

**Mémoire présenté le :**

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA  
et l'admission à l'Institut des Actuaires**

Par : BORGOMANO EDOUARD

Titre Construction et certification d'une table de maintien en incapacité

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

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

*Membre présents du jury de l'Institut  
des Actuaires*

signature

Entreprise :

Nom : APRECIALIS

Signature :

Directeur de mémoire en entreprise :

Nom : DELERIVE Philippe

Signature :


Invité :

Nom :

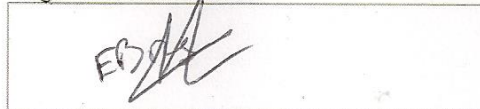
Signature :

**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 l'ISFA*

# Résumé

Les organismes assureurs, lors de la tarification et du provisionnement de contrats de type « Arrêt de Travail », sont souvent amenés à construire des tables d'expérience de maintien en incapacité et à les faire certifier par un actuaire indépendant.

Ce mémoire a pour objectif de construire une table d'expérience de maintien en incapacité tout en expliquant le rôle du certificateur lors du processus de construction. La construction de la table d'expérience est également l'occasion d'étudier et de comparer différentes méthodes de construction.

# Abstract

In order to price and reserve contracts of type « work stoppage », insurance companies used to construct temporary disablement recovery tables, which are certified by an independant actuary.

The goal of this memory is to construct a temporary disablement recovery table while explaining the job of the certifier during the construction process. The construction of the temporary disablement recovery table is also the occasion to study and to compare differents methods of construction.

# Remerciement

Je souhaite remercier ma tutrice universitaire, Esterina Masiello, pour sa disponibilité et ses relectures qui m'ont particulièrement aidé.

Je remercie mon tuteur d'entreprise et directeur général d'Apécialis, Philippe Delerive, pour son soutien et pour m'avoir offert l'opportunité de travailler sur ce sujet.

Je remercie également Raymond Tang pour son encadrement quotidien durant mon année d'alternance.

Enfin, je remercie mes collègues de travail Alicia, Mordehai et Raphael pour avoir contribué à créer une atmosphère de travail agréable et joyeuse.

# Table des matières

<b>Introduction .....</b>	<b>6</b>
<b>1. Contexte de l'étude.....</b>	<b>8</b>
1.1 La protection sociale.....	8
1.2 Risque « Arrêt de travail » .....	9
1.3 Les intervenants du marché et le cadre règlementaire .....	10
1.4 Le provisionnement des garanties incapacité.....	18
1.5 Le sujet de l'étude .....	24
<b>2 Présentation des données .....</b>	<b>25</b>
2.1 Construction et contrôle de la base de données.....	25
2.2 Statistiques sur la base de données .....	33
2.2.1 Test du Khi 2 d'indépendance .....	33
2.2.2 Etudes statistiques des variables .....	35
2.3 Conclusion.....	46
<b>3 Conception de lois de maintien .....</b>	<b>48</b>
3.1 Introduction aux notations mathématiques.....	48
3.2 Introduction au vocabulaire .....	49
3.2.1 Définition de la censure et de la troncature dans le cadre de modèle de durée .....	50
3.2.2 Définition censure et troncature dans le champ d'application du maintien en incapacité .....	50
3.3 Méthode de construction d'une loi de maintien .....	52
3.3.1 Méthode de l'estimateur actuariel .....	53
3.3.2 Estimateur de Nelson Aalen .....	54
3.3.3 La méthode de Kaplan Meier .....	57
3.3.4 Comparaison entre l'estimateur de Kaplan-Meier et d'Harrington-Fleming .....	61
3.4 Présentation de la loi de maintien obtenue selon la méthode de Kaplan-Meier.....	61
3.5 Analyse des boni mali .....	64
<b>4 Lissage.....</b>	<b>67</b>
4.1 Intérêt du lissage .....	67
4.2 Estimateur de Greenwood .....	68
4.3 Validation du lissage.....	70
4.3.1 Intervalle de confiance de la fonction de survie.....	70
4.3.2 Test khi 2 .....	71
4.4 Méthodes de lissage .....	71

4.4.1	Lissage par Moyenne Mobile .....	72
4.4.2	Lissage par la méthode de Whittaker-Henderson .....	74
4.4.3	Méthode des splines.....	82
4.4.4	Lissage sur les espérances de maintien résiduelle .....	89
4.5	Résultats des lissages.....	91
4.6	Impact sur le provisionnement.....	93
<b>Conclusion .....</b>		<b>97</b>
<b>Bibliographie.....</b>		<b>98</b>

# Introduction

« Le calcul des provisions techniques de prestations d'incapacité de travail et d'invalidité est effectué à partir des éléments suivants :

1. les lois de maintien en incapacité de travail et en invalidité indiquées aux articles 600-2, 600-3 et 600-4 du présent règlement. Toutefois, **il est possible pour une entreprise d'assurances d'utiliser une loi de maintien établie par ses soins et certifiée par un actuair e indépendant de cette entreprise, agréé à cet effet par l'une des associations d'actuaire s reconnues par l'Autorité de contrôle prudentiel et de résolution.**

... »

Art. 143-12 du règlement ANC (Autorité des Normes Comptables) 2015-11 du 26/11/2015

La réglementation française offre à l'Organisme Assureur, pour le calcul de ses provisions des contrats de type Arrêt de Travail, la possibilité d'utiliser une loi de maintien basé sur les arrêts de travail survenus à ces assurés, ce qui implique de les faire certifier et suivre par un actuair e indépendant. Cette loi de maintien mène à la construction d'une table appelé table d'expérience, facilitant le calcul des provisions. Dans le cas où l'Organisme Assureur ne fait pas appel à un actuair e indépendant, il doit alors s'appuyer sur la table règlementaire en vigueur.

Véronique Lamblé, fondatrice de la société d'actuariat conseil Aprecialis, est habilitée à certifier des tables d'expériences, notamment des tables de maintien en incapacité et est donc sollicité par des Organismes Assureurs pour la certification de leur table d'expérience. A ce titre, Aprecialis doit alors auditer les différentes étapes de construction réalisée par l'Organisme Assureur mais également s'assurer de la pertinence et de la relative prudence de la table construite.

Outre l'explication du rôle du certificateur, l'intérêt du mémoire réside dans la présentation des étapes de construction d'une table d'expérience dont notamment la conception d'une loi de maintien et le lissage qui seront l'occasion d'appliquer différentes méthodes.

L'objectif final de ce mémoire est de construire une table d'expérience, potentiellement certifiable par un actuair e indépendant, et de comparer l'écart produit en termes de provisions par rapport à l'utilisation de la table règlementaire.

Le mémoire est structuré de la façon suivante. Le chapitre 1 vise à contextualiser notre étude. Dans le chapitre 2, on présente les données servant de base à notre étude. Ensuite, le chapitre 3 a pour but de présenter et de concevoir les lois de maintiens issus de nos données. Enfin le

chapitre 4 explique l'intérêt de lisser les lois de maintien obtenus précédemment. Différentes méthodes de lissage y sont présentées. Le mémoire se termine par la comparaison en termes de provisionnement entre la loi de maintien réglementaire et la loi d'expérience.

# 1. Contexte de l'étude

## 1.1 La protection sociale

La protection sociale désigne l'ensemble des mécanismes qui permettent aux individus de faire face financièrement à certains risques sociaux, c'est-à-dire aux situations pouvant provoquer une baisse des ressources ou une hausse des dépenses, telles que :

- la santé, qui comprend la maladie, l'invalidité, les accidents du travail et les maladies professionnelles ;
- la famille, qui intègre les prestations familiales (allocations familiales, aides pour la garde d'enfants), les indemnités journalières et prestations liées à la maternité ;
- la vieillesse et la survie, qui incluent principalement les pensions et les pensions de réversion ;
- le logement, qui comprend essentiellement les allocations de logement ;
- l'emploi, qui comprend divers dispositifs liés à l'insertion et à la réinsertion professionnelle ainsi que l'indemnisation du chômage ;
- la pauvreté et l'exclusion sociale, qui regroupent des prestations diverses en faveur des personnes démunies ;
- le risque de dépendance, ce nouveau risque a été créé pour financer la prise en charge des personnes âgées ou handicapées ; actuellement, le risque dépendance est couvert par divers dispositifs qui ne constituent pas une réponse d'ensemble.

### **Quelques chiffres sur la protection sociale en France et son évolution ces dernières années**

La DREES (Direction de la Recherche, des Etudes, de l'Évaluation et des Statistiques) a publié en 2018 un rapport sur la protection sociale en France et en Europe en 2016 dans le but d'analyser les évolutions récentes des prestations sociales et des ressources les finançant, pour les différents risques (santé, vieillesse-survie, famille, emploi, logement, pauvreté-exclusion sociale). Ces dépenses sociales sont mises en perspective et comparées avec celles des autres États membres de l'Union européenne.

Selon la DREES, la France a versé 714,5 milliards d'euros de prestations sociales en 2016, ce qui représente 32,1 % du PIB (+4 points en 20 ans), ce qui en fait le pays de l'Union européenne numéro 1 en termes de dépenses en protection sociale.

Les comptes de la protection sociale agrègent les interventions :

- des régimes et organismes publics (administrations de Sécurité Sociale, régimes d'intervention sociale de l'État, ...)



- de certaines interventions de la sphère privée (régimes de la mutualité et de la prévoyance, régimes directs d'employeurs des grandes entreprises, régime d'intervention sociale des institutions sans but lucratif).

Les administrations publiques ont versé, en 2016, 91 % de la totalité des prestations.

Si les prestations de protection sociale ont progressé à un rythme soutenu entre 2000 et 2010 (+4,3 % par an en moyenne), leur croissance est nettement plus modérée depuis : elles augmentent de 2,8 % par an entre 2010 et 2014, puis de moins de 2 % par an en 2015 et 2016. Les mesures de maîtrise des dépenses, en particulier la montée en charge des réformes des retraites et le respect de l'Objectif national de dépenses d'assurance maladie depuis 2010, ont contribué au ralentissement de la croissance des prestations de protection sociale.

Les ressources de la protection sociale sont de natures diverses : cotisations sociales, impôts et taxes affectées (notamment la Contribution Sociale Généralisée [CSG]), contributions publiques... Les cotisations sociales représentent à elles seules 61% du total des ressources de la protection sociale.

En 2016, le déficit de la protection sociale se réduit de nouveau significativement et atteint son niveau le plus bas depuis la crise de 2008 (0,4 milliards en 2016 contre 3,5 milliards en 2015). Cette baisse du déficit s'explique notamment par l'amélioration de 1,9 milliard d'euros du solde du régime général de la Sécurité Sociale.

## 1.2 Risque « Arrêt de travail »

Le risque « arrêt de travail » est un risque social pouvant provoquer une baisse des ressources ou une hausse des dépenses et qui est couvert par la protection sociale.

Un arrêt de travail est une prescription médicale délivrée à un patient en cas de maladie, accident de la vie courante ou encore en cas d'accident de travail ou de maladie professionnelle, nécessitant l'interruption de l'exercice de son activité professionnelle.

Une maladie professionnelle est une altération de l'état de santé d'une personne, constaté par un médecin, causé par les différents travaux qu'elle a été amenée à effectuer au cours de sa carrière professionnelle tandis que l'accident du travail est un accident survenu à toute personne salariée par le fait ou à l'occasion du travail. Dans les deux cas, l'interruption du travail est d'origine professionnelle contrairement au cas de maladie et d'accident de la vie courante qui sont d'origine privé.

Ce risque « arrêt de travail » est couvert par des garanties « incapacité/invalidité » proposés par des intervenants présentés ultérieurement.

## **Définition de l'incapacité et de l'invalidité.**

L'incapacité est définie comme l'incapacité physique d'exercer son activité professionnelle. Elle doit être constatée par un médecin et faire l'objet d'un arrêt de travail.

Il existe 3 sorties possibles de l'état d'incapacité : le rétablissement, le décès et le passage en invalidité.

L'incapacité dure au maximum 3 ans. Au-delà de ce délai, si l'assuré est dans l'impossibilité de reprendre une activité professionnelle, l'incapacité peut être transformée en invalidité, selon la décision de la sécurité sociale.

L'invalidité fait donc généralement suite à une période d'incapacité. Il s'agit d'une incapacité prolongée qui se fait dès que l'incapacité est stabilisée.

Trois catégories d'invalidité sont possibles :

- 1ère catégorie : capable d'exercer une activité rémunérée (soit son ancien travail à temps partiel soit autre travail).
- 2ème catégorie : incapacité d'exercer une activité professionnelle.
- 3ème catégorie : invalides incapables de travailler et qui ont besoin d'une tierce personne pour les assister dans les actes quotidiens de la vie.

## **1.3 Les intervenants du marché et le cadre réglementaire**

Pour pallier au risque d'« arrêt de travail » entraînant une perte de revenus, la réglementation impose à différents intervenants d'entrer en jeu à divers degrés via des garanties « incapacité/invalidité ». La partie suivante a pour but de présenter ces différents intervenants en rappelant les obligations réglementaires auxquelles ces derniers sont soumis.

### **Les régimes de base**

#### **La Sécurité Sociale**

La Sécurité Sociale, régime général ou de base, est le premier intervenant du marché. Il prend en charge les professions autres que les indépendants et les professions agricoles. Les professions agricoles sont prises en charge par la MSA (Mutualité Sociale Agricole) et les indépendants par la Sécurité Sociale Indépendant (ex RSI). Nous détaillerons ultérieurement la prise en charge des arrêts de travail pour les professions agricoles et les indépendants pour se pencher préalablement sur le cas de la Sécurité Sociale.

Lorsqu'un individu est en état d'incapacité, la Sécurité Sociale verse des Indemnités Journalières (IJ) afin de compenser partiellement la perte de revenus causée par l'arrêt de travail.

Le montant versé dépend de l'origine de l'arrêt de travail. En effet, la Sécurité Sociale distingue les arrêts de travail d'origine professionnelle (accident de travail ou de maladie professionnelle) et ceux d'origine privé (maladie et accident de la vie courante).

### **Maladie professionnelle et accident du travail**

La Sécurité Sociale est subdivisée en différentes branches. La branche Accidents du travail – Maladies professionnelles (AT-MP) gère les risques professionnels auxquels sont confrontés les travailleurs : accidents du travail, accidents de trajet et maladies professionnelles.

Tous les soins médicaux liés à un accident du travail ou à une maladie professionnelle sont couverts sur la base et dans la limite des tarifs de la Sécurité Sociale. Les soins de ville ou encore l'hospitalisation sont couverts à 100% tandis que les prothèses dentaires et certains produits d'appareillage sont couverts à 150%.

Le montant des indemnités journalières versées par la Sécurité Sociale est calculé à partir du salaire brut du mois précédant l'arrêt de travail. Ce salaire, divisé par 30,42, détermine le salaire journalier de base. Le montant des indemnités journalières évolue dans le temps.

Pendant les 28 premiers jours suivant l'arrêt de votre travail : l'indemnité journalière est égale à 60 % du salaire journalier de base, avec un montant maximum plafonné à 202,78 € au 1er janvier 2019.

À partir du 29e jour d'arrêt de travail : l'indemnité journalière est majorée et portée à 80 % du salaire journalier de base, avec un montant maximum plafonné à 270,37 € au 1er janvier 2019.

Au-delà de trois mois d'arrêt de travail : l'indemnité journalière peut être revalorisée en cas d'augmentation générale des salaires.

### **Maladie et accident de la vie courante**

La branche Maladie assure la prise en charge des dépenses de santé des assurés et garantit l'accès aux soins. Elle recouvre les risques Maladie, Maternité, Invalidité et Décès. Les arrêts de travail d'origine privé (maladie et accident de la vie courante) sont donc couverts par la branche Maladie.

Les indemnités journalières, versées par la Sécurité Sociale, ne débutent qu'après un délai de carence de 3 jours. Le délai de carence est la période qui s'écoule entre l'ouverture d'un droit et le versement des prestations correspondantes. En d'autres termes, l'indemnisation commence à partir du 4ème jour d'arrêt de travail. Cependant, si la durée de travail entre

deux arrêts est inférieure à 48 heures, le délai de carence n'est pas appliqué car on considère qu'il s'agit d'un seul et même arrêt.

De plus, la Sécurité Sociale pose certaines conditions pour bénéficier du versement des indemnités journalières selon la durée de l'arrêt de travail.

Si la durée de l'arrêt est inférieure à 6 mois, l'assuré doit satisfaire au minimum à l'une des 2 conditions suivantes :

- avoir travaillé au moins 150 heures au cours des 3 mois précédant l'arrêt de travail,
- avoir cotisé sur un salaire au moins égal à 1015 fois le montant du SMIC horaire au cours des 6 mois précédant l'arrêt de travail.

Si l'activité est exercée de manière saisonnière ou discontinue, l'assuré doit satisfaire au minimum à l'une des 2 conditions suivantes :

- avoir travaillé au moins 600 heures ;
- avoir cotisé sur un salaire au moins égal à 2 030 fois le montant du SMIC horaire, au cours des 12 mois civils ou des 365 jours précédant l'arrêt de travail.

Si la durée de l'arrêt est supérieure à 6 mois, l'assuré doit satisfaire à la condition suivante :

- justifier de 12 mois d'immatriculation en tant qu'assuré social auprès de l'Assurance Maladie au moment de l'arrêt.

et doit satisfaire au minimum à l'une des 2 conditions suivantes :

- avoir travaillé au moins 600 heures au cours des 12 derniers mois précédant l'arrêt.
- avoir cotisé sur un salaire au moins égal à 2030 fois le montant du SMIC horaire au cours des 12 mois précédant l'arrêt.

### **Montant versé par la Sécurité Sociale**

Le montant journalier de prestations versé par la Sécurité Sociale est égal à 50 % du salaire journalier de base. Ce salaire journalier de base est calculé sur la moyenne des salaires bruts soumis à cotisations des trois derniers mois travaillés précédant l'arrêt de travail, moyenne plafonné à 1,8\* SMIC mensuel sur la base de la durée légale et non plus à hauteur du plafond de la Sécurité Sociale.

Par exemple, pour un salarié percevant un salaire mensuel de 2 000 € bruts qui est en arrêt de travail pour maladie :

Calcul du Salaire Journalier de Base (SJB) = 65,75 € (correspond à la moyenne du salaire perçu au cours des 3 derniers mois divisé par 91,25 jours, soit  $2\,000 * 3 / 91,25$ ) ;

Montant maximal de l'IJ en 2019 : 45,01 € (correspond à  $(1,8 * \text{SMIC mensuel} * 3 / 91,25) * 50\%$ );

Calcul de l'IJ maladie de base versée par l'Assurance Maladie = 50 % du SJB soit 32,87 € ; ce montant est inférieur au plafond de 45,01 € et sera donc pris en compte tel quel dans le calcul du complément de salaire ;

Les personnes ayant trois enfants à charge perçoivent au bout du 31<sup>ème</sup> jour 2/3 du salaire journalier de base. Cependant, il existe toujours un plafond de remboursement dépendant du nombre d'enfants à charge.

### **La MSA**

Pour les travailleurs agricoles, on peut noter que la MSA verse des indemnités journalières tout comme la Sécurité Sociale. Le délai de carence est de 7 jours. Les 28 premiers jours de l'arrêt maladie (moins les jours de carence) sont rémunérés à 21,33 €/jour. À partir du 29<sup>ème</sup> jour (moins les jours de carence), le montant de l'IJ sera de 28,44 €/jour.

### **La Sécurité Sociale des indépendants**

Pour les travailleurs indépendants, on peut retenir que le délai de carence est de 7 jours et que le travailleur indépendant recevra 1/730<sup>e</sup> de son revenu annuel moyen des 3 dernières années.

D'une manière générale, la Sécurité Sociale rembourse mieux ses assurés que la MSA et la Sécurité Sociale des Indépendants, même si l'indemnisation de la Sécurité Sociale reste relativement faible.

### **L'employeur**

L'accord de mensualisation de 1978, améliorée par l'article 5 de l'ANI du 11/01/2008, prévoit l'obligation pour l'employeur de maintenir le salaire de ses salariés sous certaines conditions. Par conséquent, l'employeur complète le versement des indemnités journalières versés par la Sécurité Sociale.

Selon l'accord de mensualisation, pour que le salarié puisse bénéficier de la participation financière de son entreprise à son arrêt de travail, ce dernier doit avoir au moins un an d'ancienneté dans l'entreprise. Le délai de franchise est de 7 jours (néant pour les accidents du travail ou maladies professionnelles).

En prenant en compte les indemnités versés par la Sécurité Sociale, le salarié reçoit :

- 90 % de son salaire pendant les 30 premiers jours
- 66,66 % de son salaire pendant les 30 jours suivants

Les durées d'indemnisation augmentent de 10 jours par tranche de 5 ans d'ancienneté et sont limitées à 90 jours.

Le tableau suivant (Figure 1.1) récapitule les conditions énoncées précédemment :

Ancienneté	Maintien à 90%	Maintien à 66,66%
< 1 an	-	-
1 - 5 ans	30 jours	30 jours
6 - 10 ans	40 jours	40 jours
...	...	...
> 31 ans	90 jours	90 jours

**Figure 1.1 : Indemnisation de l'employeur selon l'ancienneté**

L'employeur peut soit prendre en charge cet engagement lui-même, soit s'assurer auprès d'un organisme assureur.

L'accord de mensualisation, décrit ci-dessus, correspond à l'obligation minimum de maintien de salaire par l'employeur. L'employeur peut être amené à respecter des obligations de maintien de salaire plus généreuse vis-à-vis de son salarié selon la convention collective qu'il a signé.

#### **Les organismes habilités par l'article 1 de la loi Evin du 31 décembre 1989**

L'article 1 de la loi Evin définit les organismes habilités à mettre en œuvre un contrat de protection sociale complémentaire (également appelé contrat de prévoyance), c'est-à-dire un contrat complétant l'indemnisation de la Sécurité Sociale et de l'employeur si l'assuré est un salarié par exemple. Il en existe 3 types :

- Les institutions de prévoyance, paritaires et à but non lucratif, régies par Code de la Sécurité Sociale et du Code rural ;
- Les mutuelles, à but non lucratif régies par le Code de la Mutualité ;
- Les entreprises d'assurance, régies par le Code des Assurances.

Une distinction majeure parmi les contrats de prévoyance est le caractère collectif ou individuel du contrat.

#### Prévoyance collective

La « Prévoyance Collective » permet de couvrir un groupe de personnes qui ont un lien objectif entre elles, ce groupe étant représenté par une personne morale qui va signer le contrat.

Une entreprise peut donc souscrire un contrat de prévoyance collective pour couvrir ses salariés, généralement signé par le chef d'entreprise, si ce contrat est cohérent avec les conditions fixées par le « formalisme » en vigueur dans l'entreprise. On dénombre 4 types de

« formalismes » permettant la mise en place d'un régime de prévoyance collective : Convention Collective Nationale (CCN), accord collectif, référendum, Décision Unilatérale de l'Employeur (DUE).

### **Convention Collective Nationale (CCN)**

La CCN est conclue entre les organisations patronales et les organisations syndicales de salariés. Elle traite de l'ensemble des relations entre les employeurs et les employés c'est-à-dire l'ensemble des conditions d'emploi et des garanties sociales. La CCN peut traiter de la mise en place d'un régime de prévoyance conventionnel. Ce régime est le régime de prévoyance minimum que doit souscrire l'ensemble des entreprises qui relèvent de la CCN.

### **Accord collectif**

L'accord collectif est signé entre les organisations patronales et les organisations syndicales de salariés. Il est conclu au niveau de l'entreprise uniquement. Il traite d'un ou plusieurs sujets déterminés.

### **Référendum (au sein d'une entreprise)**

Le vote des salariés détermine le régime de prévoyance. L'accord doit être ratifié à la majorité des salariés concernés et non pas seulement à la majorité des votants.

### **Décision Unilatérale de l'Employeur (DUE)**

Elle est principalement utilisée dans les petites entreprises. L'employeur doit remettre aux salariés la liste précise des garanties (notice d'information des conditions particulières et des conditions générales). Les salariés présents dans l'entreprise avant la mise en place de la prévoyance collective ne sont pas obligés d'y souscrire s'il y a une part salariale. Mais, le régime est obligatoire pour les nouveaux employés.

### **Prévoyance individuelle**

La prévoyance individuelle, comme son nom l'indique, permet de couvrir une unique personne pour des garanties prévoyance. Toute personne est libre de souscrire à un contrat de prévoyance individuelle. Les raisons à ce choix sont diverses. Un salarié peut souhaiter compléter ses garanties prévoyance si son entreprise n'a pas souscrit de contrat de prévoyance collective ou s'il estime que les garanties de ce contrat ne sont pas suffisantes. La souscription d'un contrat de prévoyance individuelle est également intéressant pour les professions agricoles non salariées et les travailleurs indépendants. En effet, comme on a pu le constater, ces derniers sont peu remboursés par leur régime de base (MSA et Sécurité Sociale des indépendants) et ils ne peuvent pas bénéficier de maintien de salaire de l'employeur et d'une prévoyance collective.

Dans le cadre de ce mémoire, nos données proviennent exclusivement de contrats de prévoyance individuelles.

### Type de franchise

Dans les contrats de prévoyance, le versement d'indemnités journalières est généralement soumis à une période de franchise. Cette période de franchise indique la période à partir de laquelle la mutuelle, la compagnie d'assurance ou l'institution de prévoyance commence à indemniser l'assuré.

Il existe différents types de franchise que l'on va présenter ci-dessous :

- Franchise continue : nombre de jours fixes qui s'applique à chaque arrêt de travail ;
- Franchise discontinue : nombre de jours cumulés qui s'applique sur l'ensemble des arrêts de travail de l'année civile ou glissante sur une année.

### Exemple : Franchise de 30 jours

Un salarié a un arrêt de 20 jours (du 01/01 au 20/01). Le second arrêt sera de 40 jours (du 01/03 au 09/04) :

Dans le cas de la franchise continue : pas d'indemnisation pour le 1er arrêt, puis 10 jours d'indemnités pour le 2ème arrêt.

Dans le cas de la franchise discontinue en année civile : pas d'indemnisation journalière pour le 1er arrêt, puis 30 jours d'IJ après le 2ème arrêt.

- Franchise absolue : l'assureur verse des IJ uniquement pour les jours d'Arrêt de Travail qui interviennent après la date de franchise.
- Franchise relative : les jours d'Arrêts de Travail intervenant pendant la période de franchise constituent un droit aux prestations.

### Exemple : Franchise de 10 jours

Un salarié a un arrêt de 30 jours :

- Dans le cas de la franchise absolue : pas d'indemnisation pendant les 10 premiers jours, puis 20 jours d'indemnités ;

- Dans le cas de la franchise relative : 30 jours d'indemnités.



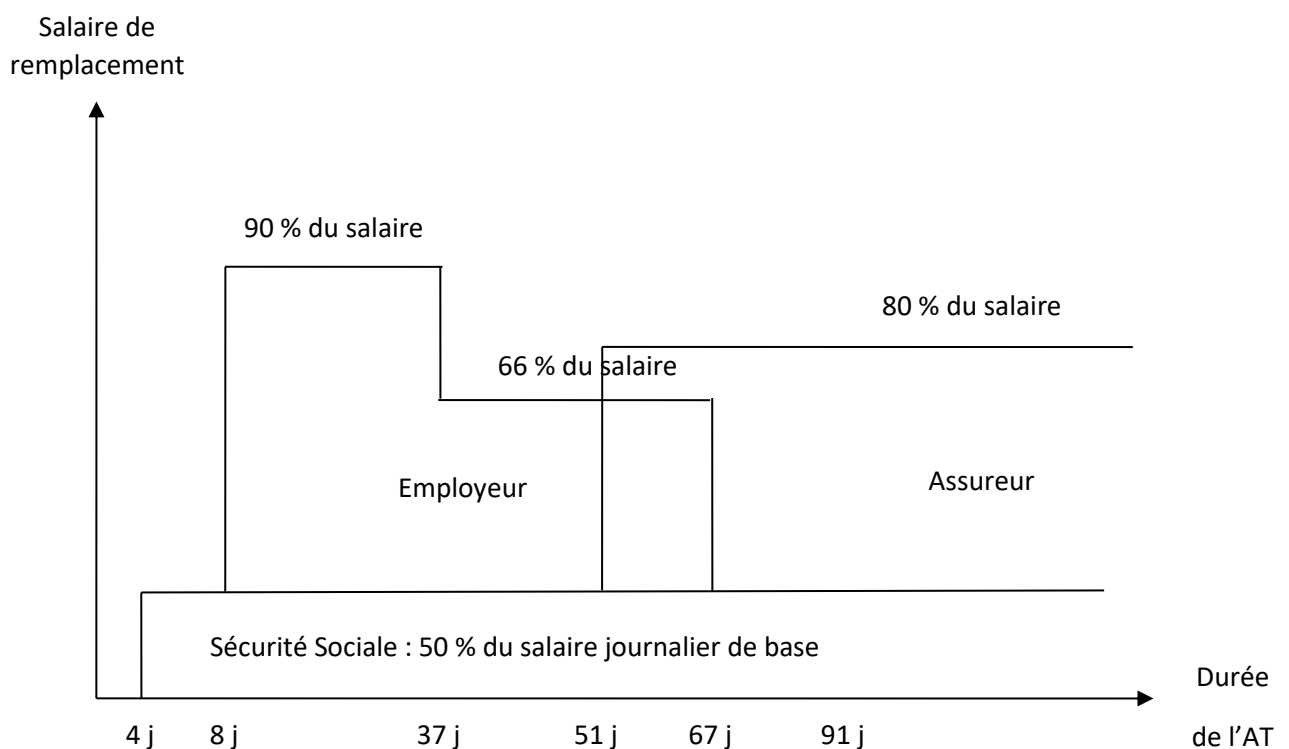
### Exemple général :

Maintenant que l'on a présenté les différents intervenants participant à l'indemnisation d'un individu en cas d'un arrêt de travail, il paraît judicieux d'illustrer l'indemnisation d'un arrêt de travail à travers un exemple :

Supposons qu'un individu soit un salarié assuré par la Sécurité Sociale. Le salarié a une ancienneté de 4 ans, et son salaire brut mensuel brut ces 3 dernières années est de 2 000 euros.

