



Université Claude Bernard – Lyon 1
**INSTITUT DE SCIENCE FINANCIERE ET
D'ASSURANCES**

**Mémoire présenté
devant l'Institut de Science Financière et d'Assurances
le 16 mai 2001
pour l'obtention
du diplôme d'Actuaire de Lyon**

Par : Mlle Aurélie GAUMET

Titre : **Construction de tables d'expérience pour l'entrée et le maintien en
incapacité**

Composition du jury des mémoires :

Membre du jury F.F.A.

M. ARNAL Pierre

Membres du jury I.S.F.A.

Mme REY Béatrice

M. PARTRAT Christian

M. PLANCHET Bernard

M. RULLIERE Didier

M. SERANT Daniel

M. TOUNISSOUX Daniel

Entreprise :

AXA France Assurance

Direction Vie Individuelle

Service Normes et Produits

Directeur de mémoire :

Mlle Nolwenn DONNART

Secrétariat :

Mme GARCIA Marie-José

Mme ANDRES Laurence

Bibliothèque :

Mme SONNIER Michèle

RESUME : CONSTRUCTION DE TABLES D'EXPERIENCE POUR L'ENTREE ET LE MAINTIEN EN INCAPACITE

Mots clés : Tables d'expérience ; entrée et maintien en incapacité ; segmentation ; double lissage ; sinistralité ; provisions ; tarif ; indemnité journalière (IJ) ; indemnité journalière en cas d'hospitalisation (IJH)

L'objet du mémoire est de fournir au groupe AXA, des tables d'expérience décrivant le risque incapacité de ses assurés ayant souscrit un contrat de prévoyance ou d'épargne comportant des garanties complémentaires en cas d'incapacité.

Les deux objectifs fixés sont les suivants: construire une première table d'entrée en incapacité et une seconde de maintien.

Pour le premier objectif, la modélisation retenue s'appuie sur une approche IARD, avec une estimation de l'espérance du nombre de sinistres sous une hypothèse de Poisson. Dans cette estimation, la durée d'exposition au risque d'incapacité est calculée au jour près, sans approximer en milieu d'année les mouvements sur le portefeuille (entrées, sorties, décès). Le volume de données étant suffisant, les tables décrivent l'entrée en incapacité selon le type de garantie (IJ ou IJH), le sexe de l'assuré, la cause du sinistre (accident ou maladie), la franchise souscrite et enfin l'âge de l'assuré.

Les tables obtenues sont lissées en fonction de l'âge par une régression polynomiale avec la méthode des moindres carrés.

Pour l'élaboration des tables de maintien en incapacité, l'estimateur retenu est celui de Kaplan Meier. Les tables sont segmentées selon le type de garantie (IJ ou IJH), la cause du sinistre (accident ou maladie) et la franchise. Des tables supplémentaires sont construites pour traduire l'impact de l'hospitalisation sur la durée des arrêts de travail.

Les tables de maintien ayant une double dimension (l'âge à l'entrée en incapacité et l'ancienneté du sinistre), un double lissage est effectué. Le premier lissage est fait à ancienneté fixée et en fonction de l'âge par une régression polynomiale. Le deuxième (à âge fixé et suivant l'ancienneté) consiste en un ajustement paramétrique à des fonctions usuelles.

Globalement, nos résultats montrent que la loi de maintien des assurés d'AXA se positionne en dessous de la table réglementaire fournie par le BCAC.

Les tables issues de ce mémoire, dont la certification est prévue, donneront à AXA un outil pour calculer les provisions et pour valider ou retarifier les garanties complémentaires en incapacité. Les segmentations poussées des tables donnent la possibilité de réévaluer la sinistralité du groupe selon sa composition (hommes/femmes, type de garantie et de franchise). Ainsi, l'influence de l'évolution du portefeuille sur la validité des tables est limitée.

L'élaboration de tables d'expérience permettra à AXA de mieux connaître la sinistralité de ses assurés en matière d'incapacité et de se doter de moyens techniques afin d'exploiter l'évolution de la réglementation en matière de provisionnement.

ABSTRACT : CONSTRUCTION OF EXPERIENCE TABLES FOR THE INCIDENCE AND THE RECOVERY OF TEMPORARY DISABLEMENT

Keywords : experience tables ; incidence, duration and recovery rates; segmentation; double graduation ; claims status ; reserves ; tariff ; daily benefits ; hospital income insurance.

The principal objective of this paper is to provide the AXA Group with a Temporary Disablement experience table for its life insurance policies underwritten with disablement riders. Two detailed objectives have been defined : to construct tables for the incidence rates and for recovery rates.

For the first objective, a typical general insurance claims incidence approach was adopted, modeling the expected number of claims as following a Poisson distribution. For this purpose, exposure to risk was calculated using the exact number of calendar days. Sufficient volumes allowed the analysis to be conducted separately for different types of risks. Risks were classified as a function of coverage type (daily benefits or hospital income insurance), gender, cause of loss (accident or disease), waiting period and attained age. The graduation of tables was performed using a polynomial regression and applying the least squares method.

The recovery rates analysis was undertaken using the Kaplan Meier approach, and were segmented as a function of coverage type (daily benefits or hospital income insurance), cause of loss (accident or disease) and waiting period. Additional tables based on claims with hospitalization were necessary to reveal the influence of hospitalization on recovery rates.

The multi dimension nature (entry age and duration of disablement) of the tables implies the need for a double graduation. The first graduation, made as a function of age while the period of disablement was held constant, was performed using a polynomial regression. The second, made as a function of the duration of the disablement, while holding the age constant, was then overlayed on the results of the first.

The final results show a globally lower claims cost than that assumed in the French prescribed table for statutory reporting.

This paper, whose certification is planned, will give AXA a tool for the revaluation of its mathematical reserves and its tariffs. The risk classification used gives the opportunity to adjust the tables for a particular application with an appropriate “blending”.

The development of experience tables will bring the AXA group a better knowledge of its disablement claims status and risks. It also provides the means to justify a revision of the level of mathematical reserves held for this business.

REMERCIEMENTS

Je tiens à remercier, tout particulièrement, Nolwenn DONNART, mon maître de stage, qui a su se rendre très disponible et me faire profiter de son expérience malgré un emploi du temps chargé.

Je remercie également Monsieur Jean ROYER pour m'avoir accueillie dans son service, Monsieur Philippe BERNARDI pour m'avoir conservée dans son équipe et permis de mener à bien mon étude, de même que Sylvie ROUJA et François BEHAEGEL pour avoir mis en place ce stage et pour leurs conseils avisés.

Mes remerciements s'adressent aussi à l'ensemble du service Normes-Produits pour la qualité et la gentillesse de leur accueil ainsi qu'à tous les membres du groupe AXA avec qui j'ai été amenée à être en contact pour l'élaboration de cette étude.

Enfin, je remercie Monsieur Daniel SERANT et Monsieur Roberto WOLFRUM pour m'avoir écoutée et m'avoir fait bénéficier de leurs conseils.

SOMMAIRE

Résumé :	2
Abstract :	3
Remerciements	4
INTRODUCTION	8
Partie I : Présentation de l'étude	10
Chapitre 1 : Le groupe AXA	10
Section 1.1 : Historique du groupe	10
Section 1.2 : Les activités du groupe	11
Section 1.3 : Axa France Assurance (AFA)	13
1.3.1 : Ses missions	13
1.3.2 : La Direction Vie Individuelle (DVI)	13
1.3.3 : Le service Normes et Produits	13
Chapitre 2 : La prévoyance complémentaire en France	14
Section 2.1 : Définitions	14
Section 2.2 : Le marché français de la prévoyance	14
2.2.1 : La Sécurité Sociale	15
2.2.2 : Les organismes assureurs	15
2.2.3 : Quelques chiffres	15
Section 2.3 : La prévoyance dans le groupe AXA :	20
2.3.1 : Le marché français de l'assurance vie individuelle pour le groupe :	20
2.3.2 : La taille du portefeuille de prévoyance étudié	20
Section 2.4 : L'incapacité en France et dans le groupe AXA	21
2.4.1 : Les prestations de la Sécurité Sociale en matière d'incapacité	21
2.4.2 : Les prestations complémentaires	21
2.4.3 : Quelques chiffres	21
2.4.4 : Chiffres AXA	22
Section 2.5 : Le cadre législatif	23
2.5.1 : La loi Évin : des provisions forfaitaires	23
2.5.2 : Arrêté du 28 mars 1996 : l'utilisation de tables	24
Section 2.6 : Intérêt des tables d'expérience	25
Partie II : Les données	27
Chapitre 1 : Présentation des données	27
Section 1.1 : Les garanties du portefeuille étudié	27
Section 1.2 : Les fichiers de données	27
Chapitre 2 : Les choix fondamentaux adoptés pour la construction des tables	30
Section 2.1 : Choix du périmètre de l'étude : la période d'observation	30
Section 2.2 : Choix des risques à évaluer	32
2.2.1 : L'exclusion des risques aggravés	32
2.2.2 : Le risque d'indemnisation	32
2.2.3 : L'indépendance des risques	32
Section 2.3 : La dimension âge et les conventions utilisées	33
Chapitre 3 : Préparation des données	34
Section 3.1 : Détection et traitement des valeurs aberrantes	34
Section 3.2 : Traitements spécifiques pour adapter les données au risque étudié	37
Section 3.3 : Volume des données après traitement	40

Partie III : Evaluation du portefeuille et du risque incapacité _____ 43

Chapitre 1 : Descriptif du portefeuille _____	43
Section 1.1 : caractéristiques des contrats souscrits _____	43
1.1.1 : les combinaisons de garantie _____	43
1.1.2 : les franchises souscrites _____	44
1.1.3 : les montants des indemnités journalières _____	45
Section 1.2 : Etude de la population sous risque _____	46
1.2.1 : Répartition Homme / Femme _____	46
1.2.2 : Répartition des assurés par année de naissance _____	46
1.2.3 : Répartition des assurés suivant leur âge à la souscription _____	47
1.2.4 : Répartition des assurés par CSP _____	47
1.2.5 : Répartition des assurés par région _____	48
1.2.6 : Les assurés ayant plusieurs contrats _____	48
Chapitre 2 : Evaluation du risque incapacité _____	49
Section 2.1 : Définitions _____	49
Section 2.2 : La répartition des sinistres _____	49
2.2.1 : Répartition des sinistres suivant leur nature _____	50
2.2.2 : Répartition des sinistres suivant leur franchise _____	50
Section 2.3 : Etude de la durée des sinistres _____	51
2.3.1 : Suivant le sexe _____	51
2.3.2 : Suivant la franchise sur le sinistre _____	52
2.3.3 : Suivant la nature _____	52
Section 2.4 : Etude de la saisonnalité _____	53
2.4.1 : La stabilité de la sinistralité dans le temps _____	53
2.4.2 : La saisonnalité _____	54
Section 2.5 : Influence des caractéristiques clients sur la sinistralité _____	55
2.5.1 : Le sexe de l'assuré en arrêt _____	55
2.5.2 : la CSP du sinistré _____	56
2.5.3 : la région du sinistré _____	56
2.5.4 : le montant d'IJ souscrit _____	57

Partie IV : Elaboration des tables d'entrée en incapacité _____ 59

Chapitre 1 : Particularité de la table d'expérience _____	59
Section 1.1 : L'utilisation de données censurées. _____	59
Section 1.2 : L'utilisation de données individuelles et journalières _____	60
Chapitre 2 : L'utilisation des tables : la tarification _____	61
Section 2.1 : 1 ^{er} modèle _____	61
Section 2.2 : 2 ^{ème} modèle _____	62
Section 2.3 : 3 ^{ème} modèle _____	63
Section 2.4 : Réduction de prime _____	64
Section 2.5 : Adaptation des tables avec les garanties souscrites _____	64
Chapitre 3 : Les estimateurs d'entrée en incapacité _____	66
Section 3.1 : 1 ^{ère} approche : le concept de taux d'entrée en incapacité _____	66
3.1.1 : Les estimateurs non paramétriques _____	67
a) L'estimateur de Kaplan-Meier _____	67
b) L'estimateur actuariel _____	70
3.1.2 : Les estimateurs paramétriques _____	71
a) L'estimateur des moments de Hoem _____	71
b) L'estimateur du maximum de vraisemblance avec hypothèse de loi d'entrée. _____	74
Section 3.2 : 2 ^{ème} approche : le modèle IARD _____	76
3.2.1 : l'estimateur _____	76
3.2.2 : l'adéquation à une loi de poisson _____	78
3.2.3 : La linéarité de λ _____	79
Section 3.3 : Choix de l'estimateur de type « IARD » _____	79

Chapitre 4 : La segmentation des tables : homogénéité de la sinistralité	80
Chapitre 5 : Les contraintes de volume	85
Section 5.1 : Estimations par classe d'âge et interpolation à "jonction souple"	85
Section 5.2 : Le glissement	87
Section 5.3 : Choix de la méthode par glissement	87
Chapitre 6 : Les tables brutes	88
Section 6.1 : adéquations au modèle	88
6.1.1 : adéquation à la loi de Poisson	88
6.1.2 : La linéarité de λ	91
6.1.3 : La stabilité dans le temps	91
Section 6.2 : les résultats bruts	92
Section 6.3 : intervalles de confiance	95
Chapitre 7 : Lissage des lois d'entrée	96
Section 7.1 : Pourquoi faire un lissage?	96
Section 7.2 : Régression non paramétrique par la méthode des noyaux	96
Section 7.3 : Régression polynomiale sur l'espérance du nombre de sinistres	98
Section 7.4 : Comparaison graphique et choix du lissage	100
<i>Partie V : Elaboration des tables de maintien en incapacité</i>	<i>102</i>
Chapitre 1 : Utilisation de données censurées	103
Chapitre 2 : Des tables de maintien adaptées au tarif et au provisionnement	104
Chapitre 3 : Les estimateurs du maintien	105
Section 3.1 : Méthode actuarielle	105
Section 3.2 : Méthode de Kaplan Meier	107
3.2.1 : Principe	107
3.2.2 : Construction	107
3.2.3 : Résultat	108
3.2.4 : Propriétés	108
3.2.5 : Méthode du BCAC	109
Section 3.3 : choix de l'estimateur de Kaplan Meier	110
Chapitre 4 : La segmentation des tables : homogénéité de la sinistralité	111
Chapitre 5 : Les contraintes de volume	116
Chapitre 6 : les tables brutes	117
Section 6.1 : La stabilité dans le temps	117
Section 6.2 : les résultats bruts	119
Section 6.3 : Comparaison avec la table réglementaire	126
Section 6.4 : les intervalles de confiance	128
Chapitre 7 : lissage des lois de maintien	130
Section 7.1 : La problématique du double lissage	130
Section 7.2 : Le lissage suivant l'âge	131
Section 7.3 : Le lissage en fonction de l'ancienneté	132
7.3.1 : lissage par Whittaker et la variante de lowrie	132
7.3.2 : Modèle de Makeham	134
7.3.3 : Régressions paramétriques par les moindres carrés	135
CONCLUSION	137
BIBLIOGRAPHIE	142

INTRODUCTION

A une époque où la couverture assurée par la Sécurité Sociale se réduit, on assiste au développement du marché privé de l'assurance des dommages corporels aussi bien à titre personnel que par des contrats de prévoyance collective.

Si tout un chacun peut se réjouir de la concurrence qui existe entre les assureurs, ces derniers se doivent d'être prudents face aux difficultés que présentent l'étude et le pilotage technique des contrats garantissant des prestations sous forme d'indemnités journalières en cas d'incapacité. Voilà les raisons pour lesquelles ils ressentent de plus en plus un besoin d'informations précises et actualisées sur le risque arrêt de travail. Valider leurs tarifs et provisions permettrait d'éviter une sous-estimation qui pourrait s'avérer lourde de conséquences sur les résultats actuels et futurs de leur activité.

D'autre part, pour satisfaire aux obligations réglementaires en matière de provisionnement, les assureurs n'ont aujourd'hui à leur disposition, à défaut de tables d'expérience certifiées, que les tables de maintien réglementaires. Ces dernières ont été proposées par le Bureau Commun des Assurances Collectives (BCAC), elles datent de 1993 et furent élaborées d'après le portefeuille de différentes compagnies d'assurance.

Dans ce contexte, La Direction Vie Individuelle d'AXA France Assurance souhaite mieux connaître le risque incapacité de son portefeuille en vue de valider de nouveaux tarifs basés sur les lois d'entrée et de maintien en incapacité propre à ses assurés. De plus, l'élaboration et la certification de tables de maintien en arrêt de travail permettraient d'ajuster au mieux le provisionnement obligatoire des prestations d'incapacité.

L'objectif de ce mémoire a donc été de proposer des statistiques sur le risque incapacité de travail d'un portefeuille AXA d'assurance individuelle de personnes afin d'établir des tables d'expérience d'entrée et de maintien en arrêt de travail.

Ce mémoire sera divisé en cinq parties, la première présentera la prévoyance complémentaire en France. La seconde traitera des données utilisées et des choix qu'elles impliquent. La troisième décrira le portefeuille et le risque étudié. Enfin, les deux dernières parties seront consacrées à l'élaboration proprement dite des tables d'expérience, d'abord celles d'entrée en incapacité puis ensuite celles de maintien.

PARTIE I :
PRESENTATION DE L'ETUDE

Partie I : Présentation de l'étude

Tout d'abord, nous allons présenter brièvement la structure du groupe AXA en France ainsi que le service dans lequel le stage a été effectué.

Et pour mieux comprendre le contexte de notre étude, nous tacherons ensuite de faire un état des lieux de la prévoyance complémentaire française aussi bien en terme de résultats qu'en terme de législation.

Chapitre 1 : Le groupe AXA

Section 1.1 : Historique du groupe

- 1817** Naissance de la première société à l'origine d'AXA : « La Compagnie d'Assurances Mutuelles contre l'Incendie dans le département de la Seine et de l'Eure ».
- 1817 à 1946** Regroupement de nombreuses petites mutuelles régionales.
- 1946 à 1978** Achat de Provinces Unies et création du groupe Mutuelles Unies au Canada.
Prise de contrôle de la Mutuelle Parisienne de Garantie en France.
- 1982** Prise de contrôle du groupe Drouot et de sa filiale d'assurance vie : La Vie Nouvelle.
- 1984** Naissance d'AXA avec Claude BEBEAR pour président.
- 1986** Prise de contrôle de la Providence et du Secours (groupe Présence).
- 1988** Lancement de la société : UNI Europe, première à naître de la restructuration.
Accord avec la Compagnie du Midi pour constituer un réseau d'Assurances français.
- 1989** Création de la société AXA Assurances, suite à la fusion des 4 sociétés : Mutuelles Unies, Assurances Groupe de Paris (AGP), Présence et Drouot.
- 1991** Rachat de la société américaine Equitable (7^{ème} assureur vie aux Etats-Unis).
- 1995** Rachat de la société australienne National Mutual (9^{ème} assureur vie aux Etats-Unis), acquisition de la société Abeille-Ré.
- 1996** Simplification de la structure de holdings du groupe AXA, disparition du holding intermédiaire Midi Participations.
Cotation à la Bourse de Wall Street, le New York Stock Exchange (NYSE), des actions de la société AXA sous forme de : American Depositary Shares (ADS).
En novembre, rapprochement avec l'UAP.
- 1997** AXA a fait part de son intention de s'implanter en Chine dès l'ouverture du marché de l'assurance chinois aux assureurs étrangers.
Fusion – absorption de la compagnie UAP par AXA.
Le groupe AXA-UAP devient l'un des tous premiers acteurs mondiaux de l'assurance.
- 1998** Restructuration des activités d'AXA et de l'UAP France créant 3 entités : AXA Assurances, AXA Courtage et AXA Conseil. Le conseil d'administration décide de soumettre au vote de l'Assemblée Générale la disparition du nom UAP en 1998.
- 1999** AXA et l'ONA s'associent pour devenir le 2^{ème} assureur marocain.
L'offre sur Guardian Royal Exchange est déclarée totalement irrévocable.

Section 1.2 : Les activités du groupe

Le groupe AXA figure parmi les leaders mondiaux de l'assurance et des services financiers. Ses activités sont soit domestiques, soit internationales (les grands risques internationaux, l'assistance et la réassurance). Il est implanté dans plus de 60 pays, sur les 5 continents et compte près de 140 000 collaborateurs à travers le monde.

1.2.1 : Situation actuelle

Chiffres d'affaires 2000 :

80 milliards d'euros (progression de 20,3% par rapport à 1999)

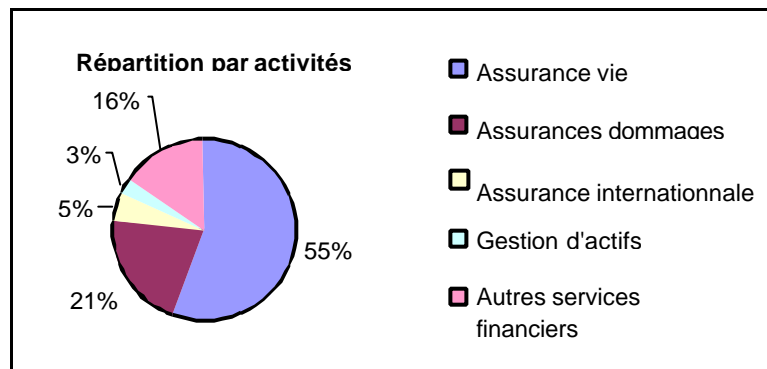
Résultat net courant pour 2000 :

2,54 milliards d'euros (progression de 24,2% par rapport à 1999)

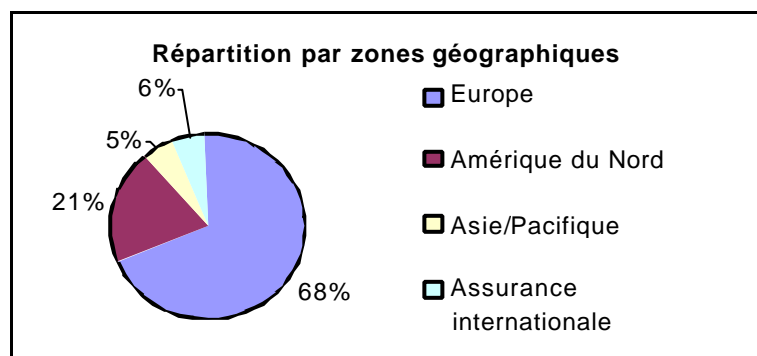
Actifs gérés :

893 milliards d'euros (progression de 14,3% par rapport à 1999)

1.2.2 : Répartition par activités



1.2.3 : Répartition par zones géographiques



1.2.4 : Les sociétés françaises du groupe AXA

AXA ASSURANCES :

Société d'assurances dommages et d'assurances de personnes pour une clientèle de particuliers et d'entreprises. Cette société distribue ses produits exclusivement par l'intermédiaire d'agents généraux.

AXA COLLECTIVES :

Comme AXA Assurances, cette société travaille dans l'assurance Dommages et l'assurance de personnes pour particuliers et entreprises. Cependant, elle se différencie par son canal de distribution puisqu'elle travaille avec les courtiers français et internationaux.

AXA CONSEIL :

Pour les sociétés regroupant les réseaux salariés, regroupement des réseaux salariés de l'UAP et d'ALPHA Assurances.

AXA CORPORATE SOLUTIONS :

Société qui englobe AXA Ré et AXA Cession. La première est une société de réassurance spécialisée dans les grands risques et la seconde contrôle l'ensemble des opérations de réassurance du groupe AXA.

DIRECT ASSURANCE :

Sa vocation est de commercialiser des contrats d'assurances dommages et de personnes aux particuliers en utilisant les méthodes de vente directe (vente par téléphone).

3 mutuelles sont intégrées au groupe AXA :

LES MUTUELLES ASSOCIEES :

- mutuelle Saint-Christophe : Société spécialisée dans l'assurance IARD des ecclésiastiques, professeurs de l'enseignement privé, etc...
- mutuelle de Marseille : Société d'assurance IARD.
- mutuelle Phocéenne : Société d'assurance vie.

AXA INVESTIMENT MANAGER et BANQUE D'ORSAY :

Gestion financière pour compte de tiers.

GIE AXA IMMOBILIER :

Gestion des actifs immobiliers du Groupe, gestion de l'épargne immobilière, ingénierie des opérations de promotions, locations...Le patrimoine géré représente 2 millions de mètres carrés.

AXA BANQUE, AXA CREDIT, AXA GESTION INTERESSEMENT :

Sociétés financières.

AXA FRANCE ASSURANCE :

C'est la société holding qui détient le capital des sociétés d'assurance opérant en France et qui regroupe les directions centrales.

Section 1.3 : Axa France Assurance (AFA)

1.3.1 : Ses missions

Les différentes directions centrales d'AFA exercent auprès de l'ensemble des sociétés du groupe en France les missions suivantes :

- Elles assurent la cohérence des politiques générales des sociétés d'assurance en respectant la décentralisation opérationnelle.
- Elles sont le support de l'ensemble des activités où l'effet de taille permet d'être plus efficace et de réaliser des économies d'échelle.
- Elles effectuent le contrôle de l'activité et des résultats des sociétés d'assurance.

Axa France Assurance comprend onze directions centrales dont la Direction Vie Individuelle au sein de laquelle a été effectué ce mémoire.

1.3.2 : La Direction Vie Individuelle (DVI)

Elle a pour vocation de mettre en œuvre les politiques techniques et juridiques attachées à la distribution, la gestion et la comptabilisation des produits d'assurance Vie à souscription individuelle et celles liées à la gestion des sociétés Vie opérant en France. La DVI s'occupe aussi de l'évaluation des provisions sur le périmètre de l'assurance Vie individuelle. Enfin, son rôle est de diffuser l'expertise en Vie.

L'étude présentée dans ce mémoire a été réalisée dans le service Normes et Produits de cette direction.

1.3.3 : Le service Normes et Produits

Comme son nom l'indique, ce service s'occupe de réaliser des «Normes », c'est à dire des études sur différentes problématiques qui touchent les sociétés Vie du groupe ; les résultats de ces études servent ensuite de référence pour tous les collaborateurs.

Ce service s'inscrit aussi dans l'univers « produits » au sens où il exerce un contrôle des produits d'assurance Vie individuelle d'Axa en France, de leur conception à leur suivi.

Suite à cette présentation du groupe AXA, nous allons maintenant nous intéresser à la prévoyance complémentaire en France.

Chapitre 2 : La prévoyance complémentaire en France

Ce chapitre va permettre de définir les termes essentiels de cette étude. Il présentera également le marché de la prévoyance en France et dans le groupe AXA et s'attachera plus particulièrement à l'assurance en cas d'incapacité de travail. Pour finir, il précisera le cadre législatif de cette étude et justifiera l'élaboration de tables d'expérience.

Section 2.1 : Définitions

◆ Les opérations de prévoyance :

Ensemble des contrats d'assurance de personnes offrant des garanties de versement de prestations en cas de survenance d'un risque : décès, incapacité de travail, invalidité ou maladie. Sont regroupées dans cette catégorie les assurances en cas de décès, les assurances de dommages corporels (maladie et accident), l'assurance dépendance et l'assurance chômage.

◆ L'assurance en cas de décès :

Contrat d'assurance souscrit soit individuellement soit par l'intermédiaire d'une entreprise ou d'une association (à l'occasion d'un emprunt par exemple) et garantissant le versement d'un capital en cas de décès avant le terme du contrat.

Cette garantie peut être complétée par des garanties de dommages corporels : prestation en cas d'invalidité ou d'incapacité consécutive à une maladie ou à un accident, ou majoration de la garantie en cas de décès accidentel.

◆ L'assurance de dommages corporels :

Contrat d'assurance souscrit individuellement ou collectivement, garantissant le remboursement de frais de soins en complément des régimes obligatoires de protection sociale, ou le versement d'un capital en cas de décès par accident, ou le paiement d'indemnité en cas d'incapacité de travail ou d'invalidité.

Les éléments cités ci-dessus peuvent être proposés selon les cas sur des contrats d'assurance vie (individuels ou collectifs), ou de santé, ou sur des contrats individuel accident dits d'assurance dommage (ou IARD).

Bien que notre étude porte uniquement sur le risque incapacité de travail, nous présenterons préalablement le marché de la prévoyance en France.

Section 2.2 : Le marché français de la prévoyance

Plusieurs acteurs animent ce marché. En effet, les couvertures de prévoyance sont assurées d'une part par le régime général de la Sécurité Sociale, d'autre part par les régimes complémentaires des organismes assureurs.

2.2.1 : La Sécurité Sociale

Depuis sa création en 1945, elle a pour objectif d'assurer la protection des individus (quel que soit leur statut professionnel) devant les conséquences financières de certains événements (maladie, accident, maternité), ou de certaines situations familiales (veuvage, ...). Elle couvre aussi depuis quelques années le risque décès. Il faut savoir que le régime de la Sécurité Sociale classe les prestations liées à l'incapacité de travail dans les frais de santé. On remarque actuellement un certain désengagement de la Sécurité sociale qui prenait en charge, en 1997, 74% des dépenses de soins et biens médicaux contre 77% en 1980.

2.2.2 : Les organismes assureurs

La Sécurité Sociale n'assure pas le remboursement de la totalité des frais de l'assuré. Celui-ci garde notamment à sa charge le ticket modérateur. Des organismes assureurs proposent donc des contrats d'assurance complémentaire.

Voici les organismes agréés (par la loi Evin) pour proposer ces produits d'assurance :

- Les sociétés d'assurance régies par le Code des Assurances, ce sont soit des SA soit des sociétés de personnes (sociétés d'assurance mutuelles).
- Les institutions de prévoyance régies par le Code de la Sécurité Sociale.
- Les mutuelles régies par le Code de la mutualité
- Les institutions régies par le code Rural

Chez les assureurs, les garanties de dommages corporels peuvent être commercialisées sur différents types de contrats :

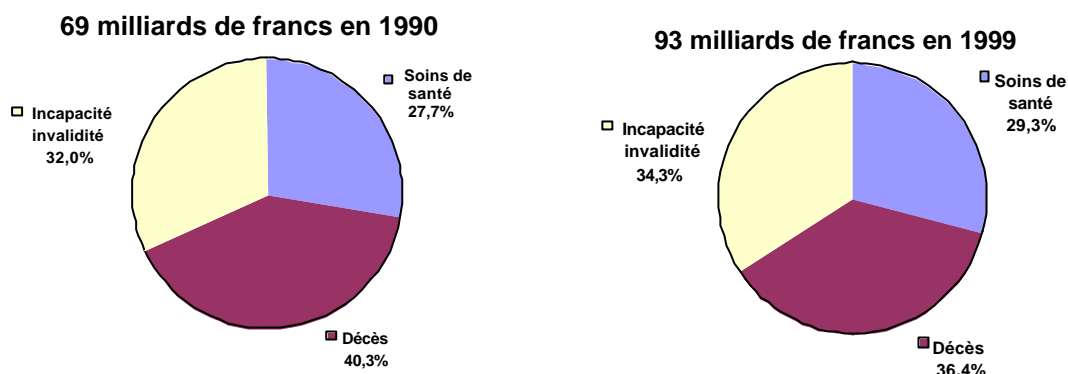
- **Les contrats de prévoyance** (individuels ou collectifs) qui proposent comme garantie principale un capital en cas de décès ou d'invalidité assimilée au décès, ainsi que des garanties complémentaires (indemnités journalières, exonération de paiement des primes en cas d'incapacité ou d'invalidité...).
- **Les contrats d'épargne** où l'on peut souscrire (facultatif) des garanties complémentaires telles que le décès accidentel, l'incapacité ou l'invalidité.
- **Les contrats de santé** qui couvrent les frais de soins et donc aussi parfois les indemnités journalières.
- **Les contrats individuels accident** isolément ou couplé avec de l'assurance de responsabilité civile et en automobile.

2.2.3 : Quelques chiffres

a) *Les opérations de prévoyance : une progression régulière*

En 1999, le montant des cotisations recueillies par les sociétés d'assurances au titre des opérations de prévoyance atteint 93 milliards de francs, en progression de + 2 % par rapport à 1998. Ces opérations incluent les assurances en cas de décès et les assurances santé et accidents (dommages corporels).

Répartition des cotisations collectées en prévoyance



Les assurances en cas de décès

Les cotisations des contrats d'assurances en cas de décès s'élèvent à 34 milliards de francs en 1999, en hausse de + 5 % par rapport à 1998. Cette croissance est un peu supérieure à celle de l'année précédente. Elle traduit une demande régulière des particuliers pour une protection complémentaire de leurs proches, qu'il s'agisse de contrats d'assurance souscrits à titre individuel, dans le cadre d'une entreprise ou à l'occasion d'un emprunt.

Les assurances en cas de décès se répartissent en deux grandes catégories :

- Les contrats à souscription individuelle, dont les cotisations représentent 10 milliards de francs. Ils peuvent être souscrits soit directement auprès d'une société d'assurances ou de ses intermédiaires (7 milliards de francs) soit dans le cadre de groupes ouverts (3 milliards de francs), essentiellement auprès de guichets d'établissements financiers
- Les contrats à souscription collective, pour 24 milliards de francs, qui comprennent sensiblement à parts égales les contrats souscrits par les entreprises pour leurs salariés et les contrats à l'occasion d'un emprunt, essentiellement immobilier.

Les assurances de dommages corporels (santé et accidents)

Les garanties offertes sont souscrites dans le cadre de contrats d'assurance santé ou accidents corporels, ou sous la forme de garanties complémentaires annexées aux contrats d'assurance vie.

