficient « invalidité » ( $pminv_{0\ an}$ ) :

$$S_{x,A_2}^{3\ an} = \left( \sum_{i=1}^{12} i \times (m_x^{i+23\ mois} - m_x^{i+24\ mois}) \right) \times N_{x-2,A_0}^{incap,1\ an} \times Salaire_{x-2}^{mensuel} \times 26\%,$$

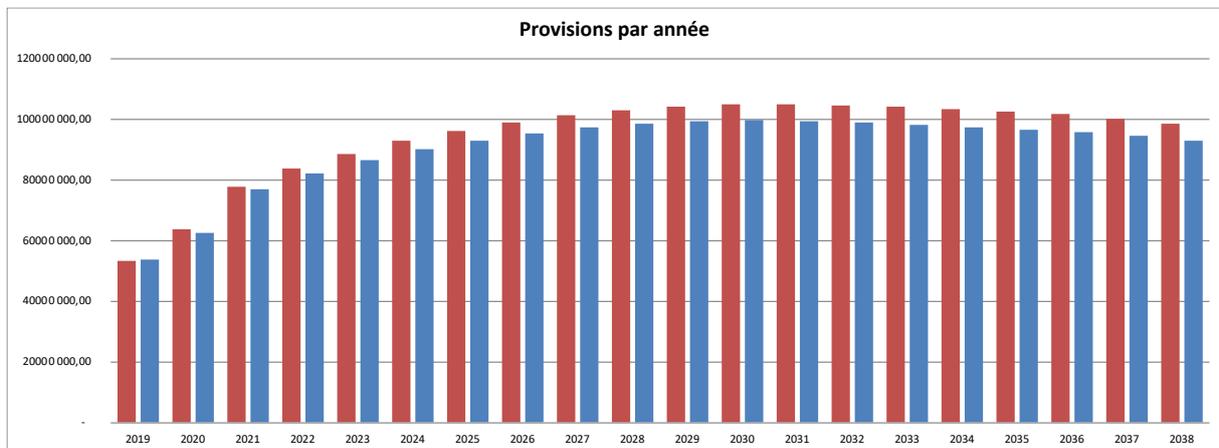
$$PM_{x,A_2}^{3\ an} = N_{x,A_3}^{inval} \times Salaire_{x-3}^{annuel} \times 26\% \times pminv_{0\ an}.$$

Pour obtenir les charges nous n'avons plus qu'à sommer les montants des prestations et des provisions.

Les résultats sont retranscrits ci-dessous :



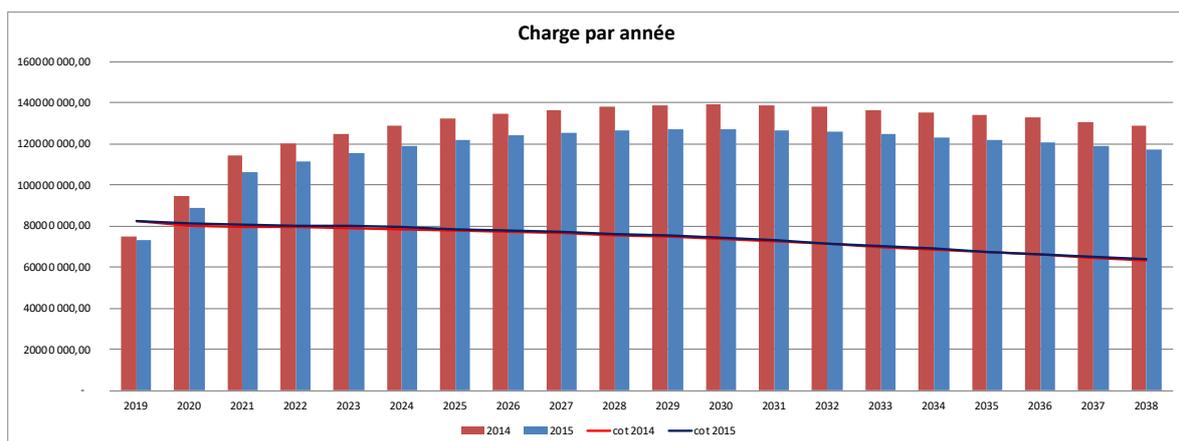
Les prestations payées pour l'incapacité à partir des tables 2014 sont bien plus élevées que celles payées à partir des tables 2015. L'accélération analysée dans ce mémoire implique donc des bonis non-négligeables pour le portefeuille de prestations de l'assureur, estimés de l'ordre de 19% sur ce portefeuille.