L'entreprise a souscrit un contrat d'incapacité / invalidité avec une franchise continue absolue de 50 jours et dont le montant de la garantie est de 80 % du salaire brut en prenant en compte la Sécurité Sociale.

Le salaire de remplacement du salarié sera le suivant (Figure 1.2) :



**Figure 1.2 : Exemple de salaire de remplacement d'un salarié**

Comme on l'a vu dans l'exemple du calcul du versement de l'IJ par la Sécurité Sociale, le salarié recevra de la part de la Sécurité Sociale : 50 % de son salaire journalier de base qui est calculé sur la moyenne des salaires bruts soumis à cotisations des trois derniers mois travaillés précédant l'arrêt de travail :

$$50\% * 2\,000 * 3 / 91,25 = 32,87 \text{ €}$$

Ceci est inférieur au montant maximal de l'IJ versé par la Sécurité Sociale : 45,01 €.

Le délai de carence de la Sécurité Sociale est de 3 jours.

Le salarié a une ancienneté de 4 ans. Par conséquent l'employeur a l'obligation de verser à son salarié, en prenant en compte les indemnités versés par la Sécurité Sociale, l'indemnité suivante :

- 90 % de son salaire pendant les 30 premiers jours ;
- 66,66 % de son salaire pendant les 30 jours suivants.

Le délai de carence étant de 7 jours.

L'entreprise a souscrit un contrat d'incapacité / invalidité avec une franchise continue absolue de 50 jours et dont le montant de la garantie est de 80 % du salaire brut – Sécurité Sociale.

### **La décomposition du risque incapacité**

Le risque incapacité se décompose en 2 risques :

- le risque d'entrée en incapacité : risque pour un assuré d'entrer en arrêt de travail.
- le risque de maintien en incapacité : risque pour un assuré, entré en incapacité, de le rester.

Cette décomposition a pour but de séparer le risque lié à la fréquence des arrêts de travail (risque d'entrée en incapacité) du risque lié à la durée des arrêts de travail (risque de maintien en incapacité).

Dans la suite de ce mémoire, on ne s'intéressera qu'au risque de maintien en incapacité.

## **1.4 Le provisionnement des garanties incapacité**

Tout organisme assureur, habilité à mettre en œuvre un contrat de prévoyance, est soumis à une réglementation concernant la constitution de ses provisions pour ses garanties incapacité.

Avant la loi Evin du 31 décembre 1989, la réglementation en matière de provisionnement n'était pas homogène parmi les trois familles d'organismes assureurs habilités à mettre en œuvre un contrat de protection sociale complémentaire.

L'article 7 de la loi Evin est venu harmoniser la réglementation en introduisant une obligation de provisionnement : « 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 a ensuite précisé 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 ; »

L'arrêté du 28 mars 1996 a fixé de nouvelles règles de provisionnement pour les garanties incapacité et invalidité. Cet arrêté a instauré des tables réglementaires servant de base pour le calcul des provisions, avec la possibilité d'utiliser des tables alternatives dites « tables d'expériences », dans le cas où ces dernières sont certifiées par un actuair e indépendant. Ces règles de provisionnement ont été revues en 2010 suite à l'arrêté du 24 décembre 2010. Les tables réglementaires ont alors été modifiées afin d'être plus en accord avec la sinistralité actuelle des arrêts de travail.

Elles ont été entièrement reconstruites en 2013, suite au report de la date de départ à la retraite, passant de 62 ans à 65 ans, afin de mieux prendre en compte les risques relatifs à ces âges, mais sont toujours non homologuées à ce jour.

L'article 143-12 du règlement ANC (Autorité des Normes Comptables) résume la législation en vigueur concernant le provisionnement des garanties incapacité :

« Le calcul des provisions techniques de prestations d'incapacité de travail et d'invalidité est effectué à partir des éléments suivants :

1. les lois de maintien en incapacité de travail et en invalidité indiquées aux articles 600-2, 600-3 et 600-4 du présent règlement. Toutefois, il est possible pour une entreprise d'assurances d'utiliser une loi de maintien établie par ses soins et certifiée par un actuair e indépendant de cette entreprise, agréé à cet effet par l'une des associations d'actuaires reconnues par l'Autorité de contrôle prudentiel et de résolution.
2. un taux d'actualisation qui ne peut excéder 75% du taux moyen au cours des vingt-quatre derniers mois des emprunts de l'Etat français, sans pouvoir dépasser 4,5% ;
3. dans le cas des rentes allouées au titre des accidents survenus à compter du 1er janvier 2013 et dont le montant est revalorisé en application de la loi n° 51-695 du 24 mai 1951 ou de la loi n° 74-1118 du 27 décembre 1974, un taux d'inflation égal à 2,25 %.

Ces dispositions ne s'appliquent pas aux prestations issues de contrats d'assurance de groupe souscrits par un établissement de crédit, ayant pour objet la garantie du remboursement d'un emprunt, ni à celles issues de contrats d'assurance couvrant des risques visés au 3° du premier alinéa de l'article L.310-1 du code des assurances. »

### **Les tables réglementaires**

Selon l'article 143-12 du règlement ANC, le calcul des provisions techniques de prestations d'incapacité de travail et d'invalidité est effectué à partir de lois de maintien en incapacité de

travail et en invalidité. La conception de ces lois de maintien sera présentée ultérieurement. Ces lois de maintien permettent d'aboutir à des tables. Ces tables sont au nombre de 3 : table de maintien en incapacité, table de maintien en invalidité et table de passage en invalidité et sont indiquées aux articles 600-2, 600-3 et 600-4 du règlement ANC. Ces tables ont été établis par le BCAC (Bureau Commun des Assurances Collectives) et constituent les tables réglementaires.

Nous ne présenterons que la table de maintien en incapacité car c'est la table assujettie à la certification dans notre étude.

### Table règlementaire de maintien en incapacité

La table règlementaire de maintien en incapacité est disponible à l'article 600-3 du règlement ANC.

La table règlementaire de maintien en incapacité (Figure 1.3) est une table à 2 dimensions (âge et mois d'ancienneté). Chaque ligne correspond à un âge et chaque colonne correspond à un mois d'ancienneté. Ci-dessous, nous avons un extrait de cette table.

Age	Mois																		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
23 ans ou moins	10000	2842	1743	1144	838	625	455	339	291	253	215	187	173	152	138	129	123	114	102
24	10000	2931	1848	1215	894	657	478	343	291	256	217	183	166	143	130	121	114	105	95
25	10000	3080	2001	1345	997	739	536	382	327	289	251	216	195	172	159	149	140	129	116
26	10000	3177	2112	1461	1087	812	591	431	372	325	285	249	226	201	186	171	161	150	137
27	10000	3251	2180	1540	1156	869	643	476	407	360	320	285	263	237	222	207	192	179	168
28	10000	3298	2243	1600	1209	915	688	524	448	400	359	322	297	270	255	238	222	210	199
29	10000	3348	2273	1640	1246	956	726	559	476	425	384	352	327	298	280	262	247	233	220
30	10000	3386	2275	1659	1264	964	744	583	494	439	396	363	338	308	287	267	252	240	227
31	10000	3388	2228	1618	1249	965	756	595	501	449	406	375	347	318	295	276	261	250	236
32	10000	3433	2238	1617	1254	975	772	612	522	468	421	388	357	325	302	279	264	252	235
33	10000	3466	2235	1627	1260	983	782	628	540	484	431	395	364	332	310	286	270	256	238
34	10000	3567	2298	1684	1321	1033	828	684	597	535	477	436	401	366	344	319	298	282	265
35	10000	3645	2331	1705	1357	1082	876	732	647	586	528	481	443	402	377	351	331	309	294
36	10000	3701	2390	1747	1390	1106	905	771	682	617	560	508	469	428	397	370	347	323	308
37	10000	3822	2458	1804	1430	1148	932	801	704	635	579	526	487	443	406	379	357	335	319
38	10000	3958	2526	1851	1479	1193	980	841	739	671	616	564	521	477	439	411	384	358	340
39	10000	4035	2600	1923	1541	1266	1055	915	807	739	680	623	572	530	486	455	427	400	381
40	10000	4073	2652	1973	1575	1303	1097	965	853	783	719	659	607	565	521	490	458	428	404
41	10000	4214	2776	2096	1680	1408	1193	1054	937	866	798	731	676	626	582	552	519	483	455
42	10000	4364	2930	2237	1814	1540	1314	1162	1039	971	895	825	764	710	666	630	593	553	521
43	10000	4473	3046	2341	1907	1633	1400	1243	1120	1045	965	892	830	774	726	691	654	614	582
44	10000	4621	3155	2417	1974	1676	1441	1282	1158	1077	1000	928	872	809	760	725	682	643	608
45	10000	4780	3318	2557	2097	1776	1529	1361	1240	1148	1069	1001	938	872	825	791	745	705	675
46	10000	4895	3392	2641	2190	1860	1609	1437	1319	1218	1132	1066	997	929	882	843	793	756	728

**Figure 1.3 : Extrait de la table règlementaire de maintien en incapacité**

La table de maintien en incapacité se lit de la façon suivante : Pour chaque âge et à 0 mois d'ancienneté en incapacité, on observe 10 000 individus qui viennent d'entrer en état d'incapacité. Pour chaque âge et pour chaque mois d'ancienneté en incapacité, on peut lire le nombre de personnes toujours en état d'incapacité.

Par exemple, à 27 ans au bout de 2 mois d'ancienneté en incapacité, on ne dénombre plus que 2 180 personnes en état d'incapacité sur les 10 000 individus observés au début de l'étude. On en déduit que la probabilité pour une personne de 27 ans d'être encore en état d'incapacité 2 mois après son entrée est de 21,8%.

Dans cette table, l'âge varie de 23 ans ou moins à 66 ans. Le nombre de mois varie de 0 à 36 mois car rappelons-le, l'état d'incapacité dure maximum 3 ans (3\*12 mois).

La base de données sur laquelle s'appuie le BCAC provient d'un panel d'organismes assureurs.

### Coefficients de provisionnement

Au titre du calcul de ses provisions techniques de prestations d'incapacité de travail et d'invalidité, l'organisme assureur doit constituer trois types de provisions :

- une provision pour rente d'incapacité en cours ;
- une provision pour rente d'invalidité en attente ;
- une provision pour rente d'invalidité en cours.

La provision pour rente d'incapacité en cours vise à provisionner le risque qu'un assuré en état d'incapacité se maintienne dans cet état. La provision pour rente d'invalidité en attente vise à provisionner le risque que cet assuré en état d'incapacité devienne invalide. Enfin, la provision pour rente d'invalidité en cours couvre le risque qu'un assuré invalide reste invalide.

Pour un assuré valide, aucun provisionnement n'est effectué.

A partir de la table de maintien en incapacité, l'organisme assureur peut calculer ses provisions pour rente d'incapacité en cours de service, c'est à dire une provision correspondant aux prestations d'incapacité à verser après l'expiration de l'exercice au titre des sinistres en cours au 31 décembre.

La formule est la suivante :

$$\begin{aligned}
 PM(x, anc) &= Coeff PM(x, anc) * \frac{I}{12} \\
 &= \frac{1}{2 * L_{x,anc}} \sum_{k=anc}^{35} \left[ \frac{L_{x,k}}{(1+i)^{(k-anc)/12}} + \frac{L_{x,k+1}}{(1+i)^{(k+1-anc)/12}} \right] * \frac{I}{12}
 \end{aligned}$$

Et où :

$i$  : taux d'actualisation

$x$  : âge de l'assuré à l'entrée dans l'état d'incapacité

$anc$  : ancienneté de l'assuré dans l'état d'incapacité (en mois)

$L_{x,k}$  : nombre d'individus entrés en incapacité à l'âge  $x$  et encore en incapacité après  $k$  mois

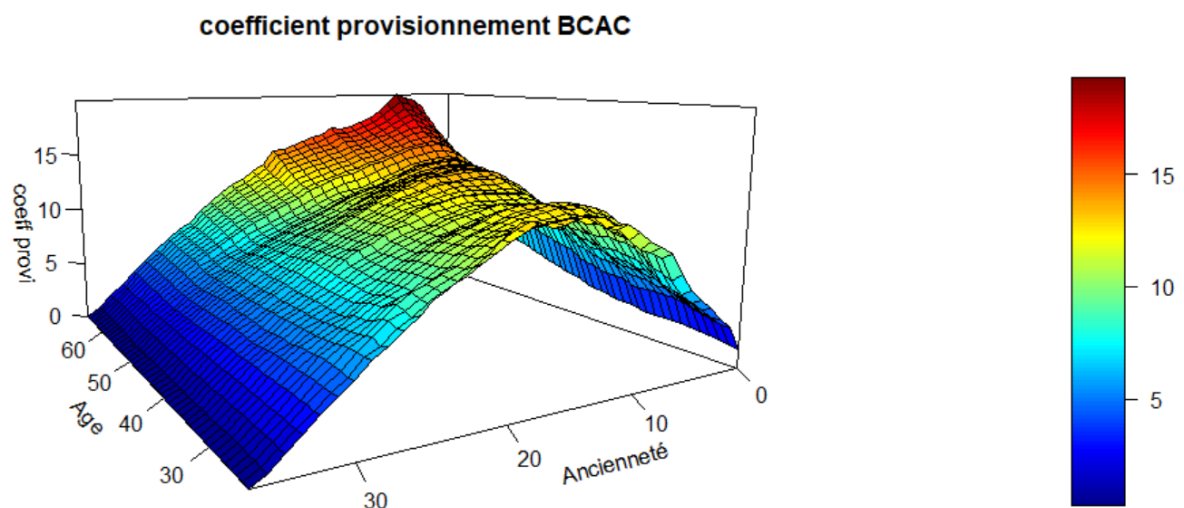
$I$  : indemnité à verser à l'assuré par an

L'organisme assureur s'appuie sur la table réglementaire pour déterminer l'ensemble des  $L_{x,k}$ , nécessaire au calcul de  $PM(x, anc)$ .

La formule représente la valeur actuelle probable de l'engagement de l'assureur vis-à-vis de la personne en incapacité.

Le coefficient de provision (*Coef*  $PM(x, k)$ ) pour un individu entré en incapacité à l'âge  $x$ , et présent dans l'état depuis  $k$  mois, représente le montant de la provision mathématique pour un euro de rente mensuel.

Le graphique suivant (Figure 1.4) représente les coefficients de provision selon la table du BCAC en vigueur, avec un taux d'actualisation nul.



**Figure 1.4 : Graphique des coefficients de provisionnement issu de la table du BCAC (taux d'actualisation nul)**

On constate que les coefficients de provisions sont les plus élevés, pour chaque âge d'entrée en incapacité, lorsque l'ancienneté est aux alentours de 18 mois, soit au milieu de la durée maximum de l'incapacité.

De par sa construction, la formule du coefficient de provisionnement prend en compte la durée potentielle de maintien dans l'état d'incapacité et la probabilité de sortir d'incapacité chaque mois.

Selon la table du BCAC, c'est aux anciennetés les plus faibles que les probabilités de sortir d'incapacité sont les plus fortes (par exemple, selon le BCAC la probabilité de sortir de l'état d'incapacité au bout d'un mois est de 59% pour un assuré de 40 ans et n'est plus que de 8% au bout de 8 mois d'ancienneté). Par conséquent, plus un assuré passe du temps en état d'incapacité, plus il a de chance de se maintenir en incapacité et donc plus il faut provisionner. Cependant, la durée maximum d'incapacité étant limité à 3 ans, plus l'assuré reste en état d'incapacité, moins sa durée d'indemnisation sera longue et donc moins l'organisme assureur aura à provisionner. D'où un pic de provisionnement atteint au milieu de la durée de maintien en incapacité maximum.

### **La table d'expérience de maintien en incapacité**

Selon la réglementation, l'organisme assureur peut se substituer à la table réglementaire en utilisant une table d'expérience « établie par ses soins et certifiée par un actuaire indépendant de cette entreprise, agréé à cet effet par l'une des associations d'actuaire reconnues par l'Autorité de contrôle prudentiel et de résolution. ».

L'utilisation d'une table d'expérience a plusieurs avantages.

#### Un risque sur mesure :

Premièrement, la table réglementaire du BCAC a été construite à partir d'observations provenant d'un large panel d'organismes assureurs. Etant donné que les populations assurées peuvent énormément varier selon l'organisme assureur ou encore le type de contrat proposé, la table réglementaire ne pourra jamais refléter exactement le risque auquel est soumis un contrat spécifique d'un organisme assureur. En d'autres termes, la table réglementaire représente un risque « moyen » mais qui ne peut être le plus adapté à une population spécifique contrairement à une table basée sur l'expérience.

#### Un positionnement stratégique :

De plus, si un organisme assureur utilise une table d'expérience alors que ses concurrents utilisent tous la table réglementaire, ce dernier peut ainsi se distinguer car il est capable de mieux provisionner et de tarifer plus justement.

#### Une segmentation possible :

Enfin, plusieurs variables peuvent expliquer des durées de maintien en incapacité différentes parmi les assurés d'un organisme assureur (par exemple la catégorie socio professionnelle).

La table de maintien en incapacité du BCAC ne prend en compte qu'une seule variable de ce type : l'âge. Avec une table d'expérience, on peut segmenter notre loi selon différentes variables pour être encore plus proche du risque auquel on est soumis. Un exemple classique est la segmentation selon l'origine de l'arrêt de travail (maladie ou accident). Très souvent, on observe une durée de maintien en incapacité significativement différente selon l'origine de l'arrêt de travail.

Quelque soit la volonté pour laquelle un organisme assureur souhaite utiliser une table d'expérience, il doit disposer d'une base de données suffisamment importante.

## 1.5 Le sujet de l'étude

L'objet de ce mémoire est de construire une table d'expérience de maintien en incapacité, en expliquant le rôle du certificateur tout au long du processus de construction.

La table d'expérience de maintien en incapacité est construite pour le provisionnement de rente d'incapacité en cours de service avec la volonté de segmenter cette table selon l'âge et le type d'arrêt de travail (maladie ou accident). En s'appuyant sur la qualification du certificateur, on expliquera si une telle table est certifiable.



## 2 Présentation des données

Les données utilisées pour ce mémoire sont celle d'une compagnie d'assurance française. Le périmètre de l'étude correspond à des individus ayant tous souscrit un contrat de prévoyance individuelle provenant de 3 produits différents proposant des garanties incapacité/invalidité.

La période d'observation est fixée du 01/01/2011 au 31/12/2015. Seules les personnes entrées en incapacité entre 21 et 70 ans font partie de l'étude.

### 2.1 Construction et contrôle de la base de données

Pour construire une table d'expérience, nous avons besoin préalablement de nous appuyer sur une base de données « propre », c'est-à-dire une base de données censée respecter certaines hypothèses statistiques et ayant subi différents contrôles de cohérence. Le rôle du certificateur intervient dès cette étape de construction de la base de données pour s'assurer de sa bonne construction.

Comme pour de nombreuses études statistiques, les modèles employés dans l'élaboration de tables d'expérience recourent à deux hypothèses fondamentales :

- l'indépendance des données individuelles ;
- le caractère d'homogénéité des données (identique distribution).

En pratique les données sont rarement homogènes néanmoins les facteurs d'hétérogénéité doivent être identifiés dans la mesure où ils feront l'objet d'un suivi. Par exemple, le sexe d'un assuré peut être un facteur d'hétérogénéité en ayant un impact sur la durée de maintien en incapacité. Or si on construit une table d'expérience sans prendre en compte ce facteur d'hétérogénéité, on doit alors faire attention à ce que le ratio homme/femme de nos assurés ne varie pas trop dans l'avenir.

#### **Détermination du champ d'utilisation de la loi d'expérience**

La première étape de la construction de notre base de données est la sélection des produits d'assurances sur lesquelles la table va être construite, en apportant une attention particulière aux conditions générales et particulières des contrats car ils sont une source d'hétérogénéité. En effet, le type de franchise, la sélection médicale ou encore le délai de carence sont des exemples d'hétérogénéité potentiel.

### Présentation des produits :

Les contrats de type Arrêt de travail provenant de 3 produits d'assurance sont retenus. Ces produits seront notés A, B et C.

En analysant les conditions générales et particulières ainsi que les notices d'information des 3 produits, on observe que les conditions d'exclusions pour les 3 produits sont les mêmes. On s'assure par la même occasion que ces 3 produits couvrent bien les mêmes risques (dans notre cas le risque de maintien en incapacité pour la garantie de type « Arrêt de travail » pour cause d'accident et de maladie).

Cependant, on constate un large panel de choix pour les assurés que ce soit pour le montant des garanties ou encore pour le choix de la franchise, également source d'hétérogénéité.

### **Contrôles de premier rang**

Après avoir ciblé et analysé les produits qui seront potentiellement provisionnés par une table d'expérience et non selon la table du BCAC, le rôle du certificateur est de comprendre la structure informatique de production des bases de données de l'assureur. Le but est de mettre en lumière les pratiques et les événements qui peuvent, a priori, impacter la qualité de la base ou conduire à des données incomplètes tel que :

- le rachat de portefeuilles
- la fusion d'entités
- les changements de systèmes de gestion
- l'évolution des processus de gestion.

En général, l'organisme assureur met à disposition du certificateur un document exposant le schéma global de production des bases de données ainsi que les événements pouvant impacter les bases de données.

Dans le cadre de notre étude, nous avons vérifié qu'aucun rachat de portefeuilles ni de fusion d'entités ne s'était produit. Suite à cette vérification, nous obtenons de par la structure informatique de production des bases de données, 2 bases qui vont nous servir pour la suite : la base des sinistres et la base des contrats.

La base des sinistres contient 9 592 859 lignes et 91 variables dont notamment les variables suivantes :

Nature de la variable
N° de sinistre
Identifiant assuré
N° de contrat
Sexe
Date de naissance
CSP
Code Produit prévoyance
Choix franchise
Année d'exercice
Date survenance sinistre
Date Arrêt de travail
Date début d'indemnisation
Date fin d'indemnisation
IJ (Indemnité journalière)
Montant indemnisé total
Nombre de jour indemnisé
Motif sortie IJ
Taux d'indemnisation

**Figure 2.1 : Extrait des variables de la base « sinistre »**

La base des contrats contient 4 023 849 lignes et 63 variables dont notamment les variables suivantes :

Nature de la variable
Identifiant assuré
N° de contrat
Sexe
Date de naissance
CSP
Garantie (clos ou en cours)
Date garantie clôture
Produit prévoyance
Choix franchise
Choix montant IJ
Montant cotisation

**Figure 2.2 : Extrait des variables de la base « contrat »**

La première étape de construction a été de limiter les deux bases (sinistre et contrat) au périmètre de notre étude, c'est-à-dire que seuls les assurés ayant souscrit une garantie « Arrêt de travail » provenant des produits A, B et C ont été conservés, en prenant l'année de vue 2015.

La base « sinistre » et la base « contrat » ont ensuite été jointes grâce à la variable « Identifiant assuré » formant ainsi une base de 1 035 838 lignes.

### **Contrôle de cohérence interne**

Suite à la jointure de nos 2 bases (sinistre et contrat), le rôle du certificateur vise à auditer l'ensemble des contrôles de cohérence interne réalisés par l'organisme assureur. Ces contrôles visent à s'assurer de la cohérence interne de la base de données constituée. Ils portent a minima sur :

- l'exhaustivité des données ;
- la bonne chronologie des dates dans les données individuelles ;
- la vraisemblance des informations contenues dans les bases (primes en rapport avec les garanties, etc.).

Les données exploitées sont soumises aux aléas de l'erreur humaine et des anomalies informatiques mais aussi des différentes migrations de données lors de changement de système informatique. Il est donc important de mettre en évidence ces erreurs et de les éliminer.

Les contrôles interne suivants ont été effectués sur notre base :

#### 1) Calcul pour retrouver des variables manquantes :

La première étape de contrôle de cohérence interne a été de chercher à retrouver le plus grand nombre d'information manquante afin d'obtenir des données exhaustives. Certaines variables sont liées entre elles et il est parfois possible de retrouver la valeur d'une variable si on dispose de la valeur de différentes variables liées à elle.

Par exemple, si pour une ligne de notre base de données, on ne dispose pas de la durée d'indemnisation mais que l'on dispose de l'ensemble des autres variables, alors on peut retrouver la durée d'indemnisation de deux façons :

- $durée\ d'indemnisation = date\ fin\ indemnisation - date\ début\ indemnisation + 1.$
- $durée\ d'indemnisation = montant\ cumulé\ IJ / montant\ IJ$

La date de début d'indemnisation, si elle est manquante, peut se retrouver si on dispose de la date de survenance du sinistre et de la franchise en calculant :

- $date\ début\ d'indemnisation = date\ de\ survenance\ sinistre + durée\ franchise$

Certaines dates de naissance manquantes ont pu être retrouvées à l'aide d'une base de données indépendante contenant la date de naissance des assurés.

Ces calculs pour retrouver des variables manquantes sont également utiles pour repérer des données aberrantes. En effet, si pour une ligne de notre base de données, on dispose de toutes les informations mais que

- $durée\ d'indemnisation \neq date\ fin\ indemnisation - date\ début\ indemnisation + 1$  ; alors on est en présence d'une donnée aberrante qui devra être supprimé ou corrigé. La suppression de l'ensemble des données aberrantes de notre base de données constitue la seconde étape de nos contrôles de cohérence interne.

## 2) Suppression de données aberrantes :

Les contrôles visant à supprimer les données aberrantes ainsi que le pourcentage d'anomalies retrouvés sont présentés dans le tableau suivant :

<b>Controles</b>	<b>% d'anomalies</b>
$date\ début\ indemnisation < date\ survenance\ sinistre$	1,62%
$durée\ d'indemnisation \neq date\ fin\ indemnisation - date\ début\ indemnisation + 1.$	0,00%
$durée\ d'indemnisation \neq montant\ cumulé\ IJ / montant\ IJ$	2,36%
$date\ début\ d'indemnisation \neq date\ survenance\ sinistre + durée\ franchise$	3,18%
$durée\ d'indemnisation > 3\ ans$	0,00%
$date\ fin\ indemnisation < date\ début\ indemnisation$	0,00%
$date\ début\ indemnisation > date\ clôture\ garantie$	0,00%
$date\ fin\ indemnisation > date\ clôture\ garantie$	0,43%
$montant\ IJ > 500\ euros^*$	0,54%

**Figure 2.3 : Tableau représentant le % d'anomalies à chaque contrôle**

\*Les sinistres dont les montants d'indemnités journalières sont supérieurs à 500 euros sont considérés comme des cas spécifiques et sont retirés de la base de données.

## 3) Suppression des observations dont la date de clôture de garantie ou de fin d'indemnisation est inférieur au 01/01/2008

La durée d'incapacité étant de 3 ans maximum et notre étude se basant sur les IJ payées entre le 01/01/2011 et le 31/12/2015, les sinistres dont la date de clôture de garantie ou de fin d'indemnisation est inférieure au 01/01/2008 n'entrent donc pas dans notre étude.

4) Suppression des dates de naissance manquantes :

On supprime chaque ligne de la base de données où l'on ne dispose pas de la date de naissance car c'est une information indispensable pour la suite de nos travaux

5) Suppression des observations d'âge non compris entre 21 et 70 ans :

L'organisme assureur souhaite construire une table d'expérience pour des personnes dont l'âge d'entrée en incapacité se situe entre 21 et 70 ans.