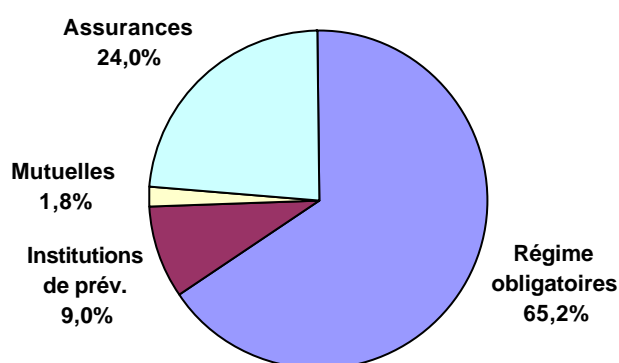
Les cotisations d'assurance santé et accidents corporels s'élèvent à 59 milliards de francs en 1999 (en affaires directes), soit un montant identique à celui de 1998. Les cotisations versées au titre de la protection complémentaire santé et accidents ont progressé très faiblement au cours des quatre dernières années. Cela résulte, d'une part, des distorsions de concurrence, principalement d'ordre fiscal, qui perdurent sur ce marché et, d'autre part, du non-renouvellement de certains contrats.

Ces cotisations se répartissent entre les garanties de remboursements de frais de soins (prestations en nature) pour 28 milliards de francs, en faible progression par rapport à 1998 - la progression des contrats individuels étant supérieure à celle des contrats collectifs -, et les garanties de versements d'indemnités (prestations en espèces) pour 31 milliards de francs, avec une croissance soutenue pour l'assurance individuelle.

b) L'indemnisation en cas de décès, d'incapacité ou d'invalidité

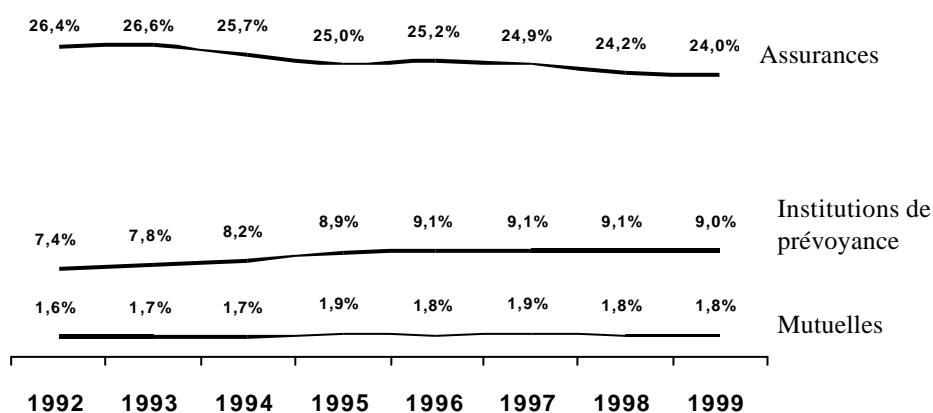
En 1999, les indemnités versées par les régimes de base pour arrêt de travail, invalidité, décès, et accidents du travail sont évaluées à 96 milliards de francs ¹. Si l'on ajoute en plus les 51 milliards de francs supplémentaires qui ont été versés cette même année par l'ensemble des organismes complémentaires, ce sont au total 147 milliards de francs d'indemnités qui ont été versées à la fois par les régimes de base et par les organismes complémentaires.

**Versements d'indemnités incapacité, invalidité et décès
147 milliards de francs en 1999**



Avec 35 milliards de francs d'indemnités versées en 1999, soit un montant en hausse de + 3,2 % par rapport à 1998, les assureurs totalisent plus des deux tiers (69 %) des indemnités versées par les organismes complémentaires et près d'un quart (24 %) de celles versées par l'ensemble des intervenants. Cependant, la part des assureurs dans le financement global a connu une diminution régulière depuis 1992 tandis que celles détenues par les autres organismes complémentaires (mutuelles et institutions de prévoyance) ont augmenté jusqu'en 1995.

**Part du financement des organismes complémentaires
dans le versement global d'indemnité incapacité – invalidité – décès**



Versements d'indemnités en cas de décès, d'incapacité et d'invalidité

¹ Estimation FFSA

(en milliards de francs)

Organismes	1995	1996	1997	1998	1999	99/98
Régimes obligatoires	86,4	87,6	88,2	91,6	96,2	5%
Organismes complémentaires	48,1	49,4	49,5	49,7	51,1	2,80%
dont sociétés d'assurances*	33,6	34,5	34,3	34,2	35,3	3,20%
dont mutuelles**	2,5	2,5	2,6	2,6	2,6	-
dont institutions de prévoyance	12	12,4	12,6	12,9	13,2	2,30%
ENSEMBLE	13.45	137	137.7	141.3	1447.3	4.20%

Sources : Cnamts et estimations FFSA

* hors contre-assurance des contrats en cas de vie

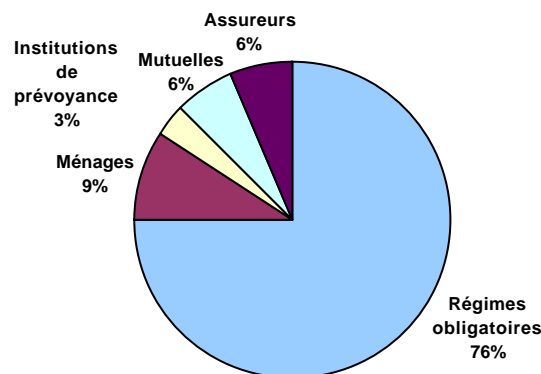
** relevant du code de la mutualité

c) Les assureurs dans l'ensemble de la protection décès-maladie-accidents

Au cours de l'année 1999, si l'on cumule à la fois les prestations versées au titre des dépenses de soins et de biens médicaux ainsi que les indemnités versées en cas de décès, d'incapacité ou d'invalidité, le montant versé par les régimes de base est évalué à 677 milliards de francs et celui des organismes complémentaires à 143 milliards de francs.

Si l'on tient compte en plus des 83 milliards de francs qui sont laissés à la charge des ménages, ce sont globalement 903 milliards de francs qui ont été versés en 1999 au titre de la protection sociale décès, maladie et accidents ².

Financement de la protection décès-maladie-accidents 903 milliards de francs en 1999



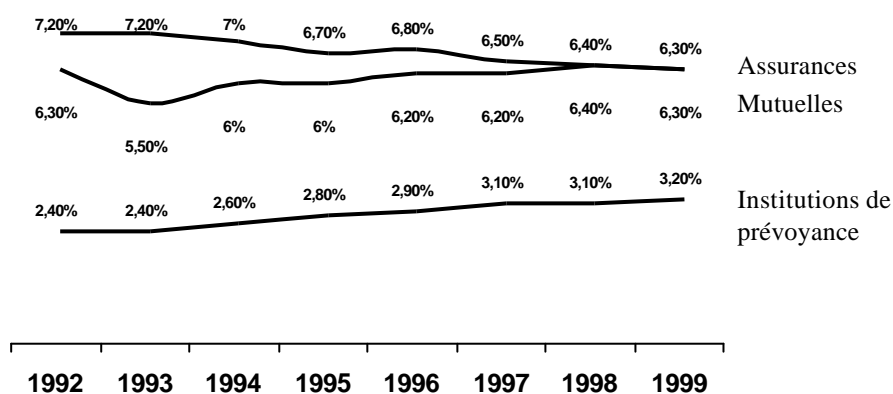
² Estimation FFSA

Prestations versées au titre de la prévoyance en 1999 par l'ensemble des organismes

	SANTÉ		DECES-INCAPACITE INVALIDITE		TOTAL	
	montant (milliards F.)	répartition	montant (milliards F.)	répartition	montant (milliards F.)	répartition
Organismes de base	581	77%	96	65%	677	75%
Organismes complémentaires	92	12%	51	35%	143	16%
- dont sociétés d'assurances	22	3%	35	24%	57	6%
Charges des ménages	83	11%	-	-	83	9%
TOTAL	756	100%	147	100%	903	100%

Avec un total de 57 milliards de francs versés en 1999, la part des assureurs atteint 6,3 % de l'ensemble de la protection sociale décès – maladie – accidents ³ soit un poids identique à celui détenu par les mutuelles (6,3 %), mais environ le double de celui des institutions de prévoyance (3,2 %)

Part de financement des organismes complémentaires Dans la protection décès-maladie-accidents



³ Estimation FFSA

Section 2.3 : La prévoyance dans le groupe AXA :

Après une présentation générale de la prévoyance en France, nous allons préciser la situation du groupe AXA dans ce domaine.

2.3.1 : Le marché français de l'assurance vie individuelle pour le groupe :

<i>en millions de Francs</i>	assurance vie individuelle (épargne + prévoyance)	dont prévoyance individuelle
Chiffre d'affaires	55 193	1 452
Prestations	38 592	895
Provisions mathématiques	374 113	1 638

sources : états comptables du 31/12/2000

2.3.2 : La taille du portefeuille de prévoyance étudié

Le volume important en nombre de contrats permet de penser que la loi de convergence de nos estimateurs des lois d'entrée et de maintien en incapacité pourront s'appliquer sur un grand nombre de segments du risque.

	volume de contrats en cours	montant des PM (en F)	Capitaux assurés (en F)
vie individuelle (épargne + prévoyance)	4 648 682	183 233 292 235	329 599 758 119

sources : états comptables au 31/12/2000

Section 2.4 : L'incapacité en France et dans le groupe AXA

Parmi tous les dommages corporels, ce mémoire ne traitera que le risque incapacité.

2.4.1 : Les prestations de la Sécurité Sociale en matière d'incapacité

La Sécurité Sociale s'engage à verser à tout cotisant salarié, obligé d'interrompre son activité professionnelle en raison d'une incapacité physique temporaire, des indemnités journalières représentant la moitié de son gain journalier dans la limite de $(1/720)^{\text{ème}}$ du plafond annuel de la Sécurité Sociale. Des limites temporelles sont aussi appliquées puisqu'il existe un délai de carence de trois jours (\approx une franchise) en incapacité et la durée maximale de versement d'indemnités est de trois ans.

Cas particuliers :

- Les assurés avec au moins trois enfants à charge reçoivent 2/3 de leur gain journalier à partir du 31^{ème} jour d'arrêt.
- En cas de maladie professionnelle ou d'accident du travail, les indemnités journalières s'élèvent à 60% du gain journalier pendant les 28 premiers jours d'arrêt et à 80% ensuite.

Le besoin d'un régime complémentaire en terme d'incapacité se justifie par le plafonnement du montant des indemnités journalières par la Sécurité Sociale.

2.4.2 : Les prestations complémentaires

Elles permettent à l'assuré arrêté de compenser ou du moins de limiter la perte de revenus subie en raison du plafonnement des prestations servies par la Sécurité Sociale.

La validation de l'état d'incapacité est soumise à l'avis d'un médecin. Il faut cependant distinguer les prestations versées au titre d'un contrat collectif obligatoire de celle versées sur une assurance individuelle facultative. En effet, le caractère obligatoire de l'assurance collective et la volonté législative de protection de l'assuré en matière de prévoyance imposent certaines règles à l'assureur. La loi Evin qui sera détaillée ultérieurement précise les modalités à respecter pour la couverture (obligatoire ou facultative) de risques sociaux.

2.4.3 : Quelques chiffres

Dans les résultats, l'incapacité est rarement dissociée de l'invalidité ou de la santé, ils forment ensemble les dommages corporels. Voici quelques chiffres concernant le chiffre d'affaires et les prestations versées au titre des dommages corporels :

(en milliards de francs)	contrats individuels			contrats collectifs			Total affaires directes		
	soins de santé	indemnités	total	soins de santé	indemnités	total	soins de santé	indemnités	total
chiffres d'affaires dommages corporels en 1998	15,7	11,6	27,3	11,8	19,8	31,6	27,5	31,4	58,9
chiffres d'affaires dommages corporels en 1999	16	12,4	28,4	11,7	19,2	30,9	27,7	31,6	59,3

sources : Commission de contrôle et FFSA(sociétés vie, dommages et mixtes)

affaires directes (en milliards de francs)	PRESTATIONS	DOTATIONS AUX PROVISIONS	TOTAL
charges des prestations dommages corporels en 1998	42,9	4,8	47,7
charges des prestations dommages corporels en 1999	43,7	4,5	48,2

sources : Commission de contrôle et FFSA(sociétés vie, dommages et mixtes)

Les assurances de dommages corporels (santé et accidents)

Les garanties offertes sont souscrites dans le cadre de contrats d'assurance santé ou accidents corporels, ou sous la forme de garanties complémentaires annexées aux contrats d'assurance vie.

Les cotisations d'assurance santé et accidents corporels s'élèvent à 59 milliards de francs en 1999 (en affaires directes), soit un montant identique à celui de 1998. Les cotisations versées au titre de la protection complémentaire santé et accidents ont progressé très faiblement au cours des quatre dernières années. Cela résulte, d'une part, des distorsions de concurrence, principalement d'ordre fiscal, qui perdurent sur ce marché et, d'autre part, du non-renouvellement de certains contrats.

Ces cotisations se répartissent entre les garanties de remboursements de frais de soins (prestations en nature) pour 28 milliards de francs, en faible progression par rapport à 1998 - la progression des contrats individuels étant supérieure à celle des contrats collectifs -, et les garanties de versements d'indemnités (prestations en espèces) pour 31 milliards de francs, avec une croissance soutenue pour l'assurance individuelle.

2.4.4 : Chiffres AXA

Sur le périmètre retenu, c'est à dire les assurés d'un portefeuille vie individuelle d'AXA, voici les volumes obtenus lorsque l'on ne considère que les contrats en cours début 2001 et possédant des garanties dommages corporels (incapacité, invalidité et décès accidentel) :

	Total
Nombre de contrats en cours avec des garanties dommages corporels	269 164
Provisions (en F) pour les dommages corporels	486 396 396

Section 2.5 : Le cadre législatif

Toute société d'assurance est soumise à une réglementation stricte qui permet de limiter les risques de faillite et ainsi de protéger les assurés.

Depuis 1990, le législateur a cherché à protéger et sécuriser les opérations de prévoyance en instaurant des règles de plus en plus exigeantes pour le provisionnement du risque incapacité / invalidité.

En effet, avant la loi évin du 31 décembre 1989, les organismes assureurs n'étaient soumis à aucune obligation en matière de provisions. Le principe de répartition imposait simplement que les primes de l'année couvrent les sinistres de l'année.

En principe, les organismes assureurs utilisaient les méthodes proposées par le guide de l'assurance de groupe du Bureau Commun des Assurances Collectives (BCAC). Ce guide définit les trois provisions suivantes :

- provision pour incapacité (pour couvrir les indemnités à verser à l'incapable).
- provision pour rente en attente (lorsque l'assuré est en incapacité il y a un risque de passage en invalidité).
- provision pour invalidité (pour couvrir la rente totale à verser à l'invalidé).

A l'époque, ces provisions étaient calculées sur la base des travaux réalisés par **M. Wetzel** dans les années 60 et la différenciation par âge était imprécise (loi de maintien des invalides entrés à 50 ans uniquement).

Cependant rien n'obligeait réellement l'assureur à utiliser cette méthode ni même à provisionner ces risques.

Dans un souci de protection de l'assuré, le législateur a donc imposé, par la loi Evin, certaines règles concernant les opérations de prévoyance complémentaire.

2.5.1 : La loi Evin : des provisions forfaitaires

Cette loi est le premier texte qui vise l'ensemble des organismes autorisés à pratiquer la prévoyance collective complémentaire.

L'article 7 de cette loi introduit notamment une obligation de provisionnement puisque «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 pour l'application de cet article précise le montant minimum à provisionner :

«

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

»

L'article 29-V de la loi précise cependant qu'il s'agit d'une mesure transitoire en attendant la mise au point de tables de provisions plus précises.

2.5.2 : Arrêté du 28 mars 1996 : l'utilisation de tables

Cet arrêté a créé l'article A 331-22 du Code des Assurances sur les provisions de prestations d'incapacité et d'invalidité. Contrairement à la loi Evin précitée, cet article s'applique aussi bien aux opérations individuelles que collectives.

Il introduit :

- Les tables «officielles » de maintien en incapacité de travail et invalidité, servant de base au calcul des provisions.
- La possibilité d'utiliser une loi de maintien certifiée.
- Le taux d'actualisation à appliquer.

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

1° les lois de maintien en incapacité de travail et en invalidité indiquées en annexe (de l'article ou de la note du BCAC du 27 août 1993).

Toutefois, il est possible pour une entreprise d'assurances d'utiliser une loi de maintien 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 la commission de contrôle des assurances ;

2° un taux d'actualisation qui ne peut excéder 75% du taux moyen des emprunts de l'Etat français calculé sur base semestrielle, sans pouvoir dépasser 4,5% . »

« les dispositions du présent arrêté sont applicables à compter du 1^{er} janvier 1997. »

Cependant, les organismes assureurs concernés ont la possibilité d'étaler sur 5 ans (soit jusqu'au 31/12/2001) la charge qui résulterait de l'utilisation des tables pour le calcul des provisions.

Par ailleurs, il ne faut pas oublier que les assureurs sont tenus à des principes de base de provisionnement rappelés dans les articles L331-1 et R 331-1 relatifs à la constitution des provisions.

Enfin, signalons qu'en matière de tarification, aucune certification n'est imposée pour l'utilisation de tables d'expériences.

Section 2.6 : Intérêt des tables d'expérience

Les tables réglementaires du Bureau Commun des Assurances Collectives (BCAC) ont été construites à partir d'observations menées sur une population particulière (les assurés collectifs des principales compagnies françaises d'assurance) ; ainsi, cette population présentait des caractéristiques précises telles que, la catégorie socioprofessionnelle, une sur-représentation de certaines tranches d'âge par rapport à d'autre, une répartition hommes/femmes particulière.

Or ces facteurs impactent directement le risque incapacité (fréquence et durée) et la population couverte par les garanties incapacité d'une compagnie d'assurance particulière ne présente pas forcément les mêmes caractéristiques.

En effet, ces garanties ont pu être proposées préférentiellement à une catégorie socioprofessionnelle par exemple.

Le niveau de sinistralité observé sur cette population pourra donc s'écarter sensiblement de celui donné par les tables réglementaires.

L'organisme assureur risque alors de se retrouver en situation de sur-provisionnement ou de sous-provisionnement en utilisant les tables réglementaires.

Une table d'expérience construite en fonction de la population couverte et de la sinistralité observée sur cette population, permettra alors d'apprécier de manière satisfaisante le risque supporté.

Par ailleurs, les tables du BCAC ont été construites selon certaines définitions de l'état d'incapacité. Or les garanties proposées par l'assureur ne correspondent pas toujours exactement à ces définitions (périodicité des garanties différente ou garanties qui se poursuivent au-delà de l'âge limite des tables du BCAC : 60 ans).

Dans ces différents cas de figure, l'outil tables réglementaires n'est pas adapté aux garanties et donc pas utilisable tel quel pour le calcul des provisions, d'où l'avantage de construire une table d'expérience à partir d'observations menées sur le portefeuille de l'assureur.

D'après ce constat, il a été décidé de construire des tables d'expérience basées sur le portefeuille des assurés d'AXA ayant souscrits un contrat comportant au moins une garantie complémentaire en cas d'incapacité. Ces dernières concerneront **l'entrée et le maintien en incapacité** uniquement.

PARTIE II :
LES DONNEES

Partie II : Les données

Une table d'expérience repose par définition sur des données issues de l'observation d'un échantillon. Ces données seront détaillées dans le premier chapitre. Le suivant exposera les choix fondamentaux adoptés pour la construction des tables. Enfin, pour obtenir des résultats fiables, un travail préliminaire de traitement des données aberrantes sera effectué dans le dernier chapitre.

Chapitre 1 : Présentation des données

Section 1.1 : Les garanties du portefeuille étudié

Dans le portefeuille étudié, on s'intéresse aux garanties comportant une prestation liée à l'incapacité:

- versement d'indemnités journalières (IJ) en cas d'incapacité que la cause soit un accident ou une maladie. La tête assurée doit être précisée car il est possible d'introduire plusieurs assurés sur le même contrat. Le montant et la durée de la garantie ainsi que la durée de la franchise sont fixés au contrat.
- versement d'indemnités journalières (IJH) en cas d'hospitalisation uniquement.
- Exonération (EXO) pour l'assuré du paiement des primes dues au titre d'une période où il était en arrêt de travail.

Section 1.2 : Les fichiers de données

Les fichiers utilisés sont sous forme de tables SAS.

1.2.1 : Leur élaboration

Pour l'élaboration de ce mémoire, des fichiers ont été extraits du système informatique de gestion. Cette extraction concernait tous les contrats (en cours ou terminés) comportant au moins une garantie incapacité, et seules les variables utiles pour notre étude étaient sélectionnées.

Au final, nous disposons, des dernières images contrat. En effet il aurait été trop long et trop volumineux d'extraire pour chaque contrat l'historique de toutes ses images.

De plus, toujours pour des raisons de volume et de traitement informatique, il existe un archivage annuel pour les contrats du portefeuille fermés depuis plus de 5 ans. Comme le dernier archivage date d'août 1999, nous n'observons la totalité du portefeuille qu'à partir de septembre 1994 (inclus), les contrats fermés avant cette date se trouvent donc dans un historique qui n'était pas disponible pour l'élaboration de ce mémoire.

1.2.2 : Leur structure

a) Construction des lignes des fichiers :

L'ordre de fabrication des fichiers est le suivant :

- Identification du contrat (n° de contrat)
- Au moins une ligne par garanties au contrat (code garantie) (une ligne seulement quand il n'y a pas de sinistre sur la garantie)
- Quand il y a des sinistres, autant de lignes par garantie qu'il y a de sinistres touchant cette garantie, en fait l'information liée à un sinistre se trouve sur chaque ligne de garantie touchée par ce sinistre.

Illustration :

- Un contrat ayant eu deux sinistres et composé de trois garanties (touchées par les sinistres), aura six lignes dans le fichier.
- Un contrat avec trois garanties n'ayant jamais eu de sinistres ouverts, aura trois lignes dans le fichier.
- Un contrat avec trois garanties et deux sinistres, dont l'un touche les deux premières garanties et l'autre seulement la première, aura quatre lignes dans le fichier.

b) Description des variables utilisées :

Voici la description des variables présentes dans les fichiers SAS, classées suivant leur niveau d'information:

Variables liées uniquement au contrat :

- le numéro du contrat (clé pour identifier un contrat)
- le code produit du contrat
- le code situation du contrat au moment de l'extraction du fichier
- la date de la situation du contrat, c'est-à-dire, la dernière date de changement de situation du contrat
- la date d'effet du contrat
- la date du terme du contrat (elle s'obtient en ajoutant la durée de la garantie principale à la date d'effet du contrat)
- la prime totale pour le contrat
- le code fractionnement de la prime (unique, annuel, semestriel, trimestriel, bimestriel ou mensuel)
- le nom du 1^{er} au 5^{ème} intervenant (le souscripteur, le 1^{er} assuré, le cas échéant le 2^{ème} et / ou le 3^{ème} assuré, et le bénéficiaire)
- le code donnant le rôle du 1^{er} au 5^{ème} intervenant, il permet d'isoler le 1^{er} assuré
- la date de naissance du 1^{er} au 5^{ème} intervenant
- le sexe du 1^{er} au 5^{ème} intervenant
- la CSP du 1^{er} au 5^{ème} intervenant
- le code postal du 1^{er} au 5^{ème} intervenant

Variables liées uniquement à la garantie complémentaire :

- le code de la garantie complémentaire
- le libellé court de la garantie complémentaire (clé choisie pour identifier la garantie)
- le libellé long de la garantie complémentaire
- le capital de base de la garantie : pour les IJ, il s'agit des indemnités journalières, et pour les exonérations, il s'agit de la prime annuelle
- le montant de la prime par garantie
- le pourcentage ou pour millage de surprime lorsque l'assuré est classé parmi les risques aggravés, médicaux, professionnels ou sportifs

Variables liées au sinistre :

- le numéro de sinistre (clé pour identifier un sinistre)
- la tête de l'assuré qui est en sinistre
- le nom du bénéficiaire
- la cause du sinistre (A = Accident ou M = Maladie)
- le code hospitalisation (l'assuré est hospitalisé : O = Oui ou N = Non)
- la date de survenance du sinistre
- la date de déclaration du sinistre
- le nombre de jours déjà réglés au titre du sinistre en cours
- le nombre de jours restants indemnisables au titre du sinistre en cours
- le nombre de sinistres antérieurs
- la date de la fin de période indemnisée au titre de la garantie
- la date de la fin de sinistre lorsque celui-ci est terminé
- le montant déjà réglé au titre du sinistre et de cette garantie
- le nombre maximum de jours indemnisables pour un sinistre
- le nombre maximum de jours indemnisables restants pour le sinistre
- date de début du 1^{er} au 20^{ème} creux (courte reprise d'activité)
- le nombre de jours pour le 1^{er} au 20^{ème} creux
- l'indicatif du type du 1^{er} au 20^{ème} creux
- la durée de la franchise pour respectivement les motifs 1 à 4 qui sont :
Motif1 = Maladie sans hospitalisation
Motif2 = Maladie avec hospitalisation
Motif3 = Accident sans hospitalisation
Motif4 = Accident avec hospitalisation
- le nombre maximum de jours indemnisables pour un contrat

Chapitre 2 : Les choix fondamentaux adoptés pour la construction des tables

Nous allons décrire les choix fondamentaux (autres que les hypothèses mathématiques), nécessaires à la construction des tables. Ces choix ont été effectués a priori en fonction de diverses contraintes pratiques ou liées à l'utilisation des tables par la compagnie d'assurance.

Par exemple:

- il faut tenir compte de l'hétérogénéité des données;
- faire en sorte que la sinistralité étudiée sur les différents segments soit stable pour permettre d'estimer la sinistralité future;
- s'adapter aux carences d'informations (les données les mieux référencées sont seulement celles qui sont nécessaires à l'émission et la gestion des contrats).

La description de ces choix est regroupée en trois points :

- choix de la période d'observation du portefeuille
- détermination du risque que la table doit modéliser
- choix de la dimension des tables et des conventions de base à appliquer.

Section 2.1 : Choix du périmètre de l'étude : la période d'observation

Le nombre d'années d'observation des assurés doit être suffisamment grand pour obtenir des volumes conséquents sans avoir à remonter vers un passé trop lointain et risquer d'intégrer des données déjà caduques.

En effet, la santé du portefeuille peut être perturbée par des épidémies. Elle a aussi la particularité d'être influencée par le contexte économique et fiscal qui accentue l'autosélection des assurés. La table d'expérience postule une stabilité de la santé des assurés par rapport à la période d'observation.

Etant donné que l'entrée en incapacité est un événement ponctuel et non un état qui se prolonge dans le temps comme le maintien en incapacité, les contraintes dans le choix de la période d'observation ne sont pas les mêmes pour ce qui est de l'entrée et du maintien. En effet, l'incapacité peut durer jusqu'à 3 ans. L'étude du maintien exige donc une assez large période d'observation contenant au moins un cycle de 3 ans pour éviter entre autre d'avoir une trop grande proportion de données tronquées ou censurées.

Périmètre des fichiers disponibles :

ils contiennent tous les contrats en cours ou terminés avec une garantie incapacité à l'exclusion de ceux fermés depuis plus de 5 ans (par rapport à la date du dernier archivage : août 1999). De plus, l'informatisation du portefeuille date de novembre 1989, ainsi, les sinistres survenus avant cette date ne sont pas complètement renseignés et donc inutilisables pour notre étude.

Suite à ces remarques, il est évident que l'observation ne peut débuter avant 1990 (manque d'information).

2.1.1 : Période d'observation pour l'entrée en incapacité

Nous savons que notre portefeuille est incomplet avant 1994 et plus précisément avant août 1994. Il est clair que ces contrats manquants modifient la fréquence réelle des sinistres et biaisent les résultats sur l'entrée en incapacité.

De plus, compte tenu des déclarations de sinistre tardives, il manque certains sinistres survenus récemment, nous ne pouvons donc pas inclure les derniers mois observés sans introduire un biais dans les résultats. Par conséquent, voici le périmètre envisagé pour l'étude de l'entrée en incapacité :

de septembre 1994 à février 2000 inclus

La période d'observation pour l'étude des taux d'entrée en incapacité dure ainsi 5 ans et demi, ce qui semble largement suffisant, nous verrons par la suite les volumes correspondants.

2.1.2 : Période d'observation pour le maintien en incapacité

Nous savons que la durée d'observation joue un rôle important puisque la proportion de censures et de troncatures en dépend. Il a donc fallu examiner dans le détail les arguments pesant pour ou contre l'inclusion de la période 1990 à 1994 dans l'observation. Nous l'avons rappelé, l'extension de la période d'observation permettrait de travailler avec un volume de données plus important et moins censuré qui conduirait à des estimations plus fiables.

Cependant, les éléments suivants remettent en cause l'inclusion de la période 1990 à 1994 :

- plus les observations sont éloignées dans le temps, plus on est soumis au risque d'avoir des données caduques car la sinistralité a pu changer en réponse à l'évolution de la médecine par exemple ou à cause d'une modification de l'environnement économique etc...
- étant donné qu'il manque une partie des contrats sur cette période, il manque aussi une partie de la sinistralité, ce qui crée un déséquilibre dans la répartition de la sinistralité sur la période totale d'observation.
- on ne connaît pas cette sinistralité manquante (ni en nombre, ni en durée), il est donc difficile de connaître l'impact qu'elle aurait eu sur la sinistralité totale. Cependant, des hypothèses sur ces contrats manquants sont envisageables. En effet, ils ont disparu de nos fichiers parce qu'ils étaient fermés (avant septembre 1994), on peut donc penser que la majorité d'entre eux avait atteint leur terme (cause normale de fermeture), les assurés sinistrés sur ces contrats étaient donc proches de leurs 65 ans. La plupart des sinistres manquants seraient ainsi des incapacités après 60 ans. Dans le cas d'une inclusion de la période 1990-1994, il faudrait donc émettre une réserve sur les durées de maintien concernant des entrées en incapacité tardives (après 60 ans).
- Nous disposons de septembre 1994 à février 2000 d'une période d'observation fiable d'une durée de 5 ans et demi, soit presque 2 cycles entiers d'incapacité (2 fois 3 ans), ce qui semble a priori raisonnable.

Il semble donc plus judicieux d'exclure la période 90-94 de notre périmètre d'étude du maintien en incapacité. Ainsi, nous serons certains de connaître la sinistralité complète du portefeuille sur la période d'observation.

Finalement, la même période d'observation est retenue : [septembre 1994, février 2000] pour l'élaboration des lois d'entrée et de maintien en incapacité.

Section 2.2 : Choix des risques à évaluer

2.2.1 : L'exclusion des risques aggravés

Notre table doit représenter la sinistralité d'un assuré payant une prime normale. Les cas atypiques pour lesquels une tarification spéciale est adoptée ne doivent pas être pris en compte. Pour créer une table applicable à des groupes homogènes d'assurés, nous conserverons des classes de risques dits "normaux". On entend par risque "normal", le risque qui peut être traité par la compagnie d'assurance comme un cas assurable dans des conditions ordinaires et présentant un risque financier proche de la moyenne. Un risque peut être accepté par l'assureur soit en risque normal, soit en risque aggravé avec une surprime. On suppose que les contrats supprimés ont une sinistralité plus élevée. Les degrés d'aggravation sont très divers et évalués par des médecins par rapport à un risque dit "normal". Nous allons exclure les risques aggravés et ne construire une table qu'à partir des risques normaux.

2.2.2 : Le risque d'indemnisation

L'assureur cherche à chiffrer ses dépenses en terme d'indemnisation de ses assurés. De ce fait, pour évaluer le risque d'entrée en incapacité, seuls les sinistres qui dépassent la période de franchise sont à retenir puisque les autres sinistres ne sont pas indemnisables. Ils ne rentrent pas dans le risque supporté par l'assureur et sont rarement connus de lui. Ces inconnues biaiserait toute étude sur l'entrée absolue en incapacité. Ainsi quand on parlera « d'entrée en incapacité » il faudra entendre « entrée en incapacité + maintien au-delà de la franchise ».

De même pour le maintien en incapacité, nous évaluerons uniquement le maintien au-delà de la franchise et non la durée probable à partir de la date de survenance du sinistre.

De plus, l'indemnisation étant liée au contrat et à la garantie, l'échéance de l'un ou de l'autre mettra fin au droit à indemnisation et de ce fait clôturera tout sinistre en cours. Or, nous souhaitons évaluer uniquement le maintien indemnisable, ainsi, toute fin de garantie sur un sinistre en cours correspondra à la fin de l'arrêt de travail.

A noter cependant que si la définition des conditions d'indemnisation (et notamment de terme) changent, la sinistralité de la table ne sera plus représentative d'une sinistralité future.

2.2.3 : L'indépendance des risques

Pour satisfaire aux conditions d'indépendance généralement exigées pour construire des estimateurs corrects ou pour tester l'adéquation aux hypothèses, nous avons décidé de construire des tables « par tête » et non par contrat. Ainsi un assuré possédant plusieurs contrats ne comptera qu'une seule fois, mais sera exposé au risque incapacité sur toutes les périodes garanties par ses contrats.

Section 2.3 : La dimension âge et les conventions utilisées

Sachant que l'âge de l'assuré est un critère généralement déterminant pour la sinistralité et que les tables réglementaires sont proposées par âge, notre objectif est de construire des tables d'entrée et de maintien qui seront fractionnées par âge.

L'âge étant la dimension principale des tables, il est important de définir les dates de changement d'âge et les âges correspondants.

Nous savons que si elle a lieu, l'évolution du tarif d'un contrat en fonction de l'âge se produit à la date anniversaire du contrat. Etant donné que les lois d'entrée et de maintien en incapacité permettront d'ajuster les tarifs, il est nécessaire pour être cohérent de choisir cette date anniversaire du contrat comme date de changement d'âge.

Notations :

Si l'on note « x » l'âge de l'assuré, le changement de classe d'âge se traduit par le passage de l'âge x à l'âge x+1. La classe d'âge x sera notée $[x, x+1[$.