Concernant les provisions mises en place par l'assureur pour faire face au risque, elles augmentent jusqu'en 2031 avant de diminuer sur les années suivantes.

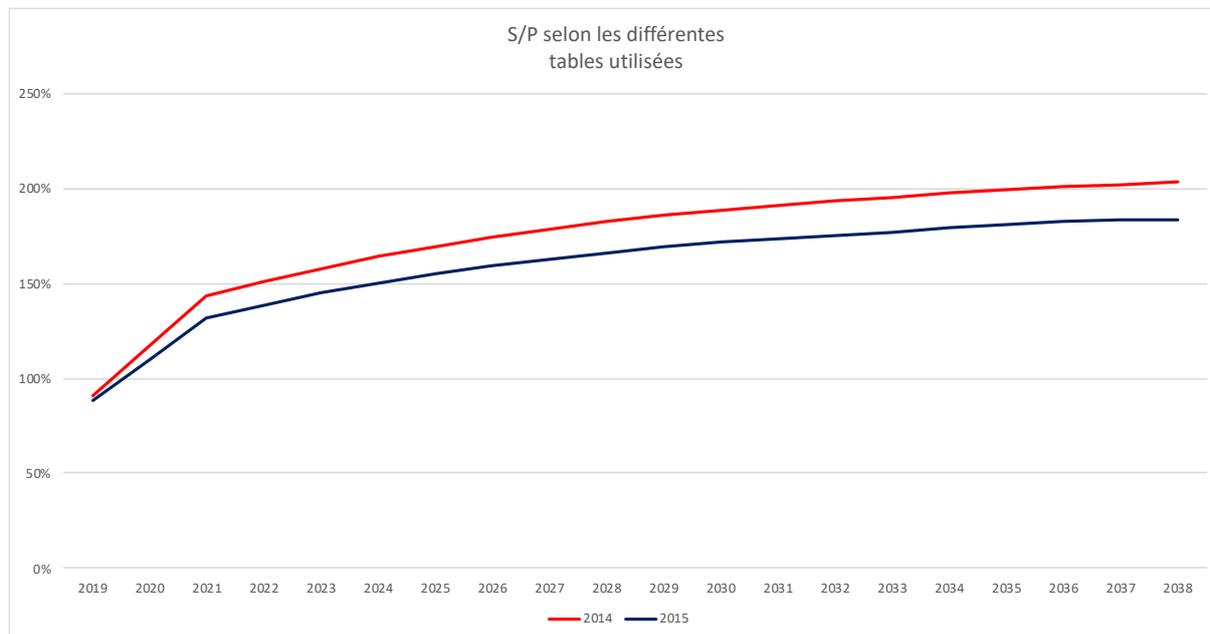
Nous expliquons ce phénomène par la croissance du nombre d'invalidités plus importante que la diminution des incapacités jusqu'à cette date, ce qui augmente les provisions à placer. La stabilisation des invalidités et la continuité de la décroissance des incapacités dans les années 2030 impliquent la baisse des provisions pour cette décennie.

Comme pour les prestations, le gain estimé de l'accélération du passage en invalidité est lui de 4% par an en moyenne pour les provisions.



En observant les charges par année, nous observons le même pic que pour les provisions.

Les résultats nous confortent dans l'idée que l'accélération du passage en invalidité permettra aux assureurs de payer moins de charges et d'engranger en plus davantage de cotisations (les tables 2015 étant plus avantageuses que celles de 2014).