6) Transformation de sinistres en « rechute »

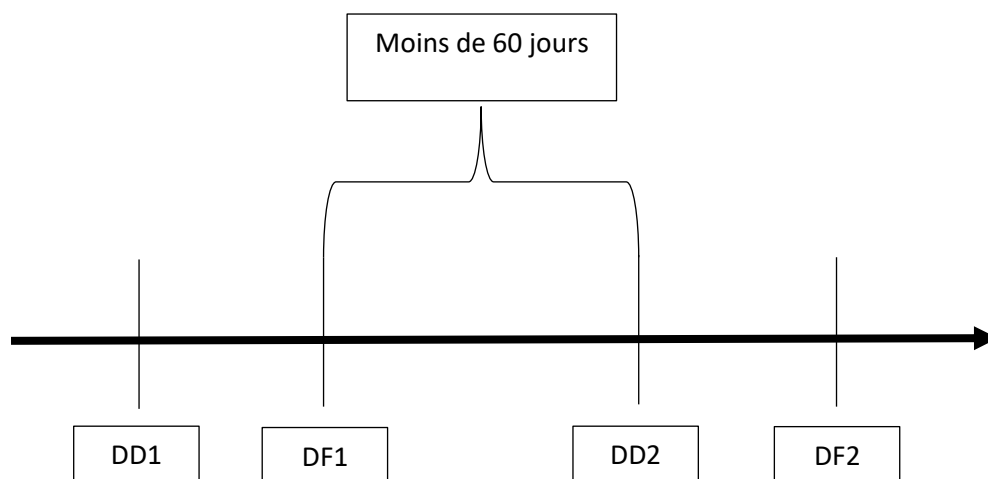
La rechute, au sens de la Sécurité Sociale, correspond soit à une aggravation spontanée de la lésion après consolidation, soit à l'apparition d'une nouvelle lésion après guérison, en dehors de toute cause extérieure. Dans tous les cas, la rechute n'est pas considérée comme un nouvel accident du travail mais un prolongement de l'accident du travail initial. Par conséquent, la durée d'incapacité causé par la rechute s'ajoute à la durée d'incapacité à l'origine de l'arrêt de travail.

Très souvent, un individu « rechute » peu de temps après être sorti de son état d'incapacité. Dans notre base de données, nous n'avons pas l'information sur la rechute. On considère pour la suite qu'une rechute correspond à un assuré ayant au moins 2 sinistres et que le délai entre le dernier jour indemnisé d'un sinistre et le 1<sup>er</sup> jour indemnisé du sinistre suivant est inférieur à 60 jours (Figure 2.4). On suppose alors qu'on fait face à un seul sinistre et on modifie la date de fin d'indemnisation pour rendre cohérente la transformation effectuée.

Soit « sinistre 1 » le 1<sup>er</sup> sinistre de l'assuré dont les dates de début et de fin d'indemnisation sont « DD1 » et « DF1 ». Soit « sinistre 2 » le 2<sup>nd</sup> sinistre de l'assuré (survenu moins de 60 jours après la fin du 1<sup>er</sup> sinistre) dont les dates de début et de fin d'indemnisation sont « DD2 » et « DF2 », alors on considère qu'on fait face à un seul sinistre où:

$$\text{Date début d'indemnisation} = DD1$$

$$\text{Date fin d'indemnisation} = DF2 - (DD2 - DF1)$$



**Figure 2.4 : Graphique expliquant une rechute**

Ce retraitement n'est pas le plus optimal car dans certains cas il s'agit bien de sinistres différents mais dans la majorité des cas il est le plus approprié d'où ce choix. Par conséquent, cela nous oblige à calculer une nouvelle date de fin d'indemnisation. Cela aboutit à avoir moins de sinistres dans notre base mais d'une durée plus longue. Etant donné que nous sommes dans un cadre où nous construisons une table de maintien en incapacité, cette transformation de sinistres est « prudente » car nous allongeons la durée de maintien en incapacité. On rappelle qu'un des rôles importants du certificateur est de s'assurer de la relative prudence de la table construite. Par conséquent, cette transformation de sinistres d'un assuré en un unique sinistre ne pose pas de problème du point de vue du certificateur.

#### 7) Suppression des doublons :

Sont considéré comme doublons deux lignes de données ayant le même numéro de sinistre. Afin de ne pas prendre en compte deux fois le même sinistre, on supprime les doublons.

L'ensemble de ces contrôles, effectués sous SAS, aboutissent à une base de données de 134 562 lignes où chaque ligne correspond à un sinistre différent unique.

### **Contrôle de cohérence externe**

Suite aux contrôles de cohérence interne, la base de données construite est censée ne plus comprendre d'anomalies et être exhaustive. Les contrôles de cohérence externe consistent à comparer les informations contenues dans la base de données avec d'autres informations de l'organisme assureur relatives aux portefeuilles considérés tel que :

- des éléments issus de l'outil de gestion ;
- des éléments comptables (prestations, nombre de contrats,...) ;
- un tableau de bord de la production mensuelle ;
- des études internes antérieures.

Ces contrôles externes permettent de qualifier la base de données et de s'assurer de son exhaustivité. Le rôle du certificateur lors de cette étape vise à s'assurer que les contrôles de cohérence externe n'aboutissent pas à des incohérences. Il peut, par exemple, demander à auditer des données qui lui semblent incohérentes. Par exemple, dans notre base de données, il existe un sinistre où la franchise est de 0.6 jours, qui est irrationnel selon le portefeuille assuré. Cette donnée aberrante aurait pu être retrouvée si un contrôle de cohérence interne adéquat avait été effectué. Grâce au travail du certificateur, ce sinistre a été audité et l'erreur a pu être annulée par l'organisme assureur qui s'est appuyé sur son outil de gestion.

### **Contrôle de cohérence temporelle**

Les contrôles de cohérence temporelle sont assez similaires aux contrôles de cohérence externe mais ne sont réalisables uniquement que si des travaux analogues ont été effectués. Les contrôles de cohérence temporelle comparent sur une plage de temps commune les données issues de deux bases s'appuyant sur des travaux similaires. Par exemple, si une précédente table d'expérience de maintien en incapacité a été produite par l'organisme assureur à partir d'une base s'appuyant sur des données de la période 2009-2013, alors on peut effectuer des contrôles de cohérence temporelle avec notre base de données s'appuyant sur la période 2011-2015. En effet, la plage de temps 2011-2013 est commune à ces 2 études.

Les contrôles de cohérence temporelle sont généralement les derniers contrôles effectués sur la base de données. D'une manière générale, l'ensemble de ces contrôles doivent être recensés, documentés et expliqués par l'organisme assureur afin que le certificateur puisse les auditer et s'assurer de la pertinence et de la relative prudence de la base de données construite.



## 2.2 Statistiques sur la base de données

Cette partie a pour but de présenter une série de statistiques descriptives de notre base afin d'en avoir une connaissance plus précise. L'intérêt de cette partie réside également dans le fait de mettre en avant des variables qui sont potentiellement source d'hétérogénéité. Pour ce faire, nous utiliserons souvent le test du Khi 2 d'indépendance.

### 2.2.1 Test du Khi 2 d'indépendance

Le test d'indépendance du Khi 2 permet de contrôler l'indépendance de deux caractères dans une population donnée.

Soient 2 variables  $X$  et  $Y$ , les valeurs possibles de  $X$  sont réparties en  $l$  classes  $A_1, \dots, A_l$ , celles de  $Y$  en  $c$  classes  $B_1, \dots, B_c$ .

Soient  $n$  observations réparties en  $l * c$  effectifs observés :  $n_{i,j}$  observations ont donné à la fois  $A_i$  et  $B_j$ , avec donc  $\sum_{i,j} n_{i,j} = n$ .

L'hypothèse testée ici est ( $H_0$ ) : "Les variables  $X$  et  $Y$  sont indépendantes".

Déroulement du test :

On crée le tableau des effectifs qui est un tableau à double-entrée. A l'intersection de la  $i$ -ème ligne et de la  $j$ -ième colonne, on écrit l'effectif  $n_{i,j}$ .

On "borde" le tableau pour calculer les effectifs marginaux :  $L_i$  est la somme des termes sur la  $i$ -ème ligne,  $C_j$  est la somme des termes sur la  $j$ -ième colonne.

		$j$		
	...	...	...	...
$i$	...	$n_{i,j}$	...	$L_i$
	...	...	...	...
		$C_j$		

On calcule les effectifs théoriques (ceux que l'on s'attend à rencontrer si  $X$  et  $Y$  étaient indépendantes) :

$$e_{i,j} = \frac{L_i * C_j}{n}$$

On calcule la valeur de la variable de test :

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}$$

On cherche la valeur critique  $\chi_a^2$  dans la table de la loi du chi-2 à  $(l - 1) * (c - 1)$  degrés de liberté.

Si  $\chi^2 < \chi_a^2$ , on accepte l'hypothèse  $H_0$ , sinon on la rejette.

En pratique, on s'appuie sur la p-value, probabilité de rejeter l'hypothèse nulle sachant qu'elle est vraie, pour rejeter  $H_0$  ou non. On fixe généralement un seuil de significativité  $\alpha$  de 5% :

- Si  $p - value < \alpha$ , on rejette  $H_0$  ;
- Si  $p - value > \alpha$ , on accepte  $H_0$ .

Vérification a posteriori des conditions d'application : il faut que les effectifs théoriques vérifient  $e_{i,j} \geq 5$  pour tous  $i, j$ .

Exemple :

Le tableau suivant (Figure 2.5) représente les salaires d'hommes et de femmes pour un total de 575 personnes

Salaire	1000 à 2000€	2000 à 3000€	3000 à 4000€	4000 à 5000€	TOTAL
Hommes	50	70	110	60	290
Femmes	80	75	100	30	285
Total	110	145	210	110	575

**Figure 2.5 : Tableau représentant le salaire d'hommes et de femmes**

L'objectif du test du Khi 2 d'indépendance dans cet exemple est de vérifier si les hommes et les femmes ont effectivement le même salaire (hypothèse nulle  $H_0$ ) ou si, au contraire, leurs salaires diffèrent.

Hypothèse nulle ( $H_0$ ) : le fait de connaître le sexe ne permet pas d'aider à deviner la tranche salariale d'un individu et inversement.

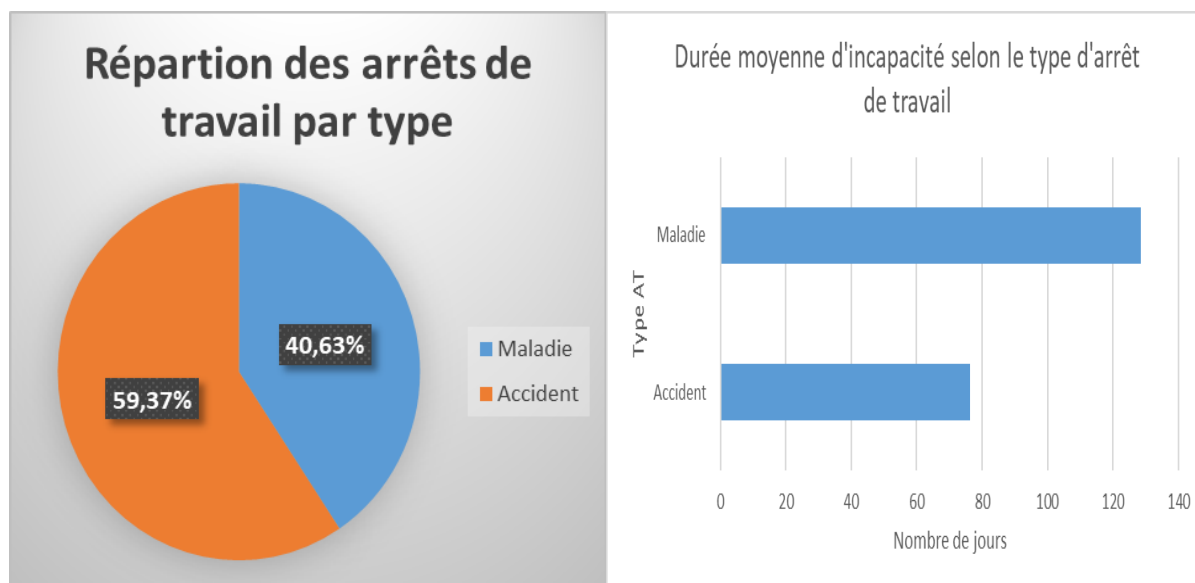
Si l'hypothèse nulle est rejetée, alors on a une relation sexe-salaire qui indiquera ici que les femmes sont moins bien rémunérées.

La p-value pour cet exemple est de 0.0005 : il y a donc un lien statistique entre le sexe et la tranche salariale car la p-value est très petite.

## 2.2.2 Etudes statistiques des variables

### Répartition et durée moyenne de maintien en incapacité selon le type d'arrêt de travail :

En figure 2.6 et 2.7, nous pouvons observer la répartition des arrêts de travail par type et la durée moyenne d'incapacité selon le type d'arrêt de travail.



**Figure 2.6 : Répartition des arrêts de travail par type**

**Figure 2.7 : Durée moyenne d'incapacité selon le type d'arrêt de travail**

Plus de la moitié des arrêts de travail de notre base de sinistre a pour cause l'accident (59,37%).

La durée moyenne d'incapacité suite à un arrêt de travail de type « Maladie » est beaucoup plus longue que pour un arrêt de travail de type « Accident ». Le type d'arrêt de travail semble être source d'hétérogénéité. La segmentation selon le type d'arrêt de travail apparaît donc logique. Un test du khi 2 d'indépendance selon le critère de Pearson a été effectué pour confirmer cette tendance.

La durée de maintien en incapacité étant une variable continue, nous avons dû construire des classes de durée de maintien en incapacité pour effectuer un test du Khi 2 d'indépendance, aboutissant ainsi au tableau suivant :

Durée de maintien (jours)	<=5	]5 ; 15]	]15 ; 30]	]30 ; 45]	]45 ; 60]	]60 ; 75]	]75 ; 90]	]90 ; 105]	]105 ; 120]	>120
Acc	2 125	15 891	16 546	10 532	6 405	4 625	3 181	2 668	1 721	11 914
Mal	2 658	6 348	10 321	8 120	4 905	3 871	2 646	2 584	1 665	16 022

**Figure 2.8 : Classes de durée de maintien en incapacité selon le type d'arrêt de travail**

Le résultat du test obtenu grâce au logiciel R est le suivant :

```

Pearson's Chi-squared test

data:  tableau
X-squared = 4890.4, df = 9, p-value < 2.2e-16

```

**Figure 2.9: Sortie graphique résultant du tableau 2.8**

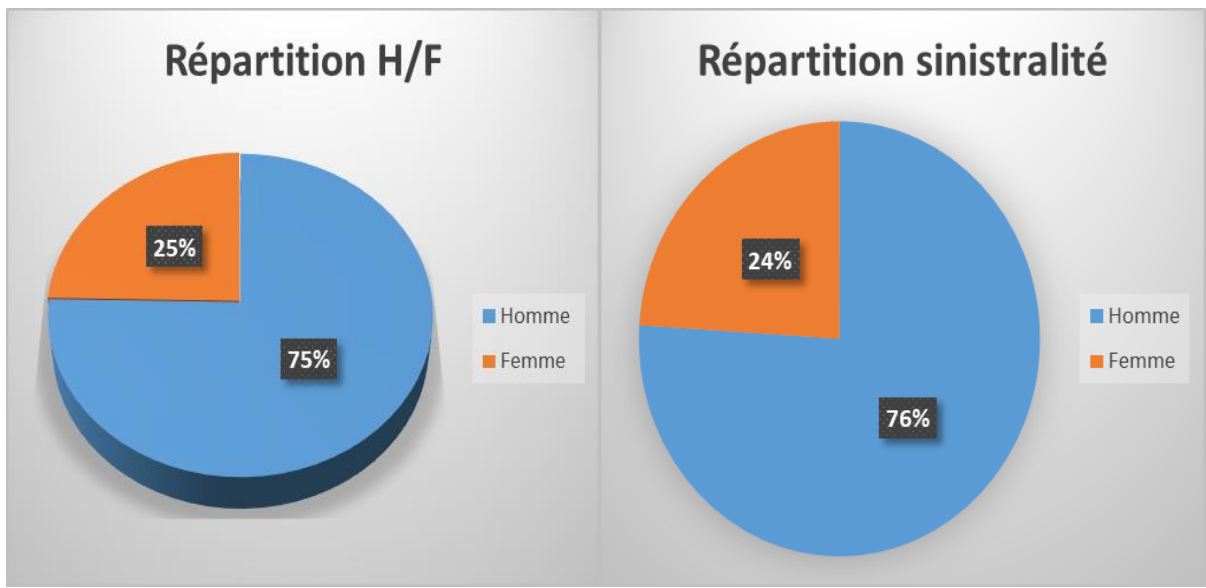
La p-value de ce test est extrêmement faible et confirme le lien entre la durée de maintien en incapacité et le type d'arrêt de travail.

Le tableau en figure 2.8 permet d'observer que les arrêts de travail de type « accident » sont beaucoup plus nombreux que les arrêts de travail de type « maladie » lorsque la durée d'arrêt se situe entre 5 et 30 jours. En revanche pour les sinistres de longue durée (>120 jours), les arrêts de travail de type « maladie » sont plus nombreux que les arrêts de travail de type « accident ».

Par la suite, on s'intéresse uniquement à la construction de la table de maintien en incapacité pour la cause « Accident ». Les statistiques présentés ci-dessous ne prennent en compte que les arrêts de travail de type « Accident ». La base de sinistre limité à l'arrêt de travail de type « accident » est composé de 80 001 sinistres.

#### Répartition des assurés homme/femme et de leur sinistralité :

En figure 2.10 et 2.11, on représente respectivement la répartition Homme/Femme de nos assurés et la répartition de leur sinistralité.



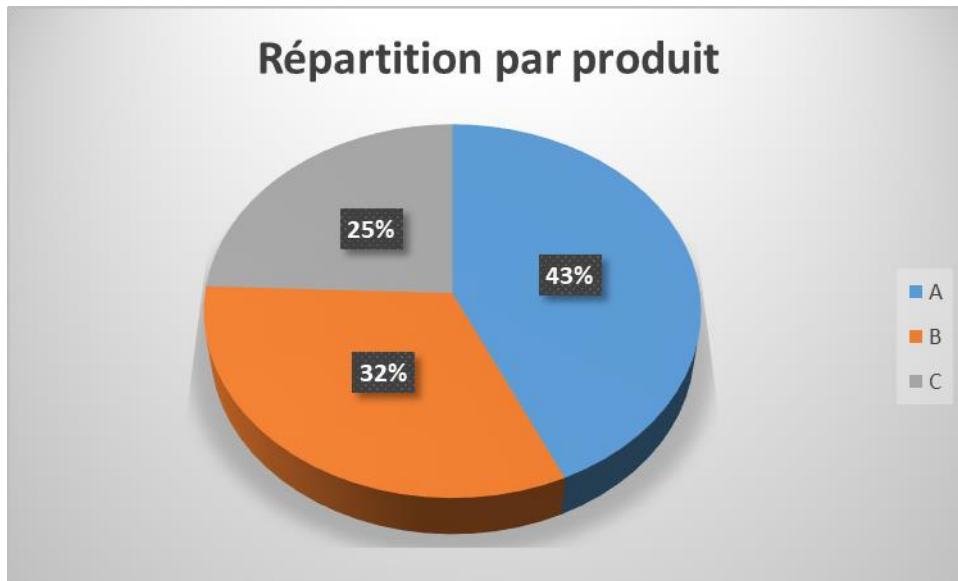
**Figure 2.10 : Répartition Homme/Femme des assurés**

**Figure 2.11 : Répartition de la sinistralité selon le sexe**

Selon la base contrat, environ 75% de nos assurés sont des hommes et 25% des femmes. Les pourcentages sont à peu près équivalents concernant la répartition de leur sinistralité dans notre base de sinistre. Un test du Khi 2 d'indépendance confirme que le sexe n'a pas d'influence sur l'entrée en incapacité ( $p - value = 0,2315$ ).

Répartition par produit :

La répartition des assurés, par produit, selon la base contrat est représentée figure 2.12.

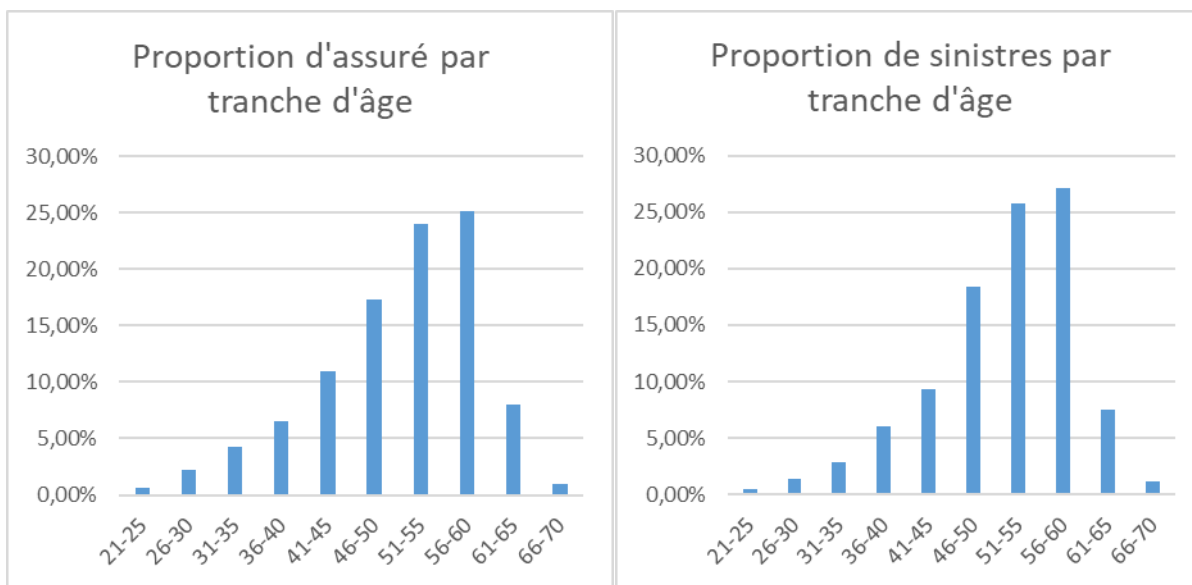


**Figure 2.12 : Répartition des assurés par produit**

Le produit d'assurance est une source d'hétérogénéité de nos données. En effet, la durée de maintien peut être différente selon le produit souscrit. Cependant, les produits A, B et C sont très modulable (par exemple quelque soit le produit, l'assuré peut choisir son niveau de franchise selon un large panel). En réalité, ces produits sont assez proches. Leur seule réelle distinction concerne les CSP qui souscrivent ces produits. Par conséquent, il paraît judicieux d'étudier la durée de maintien en incapacité selon la CSP sans distinguer le produit d'assurance souscrit.

Répartition des assurés et de leur sinistralité par tranche d'âge :

En figure 2.13 et 2.14, nous représentons respectivement la répartition des assurés et de leur sinistralité par tranche d'âge



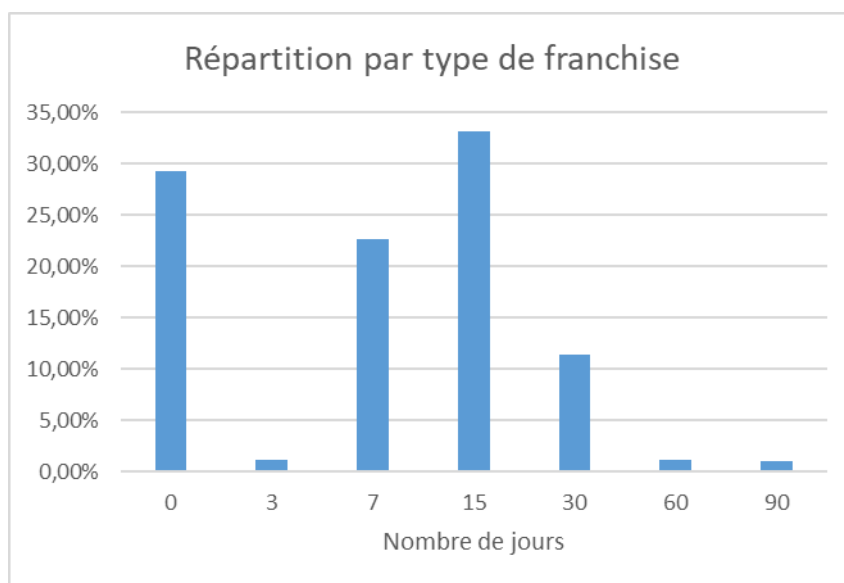
**Figure 2.13 : Proportion d'assuré par tranche d'âge**

**Figure 2.14 : Proportion de sinistres par tranche d'âge**

Les tranches d'âges extrême (21-25 et 66-70 ans) sont les moins représentées dans notre portefeuille d'assuré. En revanche les assurés entre 51 et 60 ans représentent presque la moitié de nos assurés. La proportion des sinistres est à peu près équivalente. Les assurés les plus vieux ont tendance à être légèrement sur-représentés dans la base de nos sinistres.

Répartition par niveau de franchise choisi :

En figure 2.15, on représente la répartition des assurés par type de franchise choisi.



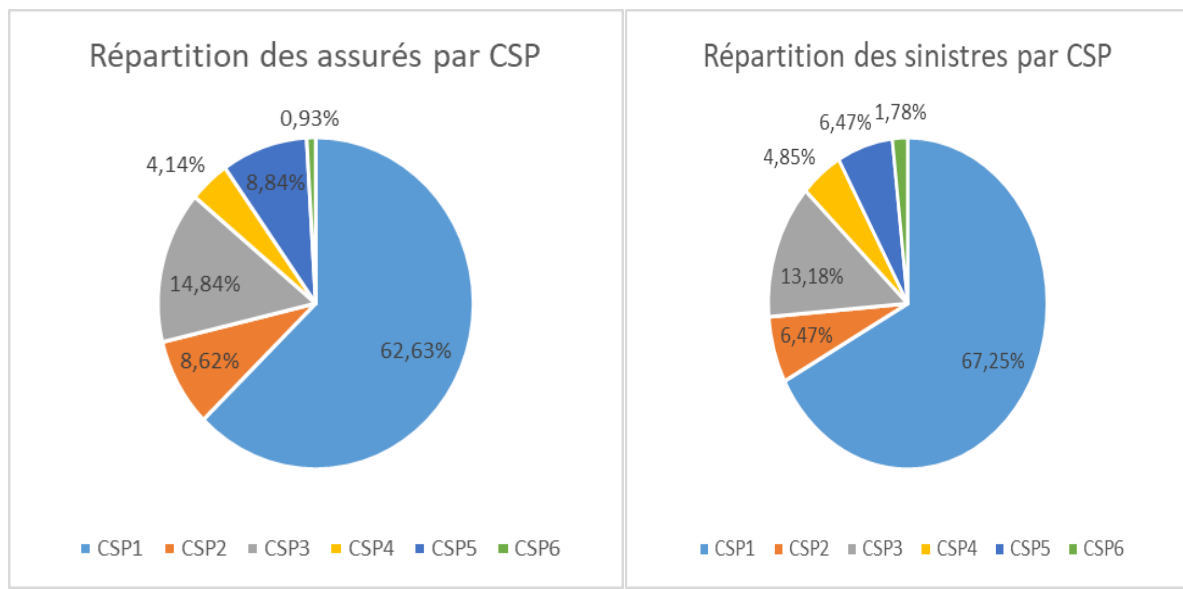
**Figure 2.15 : Répartition par type de franchise**

Pour chaque produit, l'assuré a le choix entre les niveaux de franchises suivants : 0, 3, 7, 15, 30, 60 et 90 jours.

Les niveaux de franchise les plus choisis par les assurés sont : 0, 7, 15 et 30 jours dans le portefeuille étudié.

Répartition des assurés et de leur sinistralité par catégorie socio professionnelle (CSP) :

Six catégories socio professionnelles sont représentées chez nos assurés, numérotées de 1 à 6. La répartition est donnée en figure 2.16 et 2.17.



**Figure 2.16 : Répartition des assurés par CSP**

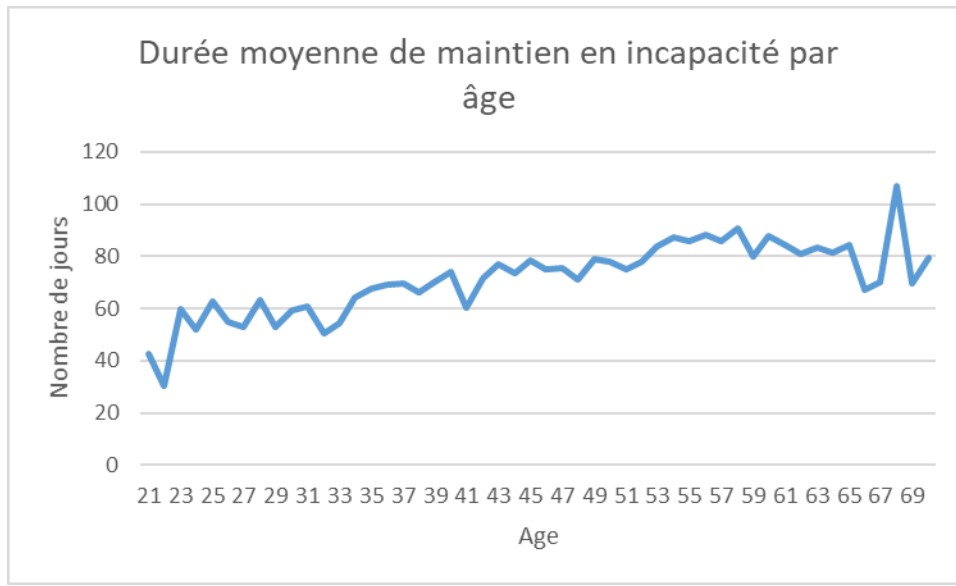
**Figure 2.17 : Répartition des sinistres par CSP**

La CSP1 est la CSP la plus représentée parmi nos assurés (environ 2/3 des assurés). Sa sinistralité est légèrement sur-représentée comparé aux autres CSP présentes dans notre base de sinistres.