Il reste maintenant à définir quel est cet âge x de l'assuré à partir de la date anniversaire du contrat.

Parmi les conventions usuelles, nous avons le choix entre :

- L'âge que l'assuré a atteint à son dernier anniversaire
- L'âge qu'il atteindra à son prochain anniversaire
- L'âge qu'il atteint dans l'année civile en cours
- L'âge atteint à sa date d'anniversaire la plus proche

Les 3 premières propositions ont le même inconvénient de permettre dans certains cas un décalage de près d'un an entre l'âge réel de l'assuré et l'âge convenu à la date anniversaire du contrat, alors que la dernière proposition entraîne au maximum un écart de 6 mois.

De plus, notre portefeuille est constitué de contrats pour lesquels le principe à la souscription est de donner à l'assuré l'âge atteint à son anniversaire le plus proche.

Il est donc logique et cohérent de choisir aussi cette convention pour élaborer nos tables.

Illustration :

Un assuré né le 10/02/1940, ayant souscrit un contrat le 20/01/1980 aura 40 ans à la souscription et entrera dans la classe d'âge suivante soit $[41, 42$ ans[le 20/01/1981.

Dans le cas litigieux où il y a exactement 6 mois entre la souscription du contrat et l'anniversaire de l'assuré, nous prendrons par convention l'âge atteint au prochain anniversaire de l'assuré, car par prudence, il vaut mieux pour l'assureur sur estimer plutôt que sous estimer l'âge réel de l'assuré.

A noter que dans nos fichiers, la date de souscription du contrat sera matérialisée par la date d'effet du contrat.

Chapitre 3 : Préparation des données

Section 3.1 : Détection et traitement des valeurs aberrantes

La majorité des systèmes de gestion informatique traitant un important volume de données présentent des données aberrantes. Il faut les traiter avant de commencer l'étude afin que nos résultats soient les plus fiables possibles. Pour cela, nous avons corrigé les erreurs en nous basant sur les connaissances de la réalité du portefeuille par le personnel d'AXA. Ce chapitre présente les faiblesses des données (et donc fixe les limites de validité des travaux effectués) par l'étude d'une part des différentes variables intervenant dans l'analyse et d'autre part des cohérences entre elles.

Pour chaque type de variables envisagé, les valeurs aberrantes sont classées suivant la correction appliquée :

- une correction dite « manuelle » où l'on remplace la valeur aberrante par sa vraie valeur récupérée sur le système source ou sur indication des gestionnaires des contrats.
- une correction dite « automatique » où la valeur aberrante est remplacée par la valeur d'une autre variable susceptible de convenir.
- une correction par une approximation à l'aide d'autres variables ou de calculs de moyennes.
- aucune correction envisagée.
- suppression de l'observation aberrante.

Voici les différents formats de variables étudiées dans les tables SAS :

- format date
- format numérique
- format caractère
- format codé : certaines variables ne prennent que quelques modalités.

Différents tests ont été effectués suivant la forme des variables et leurs relations, pour mettre en évidence d'éventuelles valeurs aberrantes.

De plus, il faut respecter un certain ordre dans les différentes corrections à effectuer car d'une part certaines variables sont liées, ce qui implique de répercuter les modifications entre elles et d'autre part, pour extraire correctement le périmètre choisi [septembre 1994, février 2000] il faut procéder en premier lieu aux vérifications et corrections liées aux dates.

Les anomalies détectées dans les différents fichiers et traitées de manière à se rapprocher de la réalité du portefeuille sont explicitées ci-dessous.

3.1.1 : Les dates

A. Liées au contrat ou aux garanties :

Date d'effet du contrat :

On détecte une anomalie lorsque la date d'effet d'une garantie complémentaire précède la date d'effet du contrat, cependant, la garantie principale de ces contrats étant une rente différée, la date d'effet du contrat a été renseignée avec la date de début de versement de la prestation.

Par une **correction automatique**, la date d'effet de ces contrats a été remplacée par la date d'effet de leur garantie la plus ancienne.

B. Date liée à l'assuré :

Date de naissance :

Cette date est parfois manquante ce qui pose un réel problème puisque l'âge de l'assuré joue un rôle essentiel dans l'étude de la sinistralité et dans l'élaboration des tables.

En fait, si cette date manque dans notre fichier, c'est qu'elle correspond à la date de référence de SAS (01/01/1960) ou qu'elle n'existe pas dans le calendrier (par exemple le 29/02/1958 alors que 1958 n'est pas bissextile!).

Pour éviter d'exclure de l'étude les assurés sans date de naissance, nous avons procédé à une **correction manuelle**.

3.1.2 : Les variables numériques

1. les franchises :

- On détecte des incohérences lorsque la durée écoulée entre la survenance du sinistre et le dernier règlement est inférieure à la franchise et qu'il y a effectivement une indemnisation ou lorsque le nombre de jours réglés coïncide exactement avec la durée écoulée entre la survenance du sinistre et le dernier règlement (sans tenir compte de la période de franchise qui n'est pas nulle au contrat). Ceci s'explique par des remises commerciales ou par le fait que toutes les garanties touchées par le sinistre n'ont pas la même franchise.
- il a été décidé dans ces cas là, d'**annuler** la franchise au contrat, puisque ce qui intéresse l'assureur, c'est surtout la durée de la période d'indemnisation.

2. le nombre total de jours indemnisés pour un sinistre par garantie :

- On détecte les cas vraiment aberrants en comparant le nombre de jours réglés avec la durée écoulée entre la fin de la période de franchise et le dernier règlement. Si le premier est supérieur au second, alors le nombre de jour réglé est ajusté **automatiquement** pour combler la différence.

3. les montants garantis :

- détection de possibles valeurs aberrantes sur les garanties IJ lorsque le montant de l'IJ est supérieur à 5000 (que ce soit en franc ou en unité de compte) ensuite, on procède à une vérification manuelle sous le système d'information pour confirmer l'erreur.

En ce qui concerne le montant de base pour les garanties exonération de prime, il doit correspondre à la prime, il faut donc d'abord regarder le fractionnement de la prime (annuel, mensuel...) puis juger d'un seuil au-delà duquel le montant serait aberrant.

Il y a aussi les cas où le montant garanti n'est pas renseigné.

- Corrections envisagées : s'il y a eu un sinistre sur la garantie en question, on peut faire une **approximation** du montant garanti en divisant le montant total réglé par le nombre de jours réglés, si un de ces paramètres n'est pas renseigné ou s'il n'y a pas eu de sinistre alors on peut calculer un **montant garanti moyen**.

Remarque :

Il est important d'avoir des montants garantis corrects car ils serviraient de poids dans l'élaboration éventuelle de tables d'entrée et de maintien par capitaux. En effet, ces tables seraient nécessaires si la sinistralité était hétérogène suivant le capital souscrit.

3.1.3 : Les autres vérifications pour détecter des incohérences

1) $Nb \text{ de jours réglés} * \text{montant de base de la garantie} = \text{montant total réglé pour la garantie} ?$

Il faut savoir que le montant de base de la garantie dont nous disposons est en fait le montant actuel, il ne coïncide donc pas forcément avec le montant utilisé lors du règlement du sinistre si celui-ci n'est pas récent. En effet, il peut y avoir eu une revalorisation de la garantie ou au contraire une réduction. Ceci explique pourquoi le test ne donne pas toujours l'égalité.

2) $Nb \text{ de jours réglés} + \text{franchise} + nb \text{ de jours restants} < nb \text{ de jours maximum par sinistre} ?$

Si on enlève la franchise, on a l'inégalité souhaitée pour tous les sinistres en cours.

Le nombre de jours maximum par sinistre correspond donc au nombre de jours maximum d'indemnisation et pas au nombre de jours de sinistre.

3) $Nb \text{ jours réglés} + \text{franchise} + nb \text{ jours restants} + nbj \text{ sin antérieur} < nb \text{ jours max par contrat} ?$

De même que précédemment, le nombre de jours maximum par contrat est en fait le nombre de jours maximum d'indemnisation en cumulant tous les sinistres du contrat.

4) *Contrats encore en cours alors que leur date de terme est dépassée :*

En considérant que c'est dû à un oubli ou un retard de gestion, ces contrats sont corrigés en modifiant leur position (passage **automatique** en position échu)

Conclusion :

Les corrections apportées aux données initiales permettent d'obtenir une base plus saine pour notre étude, mais il faut cependant garder à l'esprit que notre travail portera dans une certaine proportion sur des observations retraitées.

Section 3.2 : Traitements spécifiques pour adapter les données au risque étudié

3.2.1 : Création de la variable « fingar » : date réelle à laquelle la garantie se termine

La variable « fingar » nous sera utile pour déterminer la population sous risque à un moment donné, c'est à dire celle dont une garantie incapacité est encore en cours.

En effet, nous connaissons la dernière situation de la garantie, mais nous ne disposons pas de la date à laquelle elle est entrée dans cette situation. Ainsi, dans le cas où la garantie n'est plus en cours actuellement, nous ne savons pas toujours depuis combien de temps elle s'est terminée. Il est donc nécessaire de construire cette nouvelle variable à partir des éléments dont nous disposons.

principe de construction :

- ① Lorsque le contrat et la garantie sont tous deux encore en cours, ou lorsque la garantie est en position échue ou extinction (signe qu'elle est allée jusqu'à son terme prévu au contrat), « fingar » correspondra à la date (prévue au contrat) de fin de la garantie.
- ② Lorsque le contrat est terminé ainsi que la garantie, alors fingar sera égale à la date de situation du contrat.
- ③ Lorsque la garantie est encore en cours alors que le contrat est terminé, on considérera que c'est un oubli et on mettra encore dans fingar la date de situation du contrat.
- ④ Lorsque la garantie est terminée pour une autre raison que l'échéance et que le contrat est toujours en cours, il n'y a aucun moyen de connaître la date réelle de fin de la garantie. C'est pourquoi, nous procéderons à une *approximation* de cette date par le milieu entre le dernier jour réglé pour le sinistre le plus récent et la date d'extraction du fichier, en cas de sinistre sur cette garantie. Dans le cas contraire, fingar sera égale au milieu entre le début de la garantie et la date d'extraction du fichier.

3.2.2 : Création de la variable fra = durée de la franchise pour un sinistre

Les durées de franchises au contrat sont renseignées dans 4 variables différentes qui contiennent la franchise à appliquer suivant le motif du sinistre (accident ou maladie, avec hospitalisation ou non). La nouvelle variable « fra » donnera pour chaque sinistre la franchise réellement appliquée.

3.2.3 : Création de la variable « datefin » : date de fin du contrat

principe de construction :

Lorsque le contrat est dans une position qui signifie qu'il n'y a plus d'exposition au risque incapacité (résilié, racheté, réduit...etc), la date de situation permet de savoir depuis combien de temps il est dans cette situation.

Ainsi, « datefin » contient la date de terme du contrat si celui-ci est encore en cours ou en position « échu » et elle contient la date de situation du contrat si celui-ci est terminé pour une autre cause que le terme.

Des tests ont été réalisés pour vérifier s'il n'y avait pas d'anomalie sur cette nouvelle variable en la comparant aux dates disponibles sur le contrat :

- si datefin est antérieure à la date d'effet du contrat et s'il n'y a pas de sinistre alors on *supprime* ce contrat.
- lorsque des sinistres dépassent ou surviennent après datefin, une correction *manuelle* est effectuée pour remplacer datefin par la date de dernier règlement du dernier sinistre sur le contrat.

3.2.4 : Création des variables propres à l'assuré

Les variables propres aux intervenants (nom, date de naissance etc ...) sont renseignées pour chaque individu lié au contrat. De ce fait on dispose de 5 séries de variables (une pour chaque intervenant). Pour réaliser notre étude, il est nécessaire d'isoler, parmi les 5, la série propre à l'assuré (en utilisant le code contractant qui l'identifie).

En effet, nous souhaitons construire des tables « par tête » ce qui signifie que notre clé primaire sera : le nom + la date de naissance de l'assuré. De cette façon, les assurés ayant plusieurs contrats ne compteront qu'une fois mais seront observés sur la période cumulée de tous leurs contrats.

3.2.5 : Création de l'âge à la souscription et à la survenance

Pour notre étude, l'âge de référence sera donné par le calcul de l'âge à la souscription du contrat comme indiqué dans la convention choisie. A partir de cet âge, on détermine aussi l'âge à l'entrée en sinistre.

3.2.6 : Création des périodes d'exposition par âge

Dans le portefeuille étudié, tous les assurés ne sont pas exposés sur la totalité de la période d'observation choisie, il est donc nécessaire de déterminer la période d'exposition de chacun pour ne pas biaiser les estimations des tables.

Nous avons choisi de prendre comme début d'exposition la date d'effet de la plus ancienne des garanties complémentaires et comme date de fin d'exposition le minimum entre la date de fin du contrat et la fin de la garantie qui se termine le plus tard, tout en bornant cette période avec la période observée.

De cette manière, toute personne ayant une durée d'exposition nulle sera exclue du périmètre retenu.

Ensuite, nous avons scindé cette période d'exposition totale pour obtenir l'exposition sur chaque âge. Ainsi, nous disposons du nombre de jour exact d'exposition, pour chaque assuré et pour chacun des âges sur lesquels il est observé.

3.2.7 : Création du nombre de jours d'incapacité

Pour évaluer le maintien en incapacité, il est nécessaire de définir la durée en arrêt de travail que l'on va étudier. Ici, nous avons choisi de retenir la durée indemnisable au titre de l'incapacité c'est à dire le nombre de jours indemnisés corrigés de la différence suivante :

{fin de la période d'incapacité reconnue par le médecin} – {date de dernier règlement}

- Si la différence est positive, elle représente le nombre de jours restant à indemniser sur le sinistre, il faut donc les ajouter à ceux déjà indemnisés, sauf si l'écart est trop important (plus de 8 mois) ou si le dernier règlement est trop ancien (avant 1999). En effet, on peut expliquer une différence assez récente et peu importante par un délai de gestion dans le règlement des sinistres, les autres cas étant considérés comme des anomalies.
- Sinon, lorsque la différence est négative, il faut la retrancher. Ce cas se produit uniquement sur des garanties exonération lorsque l'incapacité est suivie d'une invalidité et que la garantie exonère aussi la prime en cas d'invalidité. Certains gestionnaires ne clôturent pas immédiatement le sinistre incapacité pour ouvrir un sinistre invalidité. De ce fait, l'exonération de prime qui est aussi garantie en cas d'invalidité continue de fonctionner. Il faut donc éliminer cette période d'indemnisation due au titre de l'invalidité et non au titre de l'incapacité.

3.2.8 : Création de la date de fin de sinistre et des indicatrices de censure et de troncature

Sur certains sinistres, la période d'incapacité ne se déroule pas en continu à partir de la fin de la période de franchise, en effet il est stipulé dans les conditions générales des contrats que « lorsqu'il s'écoule moins de 2 mois entre deux périodes d'incapacité concernant une même affectation ou un même accident, elles sont considérées comme une seule et même période ». On appelle alors creux la période d'interruption d'incapacité et on ne change pas de numéro de sinistre.

Pour faciliter le traitement des données propres au maintien en incapacité et éviter d'utiliser les variables liées aux creux qui ne sont pas très fiables, on neutralisera les creux existants en créant artificiellement la date de fin de sinistre = {date de survenance + franchise + nombre de jour d'incapacité }. Ainsi tous les sinistres sont traités en continu.

On peut alors déterminer les sinistres censurés, c'est à dire ceux dont la date de fin dépasse la période d'exposition de l'assuré (l'indicatrice de censure vaut alors 1) ainsi que les sinistres hors périmètre, c'est à dire ceux dont la date de fin ne dépasse pas le début de l'exposition, on supprimera ces sinistres.

On construit aussi l'indicatrice de troncature qui indique les sinistres tronqués, c'est à dire ceux dont la survenance est antérieure à la période d'exposition.

Remarque :

Ces indicatrices concernent la censure et la troncature comme on l'entend au niveau du maintien (*cf partie V chapitre 1*). En revanche, il faut noter que les sinistres tronqués ne feront pas partie de l'échantillon étudié pour l'entrée en incapacité puisque leur survenance n'a pas été observée sur la période retenue et qu'ils ne sont pas exhaustifs. En effet tout sinistre survenu le même jour qu'un sinistre tronqué mais n'ayant pas dépassé le début de la période d'observation a été éliminé du périmètre.

Section 3.3 : Volume des données après traitement

Les traitements présentés dans les parties précédentes modifient les volumes de données disponibles. Les tableaux récapitulatifs suivants permettent de comparer les volumes conservés par rapport aux volumes initiaux et fournissent la proportion de données utilisées qui proviennent d'une approximation sur les données initiales.

Puis, dans un dernier paragraphe, nous présenterons les volumes utilisés après modification de la structure des fichiers en observations annuelles.

3.3.1 : Les volumes traités sur les fichiers de base

a) Volumes de données supprimées :

		nombre de contrats	nombre d'arrêts
volumes initiaux		455877	61770
volumes après traitement	en absolu	450490	61317
	par rapport au volume initial	99%	99%

b) Volumes de données traitées sur le fichier final :

motif	volume par rapport au fichier final
principales approximations propres aux garanties	
remplacement des montants garantis manquants par un montant moyen	25%
approximations dans la création de la date de fin réelle de la garantie	insignifiant
principales approximations propres aux contrats	
remplacement de la date d'effet du contrat par celle de la garantie quand elle précède	0,2%
passage en position échue des contrats ayant dépassé leur terme	0,7%
corrections sur la variable date de fin du contrat que nous avons créée	insignifiant
principales approximations propres aux sinistres	
annulation des périodes de franchises non respectées	2,4%
ajustement du nombre de jours indemnisés par rapport à la durée indemnisable de la survenance au dernier règlement	2,2%

Remarques :

- Les pourcentages sont exprimés en fonction du fichier final mais aussi en se ramenant à une ligne par garantie quand on s'intéresse aux corrections propres aux garanties, ou une ligne par contrat quand on s'intéresse aux corrections propres aux contrats... etc
- Les volumes de corrections propres aux assurés sont insignifiants

Conclusion :

La proportion de données aberrantes est faible. L'utilisation de données en partie approximées ne peut donc remettre en cause la validité des fichiers utilisés. De plus, la seule approximation importante concerne les montants garantis qui ne seront pas utilisés dans notre étude.

3.3.2 : Transformation des fichiers par observation d'assurés en observation annuelle d'assurés.

Dans une seconde étape, les traitements transforment les fichiers pour tenir compte de la période d'observation qui s'écoule sur plusieurs années. En effet, un assuré peut être observé de 1994 à 2000. L'étude nécessite de travailler indépendamment âge par âge. Les fichiers séquentiels seront alors modifiés pour que l'unique séquence, qui correspond à l'observation d'un assuré pendant plusieurs années (six ans par exemple), soit décomposée en plusieurs (six par exemple) séquences à des âges différents.

Le tableau suivant fournit les volumes des fichiers après restructuration de manière à obtenir une séquence par année d'observation d'un assuré.

Volume des données exprimé en "année d'observation d'un assuré" :

nombre d'assurés dans les fichiers après traitement	430 916
nombre d'observations annuelles d'assurés dans les fichiers restructurés	1 689 671

Remarque :

Le volume est à peu près multiplié par 4. Ceci s'explique naturellement puisque la période d'observation dure environ 5 ans (de septembre 1994 à février 2000).

PARTIE III :

**EVALUATION
DU RISQUE INCAPACITE**

Partie III : Evaluation du portefeuille et du risque incapacité

Cette partie sera consacrée à la connaissance du portefeuille étudié aussi bien en terme de population et de contrat qu'en terme de sinistre. Cette évaluation est nécessaire pour savoir dans quel cas et sous quelles conditions les tables construites pourront être utilisées ultérieurement.

Ce premier aperçu de la sinistralité de nos assurés permettra aussi d'orienter la segmentation des tables.

Chapitre 1 : Descriptif du portefeuille

Dans ce chapitre, nous décrivons le portefeuille selon les caractéristiques propres au contrat puis selon les particularités des assurés.

Section 1.1 : caractéristiques des contrats souscrits

L'étude repose sur plus de 450 000 contrats (450 490 contrats pour être exact).

Le portefeuille d'incapacité correspond à presque 800 000 garanties (indemnités journalières et exonérations de prime soit 784 996 garanties en tout) .

Nous allons nous intéresser maintenant à la répartition des types de garanties sur les contrats.

1.1.1 : les combinaisons de garantie

Le tableau suivant donne la répartition des différentes garanties (IJ, IJH et EXO) sur les 2 types de contrats en portefeuille (« épargne » et « prévoyance ») :

en pourcentage	contrats de prévoyance	contrats d'épargne
que de l'exonération	36	5
que de l'ij normale	6	0
que de l'ijh	3	0
que de l'ij et de l'ijh	8	0
que de l'ij et de l'exo	18	0
que de l'ijh et de l'exo	4	0
toutes les garanties	19	0
total	95	5

Commentaires :

- 95% des contrats étudiés sont des contrats de prévoyance parmi lesquels, 36% des contrats ne sont exposés au risque incapacité qu'à travers la garantie exonération.
- 6% des contrats ne possèdent que de l'IJ toutes causes
- 3% des contrats ne possèdent que de l'IJ en cas d'hospitalisation et ne sont donc pas exposés au risque incapacité sans hospitalisation

finalement, 51% des contrats possèdent au moins une garantie IJ, 82% des contrats contiennent au moins une garantie exonération et 34 % au moins une garantie IJH.

1.1.2 : les franchises souscrites

Nous disposons en fait pour chaque garantie de quatre valeurs de franchise, une pour chacun des cas suivants :

- En cas de maladie sans hospitalisation (1)
- En cas de maladie avec hospitalisation (2)
- En cas d'accident sans hospitalisation (3)
- En cas d'accident avec hospitalisation (4)

Nous allons d'abord considérer la garantie IJ.

Parmi ces quatre franchises, nous allons nous intéresser principalement aux couples de franchises formés par la franchise (1) et la franchise (3) et noté : (1) / (3).

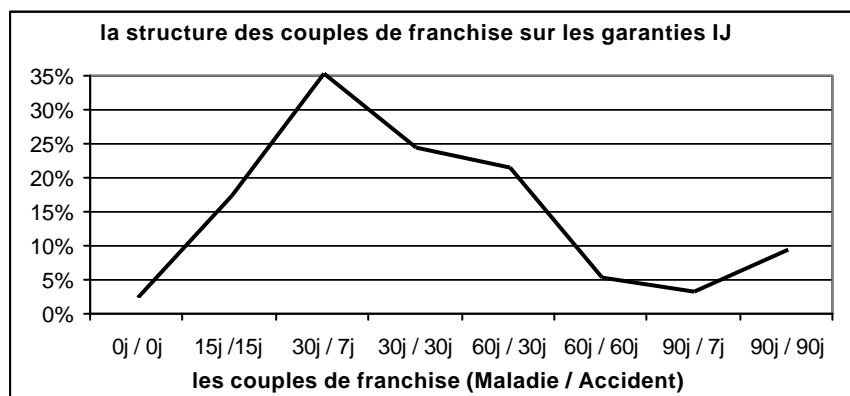
En effet, comme le montre le tableau ci-dessous, seule la cause (Maladie ou Accident) influence réellement sur la durée de la franchise.

les couples (1) / (3) de franchises sur les garanties IJ								
0j / 0j	15j / 15j	30j / 7j	30j / 30j	60j / 30j	60j / 60j	90j / 7j	90j / 90j	total
0%	15%	33%	22%	19%	3%	1%	7%	100%

* La franchise 0j/0j n'existe pas sur les IJ quand il n'y a pas d'hospitalisation

les couples (2) / (4) de franchises sur les garanties IJ								
0j / 0j	15j / 15j	30j / 7j	30j / 30j	60j / 30j	60j / 60j	90j / 7j	90j / 90j	total
7%	15%	27%	21%	19%	3%	1%	7%	100%

En fait la structure des couples (2) / (4) est quasiment identique à celle des couples (1) / (3), les couples sont identiques avec ou sans hospitalisation sauf pour 1/5 des garanties ayant 30j/7j (sans hospitalisation) et qui en cas d'hospitalisation ont une franchise nulle.



Le graphe précédent permet de visualiser la structure des couples de franchises de type (1) / (3) sur les garanties IJ.

Le couple de franchise (sans hospitalisation) principalement souscrit sur les garanties IJ est d'un mois en maladie et d'une semaine en accident.

Les couples de franchises sont répartis quasiment de la même manière sur les garanties EXO.

Quant aux garanties IJH, elles ne possèdent en fait que deux cas de franchise :

- Maladie en cas d'hospitalisation
- Accident en cas d'hospitalisation

Et finalement sur cette garantie, ces deux franchises sont toujours nulles.

1.1.3 : les montants des indemnités journalières

Voici la répartition (en %) des garanties IJ suivant leur montant et selon le sexe de l'assuré :

	homme	femme	total
en francs	%	%	%
moins de 50	2	2	2
50 à 100	33	37	34
100 à 150	40	39	40
150 à 200	11	11	11
200 à 300	10	8	9
300 à 400	3	2	2
400 à 500	1	1	1
plus de 500	1	0	0
total	100	100	100

- Les IJ se situent globalement (95%) entre 50 et 300 F et majoritairement (75%) entre 50 et 150 F la tranche principale étant de 100 à 150 F (40%).
- Les IJ sont globalement plus élevées chez les hommes que chez les femmes, cette différence peut venir d'un écart de rémunération.

Attention, 36% des contrats n'ont que de l'exonération et n'apparaissent donc pas ici où l'on ne traite que l'IJ.

La structure des montants garantis sur les IJH est semblable.

En ce qui concerne la garantie EXO, trop de montants manquent pour pouvoir faire la même étude.

Les montants moyens de chaque type de garantie sont présentés dans le tableau suivant :

	montant moyen
indemnité journalière IJ	125 F
indemnité journalière hospitalisation IJH	124 F
exonération de prime EXO	4 380 F

Section 1.2 : Etude de la population sous risque

La population étudiée comporte plus de 430 000 assurés.

Nous allons maintenant étudier les différentes caractéristiques de cette population. Ainsi, nous serons en mesure de déterminer par la suite lors de l'utilisation des tables s'il existe un risque de biais dû à l'évolution de notre population.

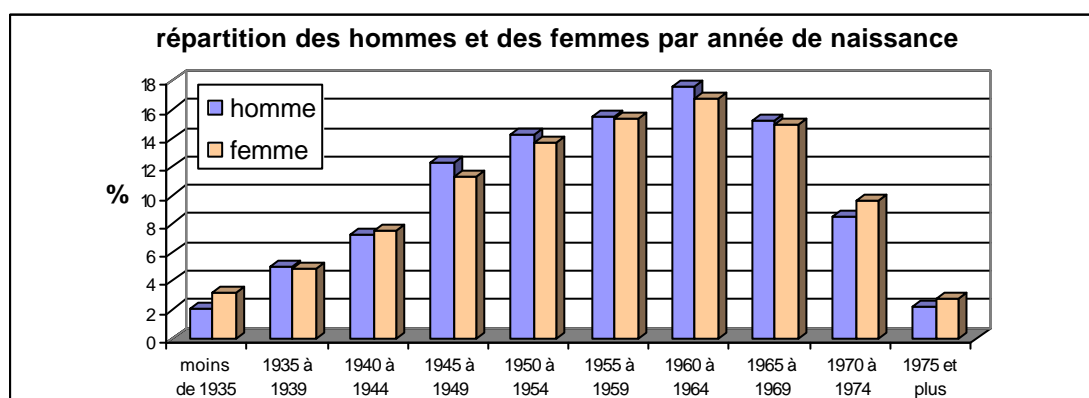
1.2.1 : Répartition Homme / Femme

Proportion d'hommes et de femmes en pourcentage	
hommes	66
femmes	34

Globalement le portefeuille est constitué d'1/3 de femmes et 2/3 d'hommes.

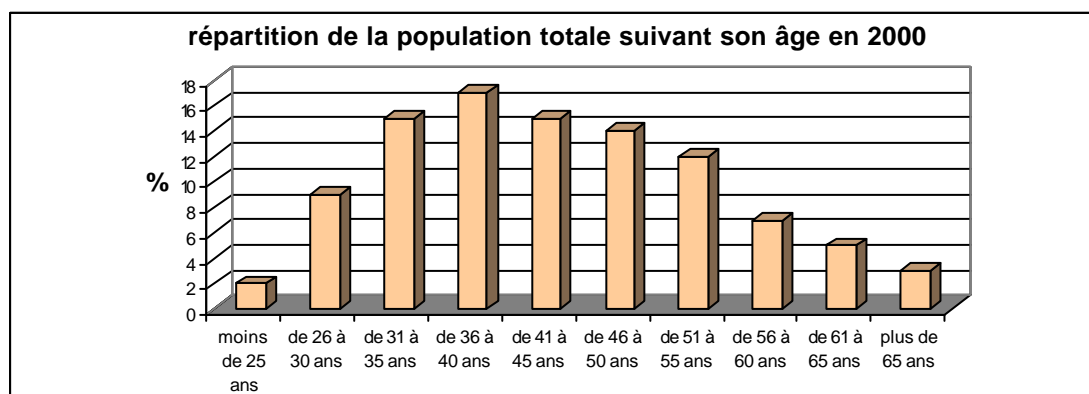
1.2.2 : Répartition des assurés par année de naissance

Les graphes ci-dessous présente la structure de la population masculine et féminine suivant l'année de naissance :



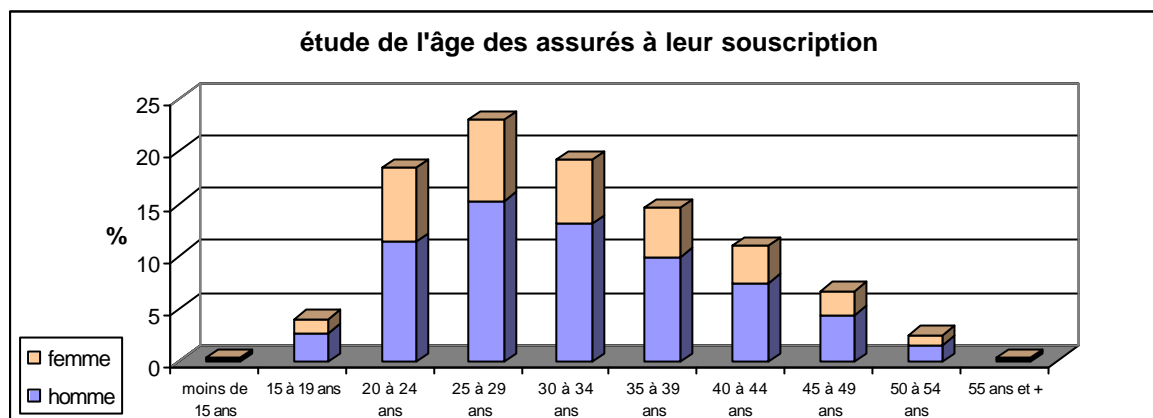
Tant dans la population masculine que féminine, les assurés sont majoritairement (73%) nés entre 1945 et 1969 ce qui signifie une forte proportion des assurés ayant entre 30 et 55 ans en 2000. La répartition sur cette période est assez homogène (de 12 à 17% par tranche de 5 ans).

La pyramide des âges de la population totale atteint son maximum sur la tranche 36-40 ans :



L'inconvénient de la structure des âges des assurés c'est que même sur un portefeuille à périmètre constant, elle évolue avec le temps, alors que la structure des âges à la souscription est fixe et définitive. C'est pourquoi nous la présentons dans le paragraphe suivant.

1.2.3 : Répartition des assurés suivant leur âge à la souscription



Les souscriptions se font principalement entre 25 et 30 ans (23%) et la fenêtre 20 à 45 ans regroupe la quasi-totalité des souscriptions (86%).

En plus de l'âge, la Catégorie Sosio-Professionnelle (CSP) des assurés est aussi une caractéristique importante de la population.

1.2.4 : Répartition des assurés par CSP

	femme	homme	total
total	%	%	%
AGRICULTEUR	2	4	3
ARTISANS-COMMERCANT	6	11	9
CADRE	2	5	4
COLLECTIVITE PUBLIQUE	0	0	0
ENTREPRISE	0	0	0
INDETERMINE	2	2	2
PROFESSION LIBERALE	3	3	3
RETRAITE	1	1	1
SALARIE	71	71	71
SANS PROFESSION	13	3	6
TOTAL	100	100	100

Globalement, le portefeuille étudié est composé à 71% de salariés tant pour les hommes que pour les femmes. Les artisans - commerçants avec 11% parmi les hommes et 6% parmi les femmes représentent la seconde CSP privilégiée en cible de clientèle. Les autres CSP sont peu représentées.

L'exposition au risque de notre population est certainement liée à sa profession, mais aussi à sa région.

1.2.5 : Répartition des assurés par région

Comme cette variable est très mal référencée sur certains fichiers, elle n'a été étudiée que sur les fichiers complets. Le tableau suivant donne la répartition des assurés selon leur région, pour le périmètre retenu qui représente 40% des effectifs.

sur 40% du portefeuille	
REGION	%
ALSACE	4
AQUITAINE	5
AUVERGNE	2
BASSE-NORMANDIE	2
BOURGOGNE	4
BRETAGNE	4
CENTRE	3
CHAMPAGNE-ARDENNE	2
CORSE	1
DOM	13
FRANCHE-COMTE	2
HAUTE-NORMANDIE	2
ILE-DE-France	8
INDETERMINE	9
LANGUEDOC-ROUSSILLON	4
LIMOUSIN	1
LORRAINE	5
MIDI-PYRENEE	3
NORD-PAS-DE-CALAIS	4
PAYS-DE-LA-LOIRE	3
PICARDIE	3
POITOU-CHARENTE	4
PROVENCE	5
RHONE-ALPES	7
Total	100

On constate que 13% des assurés étudiés sont domiciliés dans les DOM, viennent ensuite la région parisienne avec 8% des assurés et la région Rhône-Alpes avec 7%. Aucune autre tendance ne se dégage, car la répartition sur les autres régions est assez équilibrée (autour de 4% dans chaque région).