Pour autant, si comme écrit juste au-dessus, notre S/P sera meilleur grâce à l'accélération du passage en invalidité, le graphique ci-dessus nous indique tout de même que le ratio de S/P reste à corriger car trop risqué pour l'assureur.

Ce chapitre nous a donc permis de vérifier les impressions que nous avons en débutant cette étude : l'accélération a des effets positifs sur le portefeuille de l'assureur et cela peut être une des raisons qui ont poussé à simplifier les démarches pour sortir de l'état d'incapacité.

De plus si la génération des baby-boomers semble bien avoir eu une influence sur les nombres croissants d'arrêts et d'invalidités observés ces dernières années, la tendance pour les années futures serait davantage à une baisse du nombre d'assurés en état d'incapacité car la population générale française des moins de 60 ans elle-même diminue puisque les baby-boomers passent progressivement cette barre.

# Conclusion

Nous avons montré tout au long de ce mémoire et à l'aide de diverses approches que le changement de comportement des passages en invalidité et du maintien en incapacité de ces dernières années était le fruit de divers phénomènes :

- Épidémiologique tout d'abord puisque ces dernières années (2015-2018) ont traversé plus d'épidémies grippales que les années précédentes (2012-2014), ce qui a impliqué plus d'arrêts de courte durée ;
- Un surplus d'assurés puisque l'âge de départ à la retraite a augmenté progressivement entre 2013 et 2017 mais aussi car la génération des baby-boomers est arrivée dans la tranche d'âge la plus « risquée » concernant l'arrêt de travail (les 50-60 ans) au début des années 2010 et a donc impliqué un plus grand nombre d'arrêts et d'invalidités, même si le taux de passage en invalidité moyen baisse depuis 2012 ;
- Enfin une accélération du passage en invalidité, avec des démarches simplifiées pour sortir de l'état d'incapacité.

Ce dernier point a mobilisé une grande partie de notre étude puisque malgré le faible recul que nous avons, nous avons pu constater un réel changement de comportement des taux de passages entre les années 2012 à 2014 et l'année 2015.

Est-ce une tendance qui se dessine ou une simple année particulière ? Nous ne pouvons pas nous prononcer pour le moment mais nous pouvons dire que cette dernière année a vu s'accélérer considérablement (de plusieurs mois) les délais de passages en invalidité.

Cette année a également vu proportionnellement beaucoup plus d'invalidités être déclarées au cours des deux premières années que les 3 années 2012, 2013 et 2014. Ce qui a considérablement baissé le nombre de passages dans les derniers mois (puisque moins d'assuré en incapacité étaient encore présents) alors que ce sont ces mois qui ont les fréquences de passages les plus élevées. Cela nous a amené à créer une nouvelle table de maintien en incapacité et une nouvelle table de passage en invalidité. Ces tables nous ont permis de montrer les changements de provisions que devra faire l'assureur pour pallier cette accélération des passages en invalidité, si toutefois celle-ci se confirme. Ces dernières pourraient être revues à la baisse et ainsi augmenter les bonis de l'assureur.

Enfin les projections ont permis de montrer du doigt les évolutions de la population dans les années à venir et ainsi l'impact que marque la fin de la génération des baby-boomers dans le régime de prévoyance.

Ce dernier devrait logiquement voir baisser le nombre d'incapacités car le nombre d'individus diminue (les baby-boomers passant progressivement au-delà de l'âge limite pour être indemnisés).

Si toutefois le nombre d'incapacités continue d'augmenter comme c'est le cas depuis 2012 mais sans cause épidémiologique cette fois, il sera nécessaire de s'interroger sur la possible dégradation de l'environnement social des assurés et particulièrement de la tranche de travailleurs la plus âgée.

Pour donner plus de valeur à ce travail il sera nécessaire de le reprendre d'ici trois à quatre ans afin d'obtenir un recul indispensable à l'affirmation définitive des résultats obtenus ici.

De plus, il me paraît important de souligner une nouvelle fois que certaines hypothèses de ce mémoire imposent de la prudence sur les résultats obtenus dans ce travail.