Durée moyenne de maintien en incapacité en fonction de l'âge :

En figure 2.18 est représenté la durée moyenne de maintien en incapacité en fonction de l'âge.





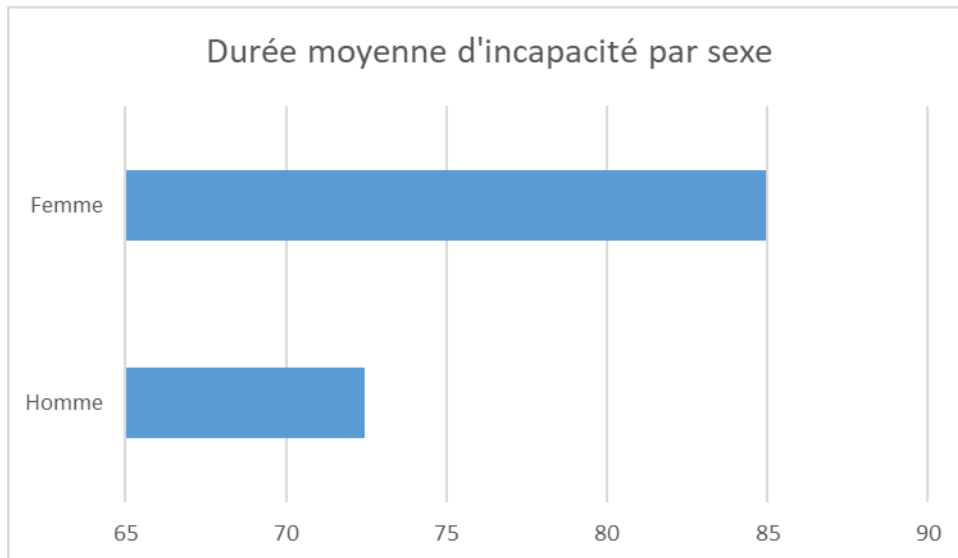
**Figure 2.18 : Durée moyenne de maintien en incapacité par âge**

Ce graphique met en évidence que l'âge a un impact sur la durée moyenne de maintien en incapacité. Plus l'âge est élevé, plus la durée moyenne d'indemnisation semble longue. Pour les assurés jeunes (de 21 à 35 ans), la durée moyenne d'incapacité ne semble pas excéder les 60 jours alors que pour les assurés les plus âgés (de 55 à 70 ans), la moyenne semble tourner autour de 80 jours. Enfin la durée moyenne d'incapacité paraît assez variable aux âges extrêmes. L'explication repose sans doute sur le faible échantillon de sinistres à ces âges rendant la moyenne assez volatile. En effet, selon le graphique 2.13 et le graphique 2.14 sur la répartition des assurés par tranche d'âge et sur la répartition des sinistres par tranche d'âge respectivement, la tranche d'âge 21-25 ans représente seulement 0,59% de nos assurés et 0,37% des sinistres, quant à la tranche d'âge 66-70 ans, elle représente 0,99% de nos assurés et 1,23% des sinistres.

La segmentation de notre table d'expérience selon l'âge semble donc évidente. Un test du Khi 2 d'indépendance a été effectué et la valeur extrêmement faible de la p-value confirme le lien entre durée de maintien en incapacité et l'âge d'entrée en incapacité.

Durée moyenne de maintien en incapacité en fonction du sexe :

En figure 2.19 est représenté la durée moyenne de maintien en incapacité en fonction du sexe.



**Figure 2.19 : Durée moyenne d'incapacité par sexe**

La durée moyenne de maintien en incapacité chez les hommes est plus faible que chez les femmes (72 jours contre 85 environ). Le sexe semble donc être un facteur d'hétérogénéité.

Le test du khi 2 d'indépendance a été effectué à partir du tableau suivant :

Durée de maintien (jours)	<=5	]5 ; 15]	]15 ; 30]	]30 ; 45]	]45 ; 60]	]60 ; 75]	]75 ; 90]	]90 ; 105]	]105 ; 120]	>120
Homme	1 835	14 154	14 579	9 165	5 512	3 866	2 642	2 216	1 446	9 565
Femme	380	2 700	2 967	2 012	1 260	1 019	694	611	377	3 001

**Figure 2.20 : Classes de durée de maintien en incapacité selon le sexe**

Le résultat est représenté en figure 2.21 :

```

Pearson's Chi-squared test

data: tableau
X-squared = 389.32, df = 9, p-value < 2.2e-16

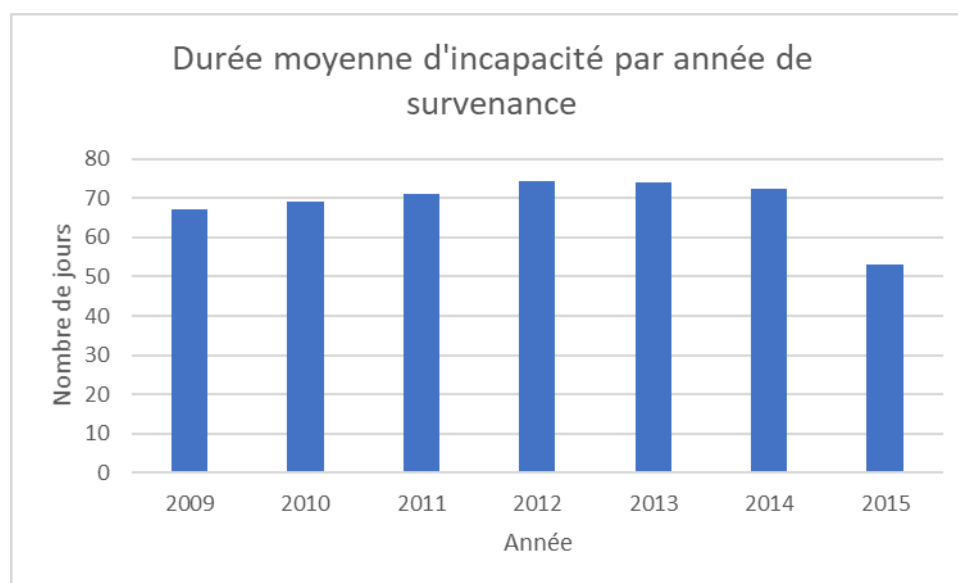
```

**Figure 2.21: Sortie graphique résultant du tableau 2.20**

Selon ce test, il existe bien un lien entre le sexe et la durée de maintien en incapacité.

### Durée moyenne de maintien en incapacité selon l'année de survenance :

En figure 2.22 est représenté la durée moyenne de maintien en incapacité par année de survenance.



**Figure 2.22 : Durée moyenne d'incapacité par année de survenance**

La durée moyenne de maintien en incapacité selon l'année de survenance est à peu près équivalente chaque année sauf pour la dernière année. Cela est logique étant donné que notre étude s'arrête le 31 décembre 2015. Par conséquent, des sinistres survenus en 2015 et toujours actifs ont dû être coupés au 31 décembre, d'où une durée moyenne plus faible cette année-là.

Un test du Khi 2 d'indépendance a été effectué à partir du tableau suivant :

Durée de maintien (jours)	<=5	]5 ; 15]	]15 ; 30]	]30 ; 45]	]45 ; 60]	]60 ; 75]	]75 ; 90]	]90 ; 105]	]105 ; 120]	>120
2009	401	3 235	3 113	1 882	1 157	815	503	440	255	1 783
2010	384	2 898	3 032	1 836	1 093	817	543	425	278	1 792
2011	374	2 756	2 814	1 713	1 058	714	513	440	292	1 757
2012	300	2 437	2 575	1 618	1 010	723	452	402	236	1 734
2013	274	2 245	2 340	1 567	862	655	474	401	235	1 653
2014	265	2 059	2 180	1 369	867	576	450	367	246	1 540
2015	333	1 825	1 971	1 353	816	595	388	312	220	963

**Figure 2.23 : Classes de durée de maintien en incapacité selon l'année de survenance**

Si on ne prend pas en compte la dernière année (2015) le résultat est le suivant :

Pearson's Chi-squared test

data: tableau

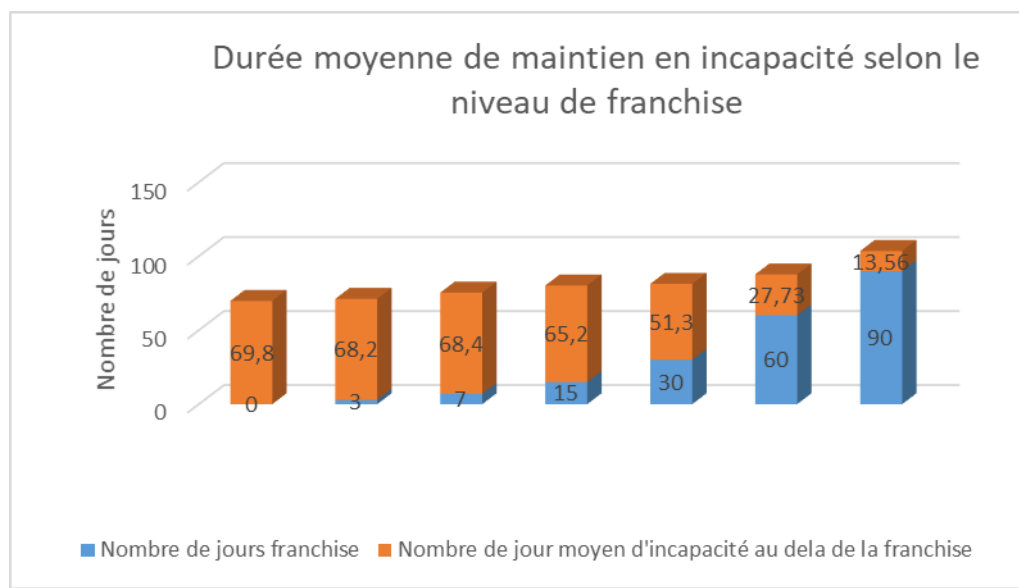
X-squared = 139.59, df = 45, p-value = 1.286e-11

**Figure 2.24: Sortie graphique résultant du tableau 2.23**

Selon le test, l'année de survenance a un impact sur la durée de maintien en incapacité. L'impact semble être toutefois plus modéré qu'avec le type d'arrêt de travail par exemple.

Durée moyenne de maintien en incapacité selon la franchise :

En figure 2.25 est représenté la durée moyenne de maintien en incapacité selon le niveau de franchise.



**Figure 2.25 : Durée moyenne de maintien en incapacité selon le niveau de franchise**

La durée moyenne de maintien en incapacité augmente avec le niveau de franchise. Cette observation peut s'expliquer de 2 façons différentes. La première est que les assurés, étant libre de choisir leur niveau de franchise, sont sans doute les mieux placés pour déterminer la durée possible de leur arrêt de travail. Par exemple, des assurés âgés sont sans doute

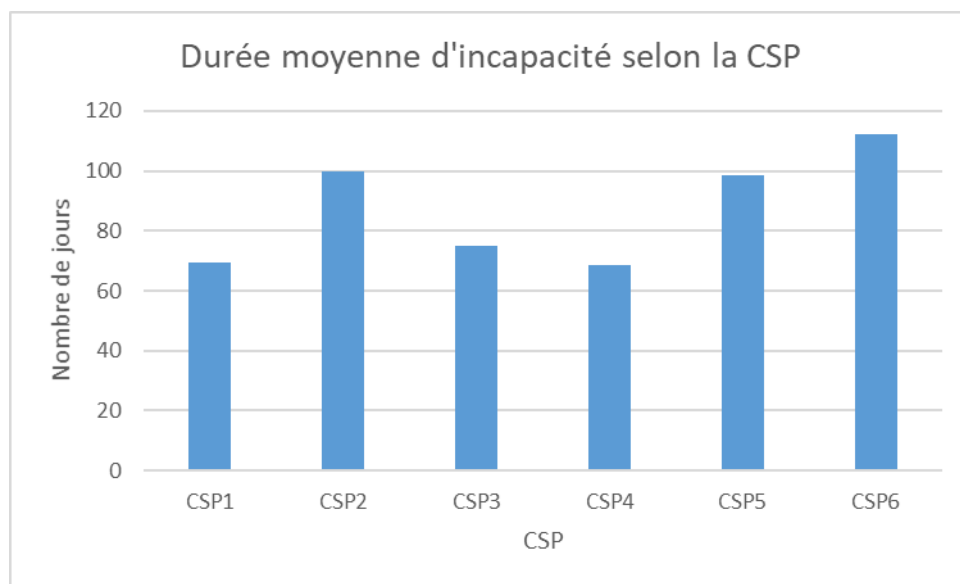
conscient que la durée de leur rétablissement en cas d'arrêt travail sera plus longue que pour un assuré plus jeune. Par conséquent, l'assuré âgé choisira un niveau de franchise plus long.

La seconde explication est qu'un assuré ne va pas déclarer un sinistre qui n'a pas dépassé son niveau de franchise car il ne recevra aucune indemnité de la part de son organisme assureur. La conséquence est que la durée moyenne réelle de maintien en incapacité est sans doute plus faible que les valeurs présentées ci-dessus car nous n'avons pas connaissance de tous les sinistres ne dépassant pas le niveau de franchise choisi. Cette conséquence ne pose pas de problème du point de vue du certificateur, qui doit s'assurer de la prudence de la table d'expérience construite, car on surestime la durée de maintien en incapacité.

En revanche, le graphique ci-dessus laisse supposer que le niveau de franchise est un facteur discriminant de nos données. Cette observation est confirmée suite au test du Khi 2 d'indépendance.

#### Durée moyenne de maintien en incapacité selon la CSP :

En figure 2.26 est représenté la durée moyenne de maintien en incapacité selon la CSP.



**Figure 2.26 : Durée moyenne d'incapacité selon la CSP**

La durée moyenne d'incapacité varie assez nettement selon la CSP. La CSP semble également être un facteur d'hétérogénéité.

Le test du Khi 2 est effectué à partir du tableau suivant :

Durée de maintien (jours)	<=5	]5 ; 15]	]15 ; 30]	]30 ; 45]	]45 ; 60]	]60 ; 75]	]75 ; 90]	]90 ; 105]	]105; 120]	>120
CSP1	1 430	11 834	12 250	7 786	4 670	3 346	2 260	1 854	1 227	7 843
CSP2	237	1 091	1 040	657	406	317	228	241	137	1 267
CSP3	244	1 750	1 997	1 303	811	570	395	335	196	1 426
CSP4	148	972	892	533	340	211	133	144	82	559
CSP5	152	1 076	1 228	790	483	375	287	223	166	1 340
CSP6	37	91	100	75	68	75	63	26	13	171

**Figure 2.27 : Classes de durée de maintien en incapacité selon la CSP**

Et le résultat en figure 2.28 confirme l'impact de la CSP dans la durée de maintien en incapacité :

```

Pearson's Chi-squared test

data:  tableau
X-squared = 821.67, df = 45, p-value < 2.2e-16

```

**Figure 2.28 : Sortie graphique résultant du tableau 2.23**

## 2.3 Conclusion

A travers les statistiques et les tests réalisés, on peut conclure que le type d'arrêt de travail, l'âge, le sexe, la CSP et le niveau de franchise sont des facteurs discriminants de nos données.

Ce résultat pourrait remettre en question la volonté de segmenter la table d'expérience uniquement en fonction de l'âge et du type d'arrêt de travail. Cependant, du point de vue du certificateur, cette décision de segmenter uniquement en fonction de l'âge et du type d'arrêt de travail n'est pas aberrante, les raisons à ce choix étant diverses.

Premièrement l'âge et le type d'arrêt sont des facteurs évident d'hétérogénéité dans la durée de maintien en incapacité et cela n'est pas intrinsèque à nos données. Les tables d'expérience sont très souvent segmentées selon ces 2 variables, d'où ce choix.

De plus, la segmentation implique de disposer d'une base de données suffisamment importante. Suite à la segmentation selon le type d'arrêt, la table de maintien en incapacité selon le type d'arrêt de travail « accident » repose sur 80 001 sinistres. Comme le montre le graphique sur la proportion de sinistre par âge, nous disposons de peu de sinistres pour les

assurés aux âges extrêmes, la seconde segmentation selon l'âge d'entrée en incapacité commence déjà à être problématique. Des retraitements pour y remédier seront présentés ultérieurement. Par conséquent segmenter de nouveau notre base de données provoquerait une « sur paramétrisation », rendant notre modélisation inefficace car soumise aux fluctuations d'échantillonnage.

Enfin, l'intérêt de segmenter notre table qu'en fonction de 2 variables est de garder une certaine simplicité de la loi de maintien construite.

Cependant, ce n'est pas pour autant qu'on doit oublier ces facteurs d'hétérogénéité que sont la CSP, le sexe ou encore la franchise. Dans le cas où la table ainsi construite est certifiée, le certificateur doit alors effectuer un suivi annuel. Ce suivi consiste à déterminer si la table est toujours en adéquation avec les données.

Par exemple, d'après nos données, la CSP2 représente 6,47% de nos sinistres et la durée moyenne d'incapacité pour cette CSP est d'environ 100 jours, sachant que la durée moyenne d'incapacité général est de 76,42 jours. Si l'année suivante la sinistralité varie fortement et que la CSP2 représente désormais 30% de la sinistralité, alors il existerait un risque de sous provisionnement car la table d'expérience ne tient pas compte de la CSP. Il y alors une inadéquation entre la table et les données.

# 3 Conception de lois de maintien

Cette partie a pour but de présenter la conception de lois de maintien s'appuyant sur la base de données présentée dans la partie précédente.

## 3.1 Introduction aux notations mathématiques

On considère une variable aléatoire  $T$  à valeurs dans  $[0; +\infty[$ , et on note dans la suite  $F(t) = P(T \leq t)$ , sa fonction de répartition. Lorsque la densité de  $T$  existe, on notera

$$f(t) = \frac{d}{dt}F(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T \leq t + h)}{h}$$

### La fonction de survie

La fonction de survie est par définition le complément à un de la fonction de répartition :

$$S(t) = 1 - F(t) = P(T > t)$$

$S$  est une fonction décroissante telle que  $S(0) = 1$  (si  $P(T = 0) = 0$ , ce que nous supposons) et  $\lim_{t \rightarrow \infty} S(t) = 0$ .

Si la durée moyenne de survie existe alors elle s'exprime simplement à l'aide de  $S$ .

$$E[T] = \int_0^{\infty} t dF(t) = - \int_0^{\infty} t dS(t) = \int_0^{\infty} S(t) dt$$

On peut le démontrer en s'appuyant sur Fubini :

$$\int_0^{\infty} S(t) dt = \int_0^{\infty} E[1_{T>t}] dt = E \left[ \int_0^{\infty} 1_{T>t} dt \right] = E \left[ \int_0^T dt \right] = E[T]$$

### La fonction de survie conditionnelle :

$$S_u(t) = P(T > u + t | T > u)$$

On s'intéresse donc à la survie d'un élément après un instant  $u + t$ , sachant qu'il a déjà fonctionné correctement jusqu'en  $u$ . En revenant à la définition de la probabilité conditionnelle on peut écrire :



$$S_u(t) = P(T > u + t | T > u) = \frac{P(T > u + t)}{P(T > u)} = \frac{S(u + t)}{S(u)}$$

La fonction de survie conditionnelle s'exprime donc simplement à l'aide de la fonction de survie.

### La fonction de hasard

La fonction de hasard est par définition :

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t)$$

Il en résulte :

$$S(t) = e^{-\int_0^t h(s) ds}$$

On note  $H(t) = -\int_0^t h(s) ds$ , la fonction de hasard cumulée, qui est telle que :

$$S(t) = e^{-H(t)}$$

Dans le cadre du maintien en incapacité, la variable aléatoire  $T$  représente la durée de maintien en incapacité. La fonction de survie  $S(t)$  représente la probabilité d'être en incapacité au-delà de  $t$ . La fonction de hasard  $h(t)$  représente le risque instantané de sortie de l'état d'incapacité.

Les méthodes de conception de lois de maintien nécessitent l'explication de deux notions fondamentales.

Les notions de « censure » et de « troncature » sont des notions communes au domaine de modèle de durée. Cependant, selon le champ d'application (entrée en incapacité, maintien en incapacité et passage en invalidité) et la littérature usuelle, la compréhension de ces notions peut porter à confusion.

On rappelle donc dans un premier temps, la définition de ces notions dans le cadre général qui est celui des modèles de durée, puis dans le champ d'application du maintien en incapacité.

## 3.2 Introduction au vocabulaire

### 3.2.1 Définition de la censure et de la troncature dans le cadre de modèle de durée

#### Censure

On parle de censure droite si  $X \geq C$  où  $X$  est la variable d'intérêt et  $C$  l'observation de la censure.

On parle de censure gauche si  $X \leq C$  où  $X$  est la variable d'intérêt et  $C$  l'observation de la censure.

Par exemple, on veut savoir à quelle âge un enfant est capable d'effectuer une certaine tâche.  $X$  représente l'âge où l'enfant est capable d'effectuer cette tâche. On observe l'enfant entre un âge  $C_1$  et un âge  $C_2$ .

Lorsque l'expérience débute, l'enfant d'âge  $C_1$  est déjà capable de l'accomplir, et pour lui  $X \leq C_1$  : il s'agit d'une censure gauche. Si à la fin de l'expérience, l'enfant n'est pas encore capable d'accomplir la tâche en question, pour lui  $X \geq C_2$ : il s'agit d'une censure droite.

#### Troncature

On dit qu'il y a troncature gauche (resp droite) lorsque la variable d'intérêt n'est pas observable lorsqu'elle est inférieure à un seuil  $c > 0$  (resp supérieure à un seuil  $C > 0$ ).

Le phénomène de troncature est très différent de la censure, puisque dans ce cas on perd complètement l'information sur les observations en dehors de la plage : dans le cas de la censure, on a connaissance du fait qu'il existe une information, mais on ne connaît pas sa valeur précise, simplement le fait qu'elle excède un seuil ; dans le cas de la troncature on ne dispose pas de cette information.

### 3.2.2 Définition censure et troncature dans le champ d'application du maintien en incapacité

La variable d'intérêt devient la durée de maintien en incapacité d'un assuré.

#### Troncature

Si on applique au sens strict la définition de la troncature vue dans le cadre de modèle de durée, on remarque qu'on est en présence de troncature gauche dans le cas où un assuré a un arrêt de travail inférieur à sa franchise. Par exemple, un assuré a souscrit un contrat Arrêt de travail avec une garantie incapacité auprès de son organisme assureur et la franchise est de 15 jours. Si l'assuré est en état d'incapacité pendant 2 jours et qu'il se rétablit ensuite, il ne va certainement pas déclarer son arrêt de travail auprès de son organisme assureur car il ne

recevra pas d'indemnité. La variable d'intérêt n'est pas observable car elle est inférieure à la franchise. C'est donc bien une troncature gauche.

Cependant, dans notre construction de table de maintien en incapacité, la franchise n'impliquera pas de troncature car on a décidé de construire une table de maintien au-delà de la franchise. Pour être plus rigoureux, on devrait donc dire que notre variable d'intérêt est la durée de maintien en incapacité dont la durée est supérieure au délai de franchise, ou alors la durée d'indemnisation tout simplement.

Dans la littérature usuelle, on parle de troncature dans le cas suivant :

- Lorsqu'un assuré entre en incapacité avant la date de début d'observation et qu'il en sort durant la période d'observation

En réalité, cela correspond clairement à la définition d'une censure gauche. En effet, on a connaissance qu'un assuré est en incapacité à la date de début d'observation sans savoir à quel moment son incapacité a débuté.

Cependant, pour la suite afin d'être en accord avec la littérature usuelle, on considèrera bien qu'un cas de ce type correspond à une troncature.

## **Censure**

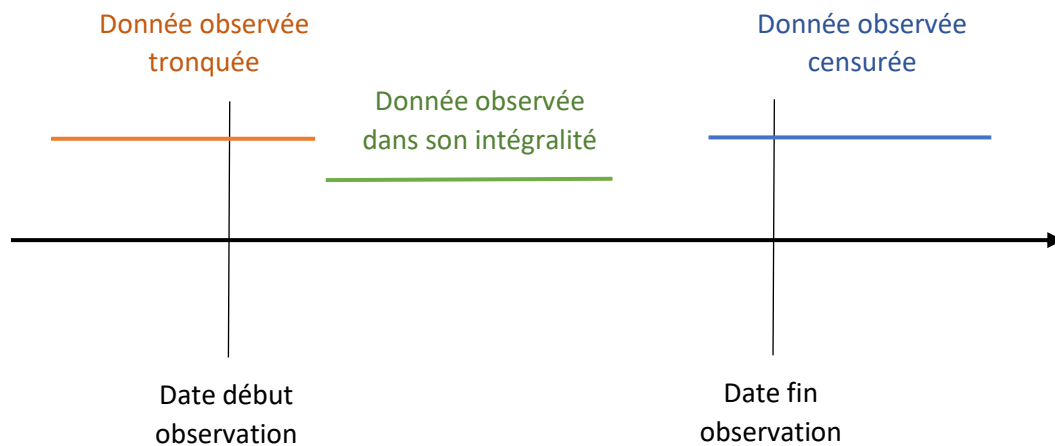
Dans la littérature usuelle, on considère qu'on est en présence de censure dans les cas suivants :

- Lorsqu'un assuré entre en incapacité pendant la période d'observation et que la fin de sa garantie incapacité intervient alors qu'il est toujours incapable.
- Lorsqu'un assuré entre en incapacité pendant la période d'observation et qu'il est toujours en état d'incapacité à la date de fin d'observation.

Ces 2 cas correspondent en réalité à des censures droites puisqu'on a l'information que l'assuré est en état d'incapacité mais on ne connaît pas la date de fin. On parlera simplement d'une censure pour être en accord avec la littérature usuelle.

## **Sortie**

On entend par sortie tout assuré qui sort de l'état d'incapacité qu'il soit tronqué, censuré ou non. En figure 3.1 est représenté un schéma récapitulatif du vocabulaire propre à la littérature usuelle.



**Figure 3.1 : Schéma récapitulatif du vocabulaire propre à la littérature usuelle**

### 3.3 Méthode de construction d'une loi de maintien

On distingue trois approches pour le calcul des lois de maintien en incapacité : l'approche paramétrique, l'approche non paramétrique et l'approche semi-paramétrique.

Dans l'approche paramétrique, les observations suivent une loi de probabilité qui dépend d'un certain nombre de paramètres. Cette approche permet d'estimer facilement les paramètres, ainsi que les intervalles de confiance et permet également la construction de tests. L'inconvénient de la méthode paramétrique est l'inadéquation pouvant exister entre le phénomène étudié et le modèle retenu.

Dans l'approche non paramétrique, aucune hypothèse sur la loi de probabilité réelle des observations n'est faite et c'est là son principal avantage. Il s'agit dès lors d'un problème d'estimation fonctionnelle, avec les ambiguïtés que cela implique – par exemple, la fonction de survie, qui est continue, sera estimée par une fonction discontinue.

L'approche semi-paramétrique est une sorte de compromis entre les deux approches précédentes.

La durée de l'incapacité étant courte (limité à 3 ans), l'approche paramétrique et semi-paramétrique semblent inadaptés. L'utilisation d'un modèle paramétrique pour l'estimation de la fonction de survie dans le cas du maintien en incapacité n'est donc pas recommandée.

Dans cette partie, nous allons présenter différentes approches non paramétriques d'estimations pour le calcul des lois de maintien en incapacité : la méthode actuarielle, l'estimateur de Nelson-Aalen et l'estimateur de Kaplan-Meier.

Nous présentons tout d'abord la méthode de l'estimateur actuariel car c'est une méthode simple permettant d'introduire les différentes notions définies précédemment.

### 3.3.1 Méthode de l'estimateur actuariel

La méthode de l'estimateur actuariel repose sur l'hypothèse que les troncatures et les censures se produisent au milieu de la période considérée.

Nous adoptons les notations suivantes :

$$d_t = \text{nombre de sorties sur la période } ]t, t + 1]$$

$$c_t = \text{nombre de censurés sur la période } ]t, t + 1]$$

$$tr_t = \text{nombre de tronqués sur la période } ]t, t + 1]$$

$$n_t = \text{population observé en } t$$

Selon l'hypothèse de l'estimateur actuariel, la population sous risque (de sortie) sur la période  $]t, t + 1]$  s'écrit :

$$exp_t = n_t + \frac{tr_t}{2} - \frac{c_t}{2}$$

Le taux de sortie conditionnel s'écrit alors sous la forme suivante :

$$q_t = \frac{d_t}{exp_t} = \frac{d_t}{n_t + \frac{tr_t}{2} - \frac{c_t}{2}}$$

La mise en application de cet estimateur est facile et peu contraignante pour les données à utiliser. Cependant, les hypothèses de construction de l'estimateur actuariel sont trop simplificatrices pour espérer obtenir des estimations proches de la sinistralité réelle de notre portefeuille.

### 3.3.2 Estimateur de Nelson Aalen

Selon PLANCHET (2005), l'estimateur de Nelson Aalen est un estimateur non paramétrique et unidimensionnel du taux de hasard cumulé.

Pour comprendre la construction de l'estimateur de Nelson-Aalen du taux de hasard cumulé, on a besoin d'introduire la notion de processus ponctuel.