Enfin, pour connaître au mieux notre portefeuille, il est intéressant de regarder la proportion de multi-détenteurs, c'est à dire ceux qui ont souscrit plusieurs contrats.

1.2.6 : Les assurés ayant plusieurs contrats

Type de contrat détenus par les multidétenteurs

nb de contrats par assurés	%
1	93
2	6
3 et plus	1

types de contrats	%
une de la prévoyance	87
une de l'épargne	2
les deux	11

Les assurés ayant souscrit une garantie incapacité sont essentiellement détenteur d'un seul contrat (93%).

La majorité des multi-détenteurs disposent de deux contrats qui sont alors deux produits de prévoyance.

Chapitre 2 : Evaluation du risque incapacité

Pour éviter toute ambiguïté, nous allons d'abord définir les termes clés liés au risque étudié, et pour avoir une première idée de la sinistralité de notre portefeuille, nous étudierons ensuite les principales caractéristiques des arrêts de travaux et des sinistrés de notre base.

Section 2.1 : Définitions

D'après les dispositions communes relatives aux garanties complémentaires du portefeuille, voici quelques définitions liées au risque incapacité :

- **Maladie :**

Toute altération de la santé constatée par une autorité médicale compétente. L'état de grossesse n'est pas assimilé à une maladie, seules les complications qui en résultent éventuellement sont garanties.

- **Accident :**

Toute atteinte corporelle, non intentionnelle de la part de l'assuré et provenant de l'action soudaine d'une cause extérieure.

- **Hospitalisation :**

Tout séjour de plus de 48 heures dans un établissement de soins public ou privé dès lors que ce séjour a pour objet le traitement médical ou chirurgical d'un accident ou d'une maladie ou l'établissement d'un bilan de santé prescrit médicalement.

- **Incapacité temporaire et complète :**

Tout état, physique ou mental de l'assuré, résultant d'un accident ou d'une maladie mettant celui-ci dans l'obligation (reconnue par AXA) d'interrompre temporairement toute activité professionnelle ou, à défaut de l'exercice d'une profession, d'observer un repos complet. Ceci explique la présence d'assurés sans profession dans notre portefeuille.

Le risque étudié est **l'incapacité temporaire et complète** comme définie ci-dessus et que nous appellerons plus simplement incapacité ou arrêt de travail.

Section 2.2 : La répartition des sinistres

Pour construire des tables par tête, nous considérerons comme un seul sinistre tout sinistre survenu à la même date sur une même tête même si cet assuré possède plusieurs contrats sur lesquels ce sinistre est indemnisé.

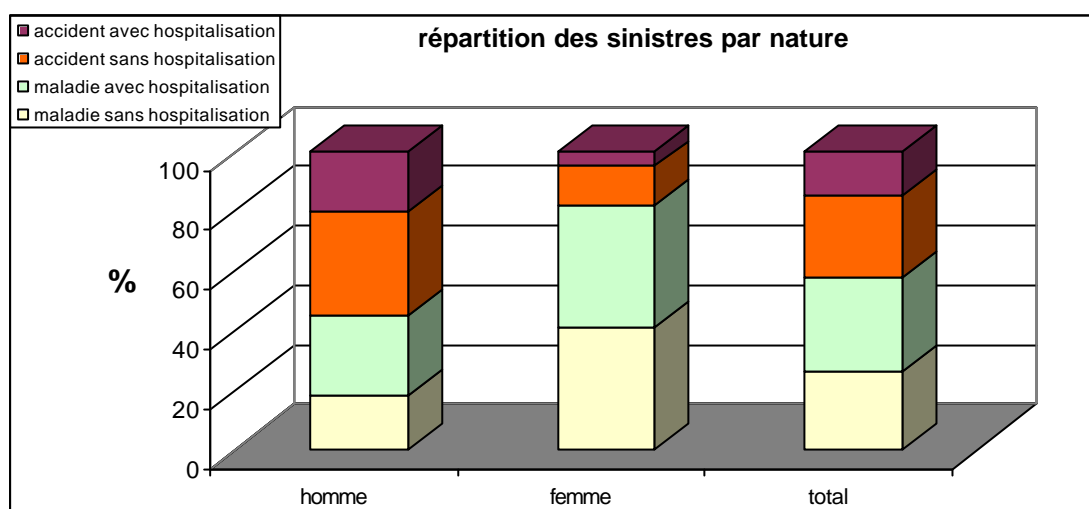
Les résultats suivants décrivent la répartition des sinistres suivant les variables descriptives propres à l'arrêt de travail (sa durée, sa franchise et sa cause).

Remarque :

les résultats présentés sont des moyennes calculées sur l'ensemble du portefeuille observé (pas de distinction par garantie, type de sinistre, âge d'entrée en sinistre ...).

2.2.1 : Répartition des sinistres suivant leur nature

En combinant la cause (accident ou maladie) avec l'existence ou non d'une hospitalisation on obtient quatre natures de sinistres dont les proportions apparaissent dans le graphe suivant :



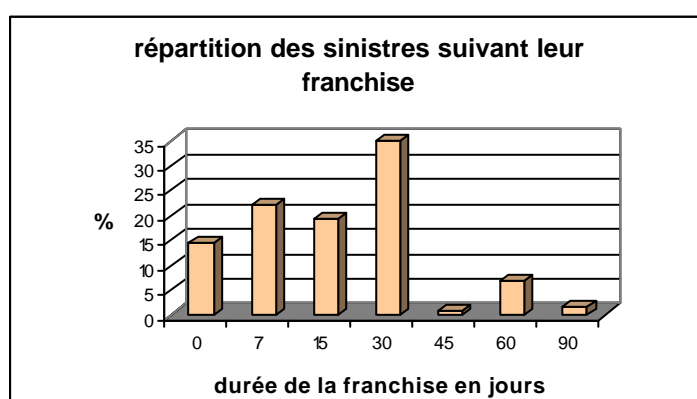
Dans l'ensemble du portefeuille, il y a presque autant d'arrêts de travail pour accident (42%) que pour maladie (58%).

Mais il en est tout autrement si l'on distingue le sexe de l'assuré. En effet, comme on pouvait le soupçonner, les femmes sont peu souvent en arrêt en raison d'un accident (20%), alors que le phénomène est inversé (en moins tranché) chez les hommes. La combinaison des deux tendances explique l'équilibre obtenu au global.

L'hospitalisation est plus fréquente sur les maladies que sur les accidents.

2.2.2 : Répartition des sinistres suivant leur franchise

Le graphe suivant indique la proportion de sinistres dans chaque durée de franchise :



Globalement les franchises ne dépassent pas 30 jours. Les sinistres du portefeuille ont principalement une franchise de 30 jours.

Nous allons maintenant nous intéresser à la durée des sinistres, un élément essentiel dans l'évaluation du risque incapacité.

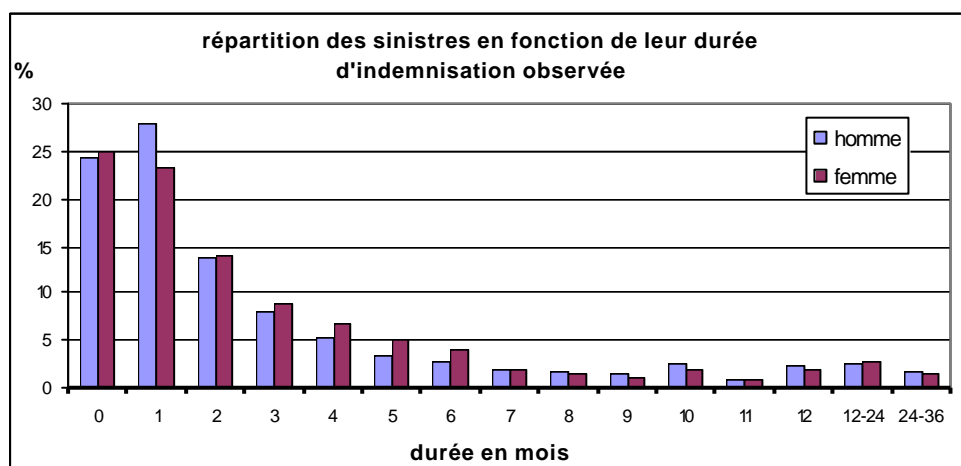
Section 2.3 : Etude de la durée des sinistres

Ici, nous étudions la durée « indemnisée » des sinistres, c'est à dire la durée au-delà de la franchise. Nous regarderons l'influence du sexe, de la franchise et de la nature du sinistre sur cette durée.

Remarques importantes:

- 1) Lorsqu'il y a plusieurs type de garanties (IJ ou EXO ou IJH) qui jouent pour un même sinistre, c'est la durée sur la garantie IJ qui prime ensuite, c'est la durée sur l'EXO. Et lorsqu'il y a plusieurs garanties IJ, c'est la durée maximum qui a été retenue.
- 2) Comme nous avons conservé tous les sinistres pour faire cette étude, nous avons mélangé les sinistres qui touchaient uniquement de l'IJH (que nous appellerons les sinistres hospitalisation) avec les autres sinistres. Donc les durées étudiées correspondent soit à des durées indemnisées d'incapacité soit dans certains cas (quand il y a uniquement une IJH sur la tête assurée) à des durées d'hospitalisation.
- 3) Lorsqu'on parle de durée indemnisée mensuelle, on entend la valeur entière la plus proche du {nombre de jours indemnisés divisé par 30,5}. Par exemple, les sinistres de moins de 15 jours sont classés dans la durée mensuelle «0 » et ceux qui durent entre 15 et 45 jours ont une durée mensuelle de « 1 ».

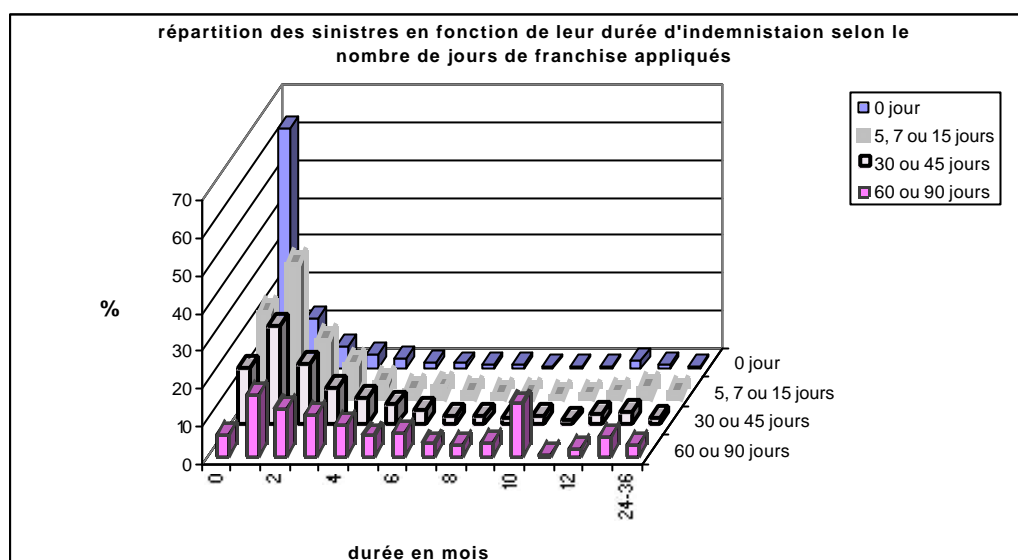
2.3.1 : Suivant le sexe



En fait, le sexe de l'assuré ne semble pas influencer particulièrement la durée des sinistres. Globalement les arrêts de travail sont courts* : 73% durent moins de 3 mois, et seulement 5% dépassent 1 an.

* le phénomène est accentué car on a vu que certaines hospitalisations sont incluses dans le périmètre étudié et comme généralement la période d'hospitalisation est assez courte on comprend pourquoi le pourcentage de sinistres indemnisés moins de 3 mois est aussi élevé.

2.3.2 : Suivant la franchise sur le sinistre



- 1) On constate que plus la franchise est longue plus la proportion de sinistre ayant une durée d'indemnisation importante augmente. Ceci est normal puisque plus la franchise est importante, plus les sinistres donnant lieu à indemnisation sont graves et donc longs.
- 2) Certains points sont à préciser à propos de la franchise 0 jour appliquée au sinistre. Pour cette franchise, la proportion de sinistres courts est vraiment très importante car les franchises IJH sont nulles et en cas d'hospitalisation certaines franchises IJ sont ramenées à 0. Par conséquent, les 60% des sinistres pour lesquels nous n'avons pas appliqué de franchises qui ont une durée moins d'un mois sont sans aucun doute des hospitalisations.
- 3) De plus parmi les sinistres sur lesquels aucune franchise n'a été appliquée, on constate une légère augmentation de la proportion pour les sinistres dépassant 10 mois d'indemnisation. Ceci s'explique par le fait que la franchise 0 jours sur une certaine garantie EXO est en fait une franchise relative de 90 jours. Par conséquent les sinistres disposant de cette fausse franchise 0 jour pour lesquels nous avons effectué une indemnisation sont relativement graves et longs puisqu'ils ont au moins dépassé 90 jours et que leur durée d'indemnisation comprend cette période de 90 jours contrairement aux autres cas où la franchise est absolue.

2.3.3 : Suivant la nature

Dans le tableau suivant figure la durée moyenne (en jours) des sinistres suivant leur nature :

	sans hospitalisation		avec hospitalisation	
	accident	maladie	accident	maladie
homme	60	158	122	123
femme	83	125	154	117

Que ce soit pour les hommes ou pour les femmes, la durée moyenne est fortement influencée par la nature du sinistre.

L'hospitalisation augmente le maintien en cas d'accident et le réduit en cas de maladie. En effet, lors d'un accident, si l'assuré est hospitalisé, c'est la preuve d'une certaine gravité, au contraire, en cas de maladie, l'hospitalisation doit correspondre à une opération, ce qui signifie plutôt une guérison.

Nous allons à présent étudier la saisonnalité des arrêts.

Section 2.4 : Etude de la saisonnalité

Jusqu'à présent, nous avons étudié uniquement les sinistres à travers leurs caractéristiques. Maintenant, au lieu de se limiter au périmètre des sinistres, nous allons les mettre en perspective avec le portefeuille correspondant. C'est à dire que l'on va s'intéresser à la sinistralité proprement dite de notre portefeuille.

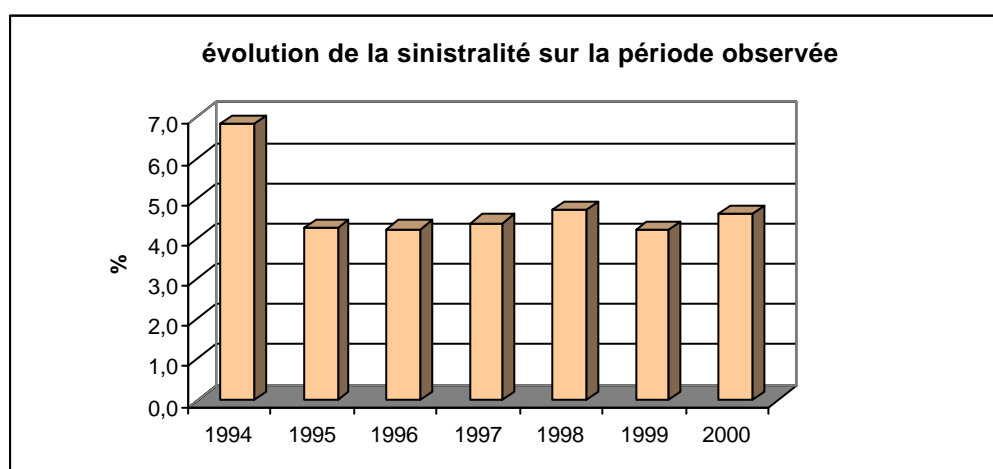
On entend ici le mot «sinistralité » au sens de risque d'entrer en incapacité représenté par le rapport (nb de sinistres /nb de contrats).

Les statistiques des paragraphes suivants sur les sinistralités sont basées sur des résultats globaux sur toute la période d'observation du portefeuille. Nous allons donc vérifier préalablement que le risque d'entrée en incapacité est stable dans le temps sur notre fenêtre d'observation.

A noter que les taux obtenus ne distinguent pas le type de garantie concernée (IJ, IJH ou EXO), la nature du sinistre (maladie / accident) et l'âge de l'assuré.

2.4.1 : La stabilité de la sinistralité dans le temps

Les graphes suivants présentent pour chaque année observée le rapport (nb de sinistres survenus / nb de contrats observés) :



La période d'observation débute en septembre 1994 par conséquent, la sinistralité est faussée car elle est calculée uniquement sur les mois d'automne / hiver où les arrêts sont plus fréquents (*cf. graphiques et explications suivantes*).

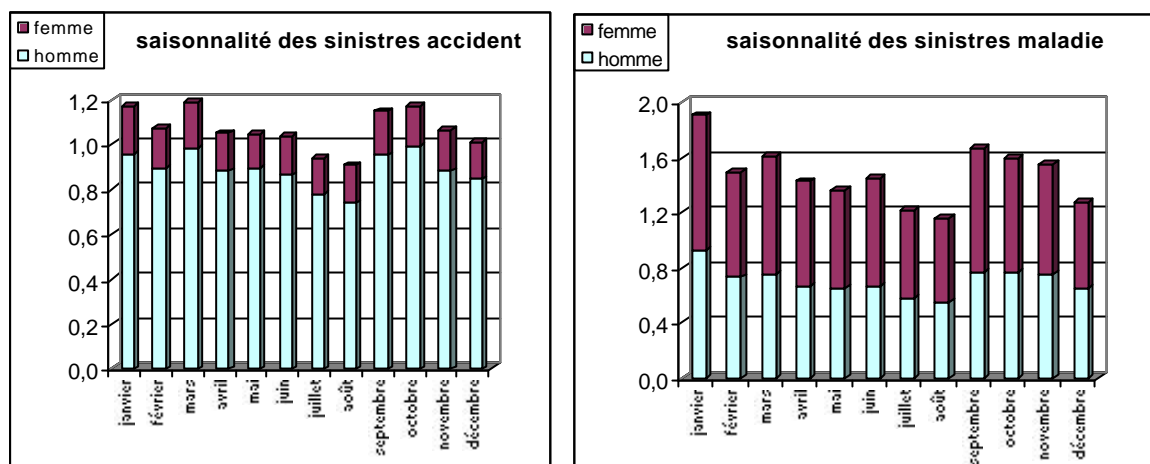
De même, en 2000, seuls les mois d'hiver (janvier et février) appartiennent à la fenêtre d'observation d'où encore le redressement de la sinistralité cette année là.

Finalement, si l'on considère uniquement les années entièrement observées soit 1995, 1996, 1997, 1998 et 1999, on constate que la sinistralité est bien constante dans le temps (environ 4,5%).

Nous pouvons maintenant regarder la saisonnalité de l'entrée en incapacité.

2.4.2 : La saisonnalité

Les graphes suivants donne l'évolution du risque d'entrer en incapacité pour accident ou pour maladie selon les mois de l'année :



Globalement, on note une plus forte sinistralité (accident et maladie) en automne / hiver.

Comme on pouvait l'imaginer a priori, les maladies surviennent plutôt pendant les saisons froides (excepté pour le mois de décembre). Un autre comportement amplifie le phénomène, en effet, les assurés ne doivent pas déclarer leur incapacité lorsqu'ils sont déjà en congé (d'été). Le pic de septembre représenterait alors la morosité du retour de vacances.

Pour les sinistres accidentels, on aperçoit une légère augmentation des fréquences en automne / hiver mais les différences mensuelles sont moins significatives.

Remarque :

On retrouve ici le faible risque d'accident chez les femmes.

Dans la prochaine section nous nous intéressons à l'influence des caractéristiques client sur la sinistralité.

Section 2.5 : Influence des caractéristiques clients sur la sinistralité

Les taux ci-après représentent les sinistralités globales pour chaque modalité de la variable descriptive analysée. C'est à dire que la sinistralité étudiée est traduite par le rapport :

$$\frac{\text{nombre de sinistres total sur la période d'observation pour la modalité } m}{\text{nombre d'assurés total sur la période d'observation pour la modalité } m}$$

cependant, les taux obtenus ne distinguent pas le type de garantie concernée (IJ, IJH ou EXO), la nature du sinistre (maladie / accident), l'âge de l'assuré, la franchise appliquée...

Ils sont par conséquent beaucoup plus grossiers que les tables car ils représentent une moyenne globale sur des risques hétérogènes (franchise...). Le manque de précision de ces taux ne permet pas de quantifier fiablement le risque, l'ordre de grandeur de ces taux n'est pas comparable aux tables d'entrée en incapacité par âge. Nous ne pouvons donc pas utiliser ces taux tels que par contre, **la comparaison de ces taux entre eux permet à défaut de la quantifier, de visualiser l'existence d'une sur- ou sous- sinistralité entre différentes populations.**

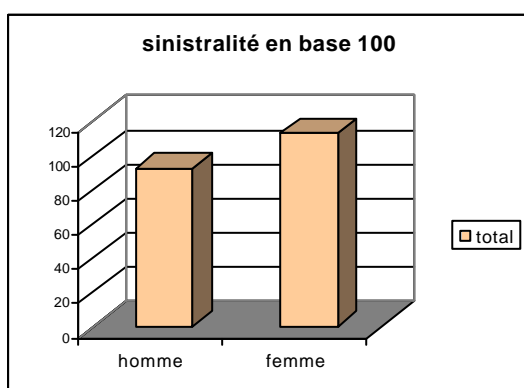
Pour faciliter la comparaison, nous prendrons comme base 100 le taux :

$$\frac{\text{nombre global de sinistres sur la période d'observation}}{\text{nombre global d'assurés sur la période d'observation}}$$

Ces premières études ont pour but d'aider à la segmentation du portefeuille pour construire les tables sur des populations homogènes.

Les variables client retenues sont le sexe, la CSP, la région et le montant souscrit.

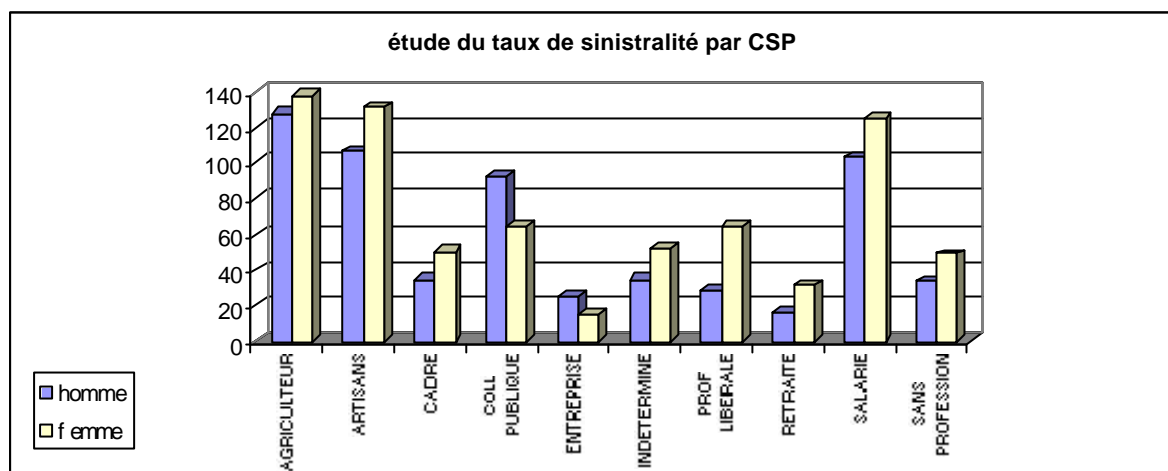
2.5.1 : Le sexe de l'assuré en arrêt



Sur l'ensemble du portefeuille, le risque d'entrer en incapacité est plus important chez les femmes que chez les hommes.

2.5.2 : la CSP du sinistré

Les résultats suivants sont toujours exprimés en base 100 et ils distinguent le sexe de l'assuré.



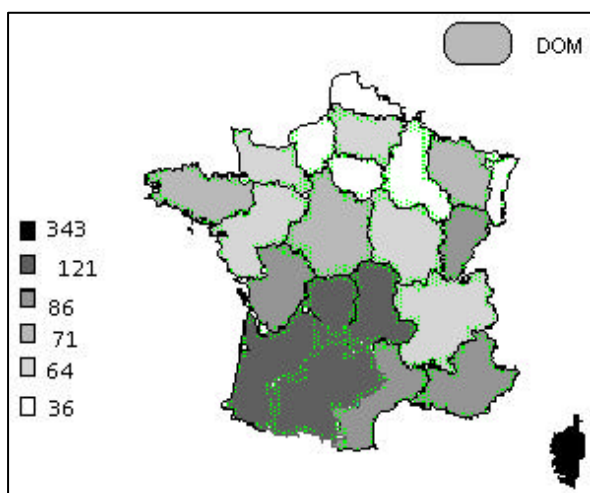
Les agriculteurs, les artisans commerçants et les salariés ont une sinistralité légèrement supérieure à la sinistralité globale du portefeuille et parmi ces trois CSP, la première présente la sinistralité la plus forte.

Par contre, les autres CSP (cadres, profession libérale, sans profession ou autres) affichent une sinistralité homogène qui est globalement bien inférieure à la sinistralité générale du portefeuille

Attention, les CSP « collectivité publique » et « entreprise » ne comportent pas assez d'assurés pour que la sinistralité soit fiable.

2.5.3 : la région du sinistré

La carte suivante représente la sinistralité par région d'origine pour la partie du portefeuille où la variable région été complètement renseignée soit 40% du portefeuille.



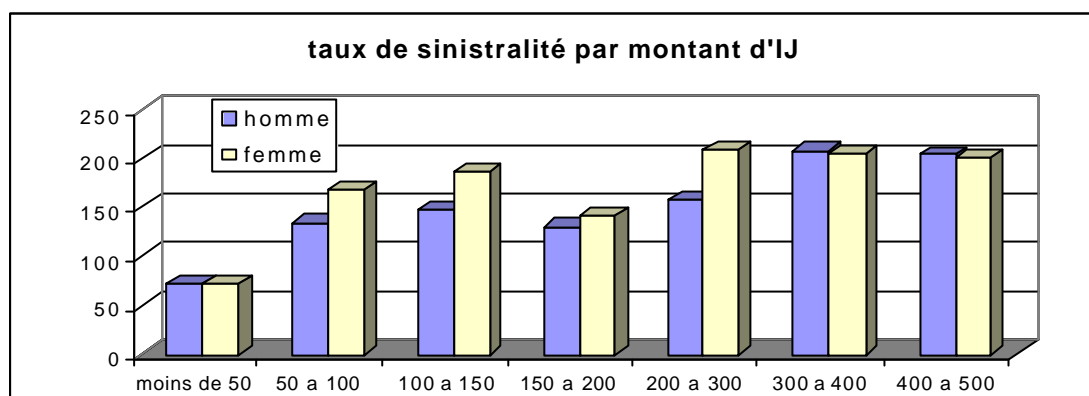
On constate **une sur-sinistralité très forte en Corse** (3.5 fois la sinistralité globale) et une sinistralité plus importante que la moyenne dans le Sud Ouest. A l'opposé, le nord, l'est et la région parisienne présentent une sous-sinistralité par rapport à l'ensemble du périmètre étudié.

2.5.4 : le montant d'IJ souscrit

Un montant moyen d'IJ est calculé pour chaque assuré et le graphe ci-après rapporte par tranches d'indemnité journalière le rapport suivant :

$$\frac{\text{le nombre total de sinistrés dans la tranche d'IJ}}{\text{le nombre total d'assurés dans la même tranche d'IJ}}$$

La base 100 reste identique à celle utilisée dans les cas précédents.



Attention, il n'y a pas assez de volume d'IJ dépassant 500 F pour pouvoir interpréter la sinistralité au-delà de 500 F elle n'a donc pas été représentée.

Globalement, la sinistralité croît avec le montant d'IJ garanti mais l'évolution n'est pas régulière et l'on constate un creux de sinistralité sur la tranche 150 à 200 F.

PARTIE IV :

**ELABORATION DES TABLES
D'ENTREE EN INCAPACITE**

Partie IV : Elaboration des tables d'entrée en incapacité

Cette partie, consacrée à la construction de la table d'expérience de la loi d'entrée en incapacité, s'articule en 7 chapitres.

- Tout d'abord seront précisées les particularités de l'estimation d'une loi à partir de données censurées et individuelles.
- Ensuite, nous souhaitons déterminer exactement l'élément à utiliser dans la formule de tarification pour traduire l'entrée en incapacité, afin d'en construire les tables représentatives. Pour ce faire, nous exposerons les différentes méthodes de tarification possibles (impliquant des approches d'estimation distinctes).
- Après quoi, nous étudierons les différents estimateurs à notre disposition, leurs propriétés et nous choisirons l'estimateur à utiliser pour notre application pratique.
- Il faudra ensuite déterminer la segmentation des tables de manière à avoir un sinistralité homogène dans chaque segment.
- Pour fiabiliser nos résultats, nous devons aussi satisfaire aux contraintes de volume avant de faire les estimations.
- Les résultats obtenus par les estimations brutes ainsi que les intervalles de confiance feront l'objet du 6^{ème} chapitre.
- Enfin, pour avoir des résultats satisfaisants, nous procéderons au lissage des tables brutes.

Chapitre 1 : Particularité de la table d'expérience

Ce chapitre décrit les caractéristiques des observations des assurés auxquels doivent s'adapter les estimateurs de la loi d'entrée.

Section 1.1 : L'utilisation de données censurées.

Nous utiliserons des estimateurs fonction des observations de la sinistralité des assurés mais une partie de ces observations est censurée. En effet, pour certains individus, toute l'information n'est pas observable on parle alors de censures. Estimer une loi d'entrée en incapacité par âge nécessite d'observer les individus d'un échantillon pendant l'année entière où ils ont l'âge x . Cependant, nous n'observons que la durée de présence dans le portefeuille ce qui n'inclut pas toujours l'année entière où l'assuré a l'âge x .

Lorsqu'il manque l'observation en début d'âge, on parle de **troncature** (censure gauche) et en fin d'âge, on parle de **censure droite**. Par abus de langage, il n'est pas toujours précisé que ces censures sont à « droite ».

Les cas de données censurées à droite sont illustrés dans les exemples suivants. :

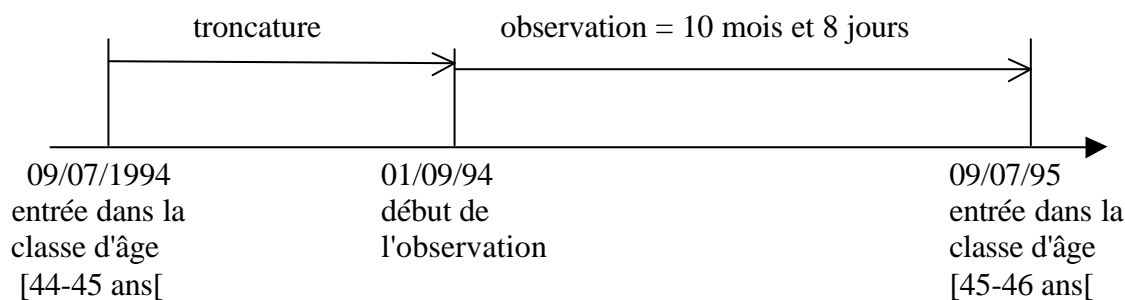
- ① Le contrat ou la garantie incapacité prennent fin avant que l'assuré ait l'âge $x+1$
- ② L'assuré rachète ou résilie son contrat ou sa garantie incapacité avant d'atteindre l'âge $x+1$
- ③ L'assuré décède avant d'atteindre l'âge $x+1$
- ④ L'assuré n'a pas atteint l'âge $x+1$ au 29/02/2000 (fin de l'observation)

NB : le cas de censure n°1 ne se produira pas ici car la classe d'âge est délimitée par 2 dates anniversaires consécutives du contrat.

Les cas d'observations tronquées se produisent quand des assurés ont souscrit leur contrat avant septembre 1994, ils appartiennent, au 01/09/1994, à une classe d'âge dans laquelle ils sont entrés avant le début de la période d'observation. L'observation sur cette classe d'âge sera par conséquent censurée à gauche.

L'exemple suivant illustre un cas de troncature :

Soit un assuré, né le 14/05/1950 qui souscrit un contrat le 09/07/1994. D'après la règle adoptée, il a 44 ans à la souscription mais nous ne l'avons pas observé l'année entière de ses 44 ans.



conclusion :

Les estimateurs de la loi d'entrée en incapacité des assurés retenus doivent pouvoir s'adapter aux censures et aux troncatures présentes dans nos observations.

Remarque :

Les censures et troncatures propres à l'étude du maintien en incapacité seront détaillées ultérieurement.

Section 1.2 : L'utilisation de données individuelles et journalières

Pour des raisons de simplification des calculs, les estimateurs sont généralement appliqués à des données groupées. Puisque ces estimateurs sont fonction du temps, ils font intervenir les dates d'entrée, de sortie et de censures au cours d'une année dans le portefeuille. Il est souvent supposé que ces mouvements se font de manière groupée au milieu de l'année. Ces hypothèses sont notamment utilisées dans les logiciels statistiques (comme SAS) estimant les lois de survie. Puisque nous disposons des dates exactes des mouvements sur le portefeuille et des moyens informatiques pour construire des estimateurs à partir de données individuelles et journalières, nous allons éviter ces approximations. L'utilisation des données individuelles sur les dates de mouvements, au lieu de grouper les nouveaux entrants, les sorties et les décès au milieu de l'année, nous donnera des estimateurs plus proches de la réalité surtout aux âges où les effectifs sont faibles et les censures fréquentes.