En effet, il n'a pas été prouvé un changement de comportement en arrêt de travail pendant l'année de survivance 2015 que pour une population très ciblée (aides à domicile). Cette rupture ne serait peut-être pas aussi marquée sur une population globale. Cependant, une étude faite sur l'ensemble des portefeuilles de Prévoyance (faite par Forsides<sup>1</sup>) montrant également une accélération du passage en invalidité tend à élargir nos résultats à d'autres types de populations.

Une suite logique à ce mémoire serait de réaliser un travail similaire sur un autre portefeuille aux caractéristiques différentes.

Enfin, certaines hypothèses desquelles dépendent les résultats n'ont pu être challengées car les données utilisées n'étaient plus disponibles. Nous avons cependant fait au mieux pour limiter au maximum les biais. Néanmoins, cette étude pourra être reprise en testant la sensibilité de certains d'estimateurs ou tables. Par exemple en utilisant une méthode de lissage par splines plutôt que celle de Whittaker-Henderson (partie III.3.1.2.C) ou en utilisant d'autres tests d'appariement que celui du Chi-2 comme celui de Wilcoxon-Mann-Whitney (partie III.3.1.1).

# Bibliographie

- 
- <sup>i</sup> *RISQUES INCAPACITÉ & INVALIDITÉ : Observations des tendances et dérives grâce aux avancées des technologies de data science*, Présentation de Forsides du 18 avril 2019 par Virginie FEMERY (groupe VYV), Pascale QUENNELLE (Forsides), et Cheick Oumar DIALLO (Forsides)
- <sup>ii</sup> *Comparaison de moyennes*, JY Baudot ([lien](#))
- <sup>iii</sup> *Use of Ranks in One-Criterion Variance Analysis*, William H. Kruskal et W. Allen Wallis, Journal of the American Statistical Association, vol. 47, no 260, 1er décembre 1952, p. 583–621 ([lien](#))
- <sup>iv</sup> *Estimation non paramétrique de la loi d'une durée de vie*, 2013 ([lien](#))
- <sup>v</sup> *Lois théoriques usuelles (Weibull)*, Photis NOBELIS ([lien](#))
- <sup>vi</sup> *Lois théoriques usuelles (Gaussiennes)*, Photis NOBELIS ([lien](#))
- <sup>vii</sup> *MODELE DE MELANGE ET CLASSIFICATION*, Danho DJROBIE ([lien](#))
- <sup>viii</sup> *Caractérisation de mélange de distributions gaussiennes*, Didier Maquin, Gilles Mourot, José Ragot, 25 mars 2014 ([lien](#))
- <sup>ix</sup> *Estimation des paramètres d'un mélange de lois normales provenant d'un modèle saut-diffusion à volatilité stochastique à deux états*, NICOLAS S. ZAMFIRESCU CONSTANTIN CHILARESCU, Journal de la société statistique de Paris, tome 139, no 2 (1998), p. 61-86 ([lien](#))
- <sup>x</sup> *Information theory and an extension of the maximum likelihood principle*, Hirotugu Akaike, Second International Symposium on Information Theory, 267-281 p, 1973
- <sup>xi</sup> *Karl Pearson's theoretical errors and the advances they inspired*, Stephen Stigler, Statistical Science, n° 23, 2008 ([lien](#))
- <sup>xii</sup> *Test de Cochran*, Jean-Philippe Auzelle ([lien](#))
- <sup>xiii</sup> *Construction de tables d'expérience pour l'entrée et le maintien en Incapacité*, Mémoire d'Amélie GAUMET, 16 mai 2001 ([lien](#))
- <sup>xiv</sup> *Individual comparisons by ranking methods*, Frank Wilcoxon, Biometrics Bulletin, 1945  
*On a test of whether one of two random variables is stochastically larger than the other*, Henry B. Mann et Donald R. Whitney, Ann. Math. Stat., 1947
- <sup>xv</sup> *Statistique des modèles non paramétriques*, Frédéric PLANCHET ([lien](#))
- <sup>xvi</sup> *Whittaker-Henderson Smoothing* ([lien](#))
- <sup>xvii</sup> *Le lissage par fonctions splines en statistique*, Philippe Bessechristine Thomas-Agnan, revue bibliographique : Statistique et analyse des données, tome 14, n° 1, 1989 ([lien](#))
- <sup>xviii</sup> *L'estimation de la densité par la méthode du noyau : méthode et outils : Note méthodologique et technique*, Magali Di Salvo, Monique Gadais, Geneviève Roche-Woillez ([lien](#))
- <sup>xix</sup> *Notions d'interpolation et d'extrapolation*, Philippe Vignoles ([lien](#))
- <sup>xx</sup> *Refonte des lois de maintien en incapacité temporaire de travail*, Mémoire de BAGUI H., 2013, ISFA ([lien](#))