Un processus ponctuel  $(N(t), t \geq 0)$  est défini comme un processus à valeurs entières adapté à une filtration  $(F_t)$  tel que  $N(0) = 0$ ,  $N(t) < \infty$  presque sûrement et tel que les trajectoires soient continues à droite, constantes par morceaux et ne présentent que des sauts d'amplitude +1.

Selon une proposition, si un processus ponctuel  $(N(t), t \geq 0)$  adapté à la filtration  $(F_t)$  est tel que  $E[N(t)] < \infty$ , alors il existe un unique processus croissant continu à droite  $\Lambda$  tel que  $\Lambda(0) = 0$ ,  $E[\Lambda] < \infty$ , et  $M(t) = N(t) - \Lambda(t)$  est une martingale.  $\Lambda$  est le compensateur de  $N$ .

Lorsque  $\Lambda$  s'écrit sous la forme  $\Lambda(t) = \int_0^t \lambda(u) du$ , alors le processus  $\lambda$  s'appelle l'intensité du processus ponctuel.

On rappelle qu'une martingale  $(M_t)_t$  est un processus adapté à une filtration  $(F_t)_t$  et est à trajectoire continues à droites avec des limites à gauche et vérifie :

$$E[|M_t|] < \infty \forall t \geq 0 \text{ et } E[M_t | M_s] = M_s \forall s \leq t$$

La décomposition  $N(t) = \Lambda(t) + M(t)$  exprime que le processus  $N$  « oscille » autour de la tendance prévisible  $\Lambda$  de sorte que la différence entre le processus d'intérêt  $N$  et sa tendance soit assimilable à un résidu, dont on maîtrise les variations. L'équation  $N(t) = \Lambda(t) + M(t)$  peut ainsi se lire comme « observations = modèle + terme d'erreur ».

Dans le cadre du maintien en incapacité, le processus ponctuel vaut 0 tant que le sujet est soumis au risque de maintien en incapacité et 1 dès que l'événement a eu lieu (sortie de l'incapacité) :

$$N(t) = 1_{X \leq t}$$

Le but est de déterminer le compensateur prévisible du processus ponctuel.

On note  $N(t-) = \lim_{u \uparrow t} N(u)$  la limite à gauche de  $N(t)$  et on s'intéresse à la loi de la variable aléatoire  $P(dN(t) = 1 | N(t-))$ , en ayant noté formellement  $dN_t = N(t + dt) - N(t)$  avec  $dt$  « petit ». La variable aléatoire  $dN(t)$  ne peut prendre que les valeurs 0 et 1.

Si  $N(t-) = 1$ , la sortie de l'état d'incapacité s'est déjà produite et le processus ne peut plus sauter. Donc :  $P(dN_t = 1 | N(t-) = 1) = 0$ .

Si  $N(t-) = 0$ , l'assuré peut potentiellement sortir de l'état d'incapacité. Donc :  $P(dN_t = 1 | N(t-) = 0) = h(t)dt$ .

On en déduit :

$$P(dN_t = 1 | N(t-)) = \begin{cases} h(t)dt \text{ avec la probabilité } 1 - S(t) \\ 0 \text{ avec la probabilité } S(t) \end{cases}$$

Puisque  $S(t)$  représente la probabilité d'être en incapacité au delà de  $t$ .

On pose alors  $\lambda(t) = h(t)1_{X \geq t}$ , produit de la fonction de hasard en  $t$  et de l'indicatrice de présence juste avant  $t$ ,  $Y(t) = 1_{X \geq t}$ . Le processus  $\lambda(t)$  est prévisible et  $Y(t) = 1$  est équivalent à  $N(t-) = 0$ . Donc  $P(dN_t = 1 | N(t-)) = \lambda(t)dt$ .

En introduisant le cas des censures, le compensateur prévisible du processus d'évènements non censurés  $N^1(t) = 1_{T \leq t, D=1}$ , s'écrit :

$$\Lambda^1(t) = \int_0^t R(u)h(u)du$$

Avec  $R(t) = 1_{T \geq t}$ , l'indicatrice de présence à risque avant  $t$ .

Dans le cas d'une population, dont on suppose que tous les individus ont la même fonction de hasard  $h$ , on associe à chaque membre de la population un processus d'évènement non censuré  $N_i^1(t) = 1_{T_i \leq t, D_i=1}$  ainsi que l'indicatrice de présence sous risque, comptabilisant les individus ni sortis d'incapacité ni censurés  $R_i(t) = 1_{T_i \geq t}$  et on construit les processus agrégés  $\bar{R}(t) = \sum_{i=1}^n R_i(t)$  et  $\bar{N}^1(t) = \sum_{i=1}^n N_i^1(t)$ . Ils comptabilisent respectivement l'effectif sous risque et le nombre d'évènements survenus non censurés.

Le processus de comptage  $\bar{N}^1(t)$  possède une intensité qui se met sous la forme  $\lambda(t) = \bar{R}(t)h(t)$

avec  $\bar{R}$  un processus observable (prévisible) et  $h$  la fonction de hasard, inconnue et à estimer.

Le fait que  $M(t) = \bar{N}^1(t) - \int_0^t \bar{R}(u)h(u)du$  soit une martingale centrée suggère de proposer  $\bar{N}^1(t)$  comme estimateur de  $\int_0^t \bar{R}(u)h(u)du$ . Mais alors le processus  $\int_0^t \frac{1_{\bar{R}(u) > 0}}{\bar{R}(u)} dM(u)$  est également une martingale et on a par construction de  $M$  :

$$\int_0^t \frac{1_{\bar{R}(u) > 0}}{\bar{R}(u)} dM(u) = \int_0^t \frac{1_{\bar{R}(u) > 0}}{\bar{R}(u)} d\bar{N}^1(u) - \int_0^t h(u)du = \int_0^t \frac{1_{\bar{R}(u) > 0}}{\bar{R}(u)} d\bar{N}^1(u) - H(t)$$

pour autant que  $t$  soit tel que  $\bar{R}(t) > 0$ . Ainsi  $\hat{H}(t) = \int_0^t \frac{1_{\bar{R}(u) > 0}}{\bar{R}(u)} d\bar{N}^1(u)$  est un estimateur naturel de  $H$ . Cet estimateur s'appelle l'estimateur de Nelson-Aalen.

Comme les processus considérés ici sont purement à sauts on peut, en notant  $\Delta\bar{N}(t) = \bar{N}(t) - \bar{N}(t-)$ , mettre cette expression sous la forme :

$$\hat{H}(t) = \sum_{\{i/T_i \leq t\}} \frac{\Delta\bar{N}(T_i)}{\bar{R}(T_i)}$$

En posant  $d(t) = \Delta\bar{N}(t)$ , le nombre de sorties d'incapacité en  $t$  et  $r(t) = \bar{R}(t)$ , l'effectif sous risque juste avant  $t$ , on peut ainsi réécrire l'équation ci-dessus sous la forme intuitive suivante:

$$\hat{H}(t) = \sum_{\{i/T_i \leq t\}} \frac{d(T_i)}{r(T_i)} = \sum_{\{i/T_i \leq t\}} \frac{d_i}{r_i}$$

On peut vérifier que cet estimateur est biaisé et sous-estime en moyenne la fonction de hasard cumulée. En effet,

$$\hat{H}(t) = \int_0^t \frac{1_{\bar{R}(u) > 0}}{\bar{R}(u)} d\bar{N}^1(u) = \int_0^t \frac{1_{\bar{R}(u) > 0}}{\bar{R}(u)} (dM(u) + \bar{R}(u)h(u)du).$$

Comme  $M$  est une martingale, il vient en prenant l'espérance des deux membres de l'équation ci-dessus

$E[\hat{H}(t)] = \int_0^t E[1_{\bar{R}(u) > 0}]h(u)du$  (car on a supposé que l'espérance d'une martingale est nulle)

Mais  $E[1_{\bar{R}(u) > 0}] = P(\bar{R}(u) > 0) = 1 - P(\bar{R}(u) = 0)$ .

On en déduit finalement :  $E[\hat{H}(t)] = \int_0^t h(u)du - \int_0^t P(\bar{R}(u) = 0)h(u)du = H(t) - \int_0^t P(\bar{R}(u) = 0)h(u)du$

Ce qui implique que :  $E[\hat{H}(t)] \leq H(t)$  : l'estimateur de Nelson-Aalen a bien tendance à sous-estimer la fonction de hasard cumulée.

On peut remarquer que l'estimateur de Nelson-Aalen du taux de hasard cumulé conduit à un estimateur naturel de la fonction de survie, en exploitant la relation  $S(t) = e^{-H(t)}$ ; on peut ainsi proposer comme estimateur de la fonction de survie :

$$\hat{S}_{HF}(t) = e^{-\hat{H}_{NA}(t)}$$

Cet estimateur est l'estimateur de Harrington-Fleming.

Comme on a montré que :  $E[\hat{H}_{NA}(t)] \leq H(t)$  et que la fonction  $g(x) = e^{-x}$  est convexe, on en déduit que :

$$E[\hat{S}_{HF}(t)] = E\left[g\left(\hat{H}_{NA}(t)\right)\right] \geq g\left(E[\hat{H}_{NA}(t)]\right) = e^{-E[\hat{H}_{NA}(t)]} \geq e^{-H(t)} = S(t)$$



L'estimateur de Harrington Fleming de la fonction de survie présente donc un biais de surestimation. En utilisant un tel estimateur, on surestime donc la durée de maintien en incapacité en moyenne.

### 3.3.3 La méthode de Kaplan Meier

La méthode de Kaplan Meier est une estimation non paramétrique et unidimensionnelle (âge par âge) de la fonction de survie, appelée également estimateur de Kaplan Meier.

L'approche unidimensionnelle de la méthode de Kaplan Meier est une limite à son utilisation mais la perte d'information que représente la non prise en compte de la loi conjointe selon les 2 dimensions du problème est faible et peu pénalisante d'après Planchet, Therond (2006).

La méthode de Kaplan-Meier rend possible la prise en compte de toutes les observations y compris les observations incomplètes (troncatures et censures). De plus, si l'on a suffisamment d'observations, cette méthode permet d'obtenir des estimations fiables. Enfin, la méthode de Kaplan-Meier possède un certain nombre de bonnes propriétés présentés ultérieurement.

Pour l'ensemble de ces raisons, l'estimation de la fonction de survie par Kaplan-Meier est la méthode la plus utilisée sur le marché.

#### Principe et construction de l'estimateur de Kaplan Meier

Selon PLANCHET (2005), la construction heuristique de l'estimateur de Kaplan Meier de la fonction de survie s'appuie sur la remarque suivante : la probabilité d'être toujours en incapacité au-delà de  $t > s$  peut s'écrire :

$$S(t) = P(T > t | T > s) * P(T > s) = P(T > t | T > s) * S(s)$$

On peut renouveler l'opération, ce qui fait apparaître des produits de terme en  $P(T > t | T > s)$  si on choisit comme instants de conditionnements les instants où se produit un événement (sortie ou censure ou troncature). On se ramène à estimer des probabilités de la forme :

$$p_i = P(T > T_{(i)} | T > T_{(i-1)})$$

Où  $p_i$  représente la probabilité d'être en incapacité à l'instant  $T_{(i)}$  sachant qu'on l'était déjà à l'instant  $T_{(i-1)}$ .

On pose :

$$q_i = 1 - p_i = P(T \leq T_{(i)} | T > T_{(i-1)})$$

Avec  $q_i$  représentant le taux de sortie conditionnel, c'est-à-dire la probabilité de sortir d'incapacité à l'instant  $T_{(i)}$  sachant qu'on était en incapacité à l'instant  $T_{(i-1)}$ .

On obtient donc :

$$S(t) = \prod_{T_{(i)} \leq t} (1 - q_i)$$

On en déduit l'estimateur de Kaplan Meier de la fonction de survie  $\hat{S}_{KM}$  en estimant  $q_i$  par maximum de vraisemblance.

Soient  $d_i$  le nombre de sortie d'incapacité à la date  $T_{(i)}$  et  $n_i$  la population sous risque juste avant la date  $T_{(i)}$ . D'après ce modèle, le nombre de sorties d'incapacité  $d_i$  suit donc une loi binomiale  $\mathcal{B}(n_i, q_i)$ . On détermine  $\hat{q}_i$  par maximum de vraisemblance.

$$L(q_1, \dots, q_m) = \prod_{i=1}^m \binom{n_i}{d_i} q_i^{d_i} (1 - q_i)^{n_i - d_i}$$

$$\ln L(q_1, \dots, q_m) = \sum_{i=1}^m \ln \left( \binom{n_i}{d_i} q_i^{d_i} (1 - q_i)^{n_i - d_i} \right)$$

$$\frac{\partial \ln L}{\partial q_i} = \frac{d_i}{q_i} - \frac{n_i - d_i}{1 - q_i} = 0$$

$$\hat{q}_i = \frac{d_i}{n_i} \quad \forall i = 1 \dots m$$

D'où

$$\hat{S}(t) = \prod_{T_{(i)} \leq t} (1 - \hat{q}_i) = \prod_{T_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

Comme énoncé précédemment, la fonction de survie selon la méthode Kaplan Meier est évaluée âge par âge. Soit  $x$ , l'âge d'entrée en incapacité :

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

Supposons maintenant qu'en  $T_{(i)}$ ,  $c_{x,i}$  têtes soient censurées et que  $tr_{x,i}$  têtes soient tronquées. On décide de se placer en temps discret avec un pas journalier.

En pratique, on est confronté à la présence d'ex aequo, on suppose alors par convention que les observations non censurées et non tronquées précèdent les observations censurées.

Le nombre de têtes restant présentes face au risque de sortir de l'arrêt de travail en  $T_{(i)}$  sera donnée par  $n_{x,i} = n_{x,i-1} - d_{x,i-1} - c_{x,i-1} + tr_{x,i-1} \quad \forall i$

Prenons un exemple pour comprendre l'impact de cette convention. Supposons qu'on débute notre observation avec 100 personnes en incapacité d'âge x. Le tableau suivant récapitule les différentes sorties, censures et troncatures pour les deux premières dates de discrétisation de notre période d'observation.

	Nb de sorties	Nb de censures	Nb de troncatures	Population
T=1	35	5	10	100
T=2	20	11	4	

$$S_x(1) = 1 - \frac{35}{100}$$

$$S_x(2) = \left(1 - \frac{35}{100}\right) * \left(1 - \frac{20}{100 - 35 - 5 + 10}\right)$$

### Propriétés de l'estimateur de Kaplan Meier

L'estimateur de Kaplan-Meier est convergent, asymptotiquement gaussien, cohérent et est un estimateur du maximum de vraisemblance généralisé. L'estimateur de Kaplan-Meier est l'unique estimateur cohérent de la fonction de survie.

### Cohérence

Un estimateur  $\hat{S}_{x,t}$  de la fonction de survie  $S_{x,t}$  est dit cohérent si pour tout t de  $\mathbb{R}^+$ , il vérifie :

$$\hat{S}_{x,t} = \frac{1}{n} \left[ \sum_{i=1}^n 1_{\{T_{x,i} \geq t\}} + \sum_{i=1}^n 1_{\{T_{x,i} < t, D_i = 0\}} \frac{\hat{S}_{x,t}}{\hat{S}_{x,T_i}} \right]$$

$\sum_{i=1}^n 1_{\{T_{x,i} \geq t\}}$  représente l'effectif des individus ni sortis, ni censurés au-delà de l'instant t.

$\sum_{i=1}^n 1_{\{T_{x,i} < t, D_i = 0\}} \frac{\hat{S}_{x,t}}{\hat{S}_{x,T_i}}$  représente le fait que les individus censurés à un instant  $T_i$  antérieur à t, survivent au-delà de t avec la probabilité conditionnelle  $\frac{\hat{S}_{x,t}}{\hat{S}_{x,T_i}}$  qui pondèrent chacun d'eux.

### GLME (Generalized Maximum Likelihood Estimator)

Définition : Soit  $\phi$  une famille de probabilités sur  $\mathbb{R}^n$  (avec la tribu borélienne) non dominée,  $\forall x \in \mathbb{R}^n$ , et  $P_1, P_2 \in \phi$ , on pose  $l(x, P_1 P_2) = \frac{dP_1}{d(P_1 + P_2)}(x)$  ; on dit alors que  $\hat{P}$  est GMLE pour P si  $l(x, \hat{P}, P) \geq l(x, P, \hat{P})$ .

Si les lois de la durée de maintien en incapacité non censurée et de la censure sont diffuses et si la famille  $\phi$  contient les lois de probabilité chargeant les points  $(T_i, D_i)$ , alors on peut montrer que  $\hat{S}_{KM}$  est GMLE pour  $S$ , autrement dit l'estimateur de Kaplan Meier est un estimateur de maximum de vraisemblance généralisé.

### Convergence et normalité asymptotique

L'estimateur de Kaplan-Meier est asymptotiquement gaussien ; précisément on a le résultat suivant :

Propriété : Si les fonctions de répartition de la survie et de la censure n'ont aucune discontinuité commune alors :

$$\sqrt{n}(\hat{S}_{x,t} - S_{x,t}) \rightarrow W_s$$

Où  $W_s$  est un processus gaussien centré, de fonction de covariance :

$$Cov(W(s), W(t)) = S(s)S(t) \int_0^{s \wedge t} \frac{dF(u)}{(1-F(u))^2(1-G(u))}$$

L'intérêt de résultats de convergence au niveau du processus lui-même plutôt que pour un instant fixé est que l'on peut en déduire des bandes de confiance asymptotique pour l'estimateur de Kaplan-Meier.

### Absence de biais dans certains cas

L'estimateur de Kaplan Meier est biaisé positivement, c'est-à-dire qu'il surestime en moyenne la vraie valeur de la fonction de survie ce qui peut s'avérer pénalisant :

$$E[\hat{S}_{KM}(t)] \geq S(t)$$

Toutefois, le lemme 3.2.1 page 99 de Fleming et Harrington (1991) donne une expression explicite de ce biais :

$$E[\hat{S}_{KM}(t) - S(t)] = E \left[ 1_{T(n) < t} * \frac{\hat{S}_{KM}(T(n))}{S(T(n))} * (\hat{S}_{KM}(T(n)) - S(t)) \right]$$

On en déduit donc que si  $t < T(n)$  ou si  $\hat{S}_{KM}(T(n)) = 0$ , alors  $E[\hat{S}_{KM}(t) - S(t)] = 0$ .

Cela équivaut à dire que l'estimateur de Kaplan Meier n'est pas biaisé jusqu'à la dernière observation ou si la dernière observation n'est pas censurée.

Pour l'ensemble de ses avantages, nous décidons d'utiliser la méthode de Kaplan Meier pour concevoir la loi de maintien de nos données.

On se place en temps discret avec un pas journalier. Pour réaliser cette méthode, on a simplement besoin du nombre de sorties, de troncatures et de censures à chaque âge et jour. On calcule donc la fonction de survie  $\hat{S}_x(t)$  pour chaque mois d'ancienneté ( $t$ ) et à chaque âge d'entrée en incapacité ( $x$ ).

On en déduit un taux de sortie conditionnelle défini de la façon suivante :

$$q(\widehat{x}, t) = \frac{\hat{S}_x(t) - \hat{S}_x(t+1)}{\hat{S}_x(t)}$$

### 3.3.4 Comparaison entre l'estimateur de Kaplan-Meier et d'Harrington-Fleming

Après transformation logarithmique, l'estimateur de Kaplan-Meier s'écrit :

$$\ln(\hat{S}_{KM}(t)) = \sum_{T_{(i)} \leq t} \ln \left( 1 - \frac{d_i}{r_i} \right)$$

Et l'estimateur de Harrington-Fleming s'écrit :

$$\ln(\hat{S}_{HF}(t)) = - \sum_{T_{(i)} \leq t} \frac{d_i}{r_i}$$

On obtient donc :

$$\ln(\hat{S}_{KM}(t)) - \ln(\hat{S}_{HF}(t)) = \sum_{T_{(i)} \leq t} \left( \ln \left( 1 - \frac{d_i}{r_i} \right) + \frac{d_i}{r_i} \right)$$

On vérifie aisément que la fonction  $f(x) = \ln(1-x) + x$  est toujours négative et donc

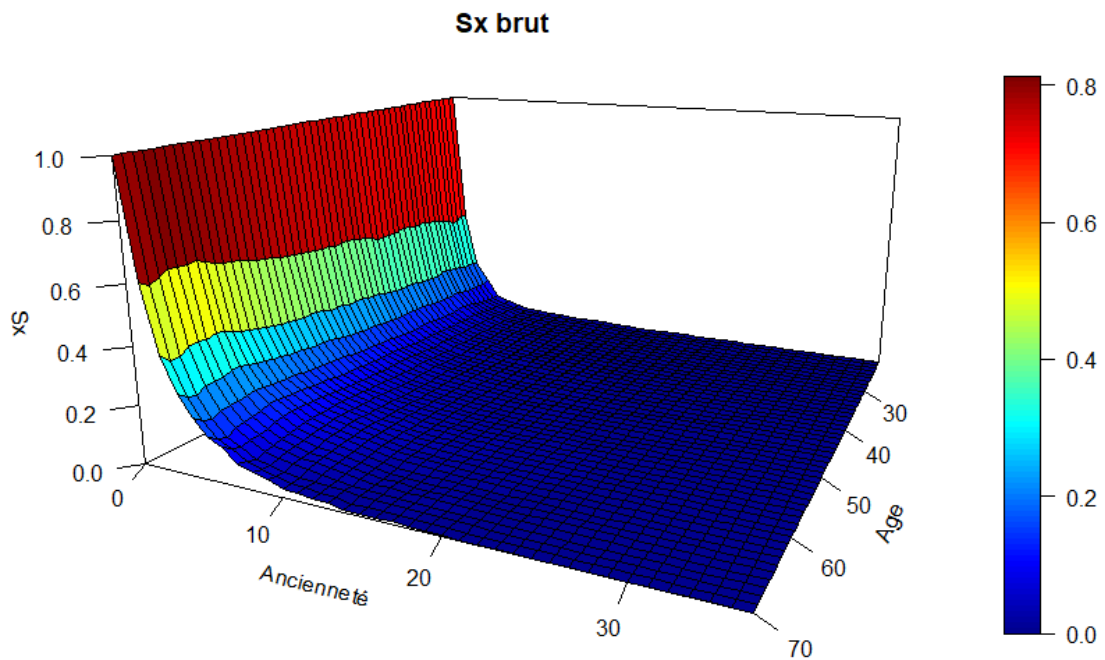
$$\hat{S}_{KM}(t) \leq \hat{S}_{HF}(t)$$

L'estimateur d'Harrington-Fleming aboutit donc toujours à une fonction de survie plus prudente qu'avec l'estimateur de Kaplan-Meier. L'estimateur d'Harrington-Fleming peut donc être intéressant à utiliser dans le cas où l'estimateur de Kaplan-Meier ne débouche pas sur une estimation suffisamment prudente de la table d'expérience.

## 3.4 Présentation de la loi de maintien obtenue selon la méthode de Kaplan-Meier

Cette sous partie a pour but de présenter la loi de maintien obtenue grâce à la méthode de Kaplan-Meier.

Graphiquement, la fonction de survie pour chaque mois d'ancienneté et à chaque âge d'entrée en incapacité prend la forme suivante (Figure 3.2) :



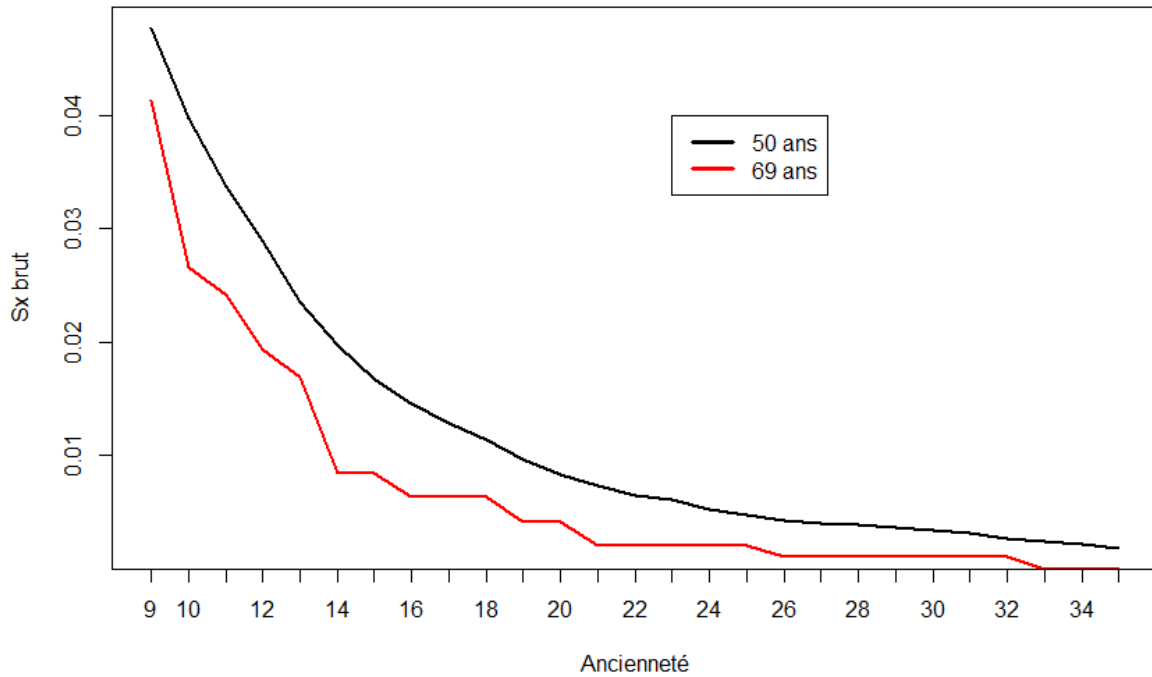
**Figure 3.2 : Représentation graphique de la fonction de survie obtenue selon la méthode Kaplan-Meier**

On peut constater que pour chaque âge d'entrée en incapacité, la fonction de survie chute très rapidement, autrement dit, la probabilité pour un assuré de se maintenir en incapacité plusieurs mois après le début de son indemnisation est très faible. Au bout de 7 mois d'ancienneté, la probabilité d'être toujours en incapacité est inférieure à 10% quelque soit l'âge d'entrée en incapacité.

Le graphique semble afficher une courbe assez lisse. Cependant, la loi de maintien contient des disparités selon l'âge d'entrée en incapacité qui ne sont pas visible sur ce graphique.

Sur le graphique suivant (Figure 3.3), on représente deux fonctions de survie à l'âge d'entrée 50 et 69 ans entre 9 et 35 mois d'ancienneté respectivement.

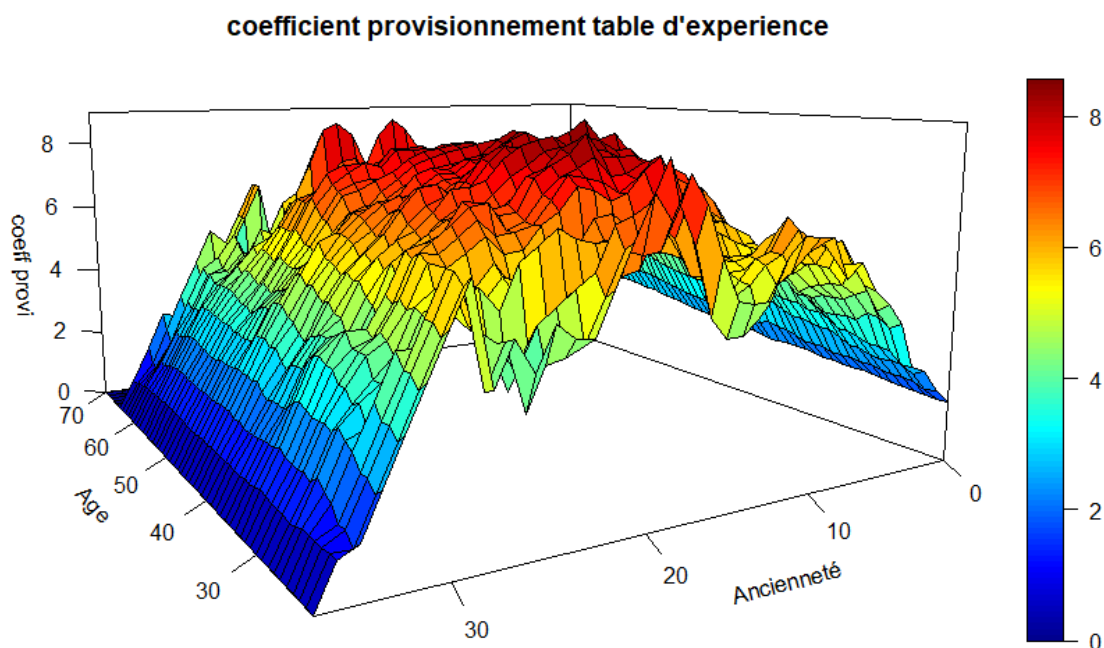
## Evolution Sx entre 9 et 35 mois d'ancienneté



**Figure 3.3**

On constate que la fonction de survie à l'âge d'entrée 50 ans est beaucoup plus régulière que la fonction de survie à l'âge d'entrée 69 ans. D'une manière générale, les fonctions de survies aux âges extrêmes (ie les âges les plus jeunes et les plus vieux) ont tendance à être plus volatile que les autres fonctions de survie.