Chapitre 2 : L'utilisation des tables : la tarification

La finalité des tables d'entrée en incapacité étant la tarification des produits comportant des garanties incapacité, il est nécessaire de connaître le modèle de tarif envisagé pour décider de ce qu'il faut réellement estimer.

Nous allons donc présenter dans ce chapitre les différentes formules de tarification envisageables.

Section 2.1 : 1^{er} modèle

$$P_x = p_1 * C_1 + p_2 * C_2 + p_3 * C_3 + \dots$$

Avec :

P_x : la prime pure associée au risque incapacité pour un assuré appartenant à la classe d'âge $[x, x+1[$

p_1 : la probabilité d'entrée au moins une fois en incapacité durant la classe d'âge $[x, x+1[$

p_2 : la probabilité d'entrée au moins deux fois en incapacité durant la classe d'âge $[x, x+1[$

p_3 : la probabilité d'entrée au moins trois fois en incapacité durant la classe d'âge $[x, x+1[$

...etc

C_1 : le capital constitutif lié à l'indemnisation d'un 1^{er} sinistre survenu sur la classe d'âge $[x, x+1[$

C_2 : le capital constitutif lié à l'indemnisation d'un 2^{ème} sinistre survenu sur la classe d'âge $[x, x+1[$

C_3 : le capital constitutif lié à l'indemnisation d'un 3^{ème} sinistre survenu sur la classe d'âge $[x, x+1[$

... etc

Ce modèle permet de tenir compte des sinistrés qui reprennent leur activité avant la fin de la classe d'âge considérée et qui sont ainsi à nouveau exposés au risque d'entrer en incapacité.

Par ailleurs, il semble légitime de penser que le capital constitutif pour un sinistre est corrélé au nombre de sinistres de l'assuré dans la classe d'âge. C'est pourquoi cette tarification dissocie le capital constitutif d'un 1^{er} sinistre de celui d'un 2^{ème} sinistre etc...

La précision de cette formule permet donc de traduire au mieux le risque supporté par l'assureur et de tarifer le cas échéant des garanties où le nombre de sinistres indemnifiables serait limité (à un ou deux ...). Mais ce type de garantie n'est pas commun sur le marché et n'existe pas dans le groupe AXA.

Il faut cependant noter que le calcul de tous ces capitaux constitutifs implique de connaître le maintien en incapacité pour un 1^{er} sinistre ainsi que celui correspondant à un 2^{ème} sinistre etc...

Ce qui oblige à construire plusieurs tables de maintien en incapacité et notamment une table associée aux doubles voir triples sinistrés (d'une classe d'âge). Cette segmentation supplémentaire réduirait considérablement le volume des données et donc la fiabilité des tables, or ces tables devront être certifiées pour être utilisées dans le provisionnement.

Section 2.2 : 2^{ème} modèle

$$P_x = p_1 * C'_1$$

Avec :

p_1 : la probabilité d'entrée au moins une fois en incapacité durant la classe d'âge $[x, x+1[$

C'_1 : le capital constitutif lié à l'indemnisation complète de tous les sinistres de l'assuré survenus dans sa classe d'âge $[x, x+1[$

Ce modèle est issu du constat suivant :

nous savons que certaines rechutes après plus de 2 mois ne se sont pas traduit informatiquement par la création d'un nouveau numéro de sinistre mais par la présence de ce que l'on a appelé un creux (reprise de travail entre 2 arrêts de travail) alors que les conditions générales des contrats du portefeuille précisent que lorsqu'il s'écoule plus de 2 mois entre 2 arrêts de travaux, il s'agit d'un nouveau sinistre. De ce fait, il est anormal par exemple de traiter différemment 2 assurés dont l'un a eu 2 sinistres consécutifs espacés de 6 mois et l'autre n'a eu qu'un seul sinistre mais avec un creux de 6 mois au milieu. Pour éviter d'avoir ce cas de figure, et pour solutionner le problème du traitement des creux, il semble donc nécessaire de regrouper en un seul sinistre tous les sinistres d'un assuré survenus dans la même classe d'âge, en prenant comme date de survenance celle du premier sinistre et comme durée le cumul des jours d'arrêt des sinistres regroupés.

Ainsi, par ce traitement préalable des données, il est possible de traiter l'entrée en incapacité sur une classe d'âge comme un événement unique et définitif tel que le décès, en calculant un unique taux d'entrée, noté $q_x = p_1$, pour la classe d'âge $[x, x+1[$.

De plus, ce traitement des sinistres permet d'aborder de la même manière les nouveaux sinistres et les rechutes de plus de 2 mois qui auraient dû entraîner l'ouverture d'un autre sinistre.

Le capital constitutif C'_1 sera en fait propre aux sinistrés de la classe d'âge et non propre aux sinistres. Son calcul passe par l'élaboration d'une table de maintien en incapacité cohérente avec la convention précédente. De ce fait, on surestime le maintien en incapacité puisqu'on remplace les petits sinistres consécutifs survenus à un même âge par un seul sinistre de durée cumulée.

Ainsi, l'utilisation de cette table de maintien pour le provisionnement conduirait à une majoration du risque incapacité. Il serait donc nécessaire de construire une table pour le tarif et une table différente pour le provisionnement, sachant qu'il est fort probable que la réglementation future impose une table unique dans les deux cas.

De plus, ce tarif mélange des sinistres de causes différentes d'où des franchises qui peuvent être différentes et donc finalement des risques différents.

Section 2.3 : 3^{ème} modèle

$$P_x = \text{Esp}(N_x) * C_x$$

N_x : la variable aléatoire nombre de sinistres incapacité survenus dans la classe d'âge $[x, x+1[$ pour un assuré.

C_x : le capital constitutif d'un sinistre survenu dans la classe d'âge $[x, x+1[$.

Ce modèle de tarification est en principe appliqué en assurance dommage où l'on raisonne en fréquence de sinistres.

Etant donnée que l'incapacité n'est pas définitive (comme l'est le décès par exemple), elle peut se reproduire plusieurs fois et notamment dans la même classe d'âge, de ce fait, il est possible de considérer l'incapacité comme un sinistre IARD.

Ainsi, le capital constitutif C_x se calcule à partir du maintien en incapacité associé à tous les sinistres survenus dans la classe d'âge $[x, x+1[$. Cette table de maintien correspondrait bien à celle envisagée pour le provisionnement d'un sinistre de la classe d'âge. Il serait ainsi possible d'appliquer la même table pour la tarification et pour le provisionnement.

Cependant, ce modèle ne distingue pas les premiers sinistres des deuxièmes ou troisièmes ... etc. En fait, les antécédents de l'assuré dans la classe d'âge ne sont pas pris en compte, et si l'on peut légitimement supposer l'indépendance des sinistres touchant des assurés différents, il n'en est pas de même pour ceux touchant un même assuré.

En dehors des modèles standard présentés ci-dessus, il faut aussi que le tarif corresponde complètement aux garanties proposées et qu'il tienne compte notamment des clauses mentionnées. Nous étudierons donc dans le paragraphe suivant l'impact sur le tarif d'une clause particulière commune à nos contrats et dans la section d'après, nous définirons les types de tables adéquates aux garanties proposées.

Section 2.4 : Réduction de prime

Dans les conditions générales de la majorité des contrats au portefeuille, figure la précision suivante :

« L'arrêt de paiement des primes entraîne la résiliation de la garantie et met fin aux prestations en cours ».

Comme cette phrase réduit l'engagement de l'assureur, il faut opérer une diminution de la prime correspondante de la manière suivante :

$$P_x - t_x * D_x$$

Avec :

P_x : prime calculée suivant le modèle de tarif choisi

t_x : taux de chute correspondant aux arrêts de paiement des primes pour la classe d'âge suivante (après $x+1$)

D_x : capital constitutif de la prestation restant à payer au-delà de $x+1$ pour les sinistres survenus sur $[x, x+1[$ et qui se prolongent au-delà de $x+1$

Il serait donc nécessaire de construire une table de maintien spécifique pour les sinistres chevauchant au moins 2 classes d'âge, mais ce ne sera pas fait dans le cadre de ce mémoire.

De toute façon, il est peu probable qu'un assuré cesse de régler ses primes alors qu'il touche encore des indemnités pour arrêt de travail et quand bien même, s'il possède la garantie exonération de prime, la mention des conditions générales n'a aucun effet.

Section 2.5 : Adaptation des tables avec les garanties souscrites

L'objectif de nos tables d'entrée est de permettre de valider ou de recalculer les tarifs de toutes les garanties complémentaires liées à l'incapacité. Dans cette optique, il faut savoir clairement comment chaque type de garantie impacte la définition même du risque d'incapacité.

Comme expliqué précédemment dans la présentation des données, il existe dans notre portefeuille trois types de garanties incapacité :

- les indemnités journalières (IJ)
- les indemnités journalières en cas d'hospitalisation (IJH)
- les exonérations de prime (EXO)

De plus, il existe des contrats sur lesquels plusieurs personnes sont assurées au niveau de la garantie principale (décès) mais parfois aussi pour l'exonération de prime. Dans ce cas, la prime est exonérée quand l'un ou l'autre des assurés est en incapacité.

De ce fait, il est impossible de traiter séparément les 2 assurés sur ces garanties, puisqu'il n'y aurait pas indépendance de l'échantillon. En effet, l'exonération ne se produit qu'une seule fois même si les 2 têtes sont sinistrées en même temps, on ne peut donc pas considérer la garantie exonération sur 2 têtes comme la somme de 2 garanties exonération chacune sur une tête.

Il est donc nécessaire d'exclure ces garanties pour construire la table principale par tête. Il est envisageable d'élaborer dans un second temps une table d'exonération sur 2 têtes.

D'autre part, il se peut qu'un assuré souscrive uniquement une garantie IJH, ainsi, il ne sera exposé qu'au risque d'entrée en incapacité avec hospitalisation. Si l'on mélange ces assurés avec le reste de la population, il est clair que l'échantillon ne sera pas complet ni homogène en terme d'exposition mais aussi en terme de sinistres puisqu'il manquera les sinistres sans hospitalisation des assurés qui n'ont qu'une garantie IJH.

Par conséquent, il faudrait construire 3 types de tables d'entrée en incapacité :

- une table principale (**IJ/EXO**) caractérisant l'entrée en incapacité toute cause sur une seule tête (avec ou sans hospitalisation) pour pouvoir tarifier les garanties IJ et EXO.
- Une table spéciale hospitalisation (**IJH**) caractérisant l'entrée en hospitalisation pour pouvoir tarifier les garanties IJH.
- Une table spéciale 2 têtes (**EXO2T**) caractérisant l'entrée en incapacité sur 2 têtes pour pouvoir tarifier les garanties EXO sur 2 têtes.

Des fichiers de données correspondants à ces 3 échantillons ont été extraits.

Remarque :

Pour tarifier les garanties complémentaires à l'âge x , il faut aussi calculer le capital constitutif d'un sinistre survenu à cet âge et donc utiliser des tables de maintien appropriées.

Par conséquent, la structure des tables d'entrée influence celle des tables de maintien. C'est à dire qu'il est nécessaire de construire au moins les 3 types de tables de maintien correspondant aux 3 types de tables d'entrée et basées sur les 3 mêmes échantillons.

A titre d'exemple voici la prime calculée à partir du 3^{ème} modèle de tarif, pour un assuré d'âge x qui aurait souscrit :

- une indemnité journalière de montant I (table associée IJ/EXO)
- une exonération de prime sur une tête de montant E (table associée IJ/EXO)
- une indemnité journalière hospitalisation de montant H (table associée IJH)

$$P_x = \text{Esp}(N_x)_{IJ/EXO} * C_x(I+E)_{IJ/EXO} + \text{Esp}(N_x)_{IJH} * C_x(H)_{IJH}$$

Où :

- $\text{Esp}(N_x)_{IJ/EXO}$ (resp $\text{Esp}(N_x)_{IJH}$) est issu de la table d'entrée IJ/EXO (resp IJH)
- $C_x(I+E)_{IJ/EXO}$ calculé à partir de la table de maintien IJ/EXO et du montant garanti $(I+E)$
- $C_x(H)_{IJH}$ calculé à partir de la table de maintien IJH et du montant garanti H

Chapitre 3 : Les estimateurs d'entrée en incapacité

Après avoir explicité les modèles de tarif envisageables, nous allons maintenant exposer les estimateurs du risque d'entrée en incapacité qui se rapportent à chaque modèle.

La 1^{ère} section traitera des estimateurs de taux d'entrée qui sont adaptés pour tarifier selon les deux premiers modèles, alors que la section suivante présentera l'estimateur IARD adapté au 3^{ème} modèle. Puis nous expliquerons pourquoi nous avons retenu ce dernier estimateur pour construire nos tables d'entrée.

Section 3.1 : 1^{ère} approche : le concept de taux d'entrée en incapacité

Il est difficile de donner un sens précis à l'appellation « taux d'entrée en incapacité » car l'incapacité est un état passager qui peut se reproduire plusieurs fois chez un même assuré à des fréquences plus ou moins importantes.

Comme cette étude doit permettre d'élaborer des tarifs qui évolueront en fonction de l'âge de l'assuré, on s'intéresse à l'entrée en incapacité par classe d'âge. Se pose alors le problème de savoir comment traiter les assurés qui ont plusieurs sinistres la même année. En effet, est ce que le taux d'entrée en incapacité dans la classe $[x, x+1[$ correspond à la probabilité d'avoir au moins un sinistre dans l'année, ou alors à la probabilité d'avoir un sinistre et un seul ?

Pour établir son tarif, l'assureur a besoin de connaître le risque d'indemnisation de ses assurés, soit en fait le risque d'entrer au moins une fois en incapacité sur la classe d'âge considérée et d'aller au-delà de la franchise.

Ce concept correspond tout à fait au 2^{ème} modèle de tarification, mais si l'on envisage le 1^{er} modèle, alors il faut estimer plusieurs taux d'entrée basés chacun sur leur propre échantillon. En effet, la formule de calcul utilise les probabilités d'entrer au moins une fois en incapacité, au moins deux fois, au moins trois fois etc...

En fait, pour calculer ces probabilités, on peut utiliser un même estimateur de taux mais en modifiant à chaque fois l'échantillon de données. Pour calculer le 1^{er} taux (la probabilité d'entrer au moins une fois), on garde l'échantillon complet, pour calculer le 2^{ème} taux (la probabilité d'entrer au moins deux fois), on enlève sur chaque âge tous les premiers sinistres de chaque assuré, pour calculer le 3^{ème} taux (la probabilité d'entrer au moins trois fois), on enlève sur chaque âge tous les premiers et deuxièmes sinistres de chaque assuré et ainsi de suite. Ce modèle est limité par la taille des échantillons obtenus.

De toute façon, que ce soit pour le 1^{er} ou le 2^{ème} modèle de tarif, il faut en fait estimer une probabilité d'entrée au moins une fois en incapacité.

Ainsi, il est possible de traiter l'entrée en incapacité sur une classe d'âge comme un événement unique et définitif tel que le décès, en calculant un unique taux d'entrée, noté q_x , pour la classe d'âge $[x, x+1[$.

Le premier type d'estimateur de la loi d'entrée en incapacité pour des données censurées que nous allons construire est appelé dans la littérature estimateur non paramétrique de Kaplan-Meier ou estimateur product-limit.

3.1.1 : Les estimateurs non paramétriques

a) L'estimateur de Kaplan-Meier

En juin 1958, E. KAPLAN et P. MEIER publient dans le journal de l'American Statistical Association un article intitulé « Non Parametric estimation from incomplete observations » dans lequel ils définissent l'estimateur « PL » de la fonction de survie $S(y)=P(Y>y)$ (PL pour Produit-Limit ainsi nommé car il est construit par produits successifs).

Rappelons brièvement en quoi consiste cet estimateur.

• *Le principe*

Après avoir découpé la période de survie observée [date où l'on atteint l'âge x , date où l'on atteint l'âge $x+1$] en sous-périodes délimitées par les dates $\{a_j, j=1,2,\dots,m\}$, la méthode de Kaplan-Meier consiste à dire que la probabilité p_x de ne pas entrer en incapacité à l'âge x , c'est la probabilité de ne pas entrer à chaque période j (1, 2, 3 ..., m) (p_j) sachant que l'on n'était pas entré à la période précédente.

$$p_x = p_1 p_2 \dots p_m \quad \text{avec } p_j = 1 - q_j$$

• *La construction*

Quand il n'y a pas de censures :

Le but de la démarche, est d'estimer Q_x la probabilité d'entrée en incapacité dans la classe d'âge $[x, x+1[$. Il est équivalent d'estimer S_x la fonction de survie sur $[x, x+1[$ sans entrer en incapacité car $q_x = 1 - S_x(1)$.

Supposons que cette loi de survie est discrète et de la forme :

$$\{(a_i, f_i) ; i=1,2,\dots,m\}, 0 \leq a_1 < a_2 < \dots < a_m \leq 1$$

les a_i étant des dates connues entre x et $x+1$ et les f_i les valeurs correspondantes prises par la fonction.

En notant X la variable durée de survie au-delà de l'âge x sans être entré en incapacité, on définit q_i la probabilité d'entrer en incapacité à la date a_i par :

$$q_i = P(X = a_i / X \geq a_i).$$

$$\text{Donc pour } t \in]0, a_m], \quad S_x(t) = \prod_{r/a_r < t} (1 - q_r) \quad (1)$$

q_i est ensuite estimé par la méthode du maximum de vraisemblance à partir de l'observation d'un n -échantillon.

Soit d_i le nombre d'entrées en incapacité à la date a_i .

A priori, le nombre d'entrées en incapacité à chaque date a_i suit une loi binomiale $B(n_i, q_i)$ avec n_i le nombre d'individus restant soumis au risque d'incapacité à la date a_i ,

$$n_i = n_{i-1} - d_{i-1} = n - d_1 - d_2 - \dots - d_{i-1}$$

d'où la fonction de vraisemblance suivante:

$$L(d_1, \dots, d_m ; q_1, \dots, q_m) = \prod_{r=1}^m C_{n_r}^{d_r} q_r^{d_r} (1 - q_r)^{n_r - d_r} \quad (2)$$

La maximisation de L s'obtient par l'annulation des dérivés partielles suivantes :

$$\frac{\partial \ln L}{\partial q_r} = \frac{d_r}{q_r} - \frac{n_r - d_r}{1 - q_r} \quad r=1,2,\dots,m$$

Ce qui conduit aux estimateurs suivants :

$$\hat{q}_r = \frac{d_r}{n_r}, \quad r=1,2,\dots,m$$

En introduisant ces estimateurs dans (1) on obtient l'expression de $\hat{S}_x(t)$:

$$\hat{S}_x(t) = \prod_{r: a_r < t} \left(1 - \frac{d_r}{n_r}\right)$$

• **Résultat**

On détermine ainsi l'estimateur de Kaplan-Meier qui admet la forme suivante:

$$\hat{Q}_x = 1 - \hat{S}_x(1) = 1 - \prod_j \left(1 - \frac{d_j}{n_j}\right) \quad (3)$$

Remarque essentielle :

Si dans (3), $d_j=0$, alors $1 - \frac{d_j}{n_j} = 1$ et ce facteur disparaît du produit ; de ce fait, dans la construction de \hat{q}_x , seuls les a_j en lesquels il y a effectivement au moins une entrée en incapacité méritent d'être considérés. Il suffit donc de retenir les diverses dates de sinistre et d'utiliser les nombres d'entrée en incapacité et les effectifs faisant face au risque d'incapacité qui leur correspondent.

• **Introduction de censures et de troncatures :**

Supposons maintenant qu'entre les dates a_{j-1} et a_j , c_{j-1} têtes soient censurées et que t_{j-1} têtes soient tronquées.

convention utilisée :

En cas d'ex æquo, une entrée en portefeuille précède une censure qui précède elle même une entrée en incapacité.

Maintenant, le nombre de têtes restant présentes face au risque d'entrer en incapacité en a_j sera donné par :

$$n_j = n_{j-1} - d_{j-1} - c_{j-1} + t_{j-1}, \quad j=1,2,\dots,m \quad \text{avec } n_0 = n, d_0 = 0 \quad (4)$$

Sous la condition d'indépendance entre le processus de censure et celui régissant les entrées en incapacité, les taux d'entrée q_1, q_2, \dots, q_m n'ont pas à être modifiés, et la précédente fonction de vraisemblance (2) est toujours valable en utilisant les n_j définis ci dessus.

Ainsi l'estimateur de Kaplan Meier \hat{Q}_x est défini d'après (3) et (4).

- **Propriétés**

L'espérance :

$$E[\hat{Q}_x] = 1 - E\left[\prod_{j=1}^m \left(\frac{n_j - d_j}{n_j}\right)\right] = 1 - \prod_{j=1}^m \left(1 - E\left[\frac{d_j}{n_j}\right]\right) = 1 - \prod_{j=1}^m (1 - q_j) = 1 - \prod_{j=1}^m p_j = Q_x$$

On en conclut que cet estimateur est sans biais.

La variance :

$$\begin{aligned} V[\hat{Q}_x] &= V[\hat{P}_x] = E[\hat{P}_x^2] - E^2[\hat{P}_x] = E\left[\prod_{j=1}^m \hat{P}_j^2\right] - E^2\left[\prod_{j=1}^m \hat{P}_j\right] = \prod_{j=1}^m E[\hat{P}_j^2] - \left(\prod_{j=1}^m E[\hat{P}_j]\right)^2 \\ &= \prod_{j=1}^m \left(\frac{p_j q_j}{n_j} + p_j^2\right) - \left(\prod_{j=1}^m p_j\right)^2 = \prod_{j=1}^m p_j^2 \left(\frac{q_j}{p_j n_j} + 1\right) - \left(\prod_{j=1}^m p_j\right)^2 = \left(\prod_{j=1}^m p_j\right)^2 \left[\prod_{j=1}^m \left(\frac{q_j}{p_j n_j} + 1\right) - 1\right] \\ &\approx p_x^2 \left(\sum_{j=1}^m \frac{q_j}{p_j n_j}\right) \text{ approximation (*)} \end{aligned}$$

L'approximation (*):

$$\begin{aligned} \prod_{j=0}^m \left(\frac{q_j}{p_j n_j} + 1\right) &= \left(\frac{q_0}{p_0 n_0} + 1\right) \left(\frac{q_1}{p_1 n_1} + 1\right) \cdots \left(\frac{q_m}{p_m n_m} + 1\right) \\ &= 1 + \frac{q_0}{p_0 n_0} + \frac{q_1}{p_1 n_1} + \cdots + \frac{q_m}{p_m n_m} + (\text{termes d'ordres 2 et plus}) \end{aligned}$$

On considérera les termes du second degré et d'ordres supérieurs comme négligeables.

- **Les inconvénients**

Dans la pratique, cet estimateur n'est pas utilisé quand il y a beaucoup de données. En effet, il demande de connaître la chronologie des entrées en incapacité sur le portefeuille ainsi que le nombre de censures entre deux mouvements. C'est un traitement lourd même avec des outils de calculs modernes.

b) *L'estimateur actuariel*

• *principe*

Comme pour l'estimateur de Kaplan Meier, q_x , le taux d'entrée en incapacité à l'âge x , est estimé par la méthode du maximum de vraisemblance à partir de l'observation d'un n -échantillon en supposant que le nombre d'entrées en incapacité à l'âge x suit une loi binomiale $B(n_x, q_x)$ avec n_x le nombre d'individus restant soumis au risque d'incapacité au début de l'âge x .

Soit d_x le nombre d'entrées en incapacité à l'âge x .

Ce qui conduit à l'estimateur suivant :

$$\hat{q}_x = \frac{d_x}{n_x}$$

De plus, on considère que toute personne censurée ou tronquée, pendant une période, est exposée au risque en moyenne sur une demi période, ce qui implique en fait l'hypothèse suivante : les censures et les troncatures sont uniformes et indépendantes sur une période.

De ce fait, l'exposition au risque d'incapacité à l'âge x devient :

$$\text{Exp}_x = n_x - \frac{c_x}{2} + \frac{t_x}{2}$$

Avec :

- c_x : le nombre d'assurés censurés à l'âge x
- t_x : le nombre d'assurés tronqués à l'âge x

• *résultat :*

$$\hat{q}_x = \frac{d_x}{n_x - \frac{c_x}{2} + \frac{t_x}{2}}$$

remarque :

cet estimateur très simplificateur n'utilise pas les données individuelles. On peut donc douter de son adéquation aux données et de sa précision.

3.1.2 : Les estimateurs paramétriques

Puisque l'estimateur de Kaplan-Meier semble assez complexe à mettre en œuvre, et puisque l'estimateur actuariel semble un peu trop simplificateur, nous avons étudié les différents estimateurs nécessitant des hypothèses sur la distribution des entrées en incapacité à un âge fixé.

Description générale des autres estimateurs.

Nous ne disposons pas d'étude précédente sur nos propres assurés nous donnant la forme paramétrique de leur loi d'entrée en incapacité mais nous connaissons des formes paramétriques qui s'appliquent à l'incapacité en générale et à d'autres tables d'expérience. Nous allons donc prendre l'hypothèse naturelle que la loi d'entrée en incapacité de nos assurés à un âge donné a une des formes paramétriques usuelles. Les estimateurs à chaque âge ont donc été construits à partir de suppositions sur l'expression de la loi en fonction du temps. Si les hypothèses paramétriques adoptées sont proches de la vraie loi d'entrée, les estimations paramétriques sont les mieux adaptées.

a) L'estimateur des moments de Hoem

L'estimateur dit de Hoem donne un résultat simple dont nous allons donner les justifications.

- **Le principe**

Il s'agit de considérer que l'assuré i n'est exposé au risque d'incapacité dans la classe d'âge considérée $[x, x+1[$ qu'entre les dates r_i (date du début d'observation de l'assuré d'âge x) et s_i (date de fin d'observation de l'assuré d'âge x par censure ou à défaut de censure, date de changement d'âge). s_i est connue (fin du contrat ou fin de la période d'observation ou résiliation). C'est à dire que l'on ne prendra en compte, pendant l'observation des assurés, que les périodes où il y avait effectivement un risque supporté par l'assureur.

- **La construction.**

① D'après le principe énoncé ci-dessus, la variable aléatoire X_i (incapable ou non) suit une Loi de Bernoulli $B_{(s_i-r_i)q_{x+r_i}}$ où q_{x+r_i} = probabilité d'entrer en incapacité sur la période qui débute en $(x+r_i)$ et qui dure (s_i-r_i) , sachant qu'au début de l'observation (à l'âge $x+r_i$) l'assuré i est en état de capacité.

Exemple:

Soit un assuré né le 06/03/1949, qui souscrit un contrat le 12/03/1980 et le rachète le 11/01/1995.

A partir du 12/03/1994, il entre donc dans la **classe d'âge [45-46[**.

Le 01/09/1994 commence l'observation de cet assuré dans la classe d'âge [45-46[. Donc **$r_i=01/09/1994-12/03/1994$ et $x+r_i=45$ ans, 5 mois et 20 jours.**

Le 11/01/1995 l'observation de cet assuré dans cette classe d'âge cesse par censure. Donc, **$s_i-r_i=11/01/1995-01/09/1994=4$ mois et 10 jours.**

- ② Les variables X_i ont une espérance $E[X_i] = {}_{s_i-r_i}q_{x+r_i}$ (1).
 Pour simplifier (1), on suppose que la probabilité pour un individu d'âge x d'entrer en incapacité entre l'âge x et l'âge $x+t$ est une fonction linéaire du temps ${}_tq_x = a + b \cdot t$ pour $t \in [0,1]$.

Or : si $t=1$, ${}_1q_x = q_x$ et si $t=0$, ${}_0q_x = 0$ donc $\boxed{{}_tq_x = t \cdot q_x}$

◆ On en déduit la propriété:

$${}_tq_x = t \cdot q_x \Rightarrow {}_t p_x = 1 - t \cdot q_x$$

par approximation, ${}_{s-t}q_{x+t} \approx {}_t p_x - {}_s p_x$ \Rightarrow $\boxed{{}_{s-t}q_{x+t} \approx (s-t)q_x}$ (2)

Par l'équation (2), l'équation (1) devient: $E[X_i] = (s_i - r_i) q_x$

- ③ Soient :

X_1, \dots, X_{n_x} , n_x variables de Bernoulli $B({}_{s_i-r_i}q_{x+r_i})$ indépendantes,

$D_x = \sum_{i=1}^{n_x} X_i$ le nombre de sinistres incapacité survenus pendant la classe d'âge $[x, x+1[$

D'après les résultats précédents, en notant Z_i , les n_x variables indépendantes telles que

$$Z_i = \frac{X_i}{s_i - r_i} \quad \text{on a : } E[Z_i] = E\left[\frac{X_i}{s_i - r_i}\right] = q_x$$

En appliquant la Loi des Grands Nombres aux variables Z_i indépendantes et uniformément distribuées, on en déduit un estimateur de q_x .

L'estimateur de Hoem est : $\hat{Q}_x = \frac{D_x}{\sum (s_i - r_i)}$

• **Propriétés de l'estimateur de Hoem**

L'espérance mathématique est:

$$E[\hat{Q}_x] = \frac{E[D_x]}{\sum_{i=1}^n (s_i - r_i)} = \frac{\sum_{i=1}^n q_x (s_i - r_i)}{\sum_{i=1}^n (s_i - r_i)} = q_x$$

Il s'agit donc d'un estimateur sans biais.

La variance est:

première méthode:

$$\begin{aligned}
 V[\hat{Q}_x] &= V\left[\frac{D_x}{\sum_{i=1}^{n_x} s_i - r_i} \right] = V\left[\frac{\sum_{i=1}^{n_x} X_i}{\sum_{i=1}^{n_x} s_i - r_i} \right] \\
 &= \frac{\sum_{i=1}^{n_x} (s_i - r_i) * q_x * (1 - (s_i - r_i) * q_x)}{\left[\sum_{i=1}^{n_x} s_i - r_i \right]^2}
 \end{aligned}$$

Seconde méthode:

D_x suit une binomiale $B\left(\sum_{i=1}^{n_x} (s_i - r_i), q_x\right)$, $\sum_{i=1}^{n_x} (s_i - r_i)$ approximée par sa valeur entière.

On en déduit la variance suivante:

$$\begin{aligned}
 V[\hat{Q}_x] &= V\left[\frac{D_x}{\sum_{i=1}^{n_x} (s_i - r_i)} \right] \\
 &= \frac{\left(\sum_{i=1}^{n_x} (s_i - r_i)\right) * q_x * p_x}{\left[\sum_{i=1}^{n_x} (s_i - r_i) \right]^2} = \frac{p_x * q_x}{\sum_{i=1}^{n_x} s_i - r_i}
 \end{aligned}$$

Nous allons maintenant exposer l'estimateur du maximum de vraisemblance obtenu avec l'hypothèse que la loi de survie suit un modèle exponentiel.

b) *L'estimateur du maximum de vraisemblance avec hypothèse de loi d'entrée.*

Nous utiliserons les mêmes notations que pour le précédent estimateur.

• *Le principe*

Cette méthode pour estimer la loi d'entrée en incapacité utilise un estimateur construit pour maximiser la vraisemblance de l'échantillon en supposant que la variable X_i suit une Bernoulli $B_{(t_i-r_i)q_{x+r_i}}$. t_i est la date de fin d'observation de l'assuré d'âge x que ce soit par censure, incapacité ou changement d'âge. La logvraisemblance fait intervenir la densité de X_i qui est la composée d'une masse de Dirac et d'une fonction de distribution entre la date observée d'entrée dans la classe d'âge et la date observée de sortie réelle de cette classe. On considère qu'un assuré sort de l'effectif dès la date de son entrée en incapacité. La date de sortie observée réelle, quand l'assuré est sinistré, est la date du sinistre.

• *Etapas de la construction*

① La densité conditionnelle d'incapacité de X_i à l'âge $x+\delta_i$ sachant que l'assuré a atteint l'âge x en état de capacité = ${}_d p_x * m_{x+\delta_i}$ où $m_{x+\delta_i}$ est le taux instantané d'entrée en incapacité au point $x+\delta_i$.

② La vraisemblance des n_x observations indépendantes et identiquement distribuées X_i

$$\text{est: } V_n = \prod_{i=1}^{n_x} ({}_t p_{x+r_i} * (m_{x+t_i})^{X_i}) \quad (3)$$

Hypothèse fondamentale:

Pour simplifier l'équation (3), on prendra comme hypothèse que la probabilité de maintien suit un modèle exponentiel.