## Lectures diverses ayant permis l'écriture de ce mémoire :

*Construction d'une loi de changement de catégorie d'invalidité et étude de l'impact sur les provisions mathématiques*, Mémoire présenté par Pierre MORLON, 15 novembre 2017 ([lien](#))

*Construction de tables d'expérience des risques incapacité et invalidité*, Mémoire de Tom LEURENT, 4 octobre 2010 ([lien](#))

---

*Élaboration d'une table d'expérience : comparaison de méthodes de lissage analytique et d'ajustement statistique*, Mémoire d'Olivier CLEMENT, 2003 ([lien](#))

*Construction d'une table de mortalité sur un portefeuille de temporaire décès*, Mémoire de Benjamin BALTESAR, 9 décembre 2013 ([lien](#))

*Introduction à l'analyse des durées de survie*, Philippe SAINT PIERRE, Février 2015 ([lien](#))

*Finite Mixture Models*, Geoffrey McLachlan et David Peel, 2000

*Classification de données mixtes par un modèle de mélange de copules Gaussiennes*, Matthieu Marbac, 16 mai 2014 ([lien](#))

*Théorie et pratique de l'assurance vie*, Michel Fromenteau et Pierre Pelauton, édition DUNOD

*Lexique de la banque et des marchés financiers*, Blanche Sousi-Roubi, édition DUNOD

*Loi n° 89-1009 du 31 décembre 1989 renforçant les garanties offertes aux personnes assurées contre certains risques*, LEGIFRANCE ([lien](#))

*Prévoyance*, Éditions législatives ([lien](#))

*La Protection Sociale*, SENAT ([lien](#))

*Notre environnement : La Sécurité Sociale, L'Assurance Maladie* ([lien](#))

*Les conséquences économiques liées au report de l'âge de la retraite*, Le Cercle de l'Épargne, 7 octobre 2016 ([lien](#))

*En 2016 : le marché de la complémentaire santé et de la prévoyance a été porté par les contrats collectifs*, Fédération Française de l'Assurance ([lien](#))

*Comment faire face à un classement en invalidité ?*, Éditions législatives, 6 décembre 2017 ([lien](#))

*Assurance invalidité au travail, comment ça marche ?*, COOVER ([lien](#))

*Article 31 de la « loi Evin » : réponses... et questions, article de GILLES BRIENS AVOCAT ASSOCIÉ, FROMONT BRIENS dans l'Argus de l'Assurance*, 4 février 2011 ([lien](#))

### Documentation sur les langages de programmation :

*Décoder des groupes dans un jeu de données grâce aux mélanges gaussiens*, Adrien Perrard ([lien](#))

*Les mélanges gaussiens*, Julien DELAFONTAINE, 29 janvier 2014 ([lien](#))

*Scatterplot3d : 3D graphics – R software and data visualization*, STHDA ([lien](#))

*Analyse des durées de vie avec le logiciel R*, Ségolen Geffray ([lien](#))

*Smooth a life table using the Whittaker-Henderson method*, MortalityTables: A Framework for Various Types of Mortality ([lien](#))

*Statistiques en sciences humaines avec R*, Jean-Herman Guay, Presses de l'Université Laval, Québec, 2012 et 2014 ([lien](#))

*Programmation et analyse statistique avec R*, Christian Paroissin, édition ellipses

*SAS l'essentiel*, Olivier Decourt, édition DUNOD