Une autre manière d'observer ces irrégularités est de représenter la courbe des coefficients de provisionnement (Figure 3.4) :



**Figure 3.4 : Représentation graphique des coefficients de provisionnement**

Cette courbe est ressemblante à la courbe des coefficients de provisionnement de la table du BCAC en vigueur que l'on avait représenté (figure 1.5) car on obtient dans les deux cas une « bosse ». Cependant les coefficients de provisionnement sont beaucoup plus irréguliers selon la table d'expérience quelque soit l'âge d'entrée en incapacité.

### 3.5 Analyse des boni mali

La construction de notre table d'expérience par la méthode de Kaplan-Meier nous a permis de calculer les coefficients de provisionnement correspondants et permet donc de calculer des provisions de type « rente d'incapacité en cours ».

L'analyse des boni mali vise à comparer les prestations réellement payées au titre du maintien en incapacité avec les prestations prédites par la loi d'expérience de maintien en incapacité.

Dans le cadre de notre étude, nous décidons de comparer les prestations réellement payées en 2015 avec les prestations prédites pour l'année 2015 par la loi d'expérience de maintien en incapacité sur le périmètre des arrêts en cours au 31/12/2014 (Figure 3.5). Cette comparaison est réitérée pour les années de survenance 2014 et 2013 (Figure 3.6 et 3.7).



Outre l'audit de cette étape, le rôle du certificateur est de s'assurer que l'analyse des boni mali conduit à un résultat relativement prudent. En effet, si la loi d'expérience de maintien en incapacité aboutit à une insuffisance des provisions par rapport aux prestations réellement payés, la loi de maintien ne pourra être certifiée.

Année		Indemnité réellement versée en 2015	Indemnité selon la table provisionnement	Ecart
2015	Pour 1 € d'indemnité mensuel	192 138,83	216 756,86	11,36%
	Montants réels	7 128 510,86	8 036 245,34	11,30%

**Figure 3.5 : Comparaison des indemnités réellement versés avec les indemnités provisionnelles pour l'année 2015**

L'année 2015 confirme la prudence de notre loi d'expérience car on aboutit à un sur provisionnement de 11,30%.

Année		Indemnité réellement versée en 2014	Indemnité selon la table provisionnement	Ecart
2014	Pour 1 € d'indemnité mensuel	217 125,94	224 523,85	3,29%
	Montants réels	8 111 462,62	8 207 417,85	1,17%

**Figure 3.6: Comparaison des indemnités réellement versés avec les indemnités provisionnelles pour l'année 2014**

Pour l'année 2014, la table d'expérience permet de légèrement sur provisionner les arrêts en cours au 31/12/2013.

Année		Indemnité réellement versée en 2013	Indemnité selon la table provisionnement	Ecart
2013	Pour 1 € d'indemnité mensuel	187 817,00	182 164,27	-3,10%
	Montants réels	7 023 583,21	6 941 582,94	-1,20%

**Figure 3.7 : Comparaison des indemnités réellement versés avec les indemnités provisionnelles pour l'année 2013**

En revanche, pour l'année 2013, notre table aboutit à un léger sous provisionnement de 1,2%.

### Conclusion

Selon l'analyse des boni mali, la table d'expérience construite semble être suffisamment prudente comparé à la sinistralité réelle de notre portefeuille. En effet, la comparaison entre les prestations réellement payées et les prestations prédites aboutit à un sur provisionnement pour les années 2015 et 2014. Ce résultat est nuancé par l'année 2013 qui aboutit à un léger sous provisionnement.

Cependant, la loi de maintien ainsi construite ne peut être certifiée par l'actuaire indépendant. En effet, la représentation graphique de la fonction de survie ainsi que celle des coefficients de provisionnement a permis de mettre en avant des irrégularités qui doivent être « corrigées » à travers une méthode de lissage présentée dans la partie suivante.

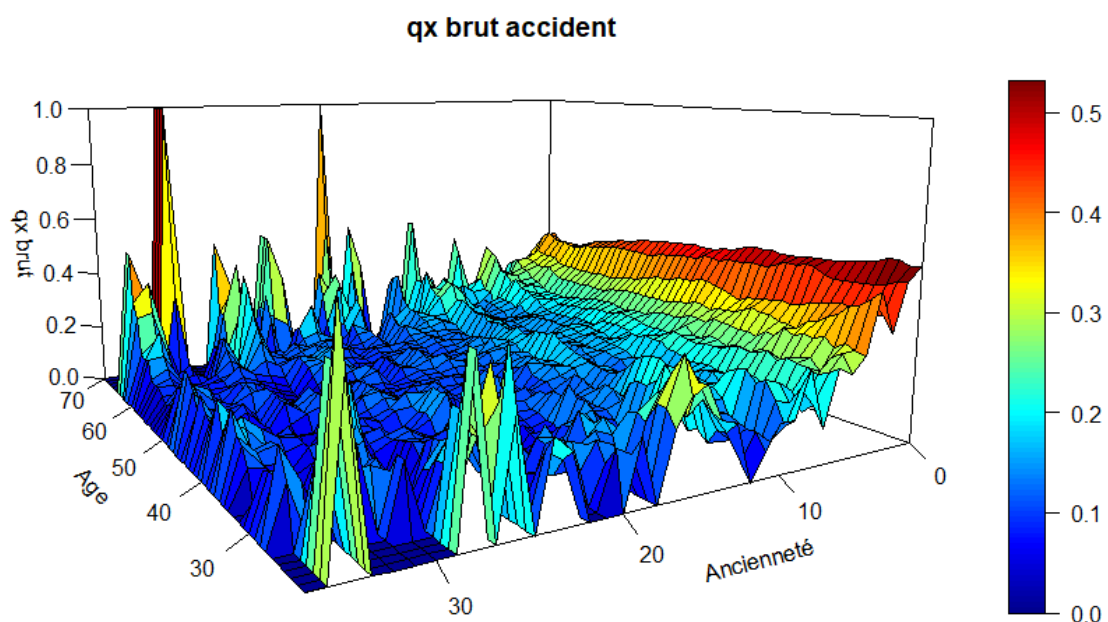
# 4 Lissage

## 4.1 Intérêt du lissage

### Analyse des taux de sorties conditionnelles

L'analyse des taux de sorties conditionnelles permet de mettre en évidence l'intérêt du lissage.

Une fois que l'on a appliqué la méthode Kaplan Meier, on obtient les taux de sorties brut représentés en Figure 4.1.



**Figure 4.1 : Représentation des taux de sortie conditionnels bruts**

On observe des petites « vagues » au centre et des « pics » sur les bords du graphique. Lors des premiers mois d'ancienneté, les taux de sorties diminuent fortement quelque soit l'âge

d'entrée en incapacité. Au bout d'un an d'ancienneté, on note des taux de sortie qui remontent légèrement pour tous les âges.

Pour les âges les plus jeunes (de 21 à 25 ans) et les plus vieux (de 65 à 70 ans), les taux de sortie ont tendance à varier de manière plus brutale que le reste de la population.

Pour les âges les plus vieux (de 66 à 70 ans), le taux de sortie atteint 1 peu avant les 36 mois d'ancienneté. En effet, on constate que selon notre base de données, l'ensemble des personnes âgées de 66 à 70 ans sont sorties de l'état d'incapacité quelques mois avant la fin de durée maximum d'incapacité (36 mois).

Cette représentation graphique de nos taux de sorties permet d'observer certaines irrégularités dans notre courbe qui n'ont sans doute pas d'explication théorique mais qui s'expliquent vraisemblablement par nos données. On parle alors de risque d'échantillonnage (i.e. le risque que l'échantillon prélevé par l'auditeur ne soit pas représentatif de la population dont il a été tiré). Ces irrégularités sont facilement repérables sur le graphique des taux de sortie bruts car elles forment des « pics » plus ou moins grand.

Par exemple, à l'âge de 70 ans et à l'ancienneté de 20 mois, on observe un taux de sortie de 1 tandis qu'à 65 ans et à ancienneté égale, le taux de sortie n'est que de 0.15 environ.

On peut supposer de manière assez certaine que ces écarts n'ont pas d'explication théorique mais sont sans doute dû à un risque d'échantillonnage. Pour remédier à ce problème, on utilise le lissage.

Le but du lissage est donc le suivant : supprimer les irrégularités de la courbe dues aux fluctuations liées à l'échantillonnage de la base de données et qui entraînent des irrégularités dans la tarification et le provisionnement.

Plus la base de données est importante, moins le risque d'échantillonnage risque d'être élevé. Parmi nos données, c'est aux âges les plus faibles et les plus élevés que la population observée est la plus faible. Par exemple, à l'âge de 70 ans, on observe une population qui ne dépasse jamais les 187 individus contre 10 320 personnes à l'âge de 54 ans. Cela peut donc potentiellement expliquer le fait d'avoir un taux de sortie égal à 1 à l'âge de 70 ans et d'ancienneté 20 mois car on observe la sortie du seul individu restant de notre population ce mois-ci.

## 4.2 Estimateur de Greenwood

Avant de lisser notre table, il est conseillé de calculer l'estimateur de la variance de l'estimateur de Kaplan-Meier appelé l'« estimateur de Greenwood » pour plusieurs raisons.

Premièrement, l'estimateur de Greenwood nous permet de mesurer la robustesse de l'estimation. Cela nous permet de cerner les zones soumises au risque d'échantillonnage.

De plus, il est également intéressant de calculer l'estimateur de Greenwood car il est convergent pour la variance asymptotique de l'estimateur de Kaplan Meier. Il permet avec la normalité asymptotique de l'estimateur de Kaplan Meier de calculer des intervalles de confiance (asymptotiques).

La formule de l'estimateur de Greenwood est la suivante :

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \gamma(t)^2$$

$$\text{Avec } \gamma(t) = \sqrt{\sum_{T(i) \leq t} \frac{d_i}{r_i(r_i - d_i)}}$$

La démonstration est la suivante :

$$\text{On rappelle que } \hat{S}_{KM}(t) = \prod_{T(i) \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

$$\text{En passant au logarithme, on obtient : } \ln(\hat{S}_{KM}(t)) = \sum_{T(i) \leq t} \ln\left(1 - \frac{d_i}{n_i}\right) = \sum_{T(i) \leq t} \ln(1 - \hat{q}_i)$$

Avec l'indépendance des variables  $\ln(1 - \hat{q}_i)$ , comme la loi de  $r_i \hat{p}_i$  est binomiale de paramètres  $(r_i, p_i)$ , on a par la méthode delta,  $V(f(x)) \approx \left(\frac{df(x)}{dx}(E(X))\right)^2 V(X)$  :

$$V(\ln \hat{p}_i) \approx V(\hat{p}_i) \left[\frac{d}{dp} \ln(\hat{p}_i)\right]^2 = \frac{\hat{q}_i(1 - \hat{q}_i)}{n_i} \frac{1}{(1 - \hat{q}_i)^2} = \frac{\hat{q}_i}{n_i(1 - \hat{q}_i)}$$

Ce qui conduit à proposer comme estimateur de la variance de  $\ln(\hat{S}(t))$  :

$$\hat{V}(\ln(\hat{S}(t))) = \sum_{T(i) \leq t} \frac{\hat{q}_i}{n_i(1 - \hat{q}_i)} = \sum_{T(i) \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

En appliquant de nouveau la méthode delta avec pour f la fonction logarithme, on obtient finalement :

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \gamma(t)^2$$

$$\text{Avec } \gamma(t) = \sqrt{\sum_{T(i) \leq t} \frac{d_i}{r_i(r_i - d_i)}}$$

Le tableau suivant (Figure 4.2) récapitule les résultats du calcul de l'estimateur de Greenwood à plusieurs âges et plusieurs mois d'ancienneté :

Age	Ancienneté (en mois)				
	1	10	19	25	32
23 ans	0,000146888	0,000904265	0,00285264	0,02991128	0,28412106
40 ans	3,87485E-05	0,000250385	0,00104265	0,00863593	0,08199094
50 ans	5,22669E-05	0,000271065	0,00101422	0,00787488	0,06422872
66 ans	0,000536025	0,00294525	0,00931195	0,06692202	0,54517797

**Figure 4.2 : Calcul de l'estimateur de Greenwood à plusieurs âges et mois d'ancienneté**

Le calcul de l'estimateur de Greenwood nous permet de conclure que :

- Les âges 66, 67, 68, 69 et 70 sont très volatils.
- Les âges 21, 22, 23, 24 et 25 sont très volatils.

Les estimations de la fonction de survie à ces âges ont donc peu de chance d'être réellement fiables. On décide alors de conserver les âges allant de 26 à 65 ans car on considère que la plage correspondante est stable. Les taux de sorties pour les âges en dehors de cette plage seront ensuite déterminés par extrapolation.

### 4.3 Validation du lissage

Comme énoncé précédemment, le lissage permet de supprimer les irrégularités de la courbe dues aux fluctuations liées à l'échantillonnage de la base de données. Cependant, le lissage doit également être suffisamment fidèle à la courbe d'origine.

Pour vérifier qu'un certain équilibre est respecté par le lissage, on peut d'abord comparer graphiquement les taux de sorties bruts et lissés. Ensuite, on doit s'appuyer sur des tests statistiques, plus rigoureux qu'une comparaison graphique, telles que les intervalles de confiance et le test du Khi 2.

#### 4.3.1 Intervalle de confiance de la fonction de survie

On rappelle que l'estimateur de Greenwood est convergent pour la variance asymptotique de l'estimateur de Kaplan Meier. Il permet avec la normalité asymptotique de l'estimateur de Kaplan Meier de calculer des intervalles de confiance (asymptotique) dont les bornes sont, pour la valeur de survie  $T_i$  :

$$\left[ S_i * \left( 1 - u_{1-\frac{\alpha}{2}} \gamma(T_i) \right); S_i * \left( 1 + u_{1-\frac{\alpha}{2}} \gamma(T_i) \right) \right]$$

Correspondant donc à

$$\left[ S_i * \left( 1 - u_{1-\frac{\alpha}{2}} \sqrt{\frac{d_1}{r_1(r_1 - d_1)} + \dots + \frac{d_i}{r_i(r_i - d_i)}} \right); S_i * \left( 1 + u_{1-\frac{\alpha}{2}} \sqrt{\frac{d_1}{r_1(r_1 - d_1)} + \dots + \frac{d_i}{r_i(r_i - d_i)}} \right) \right]$$

Pour la suite, on choisit de calculer des intervalles avec un niveau de confiance de 95%, donc :

$$u_{1-\frac{\alpha}{2}} = 1,96$$

Une fois le lissage effectué, on obtient de nouvelles valeurs de la fonction de survie ( $S_x$ ) pour chaque âge et ancienneté censées être proche des anciennes valeurs brutes obtenues tout en supprimant les irrégularités de la courbe dues aux fluctuations liées à l'échantillonnage. Si toutes les nouvelles valeurs de  $S_x$  se trouvent dans les intervalles de confiance construit précédemment, alors on peut valider le lissage effectué.

### 4.3.2 Test khi 2

La validité du lissage peut également être testée au travers d'une version ad hoc du test du Khi 2.

On effectue des regroupements sur les âges et sur les anciennetés, pour obtenir des classes bidimensionnelles. Le test du Khi 2 reste valable dans ce contexte puisque la somme de deux variables indépendantes avec une distribution du Khi 2 est une variable du Khi 2, de degré de liberté la somme des degrés de liberté. On obtient la statistique :

$$W_{k,l} = \sum_{x=1}^l \sum_{t=1}^k \frac{(D_{x,t} - \widetilde{D}_{x,t})^2}{\widetilde{D}_{x,t}}$$

Avec  $D_{x,t}$ , le nombre observé d'individus sortis entre  $t^+$  et  $t + 1$  et  $\widetilde{D}_{x,t}$  le nombre d'individus sortis entre  $t^+$  et  $t + 1$  prévu par le modèle. On peut montrer que si  $L_{x,0} \rightarrow \infty$ , alors  $W_{k,l}$  est asymptotiquement distribué comme une variable  $\chi_{k+l-2}^2$ . D'où le test consistant à rejeter l'adéquation des données brutes à notre ajustement si la réalisation de  $W_{k,l}$  est trop grande, c'est-à-dire supérieure à une valeur qui n'a qu'une probabilité  $\alpha$  d'être dépassée.

Par exemple, si on regroupe les âges en 9 classes et les anciennetés en 12 classes,  $W_{k,l}$  est distribué comme une variable  $\chi_{9+12-2}^2 = \chi_{19}^2$ . On obtient alors la région critique pour un seuil de 95%  $\{W_{k,l} > 10,1\}$ .

## 4.4 Méthodes de lissage

Il existe deux grands types de lissage (appelé également ajustement) :

- Les lissages paramétriques : qui reposent sur une hypothèse de distribution.

- Les lissages non paramétriques : qui ne reposent pas sur une telle hypothèse et qui consistent en un certain nombre de traitements à appliquer aux taux bruts.

Nous présenterons trois méthodes de lissage : Moyenne Mobile, Whittaker-Henderson et Splines.

#### 4.4.1 Lissage par Moyenne Mobile

Le lissage par moyenne mobile est une méthode non paramétrique qui est intuitive et facilement implémentable.

Le principe est de considérer que le nouveau taux de sortie lissé  $q_{x,t}^{lissé}$  est égal à une moyenne des taux bruts voisins de  $q_{x,t}^{brut}$ .

Par exemple, on peut considérer que le taux de sortie lissé  $q_{x,t}^{lissé}$  est égal à :

$$q_{x,t}^{lissé} = \frac{1}{2}q_{x,t}^{brut} + \frac{1}{2} \left( \frac{1}{8}q_{x-1,t-1}^{brut} + \frac{1}{8}q_{x-1,t}^{brut} + \frac{1}{8}q_{x-1,t+1}^{brut} + \frac{1}{8}q_{x,t-1}^{brut} + \frac{1}{8}q_{x,t+1}^{brut} + \frac{1}{8}q_{x+1,t-1}^{brut} + \frac{1}{8}q_{x+1,t}^{brut} + \frac{1}{8}q_{x+1,t+1}^{brut} \right)$$

Plus on sélectionne des points voisins à  $q_{x,t}^{brut}$  pour déterminer  $q_{x,t}^{lissé}$ , plus la table des  $q_{x,t}^{lissé}$  sera lissé.

Cependant, cette méthode présente de nombreux inconvénients qui conduisent à être prudent vis-à-vis de cette méthode.

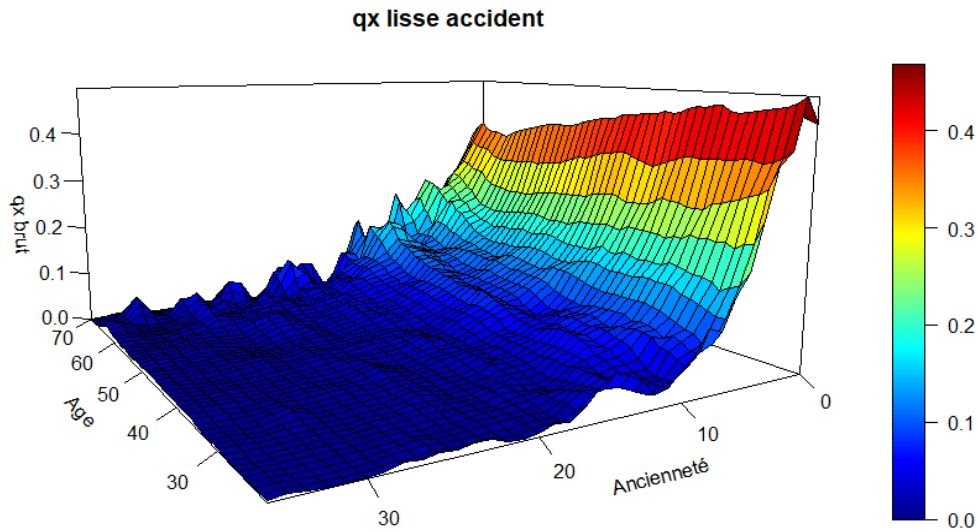
Premièrement, le choix du nombre de points à sélectionner et le coefficient attribué à chacun est arbitraire. Ensuite, cette méthode ne permet d'effectuer qu'un lissage « local » puisqu'on ne s'appuie pas sur tous les points de la table pour construire un  $q_{x,t}^{lissé}$ .

De plus, la moyenne arithmétique est sensible aux valeurs extrêmes.

Enfin, le lissage par moyenne mobile est mal adapté au traitement des bords de la table car on ne peut s'appuyer que sur un nombre plus faible de point voisins.

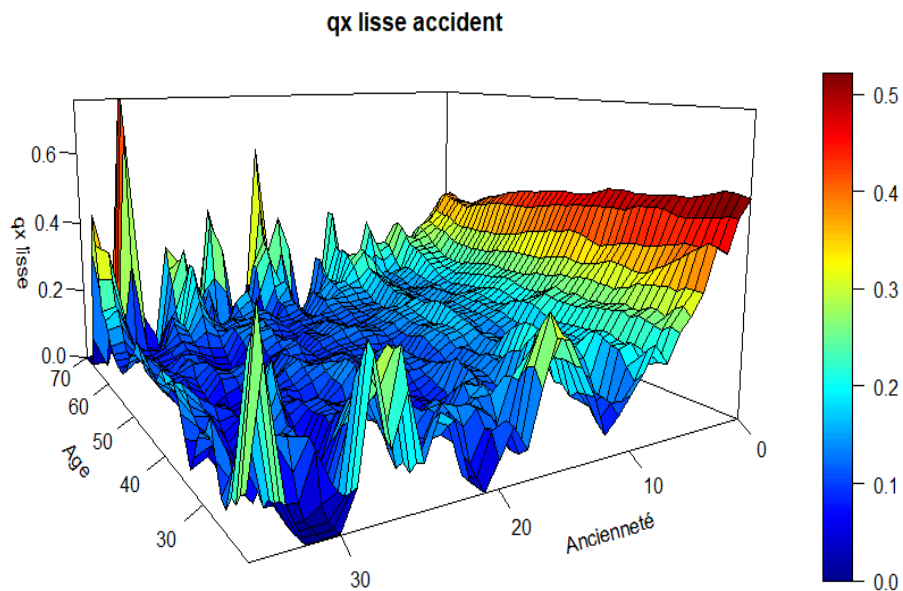
Le graphique ci-dessous (Figure 4.3) représente le lissage des taux de sortie par moyenne mobile en reprenant l'exemple ci-dessus.





**Figure 4.3 : Exemple 1 de lissage par la méthode des Moyenne Mobiles**

On constate que les taux de sorties sont beaucoup trop lisses comparé à nos taux de sorties bruts. On décide de faire un autre lissage par moyenne mobile en prenant cette fois ci moins de points voisins afin d'obtenir un lissage plus faible. Le résultat est présenté dans le graphique ci-dessus :



**Figure 4.4 : Exemple 2 de lissage par la méthode des Moyenne Mobiles**

En excluant les bords, on constate des taux de sortie qui sont plus proche de la réalité de nos taux de sortie bruts. En revanche, sur les bords, les taux de sorties sont toujours un peu trop « bruts ». Cela met en évidence les inconvénients cités précédemment de cette méthode.

De part le nombre important d'inconvénients présent dans cette méthode, nous ne retiendrons pas de lissage effectué par cette méthode.

#### 4.4.2 Lissage par la méthode de Whittaker-Henderson

La méthode de Whittaker-Henderson est une méthode de lissage non paramétrique. Elle consiste à combiner un critère de fidélité et un critère de régularité. La minimisation de ces deux critères permet d'obtenir le lissage des données.

La méthode de Whittaker-Henderson en dimension 1 se présente donc sous la forme suivante :

##### Cas de la dimension 1

On se fixe des poids  $(w_i)_{1,\dots,p}$  et on pose pour le critère de fidélité :

$$F = \sum_{i=1}^p w_i (q_i - \hat{q}_i)^2$$

Et pour le critère de régularité

$$S = \sum_{i=1}^{p-z} (\Delta^z q_i)^2$$

$z$  étant un paramètre du modèle. Le critère à minimiser est une combinaison linéaire de la fidélité et de la régularité, le poids de chacun des deux termes étant contrôlé par un second paramètre  $h$  :

$$M = F + h * S$$

La solution de ce problème d'optimisation satisfait aux conditions  $\frac{\partial M}{\partial q_i} = 0, 1 \leq i \leq p$  ; la résolution de ce système d'équations peut être effectuée au moyen de quelques manipulations matricielles. A cet effet, on pose  $q = (q_i)_{1 \leq i \leq p}$ ,  $\hat{q} = (\hat{q}_i)_{1 \leq i \leq p}$  et  $w = \text{diag}(w_i)_{1 \leq i \leq p}$  ; avec ces notations on peut écrire  $F = (q - \hat{q})' w (q - \hat{q})$  ; pour ce qui concerne le critère de régularité, si on note  $\Delta^z q = (\Delta^z q_i)_{1 \leq i \leq p-z}$ , alors  $S = (\Delta^z q)' \Delta^z q$ . Pour

détailler cette écriture, on introduit la matrice  $K_z$  de taille  $(p - z, p)$ , dont les termes sont les coefficients binomiaux d'ordre  $z$  dont le signe alterne et commence positivement pour  $z$  pair :

$$\Delta^z q_i = \sum_{j=0}^z \binom{z}{j} (-1)^{z-j} q(j+i)$$

Par exemple, pour  $z = 1$  et  $p = 3$ , on a :

$$K_1 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

On vérifie aisément que  $\Delta^z q = K_z q$ , ce qui permet finalement d'écrire le critère  $M$  sous la forme :

$$M = (q - \hat{q})' w (q - \hat{q}) + h q' K_z' K_z q$$

En développant l'expression ci-dessus, on trouve que :

$$M = q' w q - 2 q' w \hat{q} + \hat{q}' w \hat{q} + h q' K_z' K_z q$$

Ce qui conduit à :

$$\frac{\partial M}{\partial q} = 2 w q - 2 w \hat{q} + 2 h K_z' K_z q$$

La résolution de  $\frac{\partial M}{\partial q} = 0$  conduit alors à l'expression des taux ajustés :

$$q^* = (w + h K_z' K_z)^{-1} w \hat{q}$$

L'inversion de la matrice  $C = w + h K_z' K_z$  nécessite toutefois certaines précautions, car  $h K_z' K_z$  n'est pas inversible, et l'addition du terme  $w$  rend  $C$  inversible mais de ce fait l'inversion de  $C$  peut être délicate. On peut en pratique utiliser la décomposition de Cholesky de la matrice symétrique positive  $C$  pour l'inverser.

Cependant, nous sommes dans une problématique de maintien en incapacité qui est bi dimensionnelle (âge et ancienneté). Il serait faux de lisser, selon Whittaker-Henderson en dimension 1, notre série de taux de sortie par rapport à une dimension puis de lisser de nouveau par rapport à l'autre dimension car une telle approche ne capturerait pas les dépendances entre les 2 composantes des taux de sortie.

Il faut donc étendre la méthode de Whittaker-Henderson en dimension 2 pour résoudre notre problématique de maintien bidimensionnelle.

L'extension de la méthode de Whittaker-Henderson en dimension 2 ne pose pas de problème majeur. On dispose donc d'estimations  $\hat{q} = (\hat{q}_i)_{1 \leq i \leq p, 1 \leq j \leq q}$  ; le critère de fidélité se généralise immédiatement :

$$F = \sum_{i=1}^p \sum_{j=1}^q w_{ij} (q_{ij} - \hat{q}_{ij})^2$$

L'extension en dimension 2 du critère de régularité est un peu plus délicate ; on distingue d'abord la régularité verticale via l'opérateur  $\Delta_v^z$  (qui agit sur  $q_{ij}$  à  $j$  fixé vu comme une série indicé par  $i$ ) qui permet de calculer un indice de régularité verticale :

$$S_v = \sum_{j=1}^q \sum_{i=1}^{p-z} (\Delta^z q_{ij})^2$$

De la même manière, on calcule l'indice de régularité horizontale  $S_h$ , puis on pose

$$M = F + \alpha * S_v + \beta * S_h$$

Qui doit être minimisé. La résolution du problème d'optimisation s'effectue en réarrangeant les éléments pour se ramener au cas unidimensionnel. Pour cela, on définit le vecteur de taille  $p * q$ ,  $u$  tel que  $u_{q(i-1)+j} = \hat{q}_{ij}$  ; cela revient à prendre pour les  $q$  premiers éléments du vecteur  $u$  la première ligne de la matrice  $\hat{q}$ , puis ensuite les éléments de la seconde ligne, et ainsi de suite. De même on fabrique une matrice de poids en copiant sur la diagonale les lignes de la matrice  $(w_{ij})$ .

On pose donc  $w_{q(i-1)+j, q(i-1)+j}^* = w_{ij}$ .