C'est à dire que l'on fait l'hypothèse que la distribution de la probabilité (q_x) d'un individu d'âge x de tomber en incapacité entre l'âge x et l'âge $x+t$ est la suivante:

$$\boxed{{}_t q_x = 1 - {}_t p_x = 1 - e^{-m t}}$$

On trouve l'expression du taux instantané d'incapacité:

$$\begin{aligned} \text{Par définition, } m_{x+t} &= \frac{-\frac{d}{dt} l_{x+t}}{l_{x+t}} \\ \text{et } \frac{-\frac{d}{dt} l_{x+t}}{l_{x+t}} &= -\frac{d}{dt} \ln {}_t p_x \end{aligned} \quad \left. \vphantom{\begin{aligned} \text{Par définition, } m_{x+t} &= \frac{-\frac{d}{dt} l_{x+t}}{l_{x+t}} \\ \text{et } \frac{-\frac{d}{dt} l_{x+t}}{l_{x+t}} &= -\frac{d}{dt} \ln {}_t p_x \end{aligned}} \right\} \rightarrow \boxed{m_{x+t} = m = cte}$$

③ Avec un taux instantané d'entrée en incapacité constant, l'équation (3) devient:

$$V_n = m^{D_x} \prod_{i=1}^{n_x} e^{-m(ti - ri)} = e^{D_x \ln m - m \sum_{i=1}^{n_x} (ti - ri)}$$

La logvraisemblance est: $L_n = \ln V_n = D_x \ln m - m \sum_{i=1}^{n_x} (ti - ri)$

④ L'estimateur du maximum de vraisemblance minimise la logvraisemblance de l'échantillon, il est donc solution de l'équation suivante.

$$\frac{dL_n}{dm} = 0 = \frac{D_x}{m} - \sum_{i=1}^{n_x} (ti - ri)$$

$$\frac{d^2L_n}{dm^2} = -\frac{D_x}{m^2} < 0$$

On en déduit l'estimateur du maximum de vraisemblance : $\hat{m} = \frac{D_x}{\sum_{i=1}^{n_x} (ti - ri)}$

D'où, l'estimateur du maximum de vraisemblance de la probabilité d'entrer en incapacité à l'âge x :

$$\hat{Q}_x = 1 - e^{-\frac{D_x}{\sum_{i=1}^{n_x} (ti - ri)}}$$

• **Les propriétés de l'estimateur du maximum de vraisemblance**

D'après la littérature cet estimateur du maximum de vraisemblance est asymptotiquement efficace, sans biais, normal et de variance $\frac{1}{\hat{I}_n(p_x)}$ ($\hat{I}_n(p_x)$ = information de Fisher)

Section 3.2 : 2^{ème} approche : le modèle IARD

Cette approche correspond au 3^{ème} modèle de tarification.

3.2.1 : l'estimateur

- *principe*

Au lieu de retraiter les données pour pouvoir considérer l'entrée en incapacité sur une classe d'âge comme un événement unique, on pourrait tout simplement envisager de la traiter comme un sinistre d'assurance dommage tel que le sinistre automobile par exemple. En effet, de part son caractère répétitif, le sinistre incapacité pourrait être tarifié sur le modèle de l'assurance non vie en utilisant la variable N_x (nombre de sinistres sur $[x, x+1[$ pour chaque assuré) dans la formule suivante :

Prime = $Esp(N_x) * (\text{coût moyen d'un sinistre sur } [x, x+1[)$

Hypothèse fondamentale :

L'hypothèse traditionnellement retenue pour la loi de cette variable N_x est la loi de Poisson de paramètre λ_x . Dans ce cas, il est généralement admis que pour un assuré i qui ne serait observé que sur la période $[a_i, b_i]$ (avec $[a_i, b_i] \subset [x, x+1[$) en raison de troncatures et/ou de censures, la variable N_i qui lui est associé suit une loi de poisson de paramètre $\lambda_i = \lambda_x (b_i - a_i)$.

- *construction*

Nous allons maintenant déterminer l'estimateur du maximum de vraisemblance de $Esp(N_x) = \lambda_x$ il sera noter \hat{I}_x .

La fonction de vraisemblance du n-échantillon $\{N_i \approx \text{Poisson}(\lambda_i), i=1..n\}$:

$$L = \prod_{i=1}^n \frac{e^{-I_i} I_i^{N_i}}{N_i!}$$

En utilisant l'hypothèse fondamentale : $I_i = I_x (b_i - a_i)$

et après simplification, on a :

$$\ln L = -I_x \sum_{i=1}^n (b_i - a_i) + D_x \ln I_x + \sum_{i=1}^n [N_i \ln(b_i - a_i) - \ln(N_i!)]$$

avec $D_x = \sum_{i=1}^n N_i$ le nombre total de sinistres survenus sur $[x, x+1[$

équation du maximum de vraisemblance :

$$\frac{\partial \ln L}{\partial I_x} = -\sum_{i=1}^n (b_i - a_i) + \frac{D_x}{I_x} = 0$$

d'où l'estimateur du maximum de vraisemblance :
$$\hat{I}_x = \frac{D_x}{\sum_{i=1}^n (b_i - a_i)}$$

- **propriétés de l'estimateur :**

En tant qu'estimateur du maximum de vraisemblance, cet estimateur est asymptotiquement efficace.

Espérance :

$$Esp(\hat{I}_x) = \frac{\sum_{i=1}^n Esp(N_i)}{\sum_{i=1}^n (b_i - a_i)} = \frac{\sum_{i=1}^n I_x (b_i - a_i)}{\sum_{i=1}^n (b_i - a_i)} = I_x$$

Cet estimateur est sans biais.

Variance :

$$\widehat{Var}(\hat{I}_x) = Var\left(\frac{D_x}{\sum_{i=1}^n (b_i - a_i)}\right) = \frac{1}{\left[\sum_{i=1}^n (b_i - a_i)\right]^2} Var\left(\sum_{i=1}^n N_i\right) = \frac{1}{\left[\sum_{i=1}^n (b_i - a_i)\right]^2} \sum_{i=1}^n Var(N_i)$$

sachant que : $Var(N_i) = \lambda_x (b_i - a_i)$ on obtient :

$$\widehat{Var}(\hat{I}_x) = \frac{\hat{I}_x}{\sum_{i=1}^n (b_i - a_i)}$$

- **intervalle de confiance :**

Comme \hat{I}_x est sans biais :

$$\frac{\sqrt{n}[\hat{I}_x - I_x]}{\sqrt{Var(\hat{I}_x)}} \xrightarrow{L} N(0,1) \quad \text{quand } n \rightarrow +\infty$$

d'où un intervalle de confiance pour λ_x de niveau asymptotique $(1-\eta)$:

$$\left[\hat{I}_x - q_{1-\frac{\eta}{2}} * \sqrt{\frac{Var(\hat{I}_x)}{n}} ; \hat{I}_x + q_{1-\frac{\eta}{2}} * \sqrt{\frac{Var(\hat{I}_x)}{n}} \right]$$

avec $q_{1-\frac{\eta}{2}}$ = le quantile d'ordre $(1-\eta/2)$ de la loi $N(0,1)$.

3.2.2 : l'adéquation à une loi de poisson

Pour pouvoir utiliser l'estimateur ci-dessus, il faut donc préalablement vérifier pour chaque classe d'âge $[x, x+1[$, l'adéquation de la distribution de N_x à une loi de poisson.

Dans le cas d'une distribution discrète, on utilise classiquement le test d'adéquation du χ^2 .

Notations pour une classe d'âge :

- k : le nombre maximum de sinistre par assuré.
- $p_0, p_1, p_2, \dots, p_k$: les probabilités associées à la distribution de référence
- $M_0, M_1, M_2, \dots, M_k$ l'échantillon des effectifs pour chaque nombre de sinistres
- $n = \sum_{i=0}^k M_i$: population totale dans la classe d'âge considérée

Considérons alors la statistique D^2 définie comme suit :

$$D^2 = \sum_{i=0}^k \frac{(M_i - np_i)^2}{np_i}$$

Cette statistique est une mesure de l'écart aléatoire entre les effectifs réalisés et les effectifs espérés.

On sait que D^2 est asymptotiquement distribué comme une variable de χ^2_k et ceci quelle que soit la loi de N_x ; d'où le test du χ^2 :

H_0 : N_x suit une loi de poisson

On rejettera H_0 lorsque la valeur observée de D^2 est trop grande.

En fait on calculera la p-value \hat{a} (ou seuil critique) définie par :

$$\hat{a} = P(\chi^2_k > d^2)$$

Si cette p-value est trop petite (<1%), on rejette H_0 .

cas particulier des estimations :

Lorsque seule la forme de la distribution est spécifiée comme c'est le cas ici avec la loi de poisson, il est nécessaire d'estimer les paramètres de cette loi. Si m est le nombre d'estimations à réaliser, le degré de liberté du χ^2 est ainsi réduit de m . Dans le cas présent $m=1$ puisqu'il faut estimer λ le paramètre de poisson.

Il convient ici de prendre certaines précautions :

- L'estimation en question doit être l'estimation du maximum de vraisemblance effectuée au moyen des $k+1$ classes de la distribution.
- De plus, la loi de D^2 est asymptotique, mais l'on admet que $D^2 \approx \chi^2_{k-1}$ si $np_i > 1$ pour tout i (critère de Cochran). Dans le cas contraire, on procèdera à des regroupements de classes.
- Enfin, ce test ne s'applique qu'à un échantillon iid et même si nos observations sont par nature indépendantes, nous avons supposé que leur paramètre de poisson était lié à l'exposition de l'assuré.

Le test d'adéquation sera donc réalisé uniquement sur les assurés exposés totalement (1 an).

Pour les autres, il convient de vérifier l'hypothèse de linéarité du paramètre de poisson :

$$\lambda_i = \lambda_x (b_i - a_i).$$

3.2.3 : La linéarité de l

Dans l'hypothèse fondamentale, on suppose la linéarité du paramètre de poisson λ_i en fonction de la durée d'exposition.

On peut tout simplement vérifier cette linéarité graphiquement et pour chaque âge.

Pour un âge donné et pour une durée d'exposition égale à t on estime le paramètre de poisson λ_t par la moyenne empirique, soit :

$$\hat{I}_t = \frac{d_t}{n_t}$$

avec :

- d_t le nombre total de sinistres parmi les personnes exposées sur une durée t
- n_t le nombre d'assurés exposés sur une durée t

On trace ensuite le graphe des \hat{I}_t en fonction de $t \in [0,1]$ et on regarde si l'on obtient une fonction linéaire.

Maintenant que nous avons exposé les estimateurs correspondant aux modèles de tarif envisageables, nous allons expliquer pourquoi notre choix s'est porté sur l'estimateur IARD que nous venons de voir.

Section 3.3 : Choix de l'estimateur de type « IARD »

Pour déterminer notre estimateur, il faut prendre en compte non seulement les contraintes tarifaires, mais aussi les contraintes liées à nos données.

Si l'on envisage le 1^{er} modèle de tarification par exemple, il est impensable d'espérer construire les tables (d'entrée et de maintien) liées aux 2^{ème} voir 3^{ème} sinistres dans l'année. En effet, sur le portefeuille global et tout âge confondu, seulement 3347 assurés ont 2 sinistres la même année et 184 assurés en ont 3.

Quant au 2^{ème} modèle de tarification, il impose de construire deux types de tables de maintien en incapacité, un pour évaluer le capital constitutif nécessaire au calcul de la prime et un autre pour provisionner correctement les sinistres incapacité. De plus, ce tarif mélange des sinistres de causes différentes et donc des risques différents. Pour toutes ces raisons, nous éliminerons aussi ce modèle.

Il ne reste finalement que le 3^{ème} modèle, qui a priori ne pose pas de problème ni dans sa conception ni dans sa réalisation et qui impose de prendre l'estimateur de type «IARD» pour élaborer nos tables d'entrée en incapacité.

Maintenant que nous avons choisi notre estimateur, nous devons définir la segmentation de nos tables.

Chapitre 4 : La segmentation des tables : homogénéité de la sinistralité

Nous savons que pour utiliser nos tables d'entrée pour tarifier les garanties complémentaires en incapacité, il faut tenir compte de certaines contraintes pratiques et commerciales.

Tout d'abord, nous avons déjà vu que nous distinguons 3 types de tables :

- les tables liées aux garanties IJ et EXO
- les tables liées aux garanties IJH
- les tables liées aux garanties EXO sur 2 têtes

Pour des raisons d'échantillon et de priorité, nous ne nous intéresserons ici qu'aux deux premiers types de tables.

Nos résultats doivent en même temps traduire au mieux le risque supporté par l'assureur mais aussi être applicables aux futurs assurés. Cette double contrainte implique de tenir compte non seulement des variables discriminantes du risque mais aussi des variables tarifaires traditionnelles pour segmenter correctement nos tables. Nous allons lister a priori ces variables et étudier leur influence sur l'entrée en incapacité.

Section 4.1 : Les variables tarifaires traditionnelles

Hormis le montant souscrit et le cas des risques aggravés, le tarif des garanties complémentaires de dommages corporels dépend généralement de l'âge de l'assuré et de la franchise souscrite. Il est ainsi presque automatique de segmenter les tables suivant ces deux variables pour éviter une anti-sélection. En effet une compagnie qui afficherait des tarifs identiques quelle que soit la franchise souscrite par exemple risquerait d'anéantir l'antisélection. Notre segmentation doit donc au moins tenir compte de l'âge de l'assuré et des franchises souscrites. De toute façon, avoir une estimation par âge faisait partie des objectifs définis pour la construction des tables. En fait l'âge n'est pas une segmentation mais plutôt une dimension de la table.

Section 4.2 : Les variables discriminantes

Les deux variables précédentes font bien sûr partie des variables a priori discriminantes pour l'entrée en incapacité. On peut aussi y ajouter les caractéristiques client, telles que le sexe, la catégorie socio-professionnelle ou la région du domicile, les caractéristiques contrat telles que l'ancienneté du contrat ou le montant souscrit ainsi que les caractéristiques sinistre telles que la cause (accident ou maladie) ou la nature (hospitalisation ou non).

Voici finalement la liste des variables discriminantes envisageables (hormis l'âge qui est une dimension de la table) :

- la franchise souscrite
- l'ancienneté du contrat
- le montant souscrit
- le sexe de l'assuré
- la CSP de l'assuré
- la région de l'assuré
- la cause du sinistre
- la nature du sinistre

Les données disponibles nous contraignent cependant à éliminer d'office certaines variables.

En effet, la variable région n'est pas très souvent renseignée (25% des observations sont manquantes).

La CSP est plus fiable, mais il est inutile d'espérer pouvoir segmenter suivant cette variable puisque notre portefeuille est essentiellement constitué de salariés (70% des assurés) et que toute autre CSP n'est pas assez représentée pour permettre de construire une table d'expérience fiable.

Quant au montant souscrit nous ne nous en occuperons pas ici car nous avons choisi de construire dans l'immédiat une table par tête mais il n'est pas exclu d'élaborer ultérieurement une table par capitaux où le montant garanti prendra alors toute son importance.

Enfin, la nature du sinistre ne peut pas être à la base d'une segmentation des tables de type IJ (ou EXO) car il n'est pas certain que les assurés possédant uniquement une garantie IJ (ou EXO) déclarent leur hospitalisation puisque cette dernière ne modifie en rien leurs prestations. Il est donc fort probable que notre base de donnée sur les garanties IJ et EXO ne soit pas exhaustive au niveau des sinistres comportant une période d'hospitalisation. Bien entendu, le risque lié à l'hospitalisation est pris en compte dans les tables de type IJH.

Etant donné les remarques précédentes, les variables région, CSP, montant souscrit et nature du sinistre sont exclues d'office pour la segmentation.

En ce qui concerne la variable sexe, les statistiques descriptives prouvent déjà que c'est une variable très discriminante. En effet, globalement le taux d'entrée en incapacité en base 100 est de 114 chez les femmes contre seulement 93 chez les hommes (*cf la section 2.5 de la partie III*). La segmentation suivant le sexe de l'assuré est donc évidente et elle servira de base pour l'étude des autres variables.

Pour ces variables restantes : la franchise, l'ancienneté du contrat et la cause du sinistre, nous avons étudié séparément leur influence sur la sinistralité en estimant rapidement l'entrée en incapacité.

Pour certaines variables, des choix ont été faits dans la façon d'étudier cette influence.

- 1) Pour les franchises, il existe 4 valeurs par garantie correspondant aux 4 types de sinistre :
 - accident sans hospitalisation
 - accident avec hospitalisation
 - maladie sans hospitalisation
 - maladie avec hospitalisation

Une étude préalable montre que la durée de la franchise varie surtout en fonction de la cause et que l'hospitalisation modifie rarement la franchise, c'est pourquoi nous avons retenu pour notre étude les principaux couples de franchise souscrits. On entend par couple de franchise la durée en cas de maladie associée à celle en cas d'accident, soit par exemple le couple 30/7 qui signifie que l'assuré a souscrit une franchise 30 jours en maladie (sans hospitalisation) et une franchise 7 jours en accident (sans hospitalisation).

Dans notre portefeuille, on observe principalement 8 couples de franchise différents :

0/0, 15/15, 30/7, 30/30, 60/30, 60/60, 90/7 et 90/90 (les durées sont exprimées en jours).

Comme la durée de la franchise est liée à la cause du sinistre, il est impossible de regarder uniquement l'influence de la franchise sur l'entrée en incapacité indépendamment de la cause. Il a donc fallu d'abord procéder aux segmentations souhaitées sur les autres variables avant d'évaluer l'impact de la franchise sur la fréquence d'entrée.

- 2) Pour étudier l'ancienneté (durée écoulée entre la souscription du contrat et la survenance du sinistre), nous avons défini trois classes :
- de 0 à 1 an
 - de 1 à 2 ans
 - + de 2 ans

Section 4.3 : la segmentation

Pour étudier l'influence des variables envisagées (la franchise, l'ancienneté du contrat et la cause du sinistre) nous avons calculé pour chaque segmentation et par âge le nombre moyen de sinistres par assuré. Les résultats obtenus et les segmentations qui en découlent sont exposés ci-dessous.

4.3.1 : Les tables de type IJ

Contrairement à l'ancienneté, et comme supposé à priori, l'âge, le sexe, la cause et la franchise sont effectivement des variables discriminantes.

En fait, si nous avons retenu l'ancienneté du contrat c'était en supposant qu'à travers cette variable on pourrait remarquer l'influence de la sélection médicale à la souscription. Cette influence n'apparaît pas ici mais on peut l'expliquer car nous avons préalablement exclu de notre échantillon les assurés surprimés, ce qui fait que les assurés restants n'ont pas a priori de problèmes de santé particuliers et le passage à la sélection médicale ne garantit pas une sinistralité meilleure juste après la souscription en comparaison à celle quelques années plus tard (en annulant l'effet vieillissement bien sûr). En effet, nos assurés restent exposés à des maladies bénignes ou peu prévisibles, quant aux accidents, ils sont par nature inattendus.

De notre étude de sinistralité, il ressort que les hommes et les femmes n'ont pas le même comportement face au risque accident ou maladie, en effet, les femmes ont proportionnellement moins d'accidents que les hommes ce qui peut s'expliquer par une attitude plus prudente de leur part, alors qu'elles sont plus souvent en arrêt maladie surtout avant 30 ans et après 50 ans. On remarque d'ailleurs ici que l'âge intervient dans la sinistralité et plus généralement il l'aggrave.

Pour les franchises, il est évident que l'entrée en incapacité est plus fréquente sur des faibles franchises. Pour des contraintes de volumes suffisants, il est quand même préférable de regrouper certaines franchises avec une sinistralité proche. Ces regroupements varient selon la cause du sinistre.

Nous avons 3 groupes de franchises sur les accidents et 4 groupes sur les maladies.

De plus ce ne sont pas tout à fait les mêmes regroupements chez les hommes et chez les femmes.

En conclusion, on obtient 15 segmentations pour les tables IJ, elles sont précisées dans le tableau du dernier paragraphe.

NB : La franchise 0/0 ne correspond pas à une franchise nulle mais à une franchise relative 90/90 alors que toutes les autres franchises sont absolues. Ceci explique pourquoi la sinistralité sur la franchise 0/0 est proche de celle sur les franchises importantes (90/90).

4.3.2 : Les tables de type IJH

Ici, nous nous intéressons à l'entrée en hospitalisation, et contrairement à l'entrée en incapacité simple, aucune franchise n'est annoncée sur les garanties IJH. Il faut cependant savoir que par définition on appelle hospitalisation tout séjour de plus de 48 heures dans un établissement de soins. Il existe donc une franchise implicite de 2 jours commune à tous les sinistres avec hospitalisation. Cette franchise ne sera donc pas source de segmentation.

De même que pour les tables IJ, l'ancienneté du contrat n'intervient pas dans la sinistralité, il ne reste finalement que le sexe et la cause du sinistre comme variables discriminantes, ce qui conduit à 4 segmentations.

4.3.3 : Sinistralités moyennes et arbre des segmentations

Les tableaux ci-dessous précisent les segmentations faites sur les tables IJ et les tables IJH ainsi que la fréquence moyenne des sinistres pour chaque segmentation :

Tables IJ	ACCIDENT			MALADIE				
	15/15 et 30/7	30/30 60/30 et 90/7	0/0 60/60 et 90/90	15/15	30/7	30/30 et 60/30	0/0 90/7 60/60 et 90/90	
homme	15/15 et 30/7	30/30 60/30 et 90/7	0/0 60/60 et 90/90	15/15	30/7	30/30 et 60/30	0/0 90/7 60/60 et 90/90	
fréquence	6%	1,2%	0,3%	5%	3%	1,2%	0,4%	
femme	15/15 et 30/7	30/30 et 60/30	0/0 60/60 90/7 et 90/90	15/15	30/7	30/30	60/30 et 90/90	0/0 60/60 et 90/7
fréquence	2,3%	0,5%	0,1%	14%	5%	2,7%	1,1%	0,6%

Tables IJH	ACCIDENT	MALADIE
fréquence homme	1,6%	2,3%
fréquence femme	4,1%	0,4%

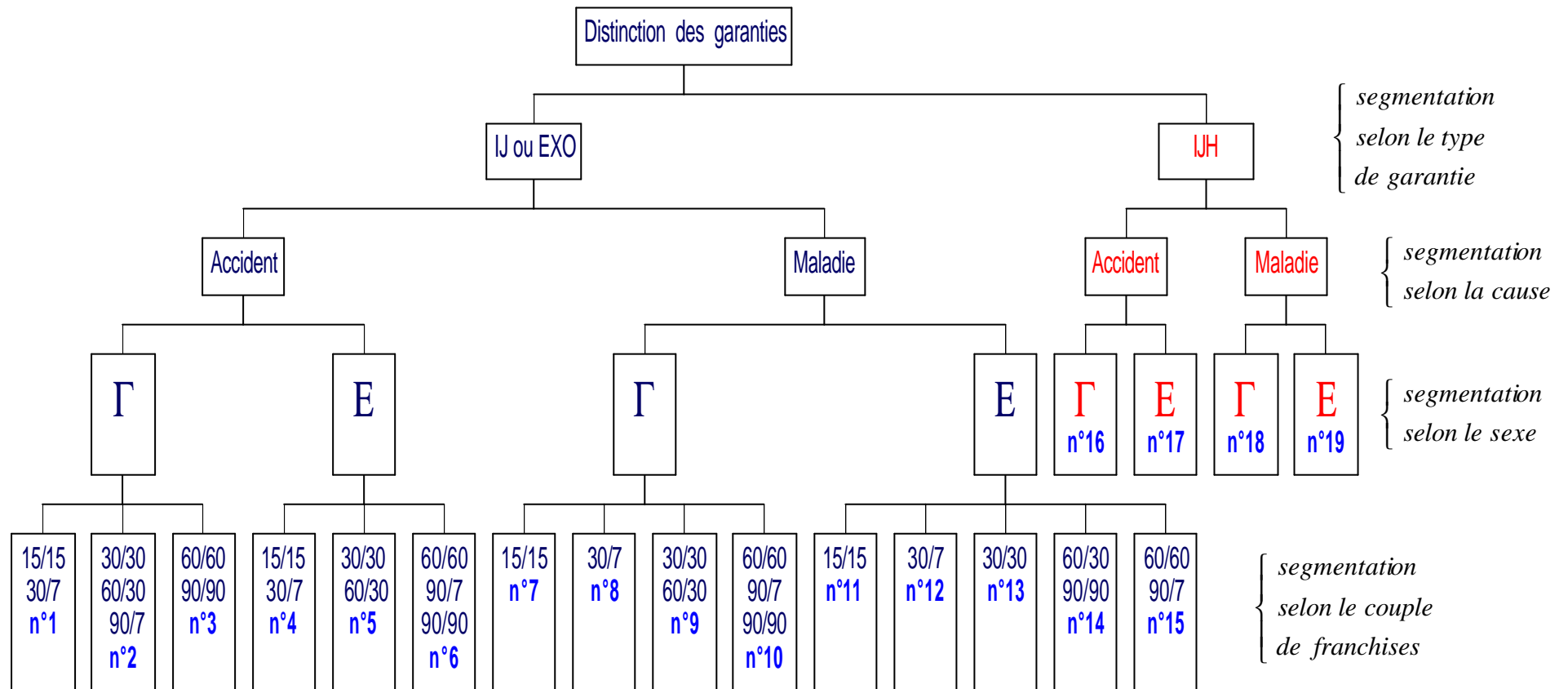
conclusion :

19 tables d'entrée en incapacité ont été élaborées. Elles traduisent chacune un risque bien précis.

Par exemple, la 1^{ère} table avec une fréquence moyenne de 6% traduit le risque d'incapacité pour cause d'accident chez les hommes ayant souscrit une garantie IJ ou EXO avec une franchise de 15 jours en accident comme en maladie ou alors une franchise de 30 jours en maladie et de 7 jours en accident.

L'arbre de segmentation ci-après permet une meilleure visualisation des segmentations conduisant aux 19 tables d'entrée en incapacité.

Les tables d'entrée



Chapitre 5 : Les contraintes de volume

Nous venons de former des groupes de sinistralité homogène suivant certaines variables discriminantes (sexe, cause et franchise). Maintenant, nous désirons obtenir pour chacune de ces segmentations une table d'entrée déclinée sur tous les âges.

Or, pour avoir des résultats fiables et robustes, il faut faire en sorte d'avoir pour chaque âge un volume de données suffisant, surtout au niveau du nombre de personnes exposées au risque d'entrée.

Ceci implique de regrouper certains âges ensemble alors qu'on désire obtenir au final des estimations brutes par âge.

Nous devons donc déterminer non seulement le volume minimum à respecter mais aussi la manière d'atteindre ce volume tout en conservant un résultat par âge.

Pour ce faire, nous avons envisagé deux méthodes :

- une estimation par classe d'âge suivie d'une interpolation à jonctions souples
- un glissement

On entend par « estimation par classe d'âge » l'estimation basée sur un échantillon qui regroupent des assurés sur des âges proches (les classes d'âge).

Après un bref exposé de ces deux méthodes, nous choisirons l'une d'elles et fixerons les volumes minimums à satisfaire.

Section 5.1 : Estimations par classe d'âge et interpolation à "jonction souple"

Au lieu de faire une estimation par âge basée sur l'échantillon propre à chaque âge, on regroupe les échantillons sur plusieurs âges et on fait à chaque fois une seule estimation valable pour tous les âges regroupés.

On obtient ainsi des estimations brutes par « paliers » (un palier correspond à un regroupement d'âges).

On veut maintenant se ramener à des estimations brutes par âge. Pour ce faire, nous allons réaliser une interpolation à jonction souple qui nécessite d'avoir des classes de même taille.

Pour faciliter la suite du raisonnement, nous considérerons que nous avons par exemple des classes d'âge quinquennales (donc des paliers de 5 ans).

La méthode d'interpolation retenue comprend deux étapes :

- le calcul des valeurs pivots
- l'application de la formule d'interpolation proprement dite

5.1.1 : calcul des valeurs pivots

La première étape consiste à ne pas faire l'interpolation sur les estimations brutes par classe mais sur des valeurs "pivots" calculées aux âges centraux de chaque classe. Ces valeurs "pivots" sont obtenues en calculant des moyennes mobiles sur le nombre de sinistres et les durées d'exposition au risque dans chaque classe d'âge quinquennale.

Le raisonnement nécessite un changement d'échelle. Nous noterons u_k et u_{k+1} les valeurs pivots correspondantes à deux classes quinquennales successives mais ces valeurs pivots sont attribuées aux âges centraux des classes distants de cinq ans.

5.1.2 : la formule d'interpolation d'Everett

Cette partie correspond à l'interpolation proprement dite.

La formule d'interpolation d'Everett est la suivante:

$$v_{x+s} = F(s) u_{x+1} + F(t) u_x$$

où:

- $0 \leq s \leq 1$ (à partir de classes quinquennales $s = (0; 0,2; 0,4; 0,6; 0,8; 1)$);
- $t = 1 - s$;
- u_x sont les valeurs pivots;
- v_{x+s} sont les interpolations par âge;
- $F(s)$ est un opérateur égal à $A(s) + B(s)\delta^2 + C(s)\delta^4 + \dots$
- $\delta u_x = u_{x+1/2} - u_{x-1/2}$
- $A(s), B(s), C(s)$ sont des fonctions polynomiales expliquées ci-dessous.

Les interpolations v_{x+s} sont des arcs polynomiaux par segments. C'est à dire qu'ils sont différents d'une classe d'âge à une autre. Ils doivent respecter certaines contraintes de continuité à leurs jonctions d'où le nom d'interpolation à jonction souple.

$$\text{Plus précisément, } v_y = \begin{cases} \dots \\ p_{x-1}(y) \quad \forall x-1 \leq y \leq x \\ p_x(y) \quad \forall x \leq y \leq x+1 \\ \dots \end{cases}$$

où les fonctions $p_x(y)$ sont des polynômes qui vérifient les propriétés suivantes:

1) les polynômes $p_x(y)$ se rejoignent aux jonctions (interpolations aux âges centraux des classes) c'est à dire :

$$v_{(x-1)+s/s=1} = v_{x+s/s=0} \text{ pour tout } x.$$

2) aux jonctions, les polynômes $p_x(y)$ ont des dérivées premières et secondes égales, c'est à dire: $v'_{(x-1)+s/s=1} = v'_{x+s/s=0}$ et $v''_{(x-1)+s/s=1} = v''_{x+s/s=0}$ pour tout x .

Ceci implique que $F(s), A(s), B(s), C(s)$ doivent être telles que:

- 1) $A(0) = B(0) = C(0) = 0$
- 2) $2F'(1) = F'(0)(2 + \delta^2)$
- 3) $A''(0) = B''(0) = C''(0)$

Une des propriétés principales d'une interpolation est le degré des polynômes qu'elle laisse invariant.

Section 5.2 : Le glissement

Cette méthode consiste à regrouper directement les âges de manière à satisfaire les contraintes de volumes tout en conservant un résultat propre à chaque âge. En effet, pour trouver l'estimation brute pour un âge donné x où les volumes sont insuffisants, nous allons regrouper les données de l'âge x avec celles des âges autour de x de telle sorte que x soit l'âge moyen pondéré de la classe d'âge et que le volume d'observations minimum soit respecté.

On entend par « âge moyen pondéré » une pondération par rapport aux expositions sur chaque âge formant la classe.

Cette opération est répétée pour chaque âge où les contraintes de volumes l'obligent, sachant bien sûr que sur les âges extrêmes, il faudra parfois se contenter de paliers (qui disparaîtront avec le lissage). Sinon, en dehors des âges extrêmes, nous obtenons bien une table par âge et non par classe d'âge comme exposé au début de la méthode précédente.

On parle de glissement car en fait pour passer d'une estimation à l'autre, on fait généralement glisser la classe d'âge de référence.

Dans l'algorithme de constitution des classes d'âge, on arrête le regroupement dès que le volume minimum est atteint et que la moyenne pondérée des âges de la classe est égale à l'âge considéré $\pm 0,4$ ans.

Section 5.3 : Choix de la méthode par glissement

Comme nous venons de le voir, il est clair que le glissement est très simple à mettre en place et facilement modulable, alors que l'interpolation est plus contraignante.

En effet, nos estimations brutes nécessitent une interpolation par âge seulement aux âges où il y a peu de données. Cependant, la méthode d'interpolation utilisée nécessite de regrouper aussi par classes d'âges certaines des données traitées par âge dans nos estimations brutes. Il faudrait donc faire une interpolation par âge sur la totalité des données et ne tenir compte des estimations interpolées que sur les âges qui appartenaient à un palier.

De plus, l'interpolation s'applique sur des paliers uniformes (même taille de classe d'âge). Or en pratique, il n'est pas évident de déterminer cette taille de palier uniforme pour toutes les segmentations et tous les âges. En effet, il faut parfois regrouper énormément (surtout aux extrémités, 10 ans par exemple) alors qu'ailleurs des classes de 2 ans suffisent. Or pour l'interpolation, il est préférable d'avoir des classes d'âge uniformes et si possible contenant un nombre impair d'années pour pouvoir définir un âge central.

Au vu de ces contraintes, nous avons choisi d'utiliser le glissement pour sa facilité d'application et d'adaptation (aux tables de maintien par exemple).

Après quelques essais, nous avons fixé le volume minimum à 2000 expositions annuelles et pour améliorer déjà la régularité de nos résultats bruts nous avons opté de plus pour un minimum de 30 sinistres par classe d'âge.

Nos tables brutes ainsi obtenues sont exposées dans le chapitre suivant.

Chapitre 6 : Les tables brutes

Le but de ce chapitre est de livrer des estimations brutes sur toutes les segmentations et pour chaque âge. Pour cela il faut tester préalablement les hypothèses nécessaires à la construction de l'estimateur.

Section 6.1 : adéquations au modèle

Le modèle choisi repose sur deux hypothèses :

- le nombre annuel espéré de sinistres par assuré suit une loi de Poisson
- le paramètre de Poisson est une fonction linéaire de la durée d'exposition au risque

Nous allons donc tester si nos données vérifient ces hypothèses.

De plus, la construction d'une table d'expérience fiable nécessite la stabilité de la sinistralité sur la période d'observation, nous rappelons ici comment nous avons vérifié cette hypothèse.

6.1.1 : adéquation à la loi de Poisson

Il faut savoir que dans cette partie, on ne s'intéressera qu'aux assurés exposés sur la totalité de l'année, sachant que pour les autres on testera la linéarité du paramètre de Poisson.