On procède de la même manière pour définir les matrices  $K_z^v$  et  $K_y^h$ . Les valeurs lissées s'obtiennent alors par :

$$q^* = (w^* + \alpha K_z^{v'} K_z^v + \beta K_y^{h'} K_y^h)^{-1} w^* u$$

Afin d'appliquer cette méthode, les paramètres suivants doivent être renseignés :

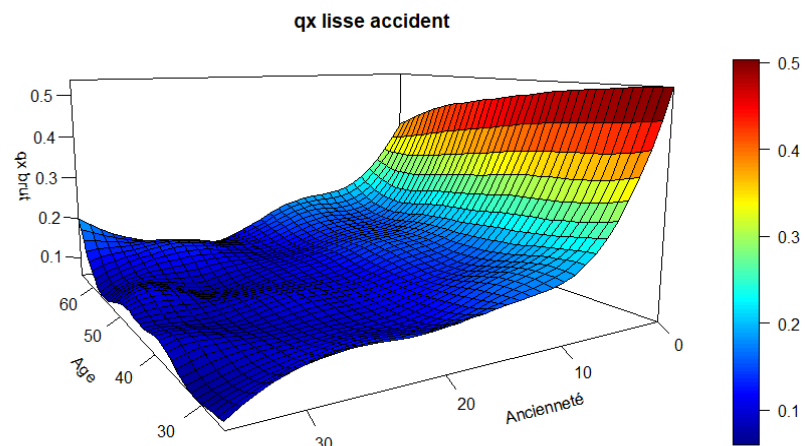
- Les ordres de régularité horizontal et vertical : permettent de définir à quelle échelle (cad à quel ordre des dérivées des taux bruts), on souhaite lisser la courbe. Ils permettent de donner une importance plus ou moins grande à la régularité.
- Les paramètres  $\alpha$  et  $\beta$  : permettent de donner une importance plus ou moins grande de la régularité selon l'ancienneté ou l'âge d'entrée.
- La matrice de poids : elle permet de réajuster les différences dues à l'échantillonnage de la base de données.

#### Choix des valeurs des paramètres :

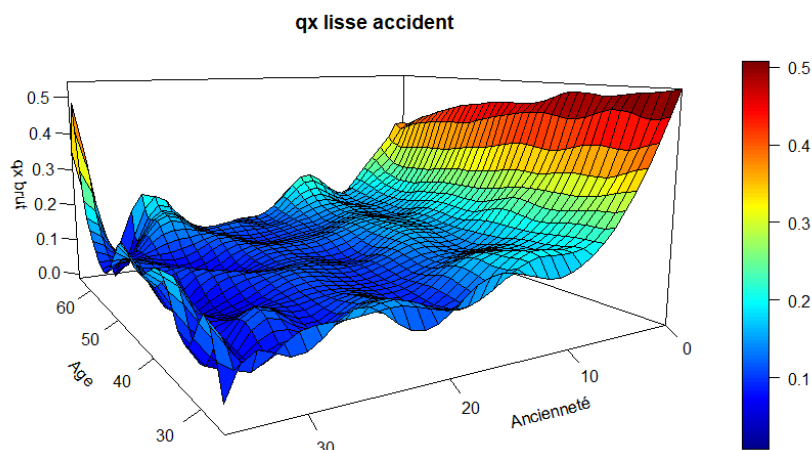
Concernant, la matrice des poids, il est conseillé de s'appuyer sur les effectifs initiaux à chaque âge en proportion du total. Ainsi âge par âge, nous affectons le même poids à chaque ancienneté. En effet, la matrice des poids correspond à une pondération du critère de fidélité. Il est donc logique d'affecter un poids plus important à un âge contenant beaucoup de personnes en arrêt plutôt qu'à un âge contenant peu de personnes en arrêt.

Concernant les autres paramètres, ils sont choisis par tâtonnement jusqu'à ce qu'on obtienne une table suffisamment lissée et validée selon différents tests statistiques présentés précédemment.

On décide d'afficher ci-dessous (figure 4.5 et 4.6) différents graphiques de taux de sorties lissés selon la méthode de Whittaker-Henderson avec des paramètres différents afin de montrer les différences de lissages obtenus. Les résultats seront discutés dans une partie ultérieure une fois que l'on aura présenté toutes les méthodes de lissage.



**Figure 4.5 : Exemple 1 de lissage par la méthode W-H selon les paramètres suivants :  
Alpha=0.5 Beta=0.5 Ordre horizontal=1 Ordre vertical=1**



**Figure 4.6 : Exemple 2 de lissage par la méthode W-H selon les paramètres suivants :  
Alpha=0.05 Beta=0.05 Ordre horizontal=4 Ordre vertical=3**

On peut observer que la courbe des taux de sortie représentés par la Figure 4.5 est beaucoup plus lisse que la courbe des taux de sortie représentés par la Figure 4.6 confirmant l'impact des quatre paramètres dans la régularité de la courbe.

## Extrapolation

Il s'agit ici de déterminer les taux de sorties lissés aux âges considérées comme volatiles et qui n'ont pas été lissés par la méthode de Whittaker-Henderson. Dans notre étude, il s'agit des âges allant de 21 à 25 ans et de 66 à 70 ans.

L'extrapolation pour le maintien en incapacité s'appuie en général sur la table du BCAC car il s'agit de la table règlementaire basée sur une étude statistique rigoureuse. Par exemple, selon Leurent (2010), deux méthodes d'extrapolation sont proposées aux âges extrêmes en lien avec la table du BCAC. Le but, à chaque fois, est de trouver une relation mathématique entre les valeurs de la table du BCAC et les valeurs de la table lissée (dans notre cas entre 26 et 65 ans). Et ensuite, grâce à cette relation en déduire les valeurs extrapolées aux âges extrêmes. Par exemple, en utilisant la méthode statistique des moindres carrées ordinaires, on cherche les  $\alpha$  et  $\beta$  minimisant :

$$\sum_{x=26}^{65} (\alpha_i + \beta_i * q_i^{BCAC}(x) - q_i(x))^2$$

Une fois obtenues, on en déduit les taux de sorties pour les âges extrêmes :

$$q_i^*(x) = \alpha_i + \beta_i * q_i^{BCAC}(x)$$

pour  $x$  allant de 21 à 25 ans et de 66 à 70 ans.

Cependant, dans notre cas, il nous est impossible de nous appuyer sur la table du BCAC car elle s'arrête à l'âge de 65 ans et notre table va jusqu'à 70 ans. De plus, en comparant graphiquement notre table lissée de taux de sortie (entre 26 et 65 ans) avec la table du BCAC, il paraît compliquer d'en déduire une relation solide entre nos deux tables.

En analysant de plus près les âges extrêmes de notre table, on a constaté que la valeur élevée de la variance de Greenwood pour ces âges s'expliquait par le faible nombre d'observations comparées aux autres âges de notre table. On a également constaté qu'aux âges les plus faibles et aux âges les plus élevés, les comportements intra étaient similaires. En effet, on représente ci-dessous, dans le premier graphique (Figure 4.7) l'évolution des taux de sortie bruts à l'âge de 21 et 22 ans et dans le second graphique (Figure 4.8), l'évolution des taux de sortie bruts à l'âge de 23 et de 50 ans.

### Evolution qx entre 1 et 36 mois d'ancienneté

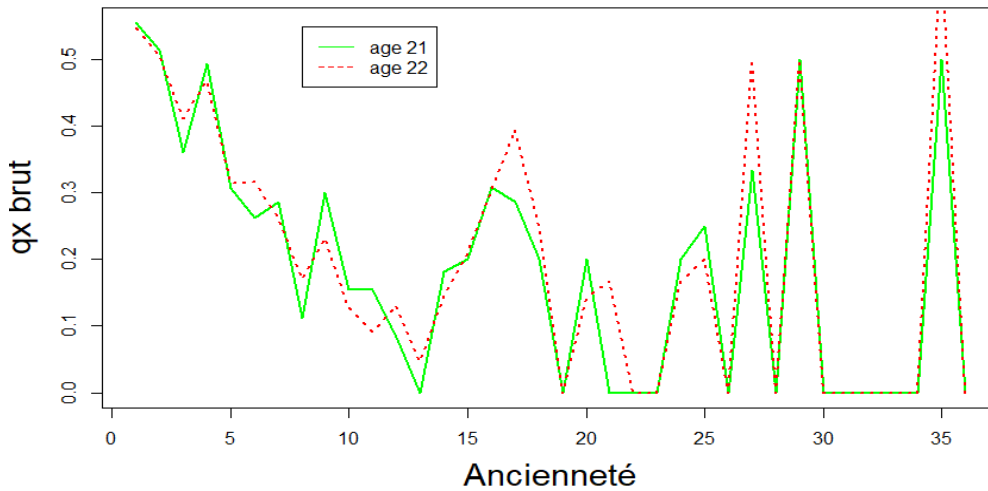


Figure 4.7 : Evolution du taux de sortie conditionnel à l'âge 21 et 22 ans

### Evolution qx entre 1 et 36 mois d'ancienneté

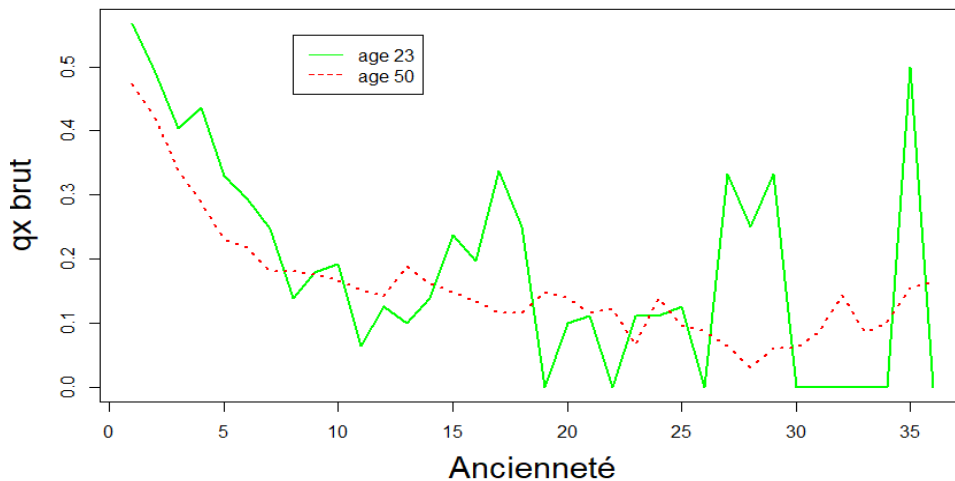


Figure 4.8 : Evolution du taux de sortie conditionnel à l'âge 23 et 50 ans

La variable ( $q_x$ ) est la plus volatile parmi les différentes variables à notre disposition ( $S_x$ ,  $L_x$ ,  $d_x$ ). Pourtant, on remarque bien une certaine relation entre le taux de sortie à 21 ans et 22 ans contrairement à 23 ans comparé à un âge beaucoup plus élevé : 50 ans. Les taux de sorties conditionnelles entre 21 et 25 ans sont visuellement proches les uns des autres. Il en est de même pour les âges entre 66 et 70 ans.

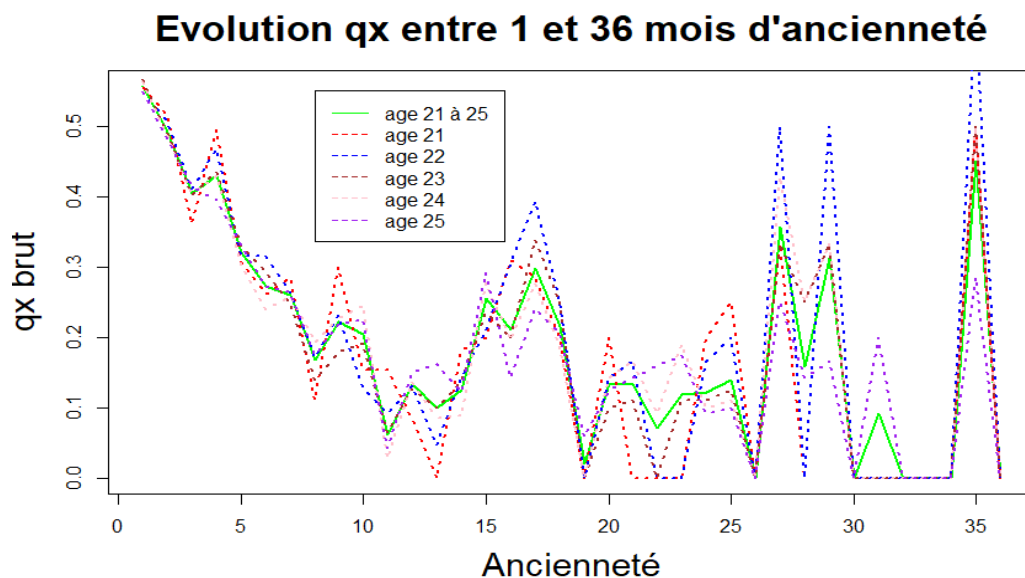
On décide donc de regrouper les âges allant de 21 à 25 ans et ceux de 66 à 70 ans. En d'autres termes, on construit une nouvelle loi de maintien commune pour les âges allant de 21 à 25 ans et ceux de 66 à 70 ans. La méthode utilisée pour cette construction est celle de Kaplan Meier. Suite à cela, on calcule de nouveau la variance de Greenwood et on constate qu'elle

baisse considérablement dans nos deux nouvelles classes d'âges (21-25 et 66-70) comparé aux valeurs trouvées précédemment lorsque ces âges extrêmes n'étaient pas regroupés. Le tableau suivant (Figure 4.9) montre bien la diminution de l'estimateur de Greenwood suite à la création des deux classes d'âges :

Age	Ancienneté (en mois)				
	1	10	19	25	32
23 ans	0,000146888	0,000904265	0,00285264	0,02991128	0,28412106
21-25 ans	7,54647E-05	0,000403107	0,00126824	0,01150517	0,09338186
66-70 ans	9,38962E-05	0,000557899	0,00193123	0,01769268	0,1462153
66 ans	0,000536025	0,00294525	0,00931195	0,06692202	0,54517797

**Figure 4.9 : Calcul de l'estimateur de Greenwood à plusieurs âges et mois d'ancienneté**

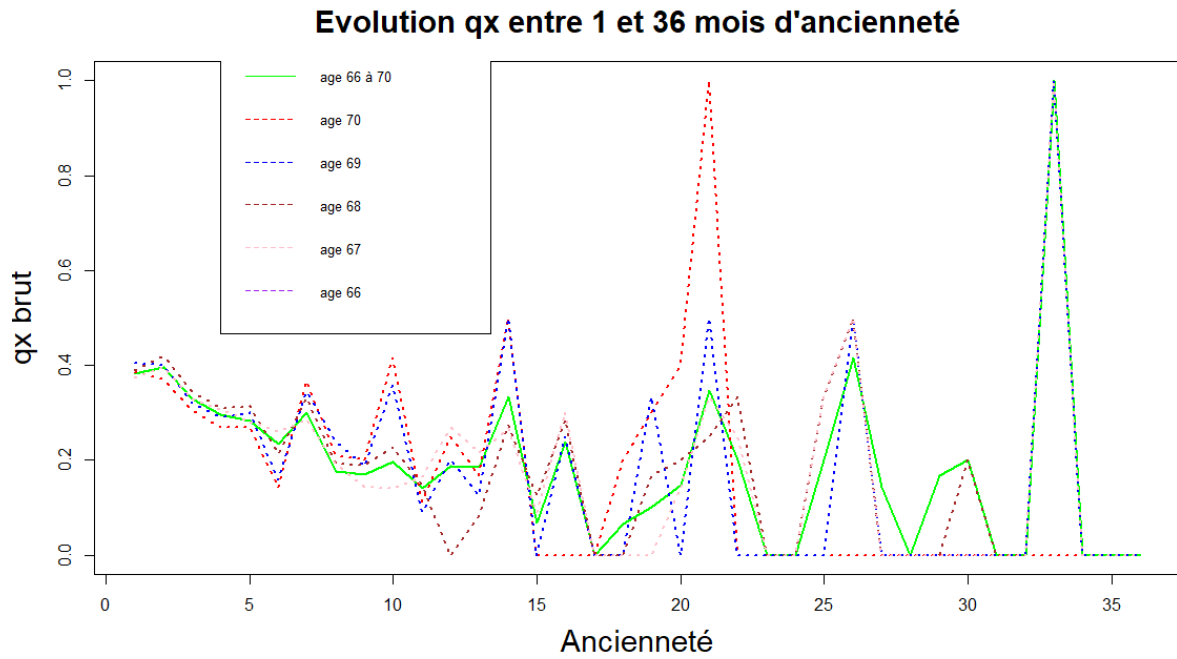
En Figure 4.10, on représente le nouveau taux de sortie conditionnel brut suite au regroupement des âges 21-25 ans ainsi que les taux de sortie conditionnels brut allant de 21 à 25 ans.



**Figure 4.10 : Evolution du taux de sortie conditionnel pour les âges allant de 21 à 25 ans et pour l'âge regroupé 21-25.**

En Figure 4.11, on représente le nouveau taux de sortie conditionnel brut suite au regroupement des âges 66-70 ans ainsi que les taux de sortie conditionnels brut allant de 66 à 70 ans.



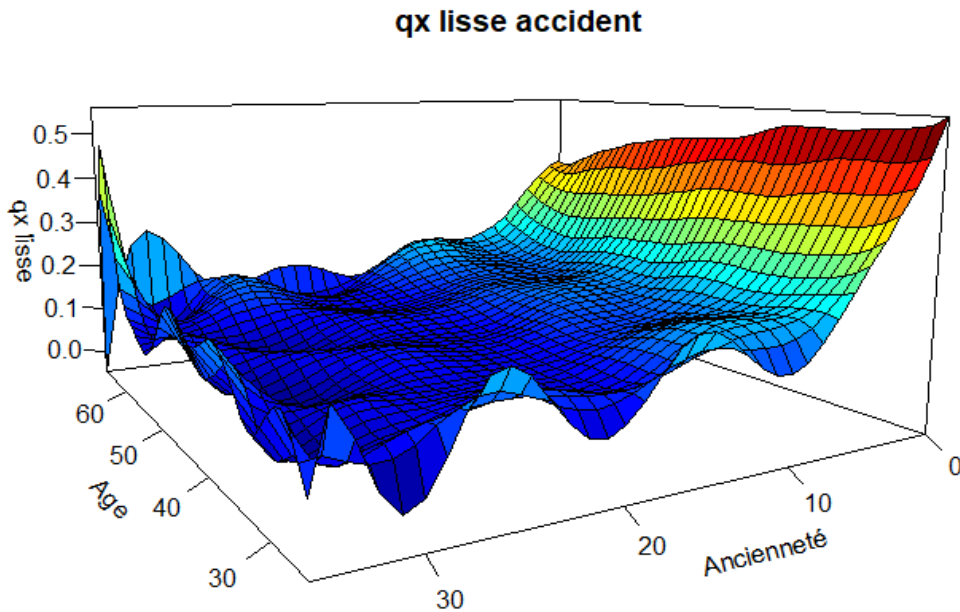


**Figure 4.11 : Evolution du taux de sortie conditionnel pour les âges allant de 66 à 70 ans et pour l'âge regroupé 66-70.**

Que ce soit en Figure 4.10 ou 4.11, on observe bien une relation des taux de sorties dans ces deux classes d'âge.

Par conséquent, on décide de lisser l'ensemble de notre table en faisant deux regroupements aux âges extrêmes.

Le graphique suivant (Figure 4.12) représente les taux de sortie lissé avec regroupement selon Whittaker-Henderson avec les paramètres  $\alpha = 0,05$  ;  $\beta = 0.05$  ;  $z = 4$  ;  $v = 3$ .



**Figure 4.12 : Lissage du taux de sortie conditionnel par la méthode WH selon les paramètres suivants :  $\alpha = 0,05$  ;  $\beta = 0,05$  ;  $z = 4$  ;  $v = 3$**

Les résultats des lissages par Whittaker-Henderson en regroupant les âges 21-25 et 66-70 seront présentés dans une partie ultérieure nommée « Résultats » (section 4.5).

#### 4.4.3 Méthode des splines

Le lissage par splines est une méthode paramétrique, c'est à dire qu'on suppose que l'ajustement repose sur une hypothèse de distribution.

Une méthode paramétrique signifie implicitement que la courbe des taux de sorties peut être représentée sur toute la plage d'ancienneté considéré par une seule fonction paramétrique. Or en réalité, comme on a pu le constater à travers la Figure 4.1 lorsqu'on a représenté les taux de sorties bruts graphiquement, il paraît très compliqué de trouver une fonction paramétrique « simple » capable de prendre en compte l'ensemble des ruptures constatées des taux bruts.

L'idée du lissage par splines est de découper la plage de la fonction à ajuster en sous intervalles, puis d'ajuster sur chaque sous intervalle une fonction « simple », en prenant des précautions pour le raccordement aux points de jonction.

Le lissage par splines s'affranchit donc de la condition restrictive de s'appuyer sur une seule fonction qui aurait été « complexe » dans notre cas, au profit de plusieurs fonctions « simples ».

Les fonctions « simples » utilisées dans le lissage par splines sont généralement des polynômes. En pratique, on considère en général des polynômes de degré 3 qui vont nous permettre de construire des splines cubiques. Le raccordement de ces arcs se fera en imposant aux points de jonction (appelé également nœuds) la continuité ainsi que l'égalité des pentes et des courbures.

Dans le cas d'un lissage par spline de la courbe des taux de sortie où l'on découpe la plage à ajuster en 2 parties seulement, on obtient 2 fonctions  $p_0(x)$  et  $p_1(x)$  et un nœud  $x_1$  tel que :

$$q_x = \begin{cases} p_0(x) & x_0 \leq x \leq x_1 \\ p_1(x) & x_1 \leq x \leq x_2 \end{cases}$$

$p_i(x)$  étant un polynôme de degré 3, avec des contraintes au nœud  $x_1$ . Le problème sous contrainte s'écrit sous la forme suivante :

$$p_0(x_1) = p_1(x_1) \quad \frac{d}{dx} p_0(x_1) = \frac{d}{dx} p_1(x_1) \quad \frac{d^2}{dx^2} p_0(x_1) = \frac{d^2}{dx^2} p_1(x_1)$$

#### En dimension 1 :

Le graphique suivant (Figure 4.13) est un exemple de lissage par spline cubique en dimension 1 du taux de sortie à l'âge de 32 ans.

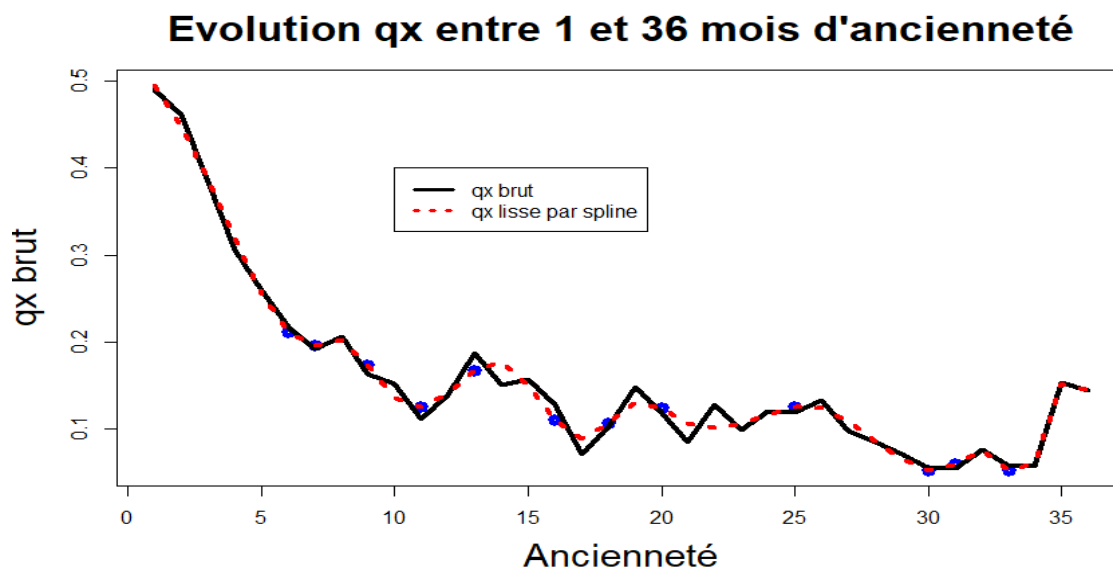


Figure 4.13 : Exemple de lissage par spline cubique en dimension 1

Dans le cas de la dimension 1, la plage à découper en sous intervalle correspond à l'ancienneté entre 1 et 36 mois. Chaque point bleu correspond à un nœud choisi selon l'allure du taux de

sortie afin de « coller » au mieux avec cette dernière. Dans notre exemple, on a sélectionné 12 nœuds, ce qui fait 13 polynômes de degré 3 construits.

## En dimension 2 :

Selon PLANCHET, WINTER (2010), la formalisation du lissage par spline en dimension 2 est un peu plus délicate qu'en dimension 1 et nécessite un minimum de formalisme, détaillé ci-après. On peut simplement retenir que l'idée est de fixer arbitrairement des nœuds pour subdiviser la surface à ajuster en zones. Ensuite, on ajuste un polynôme de degré 3 (en dimension 2) à chaque subdivision, en utilisant un critère de type moindres carrés. Enfin pour des raisons évidentes de continuité et de régularité, ce « patchwork » de polynômes de degré 3 est contraint à être de classe  $C^2$  (ou  $C^1$ ).

Formalisation

Soit  $P(t, x)$  un polynôme de degré  $n$  en fonction de  $t$  et  $x$ . Il est entièrement défini par la donnée de  $(n + 1)^2$  paramètres :

$$P(t, x) = \sum_{(i,j) \in \{0, \dots, n\}^2} a_{ij} t^i x^j$$

On note à présent  $A = (a_{ij})_{(i,j) \in \{0, \dots, n\}^2}$ , la matrice des coefficients du polynôme  $P(t, x)$ . On note également

$$T_n = \begin{pmatrix} 1 \\ \vdots \\ n \end{pmatrix} \text{ et } X_n = \begin{pmatrix} 1 \\ \vdots \\ x^n \end{pmatrix} \text{ et } X'_n = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ nx^{n-1} \end{pmatrix} \text{ le vecteur dérivé terme à terme. On obtient}$$

ainsi la représentation matricielle du polynôme :

$${}^t T_n \cdot A \cdot X_n = P(t, x)$$

Les conditions de régularité nécessitent le calcul des dérivées partielles premières et secondes de  $P(t, x)$ . On obtient facilement :

$$\frac{\partial P}{\partial t}(t, x) = \sum_{(i,j) \in \{1, \dots, n\} * \{0, \dots, n\}} i a_{ij} t^{i-1} x^j \quad \frac{\partial P}{\partial x}(t, x) = \sum_{(i,j) \in \{1, \dots, n\} * \{0, \dots, n\}} j a_{ij} t^i x^{j-1}$$

$$\begin{aligned} \frac{\partial^2 P}{\partial t^2}(t, x) &= \sum_{(i,j) \in \{2, \dots, n\} * \{0, \dots, n\}} i(i-1) a_{ij} t^{i-2} x^j \\ &= \sum_{(i,j) \in \{0, \dots, n\} * \{2, \dots, n\}} j(j-1) a_{ij} t^i x^{j-2} \end{aligned}$$

$$\frac{\partial^2 P}{\partial t \partial x}(t, x) = \sum_{(i,j) \in \{1, \dots, n\} * \{1, \dots, n\}} i j a_{ij} t^{i-1} x^{j-1}$$

Les deux premières sommes comportent  $n(n + 1)$  termes, les deux suivantes  $(n - 1)(n + 1)$  et la dernière  $n^2$  termes. Ces expressions s'écrivent simplement sous forme matricielle :

$$\frac{\partial P}{\partial t}(t, x) = {}^t T_n \cdot A \cdot X_n$$

$$\frac{\partial P}{\partial x}(t, x) = {}^t T_n \cdot A \cdot X'_n$$

$$\frac{\partial^2 P}{\partial t^2}(t, x) = {}^t T''_n \cdot A \cdot X_n$$

$$\frac{\partial^2 P}{\partial x^2}(t, x) = {}^t T_n \cdot A \cdot X''_n$$

$$\frac{\partial^2 P}{\partial t \partial x}(t, x) = {}^t T_n \cdot A \cdot X'_n$$

Les nœuds correspondent aux points intérieurs de la table. Ces points doivent être choisis le plus judicieusement possible en fonction de la forme et de l'allure générale de la surface à représenter. Par ailleurs, si le nombre de nœuds choisis est trop faible, cela conduit à un lissage peu fidèle et inversement si le nombre de nœuds choisis est trop élevé, cela produit un lissage trop fidèle et donc irrégulier.

Pour plus de simplicité, on ne définit ici que des droites horizontales et verticales, distinctes des bords de la table. Les intersections de ces droites forment alors les nœuds.

Si on définit  $h$  droites horizontales et  $v$  droites verticales, on obtient  $h v$  nœuds, et la table est subdivisée en  $(h + 1)(v + 1)$  parties. On indice alors ces parties avec  $k = 0, \dots, h$  et  $l = 0, \dots, v$ , ce qui conduit à  $(h + 1)(v + 1)$  polynômes à estimer et à recoller.

Dans notre problématique d'ajustement, la fonction brute est connue et donc représentable graphiquement. Une manière de choisir les nœuds est alors l'analyse préalable des données brutes. Les grandes tendances pourront être conservées avec un maillage épais ; néanmoins, certaines irrégularités (les pics en particulier) seront effacées. Pour éviter ceci, il faut encadrer ces irrégularités par la plus petite maille possible.

Une approche possible est donc de considérer que plus la variation de la fonction à ajuster est brutale, plus le maillage doit être petit si l'on veut conserver ces variations.

Cette approche empirique de sélection des nœuds peut être remplacée par une approche de sélection automatisé que nous présenterons par la suite.

### Conditions limites :

Pour des raisons de continuité, le raccordement entre les splines est soumis à des contraintes sur les dérivés des polynômes.

### Contraintes horizontales :

Si  $P^1$  et  $P^2$  sont adjacents horizontalement alors leurs valeurs et leurs dérivées partielles premières et secondes en fonction de  $t$  doivent être égales, ceci pour tout âge variant sur l'ensemble de définition induit par le spline. En notant  $(a_{i,j})$  et  $(b_{i,j})$  les paramètres des splines adjacents, ces contraintes s'écrivent :

$$\sum_{j=0}^n \sum_{i=0}^n a_{ij} t^i x^j = \sum_{j=0}^n \sum_{i=0}^n b_{ij} t^i x^j \quad \text{et} \quad \sum_{j=0}^n \sum_{i=0}^n a_{ij} t^i x^j = \sum_{j=0}^n \sum_{i=1}^n i b_{ij} t^{i-1} x^j$$

Avec  $t$  fixé et pour  $x$  variant dans l'ensemble de définition du spline. Cela conduit à écrire :

$$\begin{aligned} & \sum_{j=0}^n x^j \sum_{i=0}^n (a_{ij} - b_{ij}) t^i = 0 \\ \Leftrightarrow & \sum_{i=0}^n (a_{ij} - b_{ij}) t^i = 0 \quad \forall j = 0, \dots, n \Leftrightarrow {}^t T_n \cdot (A - B) = (0, \dots, 0) \end{aligned}$$

Et

$$\begin{aligned} & \sum_{j=0}^n x^j \sum_{i=0}^n (a_{ij} - b_{ij}) i t^{i-1} = 0 \\ \Leftrightarrow & \sum_{i=1}^n (a_{ij} - b_{ij}) i t^i = 0 \quad \forall j = 0, \dots, n \Leftrightarrow {}^t T_n \cdot (A - B) = (0, \dots, 0) \end{aligned}$$

Si  $t \neq 0$ , ces équations forment un système libre de  $2(n+1)$  équations avec  $2(n+1)^2$  paramètres.

### Contraintes verticales :

De même, si  $P^1$  et  $P^2$  sont adjacents horizontalement alors leurs valeurs et leurs dérivées partielles premières et secondes en fonction de  $x$  doivent être égales, ceci pour tout ancienneté variant sur l'ensemble de définition induit par le spline. Avec un raisonnement similaire à celui décrit supra, on obtient le système :

$$\begin{aligned} & \sum_{j=0}^n (a_{ij} - b_{ij}) x^j = 0 \quad \forall i = 0, \dots, n \Leftrightarrow (A - B) X_n = {}^t (0, \dots, 0) \\ & \sum_{j=1}^n (a_{ij} - b_{ij}) j t^j = 0 \quad \forall i = 0, \dots, n \Leftrightarrow (A - B) X'_n = {}^t (0, \dots, 0) \end{aligned}$$

### Contraintes aux nœuds : âge et ancienneté fixes

On cumule les deux contraintes précédentes. Par ailleurs, en un nœud, il se trouve non plus deux mais quatre polynômes adjacents.

### Contraintes en bord de table

En bord de table, il n'y a pas de contrainte de régularité, puisque le polynôme n'est pas raccordé à un autre.

### Choix du critère d'optimisation

On combine un critère de fidélité et un critère de régularité, comme décrit supra.

#### Critère de fidélité :

Ici,  $SP$  représente la surface des splines raccordés. Les polynômes ajustés doivent être le plus fidèle possible à la surface brute. Pour ceci, on utilise un critère des moindres carrés pour chaque spline :

$$\sum_x \sum_t \omega_{x,t} (SP(t, x) - q_{x,t})^2$$

Où  $\omega_{x,t}$  est le poids associé au point  $q_{x,t}$ . Notons que la sommation doit inclure les bornes, c'est-à-dire les points se trouvant sur les droites définissant l'emplacement des nœuds.

#### Critère de régularité

Comme dans le lissage de Whitaker-Henderson, on introduit un critère permettant de contrôler la régularité du spline. La fidélité à elle seule n'est pas suffisante pour l'obtention d'une courbe régulière. En effet, outre le choix du nombre de nœuds, plus une courbe est fidèle aux données brutes, moins elle sera régulière. Le meilleur moyen de contrôler la régularité est de minimiser la dérivée seconde des polynômes. Avec la norme  $L^2$ , cette contrainte s'écrit :

$$\int \int \lambda(s, y) |D^2 P(s, y)|^2 ds dy$$

Où  $\lambda(x, t)$  est une fonction continue par morceau, valant le poids associé au point  $q_{x,t}$  pour la régularité et  $D^2 P(s, y)$  la matrice Hessienne de la fonction polynomiale par morceau calculé au point  $(s, y)$ .

### Résolution

Le système à résoudre combine finalement la fidélité aux données brutes, la régularité de la surface ajustée et les contraintes aux nœuds. On combine les deux premières contraintes pour former la fonction à minimiser. La troisième constituera les contraintes du problème d'optimisation. On obtient alors le critère d'optimisation suivant :

$$\alpha \sum_x \sum_t \omega_{x,t} (SP(t, x) - q_{x,t})^2 + (1 - \alpha) \int \int \lambda(s, y) |D^2 P(s, y)|^2 ds dy$$

Le paramètre  $\alpha \in [0,1]$  permet de privilégier soit la régularité, soit la fidélité.

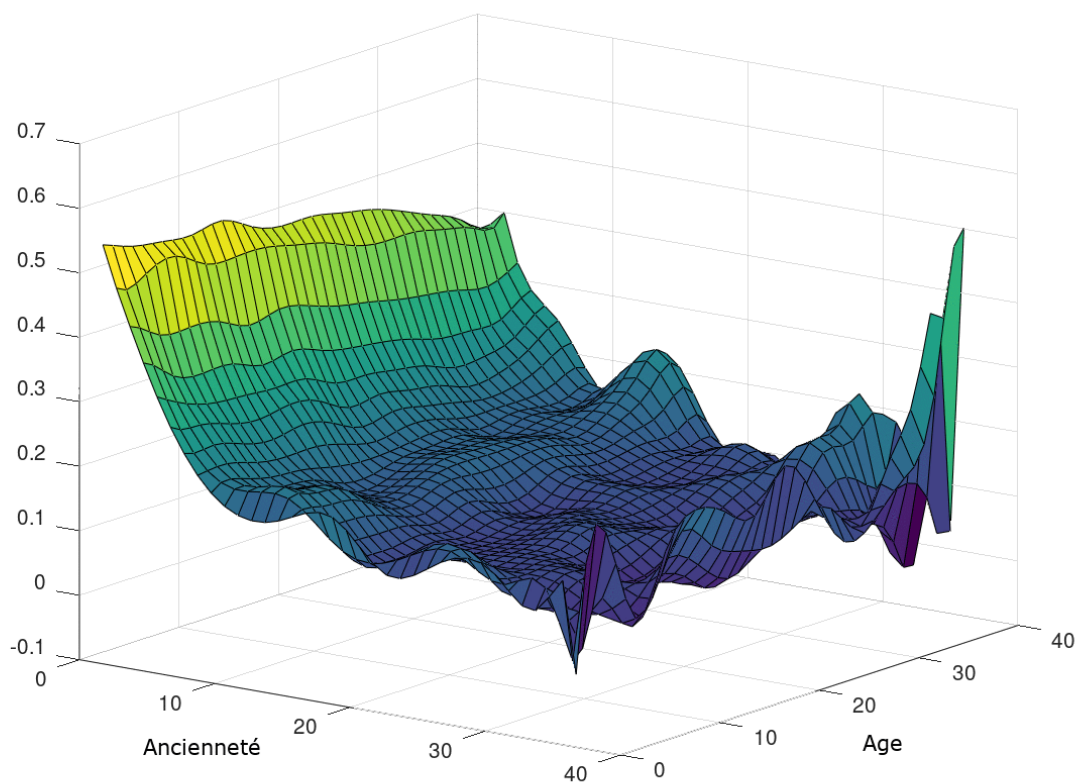
Il se trouve que la résolution de ce système est possible sur GNU Octave (logiciel gratuit comparable à Matlab), plus précisément grâce à la fonction « csaps ».

L'avantage de cette fonction est que les nœuds sont déterminés automatiquement par l'algorithme.

La matrice des poids utilisée est la même que pour la méthode Whittaker-Henderson, c'est-à-dire que la matrice des poids correspond aux effectifs initiaux à chaque âge en proportion du total.

La fonction continue par morceaux qui pondère le critère de régularité  $\lambda(x, t)$  est égale à 1.

Le graphique ci-dessous (Figure 4.14) représente le lissage des taux de sorties selon la méthode des splines pour  $\alpha = 0.0005$ .



**Figure 4.14 : Exemple de lissage par spline cubique en dimension 2**

En s'appuyant sur une valeur  $\alpha = 0.0005$ , on obtient un lissage assez similaire à la Figure 4.12.



Les splines, de par leur structure paramétrique, sont très pratiques pour faire des interpolations ou des extrapolations.

Par ailleurs, les splines fournissent une vraie surface, contrairement à la méthode de Whittaker Henderson qui ne fournit qu'un maillage. Avec une table estimée de la sorte, il est donc possible de calculer des provisions mathématiques dans le cas continu.

Les résultats des différents lissages effectués par la méthode des splines seront discutés dans la partie « Résultat » (section 4.5).

#### 4.4.4 Lissage sur les espérances de maintien résiduelle

L'idée du lissage sur les espérances de maintien résiduelle est de lisser non plus les taux de sorties mais sur les espérances de maintien résiduelles. La méthode de lissage utilisé dans le cadre de ce mémoire est Whittaker-Henderson mais on peut très bien utilisée une autre méthode tel que celle des splines par exemple.

L'espérance de maintien résiduelle s'écrit sous la forme suivante :

$$\ddot{e}_x(t) = \sum_{k=t}^{\infty} k-t p_{x,t} \sum_{k=t}^{\infty} \frac{S_{x,k}}{S_{x,t}}$$

Les espérances de maintien résiduelles correspondent à des provisions mathématiques calculées avec un taux d'intérêt nul.

L'intérêt de lisser les espérances de maintien résiduelles est de travailler avec une grandeur plus proche des provisions mathématiques.

On peut également écrire les  $\ddot{e}_x(t)$  sous la forme suivante :

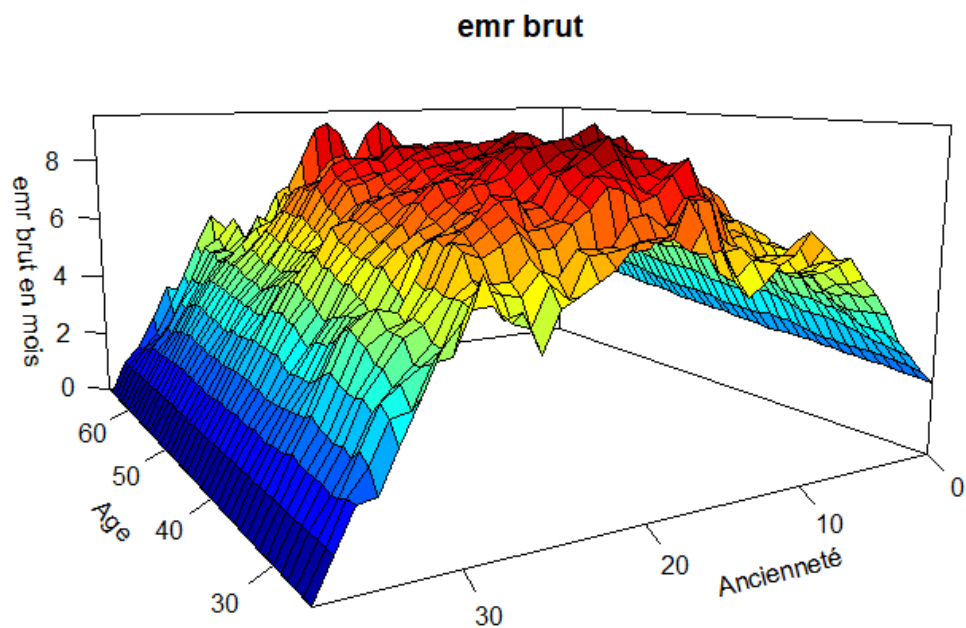
$$\ddot{e}_x(t) = \sum_{k=t}^{AncMax-1} \prod_{i=t}^k p_{x,i} \text{ pour } t = 0, \dots, AncMax - 1$$

On obtient alors :

$$p_{x,t} = \frac{\ddot{e}_x(t) - 1}{\ddot{e}_x(t+1)} \text{ pour } t = 0, \dots, AncMax - 2 \text{ et } p_{x,AncMax-1} = \ddot{e}_{x,AncMax-1} - 1$$

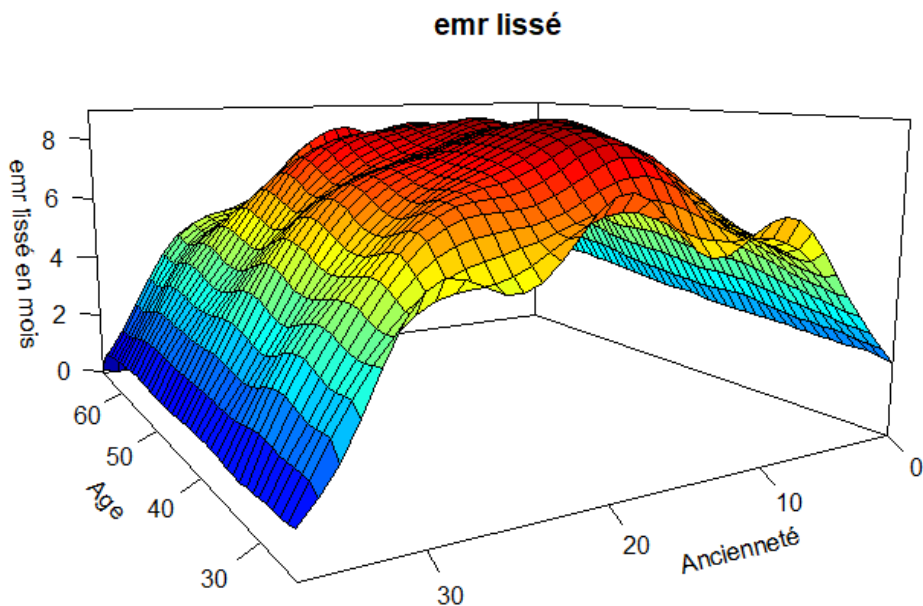
Ce qui nous permet de calculer les  $q_{x,t}$  et de comparer avec les résultats obtenus précédemment.

Le graphique suivant (Figure 4.15) représente les espérances de maintien résiduelle « brutes » en fonction de l'âge et de l'ancienneté.



**Figure 4.15 : Graphique des espérances de maintien résiduelle brutes**

En Figure 4.16, on représente le lissage des espérances de maintien résiduelle par la méthode de Whittaker-Henderson avec un regroupement pour les âges 21-25 et 66-70 avec les paramètres suivants :  $\alpha = 0,05$  ;  $\beta = 0,05$  ;  $z = 3$  ;  $\nu = 3$



**Figure 4.16 : Exemple de lissage des espérances de maintien résiduelle**

On observe la disparition des petits « pics » mais également une allure générale inchangée comparé à la Figure 4.15.

## 4.5 Résultats des lissages

Suite à la présentation des différentes méthodes de lissage et des tests statistiques permettant de valider ou d'infirmer ces lissages, nous avons effectué de nombreux tests. Les résultats sont présentés pour chaque méthode dans cette partie.

Concernant le test du Khi 2, nous avons regroupé les âges en 9 classes et les anciennetés en 8 classes. La valeur critique pour un seuil de 95% est alors de 5,89.

### Méthode Whittaker-Henderson :

Le tableau suivant résume différents tests effectués selon la méthode de lissage de Whittaker-Henderson avec à chaque fois des paramètres  $(\alpha ; \beta ; z ; v)$  de valeurs différentes.

La proportion de valeurs de la fonction de survie lissé qui se situent dans l'intervalle de confiance (Proportion IC) ainsi que la valeur de la statistique du test du Khi 2 sont présentés pour chaque combinaison de valeurs des paramètres du lissage.

	Paramètre $(\alpha ; \beta ; z ; v)$	Proportion IC	Test Khi 2
WH 1	(0.1 ; 0.1 ; 2 ; 2)	87,7%	58,78
WH 2	(0.1 ; 0.1 ; 3 ; 3)	89,4%	24,44
WH 3	(0.05 ; 0.05 ; 3 ; 3)	100%	11,53
WH 4	(0.05 ; 0.05 ; 4 ; 3)	100%	4,32

**Figure 4.17 : Résultats des tests selon la méthode de lissage Whittaker-Henderson**

Au vu des résultats, seul le lissage WH4 de paramètre  $\alpha = 0,05 ; \beta = 0.05 ; z = 4 ; v = 3$  est validé selon le test du Khi 2. De plus, pour ce lissage, l'ensemble des valeurs de la fonction de survie lissé se situent dans les intervalles de confiance. Le graphique des taux de sortie correspondant à ce lissage est représenté selon la Figure 4.12. On décide donc de retenir ce lissage pour la suite.

Le lissage WH1 et WH2, ne sont pas concluant d'après les tests statistiques effectués. Les taux de sorties conditionnels sont beaucoup trop lissés et pas assez représentatif des taux de sorties conditionnels initiaux.

### Méthode des splines :

Le tableau suivant résume différents tests effectués selon la méthode de lissage des splines avec des valeurs différentes pour le paramètre  $\alpha$ . De nouveau, la proportion de valeurs qui tombent dans l'intervalle de confiance et la valeur de la statistique du test du Khi 2 sont présentés en Figure 4.18.

	Paramètre $\alpha$	Proportion IC	Test Khi 2
SPL 1	0.1	65,3%	204,54
SPL 2	0.01	76,4%	138,62
SPL 3	0.001	88,95%	36,78
SPL 4	0.0005	100%	5,67

**Figure 4.18 : Résultats des tests selon la méthode de lissage des splines**

Au vu de résultats, seul le lissage SPL 4 est validé selon le test du Khi 2 et le test des intervalles de confiance de la fonction de survie. Le lissage SPL4 est représenté par la figure 4.14. On constate une forte similitude graphique entre le lissage WH4 et le lissage SPL4. Cependant, le lissage SPL4 comporte une erreur qui n'a pas été détecté par les 2 tests de validation : 2 valeurs de taux de sorties lissés, à des âges et des anciennetés élevées, sont inférieures très légèrement à 0. En conséquence, la fonction de survie augmente très légèrement à ces 2 endroits.

#### **Lissage des espérances de maintien résiduelle :**

Le lissage des espérances de maintien résiduelles a été effectué selon la méthode de Whittaker-Henderson. Les résultats sont présentés en Figure 4.19.

	Paramètre ( $\alpha ; \beta ; z ; v$ )	Proportion IC	Test Khi 2
WH 1	(0.1 ;0.1 ;2 ;2)	89,3%	47,58
WH 2	(0.1 ;0.1 ;3 ;3)	92,5%	17,49
WH 3	(0.05 ;0.05 ;3 ;3)	100%	4,73

**Figure 4.19 : Résultats des tests selon la méthode de lissage Whittaker-Henderson sur les espérances de maintien résiduelle**

Le lissage WH 3, figure 4.16, est validé selon les 2 tests statistiques à notre disposition. On décide donc de retenir ce lissage pour la suite. L'utilisation des espérances de maintien résiduelle plutôt que des taux de sorties conditionnels a permis d'obtenir un lissage plus rapidement validé par les tests statistiques.

#### **Conclusion**

La méthode de Whittaker-Henderson et des splines ont donné des résultats assez similaires. Cependant, on n'a pas réussi à obtenir un lissage par la méthode des splines validé par les 2 tests statistiques et ne contenant pas d'incohérence. Par conséquent, on a décidé de retenir 2 lissages se basant sur la méthode de Whittaker-Henderson : l'un s'appuyant sur les taux de sorties conditionnelles, l'autre sur les espérances de maintien résiduelle.

## 4.6 Impact sur le provisionnement

La modification des taux de sorties conditionnelles, désormais lissés, a un impact sur le montant des provisions que l'assureur doit constituer. On décide donc de calculer de nouveau les provisions de type « rente d'incapacité en cours » en prenant en compte cette modification.

Le tableau suivant (Figure 4.20) compare les prestations réellement payées en 2015 avec les prestations prédites pour l'année 2015 par la loi d'expérience de maintien en incapacité, corrigée suite au lissage, sur le périmètre des arrêts en cours au 31/12/2014

Année		Indemnité réellement versée en 2015	Indemnité selon la table provisionnement lissée	Ecart
2015	Pour 1 € d'indemnité mensuel	192 138,83	217 471,93	11,65%
	Montants réels	7 128 510,86	8 134 377,92	12,36%

**Figure 4.20 : Comparaison des indemnités réellement versées avec les indemnités prévisionnelles pour l'année 2015**

Pour l'année 2014, les résultats sont les suivants (Figure 4.21) :

Année		Indemnité réellement versée en 2014	Indemnité selon la table provisionnement lissée	Ecart
2014	Pour 1 € d'indemnité mensuel	217 125,94	228 347,68	4,91%
	Montants réels	8 111 462,62	8 310 902,06	2,40%

**Figure 4.21 : Comparaison des indemnités réellement versées avec les indemnités prévisionnelles pour l'année 2014**

Et pour l'année 2013 (Figure 4.22) :

Année		Indemnité réellement versée en 2013	Indemnité selon la table provisionnement lissée	Ecart
2013	Pour 1 € d'indemnité mensuel	187 817,00	185 928,67	-1,02%
	Montants réels	7 023 583,21	6 967 359,60	-0,83%

**Figure 4.22 : Comparaison des indemnités réellement versées avec les indemnités prévisionnelles pour l'année 2013**

#### **Conclusion de l'impact du lissage sur le provisionnement :**

De manière générale, le lissage n'a pas fortement modifié le montant total d'indemnisation prévu selon la table de provisionnement lissé. A chaque fois, le lissage a abouti à une très légère augmentation du montant à provisionner.

Lors de la comparaison entre les prestations réellement payées et les prestations prédites issus d'une table de provisionnement brute, on avait conclu que la table d'expérience était certainement suffisamment prudente pour être certifiée. Suite au lissage, les prestations prédites ont légèrement augmenté. Donc on peut toujours conclure à une prudence suffisante de la nouvelle table d'expérience lissée.

Pour conclure notre étude, il peut être intéressant d'étudier le cas où la table d'expérience construite n'est pas certifiée, l'assureur est alors obligé d'utiliser la table règlementaire du BCAC, et d'observer l'écart de prestation prédit avec le cas où la table d'expérience est utilisable.

Comme expliqué au début du mémoire, la table règlementaire du BCAC a été revue en 2013 suite au report de la date de départ à la retraite, passant de 62 ans à 65 ans, afin de mieux prendre en compte les risques relatifs à ces âges, mais elle est toujours non homologuée à ce jour. Par conséquent la table règlementaire du BCAC en vigueur date de 2010.

En Figure 4.23, 4.24 et 4.25, nous comparons les indemnités réellement versées en 2013, 2014 et 2015 respectivement aux indemnités prédites selon les tables BCAC 2010 et BCAC 2013.

Année		Indemnité réellement versée en 2015	Indemnité prédite selon la table BCAC 2010	Indemnité prédite selon la table BCAC 2013
2015	Pour 1 € d'indemnité mensuel	192 138,83	482 104,94	401 349,43
	Montants réels	7 128 510,86	18 059 244,91	15 549 320,93

**Figure 4.23 : Comparaison des indemnités réellement versées avec les indemnités prévisionnelles de la table règlementaire pour l'année 2015**

Année		Indemnité réellement versée en 2014	Indemnité prédite selon la table BCAC 2010	Indemnité prédite selon la table BCAC 2013
2014	Pour 1 € d'indemnité mensuel	217 125,94	537 829,01	452 407,97
	Montants réels	8 111 462,62	20 138 450,39	17 655 780,56

**Figure 4.24 : Comparaison des indemnités réellement versées avec les indemnités prévisionnelles de la table règlementaire pour l'année 2014**

Année		Indemnité réellement versée en 2013	Indemnité prédite selon la table BCAC 2010	Indemnité prédite selon la table BCAC 2013
2013	Pour 1 € d'indemnité mensuel	187 817,00	464 045,34	378 219,11
	Montants réels	7 023 583,21	17 463 928,42	14 136 745,23

**Figure 4.25 : Comparaison des indemnités réellement versées avec les indemnités prévisionnelles de la table règlementaire pour l'année 2013**

Cette dernière étude nous permet d'observer que les tables BCAC 2010 et 2013 sur provisionnent très largement les indemnités à verser au titre de la rente d'incapacité en cours. Cette observation met en évidence la nécessité pour l'organisme assureur d'utiliser une table d'expérience de maintien en incapacité et non la table réglementaire du BCAC.



# Conclusion

L'objet de ce mémoire était de construire une table d'expérience de maintien en incapacité tout en expliquant le rôle du certificateur lors de la construction.

La présentation du contexte de notre étude a permis de rappeler dans quelle situation et dans quel intérêt un Organisme Assureur a la possibilité de faire certifier une table d'expérience de maintien en incapacité. La contextualisation de notre étude a également été l'occasion de présenter le calcul des provisions auquel est soumis l'Organisme Assureur.

Suite à cela, nous avons consacré une partie de ce mémoire à la présentation des données. Outre l'audit du traitement des données, on a pu observer que le certificateur avait le devoir de s'assurer de la pertinence et la relative prudence dans la construction d'une base de donnée « propre ». Plusieurs contrôles de cohérences et des tests statistiques ont été réalisés dans ce sens. Ces derniers ont mis en évidence le caractère hétérogène de plusieurs variables parmi nos données dont notamment l'âge d'entrée en incapacité, le sexe ou encore la catégorie socio professionnelle. On a alors décidé de segmenter nos données uniquement en fonction de l'âge d'entrée en incapacité notamment afin d'éviter une « sur-paramétrisation » de nos données et également afin de conserver une certaine simplicité dans la construction de la loi de maintien. Les autres variables hétérogènes devront alors faire l'objet d'un suivi en cas de certification de la table d'expérience.

A partir de la base de données retraitée, on a alors présenté différentes méthodes de conception de lois de maintien dont la méthode de Kaplan-Meier, que l'on a décidé de retenir pour l'ensemble de ses bonnes propriétés. On a alors obtenu une table d'expérience « brute », c'est-à-dire une table présentant des irrégularités dues au risque d'échantillonnage. Pour pallier à ce problème, on a alors du lisser cette table. Différentes méthodes de lissage ont été présentées. Seule la méthode de Whittaker-Henderson nous a donné des résultats satisfaisants.

Enfin, une analyse des boni-mali a été effectuée. Une analyse des boni-mali vise à comparer les indemnités réellement versées par l'Organisme assureur avec les indemnités prédites selon la table d'expérience lissée. Cette analyse des boni-mali a permis de conclure à une relative prudence de la table d'expérience construite car pour deux années de survenance sur trois, on aboutit à un « sur-provisionnement » de la table d'expérience et pour la dernière année, on a abouti à un très léger « sous-provisionnement ». Une dernière étude a été produite avec pour but de comparer nos résultats avec la table réglementaire du BCAC. Cette dernière étude met en évidence la nécessité pour l'Organisme Assureur d'utiliser une table d'expérience de maintien en incapacité car la table règlementaire du BCAC aboutit à un très fort « sur-provisionnement ».

# Bibliographie

BAGUI (2013) « Refonte des lois de maintien en incapacité temporaire de travail », Mémoire ISFA.

DREES (2018) « La protection sociale en France et en Europe en 2016 » Résultats des comptes de la protection sociale, édition 2018, Dossier de Presse.

FETOUI (2015) « Impact de l'utilisation des tables d'expérience sur le provisionnement en prévoyance », Mémoire Université Paris-Dauphine.

FINAS (2009) « Modélisation du risque de maintien en incapacité », Mémoire ISUP.

FLEMING, HARRINGTON (1991) « Counting Processes and Survival Analysis », John Wiley & Sons.

GAUMET (2001) « Construction de tables d'expérience pour l'entrée et le maintien en incapacité », Mémoire ISFA.

GUGUMUS (2009) « Modélisation de l'incapacité temporaire et de l'invalidité en prévoyance collective », Mémoire DUAS.

LEURENT (2010) « Construction de tables d'expérience des risques incapacité et invalidité », Mémoire DUAS.

PLANCHET (2005), Modèles de durée. Support de cours ISFA.

PLANCHET, THEROND (2006) « Modèles de durée – Applications actuarielles », Economica.

PLANCHET, WINTER (2010) « L'utilisation des splines bidimensionnels pour l'estimation de lois de maintien en arrêt de travail », Bulletin Français d'Actuariat, Vol 7, n°13.

ROUSSEL, ZAOUI (2014) « Loi expérimentale de maintien en incapacité temporaire avec un pas quotidien », Mémoire CEA.

THEROND (2010) « Le point de vue du certificateur », Présentation.

WINTER (2005) « Méthodes bidimensionnelles pour l'ajustement de lois de maintien d'expérience en arrêt de travail », Mémoire ISFA.