Pour valider l'adéquation à la loi de Poisson, nous allons tout d'abord regarder graphiquement si cette loi est adaptée à notre portefeuille, puis nous vérifierons si la principale propriété de la loi de Poisson (espérance = variance) est vérifiée par nos données, enfin nous testerons l'hypothèse de loi par le χ^2 d'adéquation.

a) Adéquation graphique

Nous souhaitons vérifier si la structure de nos données correspond bien à la structure de la loi de Poisson. Pour ce faire, nous avons calculé pour chaque table et chaque âge nos effectifs par classe de sinistre, c'est à dire :

- le nombre d'assurés sans aucun sinistre dans l'année
- le nombre d'assurés avec 1 sinistre dans l'année
- le nombre d'assurés avec 2 sinistres dans l'année
- le nombre d'assurés avec 3 sinistres et plus dans l'année

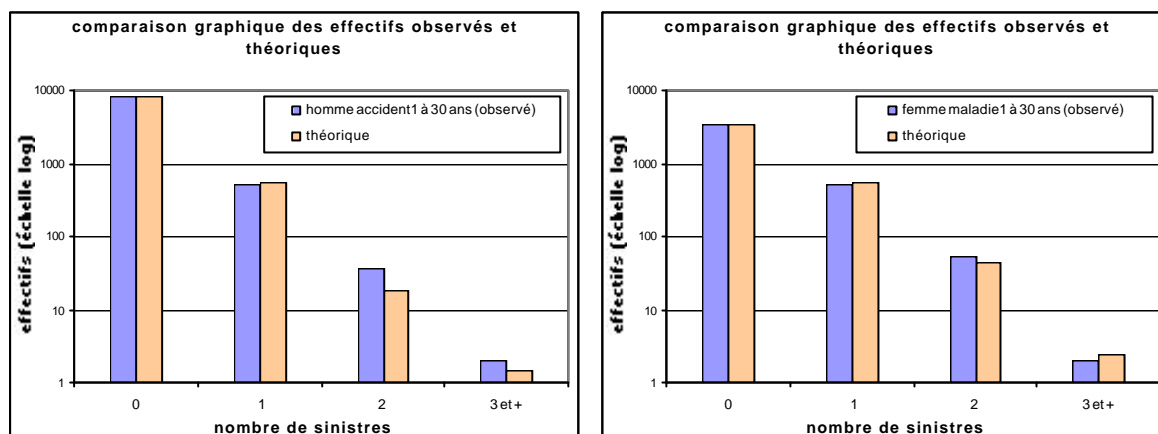
et nous avons comparé graphiquement ces effectifs observés aux effectifs théoriques obtenus avec la loi de poisson (dont le paramètre est estimé par l'espérance empirique).

Etant donné le nombre de tables (19) et le nombre d'âges sur chaque table (41) nous ne pouvons pas tout visualiser. Nous avons donc retenu uniquement l'âge le plus représentatif soit 30 ans.

Nous présenterons ici les résultats obtenus sur les deux tables IJ également les plus significatives en terme de volume c'est à dire les tables suivantes :

- Maladie1 chez les femmes
- Accident1 chez les hommes

Vu l'écart important entre les effectifs des différentes classes de sinistres, nous avons opté pour une échelle logarithmique qui permet une plus grande précision dans la comparaison. Voici les deux comparaisons graphiques retenues :



Au vu de ces graphes, nous pouvons évidemment affirmer que la structure de la loi de Poisson est tout à fait adaptée à notre portefeuille d'assurés.

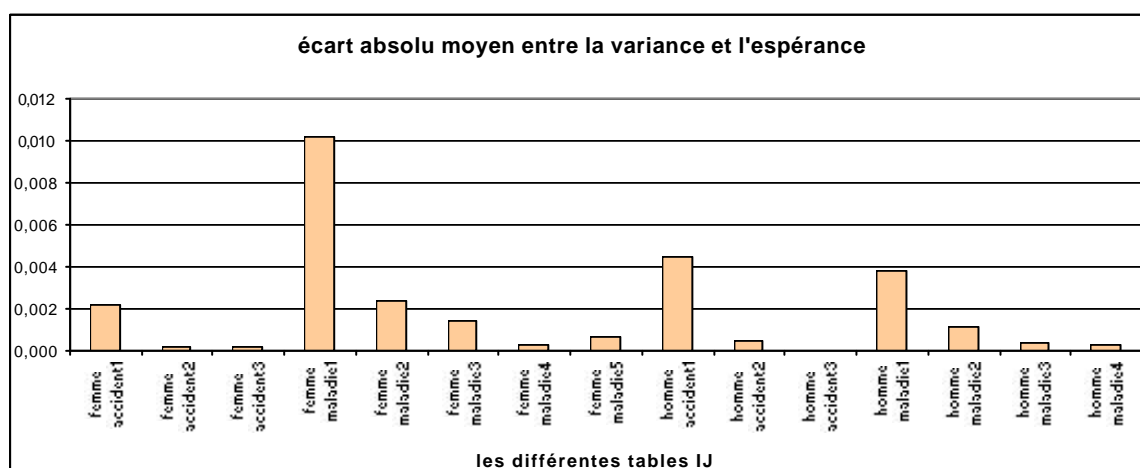
Il convient maintenant de vérifier que notre échantillon respecte la caractéristique de la loi de Poisson.

b) La caractéristique de Poisson

Une variable suivant une loi de Poisson est principalement caractérisée par l'égalité entre son espérance et sa variance.

Sur les mêmes échantillons que précédemment, nous avons donc calculé l'espérance et la variance empirique pour comparer la différence : variance - espérance.

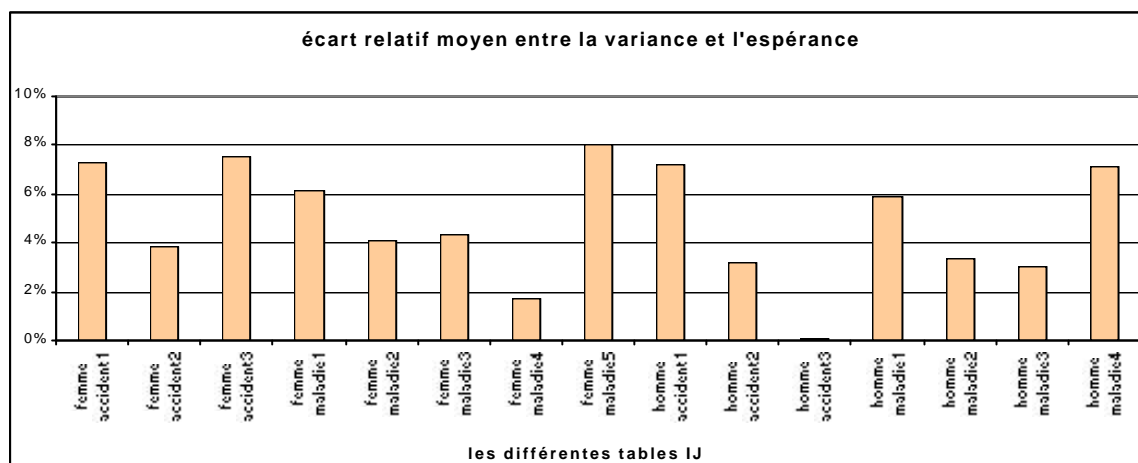
Voici ce que l'on obtient comme écart absolu moyen (sur tous les âges) suivant les tables :



La différence étant positive, c'est la variance qui est plutôt supérieure à l'espérance.

L'écart moyen est généralement faible (< 0,2%), mais on ne peut rien en conclure car l'espérance et la variance prennent aussi des valeurs très faibles. Nous allons donc plutôt regarder l'écart moyen relatif : (variance – espérance) / variance.

Voici le graphe obtenu :



Globalement, l'écart moyen relatif ne dépasse pas 6%. Nous pouvons ainsi conclure que nos échantillons satisfont la caractéristique de Poisson.

Pour être rigoureux, il faut quand même réaliser le test d'adéquation du χ^2 .

c) le test d'adéquation du χ^2

Nous avons constitué 3 classes :

- 0 sinistre dans l'année
- 1 sinistre dans l'année
- 2 sinistres et + dans l'année

De ce fait, nous avons comparé la valeur du « d^2 observé » avec les valeurs prises par un χ^2 à un seul degré de liberté pour en déduire la p-value.

Les résultats de ce test ne sont pas parfaits sur tous les âges et toutes les tables.

Lorsque le d^2 observé est trop grand, on remarque quand même une très bonne adéquation au modèle pour estimer la probabilité de ne pas avoir de sinistre dans l'année, cependant, ce modèle surestime un peu le risque d'avoir un seul sinistre de ce fait, la probabilité d'avoir 2 sinistres ou + est légèrement sous estimée par ce modèle (cf graphiques page précédente).

Nous choisissons de conserver cette modélisation malgré tout car la loi de Poisson est la seule à avoir une forme correspondant aussi bien à la structure de nos données.

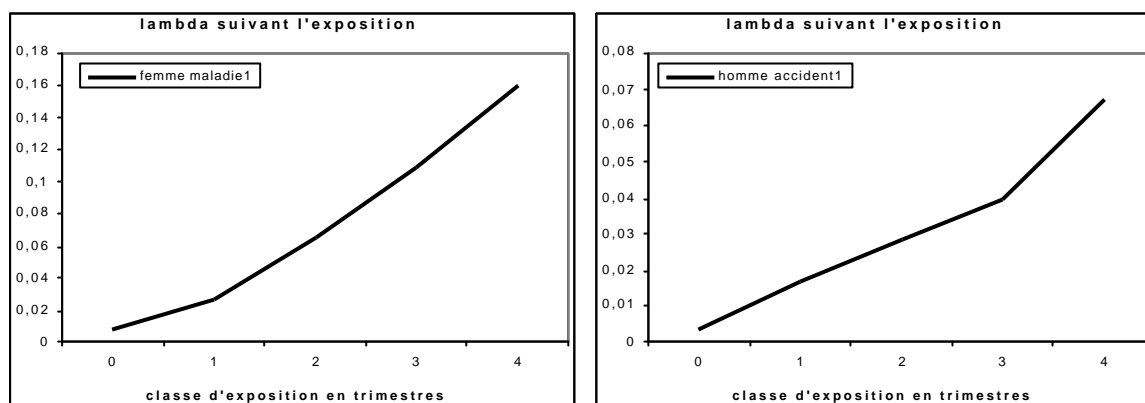
Nous sommes conscients de l'existence de certaines déviations par rapport aux valeurs exactes de cette loi (obtenues pour un sinistre et plus), mais ces petits décalages se compenseront dans le calcul de l'espérance du nombre de sinistres, qui est en fait le seul résultat, concernant l'entrée en incapacité, utilisé dans la formule de tarification.

Nous allons maintenant vérifier la seconde hypothèse liée au modèle, c'est à dire la linéarité du paramètre de Poisson : λ .

6.1.2 : La linéarité de λ

Dans toutes les tables, nous avons constitué pour chaque âge des classes trimestrielles d'exposition et estimé sur chacune d'elles la valeur de λ afin de tracer le graphe de $\lambda(t)$ où t est la classe d'exposition. Ceci représente donc 41 graphes par tables. Ne pouvant évidemment pas exposer les résultats pour chaque âge, nous avons retenu l'âge de 30 ans et les deux tables IJ maladie1 pour les femmes et accident1 pour les hommes.

Voici les graphes obtenus pour les deux tables retenues :



La linéarité de λ dans ces deux cas est aisément acceptée.

Cependant, sur certains âges, cette linéarité n'est pas obtenue, mais ceci s'explique par un manque de volume dans les classes d'exposition inférieure à un an. En effet, pour chaque assuré l'exposition annuelle partielle ne se rencontre que sur les âges extrêmes de sa période d'observation. Une étude sur les expositions suivant les segmentations et les âges révèle que généralement :

- 60% des assurés sont observés totalement (365 jours)
- 10% des assurés sont exposés de 9 à 12 mois (moins de 360 jours)
- 10% des assurés sont exposés de 6 à 9 mois
- 10% des assurés sont exposés de 3 à 6 mois
- 10% des assurés sont exposés moins de 3 mois

Il est clair que dans certaines classes d'exposition, le volume d'assurés est insuffisant pour donner de bons résultats au niveau de la linéarité de λ .

Avec ces explications, on peut dire que globalement l'hypothèse de linéarité de λ est vérifiée.

6.1.3 : La stabilité dans le temps

Allant de septembre 1994 à février 2000, notre période d'observation dure 5 ans et demi. La stabilité de la loi d'entrée sur cette période est une hypothèse nécessaire à la construction de nos tables. Nous avons validé cette hypothèse dans la section 2.4 de la partie précédente puisque nous avons mis en évidence une sinistralité constante sur les années entières de la période d'observation (1995, 1996, 1997, 1998 et 1999).

Après avoir vérifié toutes les hypothèses, nous pouvons maintenant exposer les tables obtenues grâce aux estimations brutes.

Section 6.2 : les résultats bruts

6.2.1 : les ordres de grandeur des résultats

Nous avons construit en tout 19 tables d'entrée dont 15 de type IJ et 4 de type IJH.

Pour évaluer les ordres de grandeur de nos résultats, nous allons donner pour chacune des tables le nombre espéré de sinistres à 30 et à 50 ans.

Voici les ordres de grandeur obtenus :

tables IJ	femmes									hommes					
	accident			maladie						accident			maladie		
	seq1	seq2	seq3	seq1	seq2	seq3	seq4	seq5	seq1	seq2	seq3	seq1	seq2	seq3	seq4
Espérance de sinistre à 30 ans	1,8%	0,4%	0,2%	14,7%	6,7%	3,9%	1,4%	1,3%	6,2%	1,8%	0,5%	4,6%	1,7%	1,0%	0,3%
Espérance de sinistre à 50 ans	3,0%	0,5%	0,2%	16,0%	5,7%	2,5%	1,1%	0,6%	4,7%	1,0%	0,3%	9,7%	3,7%	1,7%	0,7%

tables IJH	femmes		hommes	
	accident	maladie	accident	maladie
Espérance de sinistre à 30 ans	0,4%	4,0%	1,6%	2,1%
Espérance de sinistre à 50 ans	0,7%	5,4%	1,7%	4,1%

Ces pourcentages, extraits de nos tables, correspondent au nombre espéré de sinistres incapacité pour un assuré à l'âge de 30 ans ou de 50 ans. Pour éviter d'alourdir les commentaires des tableaux, au lieu de l'expression « nombre espéré de sinistres incapacité », nous utiliserons l'expression « fréquence d'incapacité » pour les tables IJ et « fréquence d'hospitalisation » pour les tables IJH.

Nous n'allons pas commenter ici l'influence de l'âge puisque nous ne voyons que deux âges pour chaque table. Nous aurons une vision sur tous les âges dans le paragraphe suivant.

On constate que pour comparer la fréquence d'incapacité des hommes avec celle des femmes, il faut tenir compte de la cause du sinistre.

En effet, à 30 ans sur les tables IJ, les femmes atteignent une fréquence d'incapacité pour cause de maladie de 15% contre seulement 5% au maximum pour les hommes, alors que pour les accidents, ce sont les hommes (avec 6%) qui dépassent les femmes (2%). Il en est de même sur les tables IJH.

Globalement, la fréquence de maladie est plus importante que la fréquence d'accident. Ceci s'explique peut-être par le fait qu'une seule et même maladie peut fréquemment s'accompagner de rechutes (plus ou moins espacées) et donc entraîner plusieurs arrêts de travail alors qu'un accident provoque généralement à un seul arrêt.

Bien sûr, la fréquence d'incapacité diminue lorsque la franchise augmente. En effet, la franchise croît avec le numéro de la segmentation alors que la fréquence d'incapacité diminue lorsque la segmentation augmente.

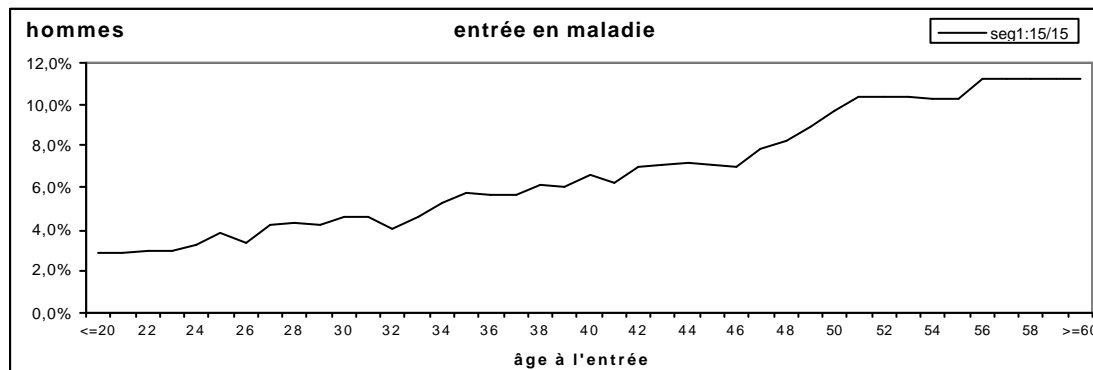
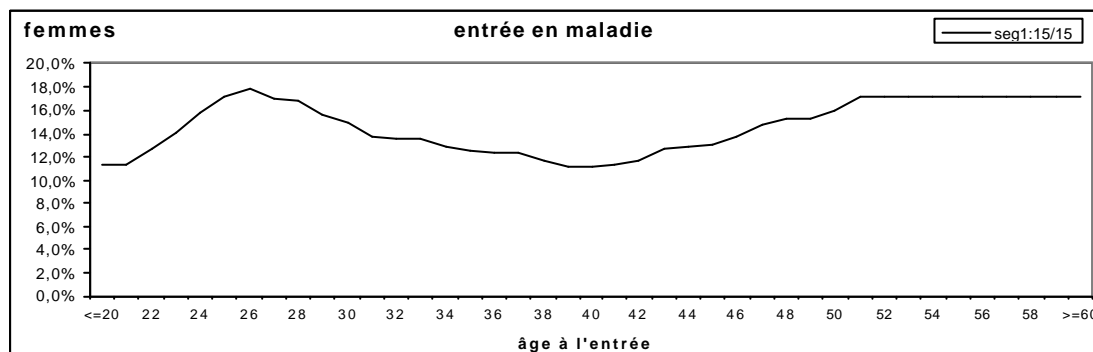
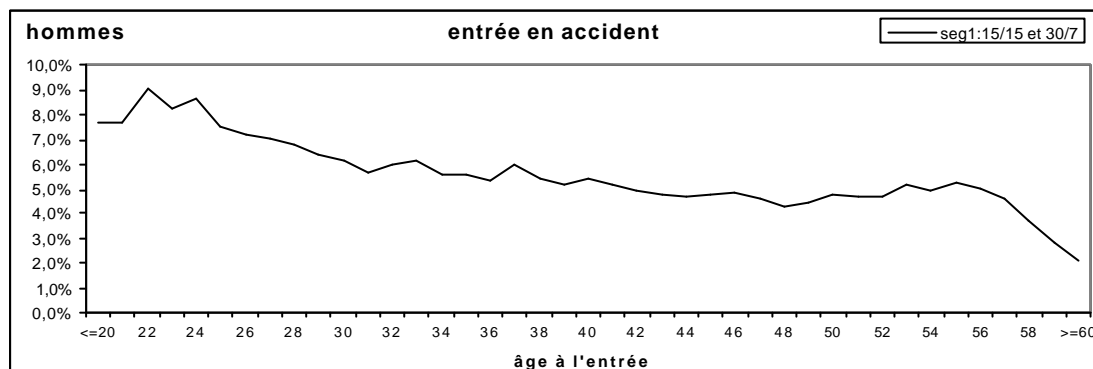
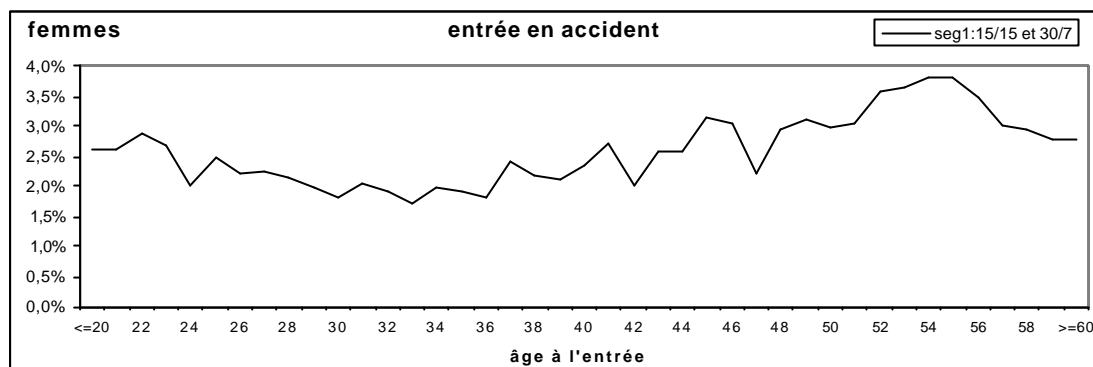
Enfin, la fréquence d'hospitalisation est plus faible que celle d'incapacité, ce qui est plausible puisque le risque est plus restreint.

Le paragraphe suivant est consacré aux tables dans leur totalité.

6.2.2 : Allure générale des courbes d'entrée en incapacité

Chacune des 19 tables construites est déclinée par âge de 20 à 60 ans. Pour visualiser et interpréter nos résultats, nous avons tracé les courbes correspondantes. Mais pour plus de lisibilité, nous ne présenterons ici que les segmentations des tables IJ correspondants aux franchises les plus basses (celles que l'on a appelé « seg1 »), car ce sont les plus représentatives en terme de volume.

Voici les 4 courbes retenues :



Contrairement à ce que l'on imagine a priori, ici, le risque d'incapacité n'est pas toujours croissant avec l'âge de l'assuré.

Cette tendance est d'ailleurs complètement inversée en ce qui concerne l'entrée en accident chez les hommes. En effet, la courbe correspondante est globalement décroissante.

On peut affirmer que l'âge est un facteur aggravant seulement pour l'entrée en maladie chez les hommes où l'on observe une fonction croissante (de 3% à 10%) quasiment linéaire.

On peut expliquer l'opposition précédente par la cause même du sinistre. En effet, beaucoup de maladies sont plus fréquentes avec l'âge alors que les accidents sont liés à l'exposition au risque et celle-ci diminue certainement en vieillissant. Avec l'âge, les gens changent souvent de rythme de vie, ils ont besoin de plus de calme, d'une vie plus posée, ils ont aussi peut-être plus conscience de la mort et donc appréhendent toute atteinte corporelle. Tout ceci se traduit par un comportement plus prudent et donc un risque d'accident plus faible.

Chez les femmes, il n'y a pas de tendance remarquable (les courbes ne sont pas monotones).

En fait, le risque maximum est atteint sur les âges extrêmes :

- entre 20 et 30 ans
- après 50 ans

Entre 30 et 50 ans, on observe un creux de sinistralité. Ce creux est très net en maladie et plus flou en accident. On retrouve ici le caractère soudain et chaotique de l'accident.

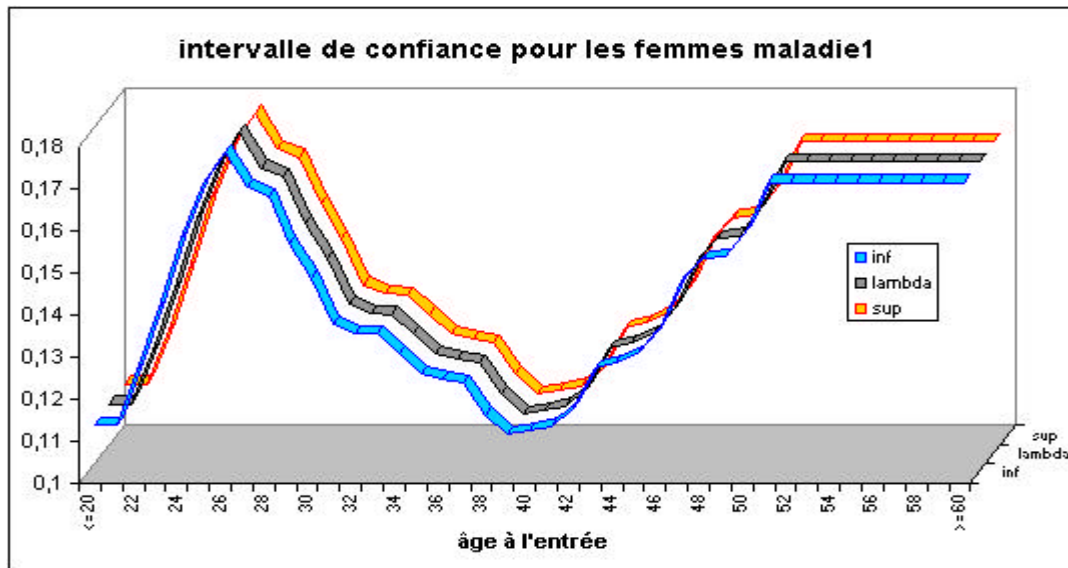
Le pic de sinistralité en maladie chez les femmes autour de 25 ans est peut-être lié à la première grossesse. En effet, même si la grossesse en elle-même n'est pas considérée comme une incapacité, toutes les complications qui peuvent en résulter sont garanties par l'assureur.

Section 6.3 : intervalles de confiance

Ici, nous avons choisi un niveau de confiance à 95%, d'où l'intervalle de confiance pour λ_x :

$$\left[\hat{I}_x - 1,96 * \sqrt{\frac{\hat{I}_x}{n \sum_{i=1}^n (b_i - a_i)}} ; \hat{I}_x + 1,96 * \sqrt{\frac{\hat{I}_x}{n \sum_{i=1}^n (b_i - a_i)}} \right]$$

Voici par exemple le graphe en 3D obtenu pour la table IJ femme maladie1 :



Si on a fait un graphe en 3D c'est parce que la largeur de l'intervalle de confiance est tellement faible qu'il est invisible sur des courbes plates. Cet intervalle de confiance est aussi extrêmement petit sur les autres tables, ce qui valide en fait nos résultats bruts.

Par manque de lisibilité nous n'exposerons pas les intervalles des autres tables.

Cette faible largeur s'explique en fait par la structure même de l'intervalle. En effet, la largeur de l'intervalle à l'âge x est donnée par :

$$L = 2 * 1,96 * \sqrt{\frac{\hat{I}_x}{n \sum_{i=1}^n (b_i - a_i)}} \approx \frac{4\sqrt{\hat{I}_x}}{n}$$

\hat{I}_x étant de l'ordre de 5%, on a $\sqrt{\hat{I}_x} \approx 20\%$

donc finalement la largeur de l'intervalle est de l'ordre de $1/n$ avec n le nombre d'assurés exposés à l'âge x .

Or par construction $n > 2\ 000$ d'où un intervalle de confiance qui ne dépassera pas $0,5\text{‰}$ en largeur.

En fait, on perçoit ici tout l'intérêt d'avoir des volumes suffisants puisque la largeur de l'intervalle de confiance et donc la fiabilité des résultats en dépend directement.

Malgré ces intervalles de confiance satisfaisants, nous ne pouvons pas nous contenter des tables brutes, il faut maintenant les lisser.

Chapitre 7 : Lissage des lois d'entrée

Dans cette partie, nous traiterons les estimations brutes précédentes afin de les rendre plus "régulières". Pour passer des premières estimations aux estimations lissées de l'entrée en incapacité des assurés d'AXA, nous comparerons deux méthodes de lissage, une paramétrique et une non paramétrique pour ensuite choisir la mieux adaptée à nos données.

Section 7.1 : Pourquoi faire un lissage?

On suppose que la loi d'entrée en incapacité sous-jacente de nos observations est régulière et continue. Ceci correspond à l'intuition et à l'observation de la progression de l'entrée en incapacité aux différents âges. De plus, en pratique, il faut que les tarifs (déterminés par la table d'entrée) progressent de manière régulière sans présenter de chutes ou de pics à certains âges. Les estimations brutes ne répondent pas à ces caractéristiques. En effet, les estimations brutes initiales ont été obtenues âge par âge indépendamment les unes des autres. Il faut donc procéder à un lissage pour réintroduire les relations qui existent entre les arrêts de travail à chaque âge et obtenir ainsi la régularité souhaitée.

Il existe plusieurs types de lissages qui peuvent se classer en deux grandes familles : les lissages paramétriques et les lissages non paramétriques. Nous avons choisi d'en étudier un de chaque type avant de choisir le lissage le mieux adapté à nos estimations.

Section 7.2 : Régression non paramétrique par la méthode des noyaux

Notations :

- soit $X = (x_1, x_2, \dots, x_n)$ un vecteur de n points. Dans notre étude, x_i représente l'âge à l'entrée en arrêt de travail.
- Soit $Y = (y_1, y_2, \dots, y_n)$ un vecteur de n points «bruts ». Ici, y_i représente l'estimation brute annuelle d'entrée en incapacité, pour chaque segmentation, cette variable dépend de l'âge à l'entrée, soit x_i .

Pour obtenir la courbe de régression de Y sur X par une méthode non paramétrique, nous allons estimer $f(x) = \text{Esp}(Y / X = x)$ connaissant $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ réalisation i.i.d de (X, Y) , et ceci à l'aide de l'estimateur de densité de Nadaraya-Watson :

$$\hat{f}_n(x) = \frac{\sum_{i=1}^n Y_i * K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

avec :

- $K(u) = \text{noyau d'Epanechnikov} = 0.75*(1-u^2)$
- $n = \text{nombre d'âges considérées}$
- $h = \text{bandwidth (déterminé par la suite)}$

Le choix de h va être effectué par validation croisée (CV : Cross Validation), c'est à dire que h minimise l'expression suivante :

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - f_{n;-i}(X_i))^2$$

avec :

$$\begin{aligned} \hat{f}_{n;-i}(X_i) &= \frac{\sum_{j=1; j \neq i}^n Y_j * K\left(\frac{X_j - X_i}{h}\right)}{\sum_{j=1; j \neq i}^n K\left(\frac{X_j - X_i}{h}\right)} \\ &= \frac{\sum_{j=1}^n Y_j * K\left(\frac{X_j - X_i}{h}\right) - Y_i * 0.75}{\sum_{j=1}^n K\left(\frac{X_j - X_i}{h}\right) - 0.75} \quad \text{car } K(0) = 0.75 \text{ par définition de } K(u) \end{aligned}$$

$CV(h)$ est une fonction paire, il suffit donc de faire varier h sur $]0, \max(X_i) - \min(X_i)[$ pour trouver le minimum de $CV(h)$ est donc le h optimal noté « h^{CV} ».

L'estimateur de régression par validation croisée s'écrit alors :

$$\hat{f}_n^{CV}(x) = \frac{\sum_{i=1}^n Y_i * K\left(\frac{X_i - x}{h^{CV}}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h^{CV}}\right)}$$

Nous allons maintenant étudier une méthode de lissage paramétrique.

Section 7.3 : Régression polynomiale sur l'espérance du nombre de sinistres

Généralement, les modèles paramétriques lissent des taux instantanés en faisant l'hypothèse qu'ils suivent une loi connue. La famille de fonctions habituellement utilisée pour un lissage paramétrique de taux instantanés est la famille de Weibull et plus particulièrement la fonction de la forme :

$\mu_t = a + b \cdot \exp(-ct)$ appelée fonction de Makeham.

Cependant, nous avons précédemment opté pour une modélisation de type IARD pour estimer l'entrée en incapacité il ne serait donc pas très judicieux pour le lissage de revenir à un modèle de taux d'entrée liés à une unique probabilité d'entrée en incapacité.

Le lissage paramétrique des taux instantanés n'est donc pas adapter dans le cas présent.

De ce fait, il vaut mieux envisager une régression polynomiale sur les estimations brutes du nombre espéré de sinistres.

La méthode de régression par polynôme est souvent utilisée pour sa facilité d'exécution et la qualité des résultats obtenus.

• Construction

- soit $X = (x_1, x_2, \dots, x_n)$ un vecteur de n points. Dans notre étude, x_i représente l'âge à l'entrée en incapacité.
- Soit $Y = (y_1, y_2, \dots, y_n)$ un vecteur de n points « bruts ». Ici, y_i représente l'espérance annuelle brute d'entrée en incapacité à l'âge x_i .

Nous souhaitons réaliser une régression polynomiale de Y sur X .

- soit le polynôme de régression de degré p : $\hat{y}_i = a_0 + a_1 * x_i + a_2 * x_i^2 + \dots + a_p * x_i^p$ pour $i=1, \dots, n$; avec \hat{y}_i la valeur de y_i lissée. X, \dots, X^p sont les vecteurs explicatifs de Y .

Les inconnues de cette méthode sont donc les coefficients a_0, a_1, \dots, a_p de ce polynôme. Pour les déterminer on écrit :

$$\hat{Y} - Y = \begin{bmatrix} \hat{y}_1 - y_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{y}_n - y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdot & \cdot & x_1^p & y_1 \\ 1 & x_2 & x_2^2 & \cdot & \cdot & x_2^p & y_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 & \cdot & \cdot & x_n^p & y_n \end{bmatrix} * \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_p \\ -1 \end{bmatrix} = X.B$$

On résout cette équation par la méthode des moindres carrés, c'est à dire que l'on cherche les coefficients a_0, a_1, \dots, a_p du polynôme qui minimisent la grandeur SSE (Sum Square Errors) ci-dessous :

$$SSE = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = (\hat{Y} - Y)' (\hat{Y} - Y) = B' . X' . X . B$$

Pour minimiser SSE, il suffit d'annuler ses p+1 dérivés partielles :

$$\frac{\partial(SSE)}{\partial a_k} = 0 \quad \text{pour } k=0,1,\dots,p$$

Remarque : dans notre étude X=(20, 21, ...,60) vecteur de 41 points.

• **Choix du degré du polynôme :**

Ce choix ne peut être fait a priori ; il est bien évident que plus le degré du polynôme sera élevé, plus la régression sera précise, en contre partie d'une perte de qualité du lissage. Pour autant, il est inutile d'augmenter le degré du polynôme si cela n'est pas justifié par la volonté d'améliorer la régularité d'une partie de la courbe brute. Pour cela, nous devons établir les critères qui vont permettre de choisir le degré optimal :

- Un critère visuel : on n'augmente le degré du polynôme que si cela améliore sensiblement la régularité de la courbe brute.
- Un critère statistique : le coefficient de détermination appelé communément $R^2 \in [0,1]$, permet d'exprimer le pourcentage de variance empirique de la variable Y expliqué par le modèle postulé. Plus le R^2 est proche de 1, meilleure est la modélisation. Cependant, d'un point de vue purement algébrique, il est possible de gonfler artificiellement la valeur de R^2 en ajoutant des variables explicatives et donc en augmentant le degré du polynôme. Pour tenir compte de cette perversion algébrique, nous allons plutôt nous intéresser au coefficient R^2 ajusté défini ci-après. On considèrera que la modélisation par un polynôme de degré k+1 est sensiblement meilleure que celle avec un degré k, si l'augmentation du R^2 ajusté est significative.

En conservant les notations précédentes, on peut dresser le tableau d'analyse de la variance :

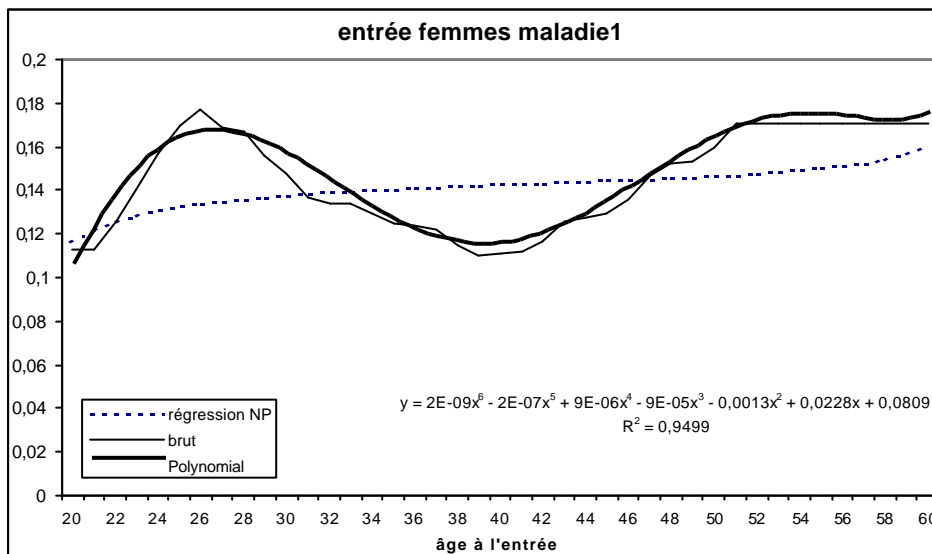
Source de variation	Somme des carrés	Degrés de liberté	Somme des carrés (moyenne)
Due au modèle	$\sum_{t=1}^n (\hat{y}_t - \bar{y})^2$	p	$\frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{p}$
Due aux erreurs	$\sum_{t=1}^n (\hat{e}_t)^2 = \sum_{t=1}^n (\hat{y}_t - y_t)^2$	n-p-1	$\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n - p - 1}$
total	$\sum_{t=1}^n (y_t - \bar{y})^2$	n-1	$\frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n - 1}$

Ainsi : $R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$ et le R^2 ajusté : $R_a^2 = 1 - \left(\frac{n-1}{n-p-1} \right) * (1 - R^2)$

Section 7.4 : Comparaison graphique et choix du lissage

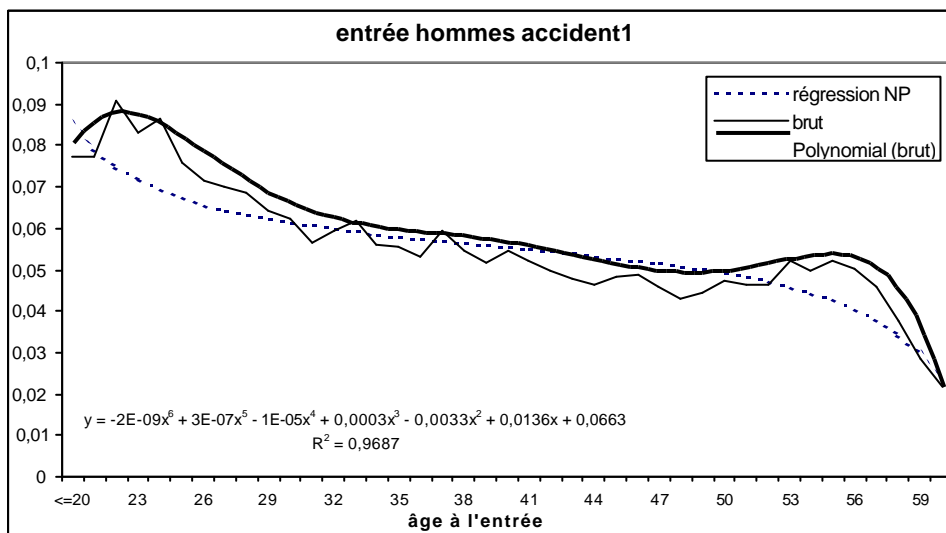
Le modèle d'estimation retenu exclu toute hypothèse de loi communément appliquée sur des taux de sortie (en décès ou en invalidité). Ainsi, le choix se réduit à deux modèles de lissage seulement. Etant donné que le premier a l'avantage d'être non paramétrique et que le second permet de jouer sur le degré du polynôme, il est difficile de les départager a priori. Nous les avons donc comparés graphiquement sur 2 segmentations représentatives avant de prendre notre décision.

Voici les 2 graphes obtenus pour les tables IJ des hommes accident1 et des femmes maladie1:



Valeur de R^2 ajusté en fonction de p le degré du polynôme

p	R2	R2 ajusté
2	0.39	0.36042
3	0.43	0.38627
4	0.9	0.88911
5	0.93	0.9216
6	0.95	0.94106



Valeur de R^2 ajusté en fonction de p le degré du polynôme

p	R2	R2 ajusté
2	0.83	0.82495
3	0.87	0.85503
4	0.92	0.90867
5	0.92	0.91074
6	0.97	0.96318

Au regard des graphiques, le lissage le mieux adapté est la régression polynomiale dont nous choisirons le degré suivant l'évolution du R^2 ajusté.

Après avoir élaboré les tables d'entrée en incapacité, nous avons construit celles du maintien. Cette élaboration est détaillée dans la partie suivante.

PARTIE V :

**ELABORATION DES TABLES DE
MAINTIEN EN INCAPACITE**

Partie V : Elaboration des tables de maintien en incapacité

L'arrêté du 28 mars 1996, qui institue à compter du 01/01/1997 des règles de provisionnement des garanties d'incapacité et d'invalidité, impose aux organismes assureurs d'utiliser les tables de maintien et de passage fournies en annexe. Il laisse cependant à l'assureur la possibilité de se servir de ses propres tables d'expériences pourvu qu'elles soient certifiées par un actuair indépendant et agréé à cet effet. Etant donné que les tables proposées par l'arrêté ont été élaborées sur une population d'assurés collectifs, elles ne sont pas représentatives du portefeuille d'assurés individuel. L'élaboration de tables d'expérience sur le maintien en incapacité propre au portefeuille d'assurés d'AXA permettrait donc de connaître le risque réel supporté par l'assureur qui pourrait ainsi, après certification, provisionner au plus juste ses sinistres incapacité.

Les tables de maintien en incapacité se présentent habituellement sous la forme d'un tableau à double entrée :

Nombre de périodes écoulées depuis l'arrêt de travail = ancienneté du sinistre

âge \ période		Nombre de périodes écoulées depuis l'arrêt de travail = ancienneté du sinistre												
		0	1	2	3	4	5	6	7	.	.	.	n-1	n
âge à l'arrêt de travail	30 ans	10000	8103	6239	.	.								
	31	10000	8082	6270	.									
	.	10000	.	.	.									
	.	10000	.	.	.									
	64	10000	9062	8342	.									
		l_0	l_1	l_2	.	.								l_n

où l_t représente pour chaque tranche d'âge le nombre de personnes restant en incapacité au bout de t périodes, avec comme base $l_0 = 10\ 000$ personnes sinistrées.

Pour élaborer cette table à deux dimensions (âge à l'arrêt de travail et ancienneté du sinistre) il faudra donc déterminer le nombre et la durée des périodes à envisager. Ce choix dépend du volume de données disponibles pour chaque durée de sinistre. Les tables réglementaires sont proposées avec un pas mensuel jusqu'à atteindre 3 ans, durée maximum d'incapacité définie par la Sécurité Sociale. Il faut signaler que nous possédons des observations journalières et des volumes de sinistres qui pourraient laisser envisager une table journalière.

Pour élaborer nos tables d'expérience, nous allons estimer la loi de maintien en incapacité des assurés d'AXA.

En terme de probabilité, cette loi correspond à la fonction de survie $S(t)$ qui représente la probabilité que la durée d'incapacité dépasse la durée t , d'où les notations suivantes :

- L'origine des temps est fixée à l'entrée en incapacité donc t désignera à la fois une durée et une date
- Y est la variable durée de maintien en incapacité
- $S(t) = P(Y \geq t)$ est la loi de maintien
- $l_t = l_0 * S(t)$: le nombre de personnes encore en incapacité en t , sur une base de l_0 sinistrés.
- $Q_t = P(Y \leq t+1 / Y > t)$ est le taux conditionnel de sortie d'incapacité avant $t+1$ sachant qu'on est en incapacité en t

En fait $Q_t = 1 - \frac{S(t+1)}{S(t)}$.

Chapitre 1 : Utilisation de données censurées

De même que pour l'étude de l'entrée en incapacité, nous sommes ici aussi confrontés au problème des observations censurées ou tronquées.

En effet, pour élaborer des tables d'expérience en maintien, il faudrait pouvoir observer le maintien de chaque sinistré dans sa totalité. Or, notre échantillon est limité à la période d'observation allant de septembre 1994 à février 2000.

De ce fait, il est possible que certains sinistrés ne soient pas observés sur la durée totale de leur incapacité.

Section 1.1 : Les troncatures

Dans le cas des assurés qui étaient déjà en arrêt de travail au 1^{er} septembre 1994, le début de leur incapacité n'étant pas observé, leur maintien est tronqué. Leur observation ne commence qu'à partir d'une certaine ancienneté en sinistre (non nulle).

Nous insistons sur le fait que dans notre étude, la franchise n'est pas considérée comme une troncature puisqu'en fait nous avons choisi de construire des tables de maintien au-delà de la franchise.

Section 1.2 : Les censures

Les assurés encore en incapacité le 29 février 2000 (fin de la période d'observation) vont être censurés à droite car on n'observe pas la durée totale de leur incapacité.

De même, l'échéance de la garantie ou du contrat lorsque l'assuré est encore en incapacité ainsi que toute limitation contractuelle de l'indemnisation seront considérées comme une censure puisque l'assuré sort de l'observation sans pour autant reprendre son activité.

Par contre, le décès sera considéré comme une sortie d'incapacité puisque celui-ci met fin à tout arrêt de travail. De toute façon, en cas de décès, il n'existe pas d'inconnue sur la durée réelle de l'incapacité, celle-ci est entièrement observée. Le décès est donc une vraie sortie d'incapacité et non une censure.

Toute résiliation de la garantie ou du contrat pendant une incapacité indemnisée étant peu probable, ce cas sera traité comme une vraie sortie et non une censure.

Par la suite, on appellera « sortie d'incapacité » la reprise d'activité ou le passage en invalidité ou le décès.

Chapitre 2 : Des tables de maintien adaptées au tarif et au provisionnement

Comme nous l'avons déjà signalé, l'élaboration de tables de maintien associée à celle des tables d'entrée permet de tarifier au plus juste les garanties complémentaires. Par conséquent, les tables de maintien doivent être en adéquation avec les tables d'entrée pour tarifier correctement. Ceci implique donc de construire trois types de tables :

- 1) le maintien en incapacité évalué sur des garanties IJ ou EXO sur une tête (table IJ/EXO)
- 2) le maintien en hospitalisation évalué sur des garanties IJH (table IJH)

Par ailleurs, on souhaite constituer des classes de risque homogène. Or d'après les statistiques descriptives sur la durée des sinistres, on remarque que généralement, l'existence d'une période d'hospitalisation pendant l'incapacité amplifie l'arrêt accident et réduit l'arrêt maladie. Pour pouvoir provisionner cette modification du risque lorsque l'on apprend que l'assuré est hospitalisé, il faudrait donc disposer d'une table de maintien en incapacité spécifique à l'hospitalisation (3^{ème} table), c'est à dire une table basée uniquement sur l'observation de la durée des arrêts comportant une période d'hospitalisation.

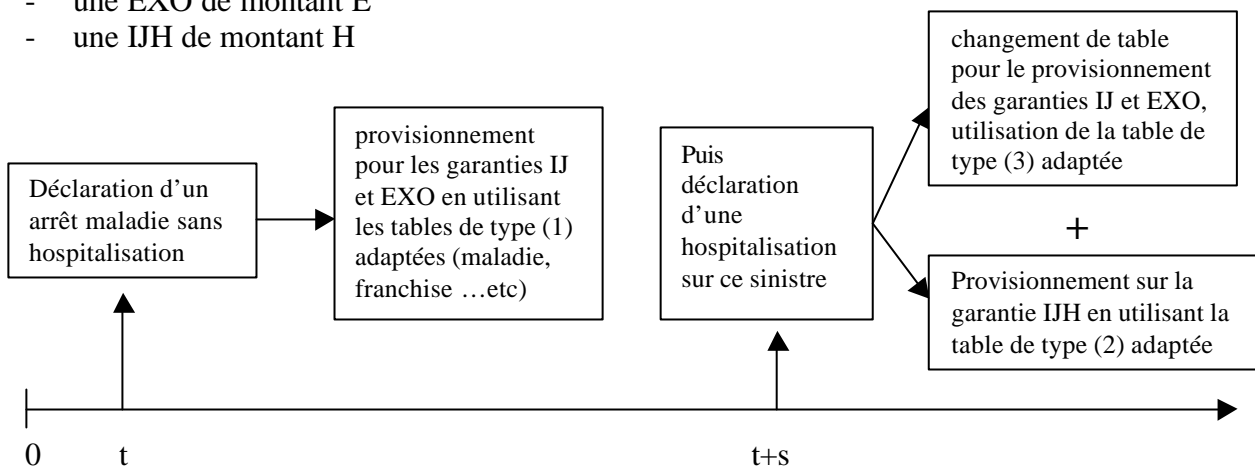
NB :

Attention à ne pas confondre cette table spéciale hospitalisation qui est une table de maintien **en incapacité** avec la table de maintien **en hospitalisation** (2) basée uniquement sur la durée de la période d'hospitalisation. En fait cette 3^{ème} table est construite à partir d'une sous population de la 1^{ère} table (table IJ/EXO). Ce sous-groupe contient uniquement les sinistres incapacité (sur une garantie IJ ou EXO) comportant une période d'hospitalisation.

Exemple de processus de provisionnement pour le sinistre suivant :

Garanties au contrat :

- une IJ de montant I
- une EXO de montant E
- une IJH de montant H



Chapitre 3 : Les estimateurs du maintien

Contrairement à l'invalidité, l'incapacité ne s'adapte pas forcément aux modélisations usuelles du décès. En effet, l'incapacité est limitée à 3 ans en durée et elle doit se produire pendant l'activité professionnelle soit entre 20 ans et 65 ans environ. Ces deux limites n'existent pas pour le décès. Par conséquent, nous avons choisi d'envisager uniquement des estimateurs non paramétriques pour éviter le risque d'une hypothèse de loi inadaptée.

Ces estimateurs sont peu nombreux et ceux qui tiennent compte des censures et troncatures sont encore plus rares.

Notre présentation se limitera donc à deux estimateurs :

- l'estimateur actuariel
- l'estimateur de Kaplan-Meier

Section 3.1 : Méthode actuarielle

3.1.1 : principe

Cette méthode s'appuie sur l'hypothèse (a) que les troncatures et les censures (concernant le maintien) se produisent au milieu de la période considérée.

En fait, on se ramène au calcul de l'estimateur actuariel du taux conditionnel de sortie Q_t défini précédemment comme la probabilité de sortir d'incapacité sur la période $]t, t+1]$ sachant qu'on est en incapacité en t .

3.1.2 : construction

Notations :

$SORT_t$: le nombre de sorties d'incapacité sur la période $]t, t+1]$

$CENS_t$: le nombre de censurés sur la période $]t, t+1]$

$TRON_t$: le nombre de tronqués sur la période $]t, t+1]$

OBS_t : la population observée en t

$SRIS_t$: la population sous risque (de sortie) sur la période $]t, t+1]$

Avec l'hypothèse (a), on obtient :

$$SRIS_t = OBS_t + \frac{TRON_t}{2} - \frac{CENS_t}{2}$$

Or Q_t est estimé trivialement par $\hat{Q}_t = \frac{SORT_t}{SRIS_t}$.

3.1.3 : Résultat

$$Q_t = \frac{SORT_t}{OBS_t + \frac{TR ON_t}{2} - \frac{CENS_t}{2}}$$

Par définition : $Q_t = 1 - \frac{S(t+1)}{S(t)}$ et $S(t) = \frac{l_t}{l_0}$.

Donc $l_{t+1} = l_t * (1 - Q_t)$ et l'on peut ainsi construire la table de maintien de proche en proche en partant de $l_0 = 10\ 000$.

Remarque :

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.

Section 3.2 : Méthode de Kaplan Meier

3.2.1 : Principe

Dans le calcul de l'estimateur de Kaplan Meier des taux d'entrée en incapacité nous avons introduit l'estimateur de la fonction de survie en état de capacité, cet estimateur peut s'appliquer pareillement à la fonction de maintien en état d'incapacité $S(t)$ en remplaçant l'événement « entrée en incapacité » par « la sortie d'incapacité ».

3.2.2 : Construction

On rappelle que l'origine des temps correspond à la date d'entrée en incapacité indemnisable soit : (la date de survenance du sinistre) + (la durée de la franchise).

De ce fait, toutes les durées correspondront aussi à des dates et inversement.

Notations :

- (Y_1, Y_2, \dots, Y_n) : le n-échantillon de la variable aléatoire Y durée de maintien en incapacité
- (C_1, C_2, \dots, C_n) : le n-échantillon des censures (à droite) correspondantes ($C_i \geq Y_i$ signifie qu'il n'y a pas de censure et $C_i < Y_i$ qu'il y a censure)
- $T_i = \text{Min}(Y_i, C_i)$
- $(T_1', T_2', \dots, T_n')$: l'échantillon des T_i ordonnés par ordre croissant
- $d(t)$: nombre de sortie d'incapacité (par reprise d'activité, passage en invalidité, décès ou fin de la garantie) à l'instant t
- $n(t)$: nombre d'incapables observables en t (et donc exposés au risque de sortie en t).

L'estimation de Kaplan-Meier (KM) obtenu précédemment (cf partie IV chapitre 3) est la suivante :

$$\hat{S}_{KM}(t) = \prod_{T_i' < t} \left(1 - \frac{d(T_i')}{n(T_i')} \right)$$

La fonction correspondante est une fonction en escalier qui part de 1 et décroît à chaque sortie d'incapacité observée.

En présence de troncatures, on calcule $n(T_i')$ en conséquence soit :

$$n(T_i') = n(T_{i-1}') - d(T_{i-1}') - c(T_i') + t(T_i')$$

Avec :

- Par convention, si la troncature, la censure et la sortie se produisent simultanément, on conviendra que l'entrée précède la censure qui précède elle-même la sortie.
- $c(T_i')$ le nombre de censures sur $]T_{i-1}', T_i']$ (T_i' inclus à cause de la convention)
- $t(T_i')$ le nombre de troncatures sur $]T_{i-1}', T_i']$ (T_i' inclus à cause de la convention)

3.2.3 : Résultat

$$\hat{S}_{KM}(t) = \prod_{T_i' < t} \left(1 - \frac{d(T_i')}{n(T_{i-1}') - d(T_{i-1}') - c(T_i') + t(T_i')} \right)$$

3.2.4 : Propriétés

En notant $\hat{S}(t)$ l'estimateur associé à l'estimation $\hat{S}_{KM}(t)$ on sait que :

- $\hat{S}(t)$ est l'unique estimateur cohérent de la loi de maintien en incapacité
- Il est maximum de vraisemblance, sans biais
- Pour $n \rightarrow +\infty$, $\hat{S}(t)$ converge presque sûrement vers $S(t)$ on a donc un estimateur correct
- Pour $n \rightarrow +\infty$, $\sqrt{n}[\hat{S}(t) - S(t)]$ est asymptotiquement gaussien

Variance :

La littérature présente généralement l'estimateur de Greenwood pour estimer la variance asymptotique de \hat{S} :

$$\widehat{Var\hat{S}}(t) = \hat{S}_{KM}^2(t) \sum_{T_i' \leq t} \frac{d(T_i')}{n(T_i')[n(T_i') - d(T_i')]}$$

Compte tenu de cette formule, en utilisant l'approximation gaussienne, il est possible de construire pour $S(t)$ des intervalles de confiance.

Inconvénient :

L'application pratique de cet estimateur suppose l'utilisation de la chronologie exacte des sorties d'incapacité sur le portefeuille ainsi que le nombre de censures et de troncatures entre deux mouvements. Cette hypothèse est très exigeante en terme de précision et de traitement des données, mais comme nous l'avons déjà signalé, nous avons à notre disposition des observations individuelles et journalières. Par conséquent, cette exigence est satisfaite.

Il faut savoir que lorsque les données disponibles ne sont pas satisfaisantes (en terme de volume ou de précision) pour permettre de construire l'estimateur de Kaplan Meier, il est possible d'utiliser l'estimateur qui a servi pour l'élaboration des tables fournies par le BCAC. Nous allons exposer très rapidement le principe et les résultats de cette méthode.

3.2.5 : Méthode du BCAC

Cette méthode présentée par Roberto WOLFRUM (1992), propose une version discrétisée de l'estimateur de KM. Cet estimateur fut en fait introduit par TURNBULL en 1976.

Le principe :

Au lieu de raisonner en temps continu et d'avoir la valeur de la fonction de survie à n'importe quelle date, on fait une estimation par période. Par exemple dans le cas de la table réglementaire, la période retenue est le mois.

La construction :

On considère m périodes ($m=36$ pour une période mensuelle), et on cherche à estimer $s=(s_1, s_2, \dots, s_m)$ par le maximum de vraisemblance, s_i étant la probabilité que l'assuré sorte d'incapacité sur la $i^{\text{ème}}$ période.

On obtient ensuite la valeur de la fonction de survie sur chaque période j par la formule :

$$S(j) = 1 - \sum_{i=1}^j s_i$$

Cependant, la solution des équations du maximum de vraisemblance n'est pas calculable algébriquement, il faut donc appliquer un algorithme itératif pour approcher la solution théorique. Après simplifications, on obtient la forme opérationnelle de l'estimateur.

La forme opérationnelle de l'estimateur :

Notations :

➤ *Pour une période i quelconque :*

- n_i : nombre d'individus sortis (de l'incapacité) à la $i^{\text{ème}}$ période ;
- λ_i : nombre d'individus censurés de la période i ;
- B_i : nombre d'individus tronqués de la période i ;

N_0 désigne le nombre d'individus observés depuis leur date d'entrée en incapacité.

Remarque :

Pour avoir une vraie distribution de probabilité il faut créer artificiellement une dernière

période $(m+1)$ telle que $S(m+1)=0$ soit $s_{m+1} = 1 - \sum_{i=1}^m s_i$.

➤ *Pour la période $m+1$ créée artificiellement :*

$$n_{m+1} = B_{m+1} = \lambda_{m+1} = 0$$

Finalement, on obtient l'algorithme de calcul suivant :

1) initialisation de la distribution de probabilité par exemple pour la distribution uniforme on prend : $s_j^0 = \frac{1}{m+1}$, pour $j=1, \dots, m+1$

2) et pour l'itération $h+1$: $s_j^{h+1} = \frac{\left(n_j + \frac{I_1}{S(1)} + \dots + \frac{I_j}{S(j)} + \dots + \frac{B_m}{S(m)} \right) s_j^h}{N_0 + \frac{B_1}{S(1)} + \frac{B_2}{S(2)} + \dots + \frac{B_m}{S(m)}}$, pour $j=1, \dots, m+1$

Pour obtenir la solution $s=(s_1, s_2, \dots, s_{m+1})$ on applique l'algorithme itératif jusqu'à ce que l'amélioration soit minime selon un critère à définir.

L'interprétation :

Le numérateur représente toutes les sorties, y compris celles qui sont estimées, de la période j . Le dénominateur représente tous les individus (estimation de ceux qui sont sortis avant d'avoir pu être observés, effectivement sortis pendant la fenêtre d'observation et estimation de ceux qui sortiront après la fin de la période d'observation) qui prennent part ou qui «auraient pu prendre part, si on avait pu les observer », à l'étude.

L'utilisation :

Cette formule est facile d'utilisation puisqu'il suffit de classer les données dans un tableau du type suivant :

Périodes	Sorties	Censures	Troncatures
1	n_1	λ_1	B_1
.	.	.	.
.	.	.	.
j	n_j	λ_j	B_j
.	.	.	.
.	.	.	.
m	n_m	λ_m	B_m
m+1	0	0	0

Section 3.3 : choix de l'estimateur de Kaplan Meier

Etant donné que l'estimateur actuariel est trop simplificateur et peu précis, il reste à choisir entre l'estimateur de Kaplan Meier et celui du BCAC, sachant que ce dernier n'est en fait qu'une version discrétisée de celui de KM.

Ces deux estimateurs sont maximum de vraisemblance cependant, l'estimateur de KM est plus proche des données individuelles dont nous disposons et donc plus précis.

Finalement, même si sa mise en œuvre semble a priori un peu plus complexe lorsque le volume de données est important, nous avons choisi la méthode de Kaplan Meier pour estimer la loi de maintien en incapacité de notre portefeuille.

Chapitre 4 : La segmentation des tables : homogénéité de la sinistralité

Pour utiliser nos tables de maintien dans la tarification et le provisionnement des garanties complémentaires en incapacité , il faut tenir compte de certaines contraintes pratiques et commerciales suivantes :

Tout d'abord, il faut construire 3 types de tables :

- les tables de maintien en incapacité liées aux garanties IJ et EXO
- les tables de maintien en incapacité en présence d'une période d'hospitalisation
- les tables de maintien en hospitalisation (liées aux garanties IJH)

Nos résultats doivent en même temps traduire au mieux le risque supporté par l'assureur mais aussi être applicables aux futurs assurés. Cette double contrainte implique de tenir compte non seulement des variables discriminantes mais aussi des variables tarifaires traditionnelles pour segmenter correctement le risque étudié. Nous avons déjà listé ces variables dans l'étude de la segmentation des tables d'entrée en incapacité, nous allons maintenant étudier leur influence sur le maintien en incapacité.

Les variables analysées sont les suivantes :

- le sexe
- la franchise
- l'ancienneté du contrat
- la cause du sinistre
- l'âge

De toute façon, avoir une estimation par âge fait partie des objectifs définis pour la construction des tables. En fait l'âge n'est pas une segmentation mais plutôt une dimension de la table.

De plus, nous avons étudié le pouvoir discriminant des variables en même temps sur l'entrée et le maintien en incapacité. Or pour l'entrée en incapacité nous avons décidé que la segmentation suivant le sexe de l'assuré servirait de base pour l'étude des autres variables. De ce fait, pour le maintien, le sexe constituera aussi la segmentation par défaut.

Pour les variables restantes : la franchise, l'ancienneté du contrat et la cause du sinistre, nous avons étudié séparément leur influence sur le maintien en calculant pour chaque segmentation et chaque type de table la durée moyenne des sinistres.

Nous allons maintenant analyser les résultats obtenus pour chaque type de table.

Section 4.1 : Les tables de type IJ

On constate que le sexe et l'ancienneté n'interviennent pas ou peu sur le maintien en incapacité.

Nous avons expliqué lors de l'entrée en incapacité pourquoi l'ancienneté n'était pas discriminante, il en est de même pour le maintien.

Ici, le maintien des hommes et des femmes ne se distingue pas vraiment. En effet, si l'on regarde par exemple à travers l'étude de la variable cause, à éléments comparables c'est à dire à cause et âge identiques, on remarque que l'écart de durée moyenne dépassent rarement 15 jours. Par exemple, pour le maintien en maladie, le détail des résultats est présenté dans le tableau suivant :

toutes les valeurs sont en jours sauf pour la différence relative

âge	<30	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	>50	total
homme	73	93	84	88	99	106	107	105	117	95	116	128	117	127	160	147	160	162	186	188	166	182	191	133
femme	86	88	88	92	99	109	115	120	123	147	132	133	145	154	139	131	150	163	143	159	155	153	187	117
dif F-H	12	-5	4	5	0	3	8	15	6	52	16	5	28	27	-21	-17	-9	1	-43	-29	-11	-29	-4	-16
dif relative	14%	-6%	4%	5%	0%	3%	7%	12%	5%	35%	12%	4%	19%	18%	-15%	-13%	-6%	1%	-30%	-18%	-7%	-19%	-2%	-13%
différence absolue moyenne :	0																							
différence relative moyenne :	1%																							

En fait, d'après nos résultats, les hommes et les femmes n'ont pas le même comportement face au risque d'entrée en incapacité mais une fois sinistrés, ce qui joue le plus sur la durée du sinistre, c'est la cause du sinistre et non le sexe de l'assuré. Pourtant, certains pourraient penser que suite à une atteinte corporelle les femmes sont a priori moins résistantes que les hommes. En fait, si leur maintien est sensiblement identique à celui des hommes c'est peut-être que deux phénomènes se compensent : les femmes seraient physiquement plus fragiles mais on peut penser qu'elles prennent mieux soin d'elles en cas d'atteinte corporelle et donc améliorent leur guérison.

Comme supposé à priori, l'âge, la cause et la franchise sont effectivement des variables discriminantes.

Pour les franchises, il est évident que plus la franchise est importante, plus les sinistres qui dépassent ces franchises risquent d'être graves et donc longs. Cependant, pour atteindre des volumes suffisants, il est quand même préférable de regrouper certaines franchises avec une sinistralité proche.

NB : La franchise 0/0 ne correspond pas à une franchise nulle mais à une franchise relative 90/90 alors que toutes les autres franchises sont absolues. Les sinistres déclarés sur cette franchise ont donc dépassé 90 jours, ceci explique pourquoi la sinistralité sur la franchise 0/0 est proche de celle sur les franchises importantes (90/90).

Section 4.2 : Les tables de type IJ avec hospitalisation

Les remarques suivantes se basent à la fois sur les résultats de la partie III concernant la durée des sinistres et sur l'étude de l'influence des variables.

Nous avons étudié l'influence des variables suivantes :

- la cause du sinistre
- le sexe de l'assuré
- l'âge à l'entrée en incapacité

Pour analyser l'influence de la variable cause, voici le tableau des durées moyennes d'incapacité (en jours) exposé dans la partie III :

	sans hospitalisation		avec hospitalisation	
	accident	maladie	accident	maladie
homme	60	158	122	123
femme	83	125	154	117

On remarque que sur les sinistres accidents, le maintien moyen augmente de 2 mois environ en cas d'hospitalisation. On peut l'expliquer par le fait que lors d'un accident, si l'assuré est hospitalisé, c'est la preuve d'une certaine gravité.

Au contraire, on voit que l'arrêt maladie est en moyenne réduit d'un mois lorsqu'il y a hospitalisation. Pour l'expliquer, on peut dire qu'en cas de maladie, l'hospitalisation correspondrait plutôt à une opération puisqu'il n'y a hospitalisation au sens de l'assureur qu'au-delà de 48 heures dans un établissement médical, et de ce fait l'hospitalisation pour opération serait synonyme de guérison.

La cause de l'incapacité joue donc ici un rôle prépondérant puisqu'elle détermine l'impact de l'hospitalisation sur le maintien.

En ce qui concerne le sexe de l'assuré, de même que pour les tables IJ, il n'intervient pas réellement dans la sinistralité. En effet, la durée moyenne des sinistres avec hospitalisation est similaire chez les hommes et chez les femmes (121 jours en moyenne pour les deux).

Par conséquent, nous allons construire des tables par âge segmentées uniquement suivant la cause du sinistre : accident ou maladie.

Section 4.3 : Les tables de type IJH

Ici, nous nous intéressons au maintien en hospitalisation étudié à travers les garanties IJH.

Les variables analysées sont :

- la cause du sinistre
- le sexe de l'assuré
- l'ancienneté du contrat
- l'âge à l'entrée en incapacité

Nous n'avons pas étudié la franchise puisque sur toutes les garanties IJH la franchise est nulle. Il existe simplement une franchise implicite de 2 jours due à la définition de l'hospitalisation.

De même que pour les tables IJ, l'ancienneté et le sexe n'interviennent pas dans la sinistralité.

Par contre, comme nous venons de l'expliquer ci dessus, la cause du sinistre est la principale variable discriminante en cas d'hospitalisation et l'âge aggrave le maintien.

Finalement pour la segmentation nous ne retiendrons que la cause du sinistre et la table sera construite par âge.

Dans la section suivante un tableau résume l'ensemble des segmentations retenues.

Section 4.4 : Maintien moyen et arbre des segmentations

Tables IJ	ACCIDENT			MALADIE				
franchises	15/15 et 30/7	30/30 60/30 90/7 0/0 60/60 et 90/90			15/15 et 30/7	30/30 60/30 90/7 0/0 60/60 et 90/90		
maintien moyen	71 jours	118 jours			104 jours	160 jours		

Tables IJ quand hospi	ACCIDENT	MALADIE	Tables IJH	ACCIDENT	MALADIE
maintien moyen	126 jours	117 jours	maintien moyen	13 jours	12 jours

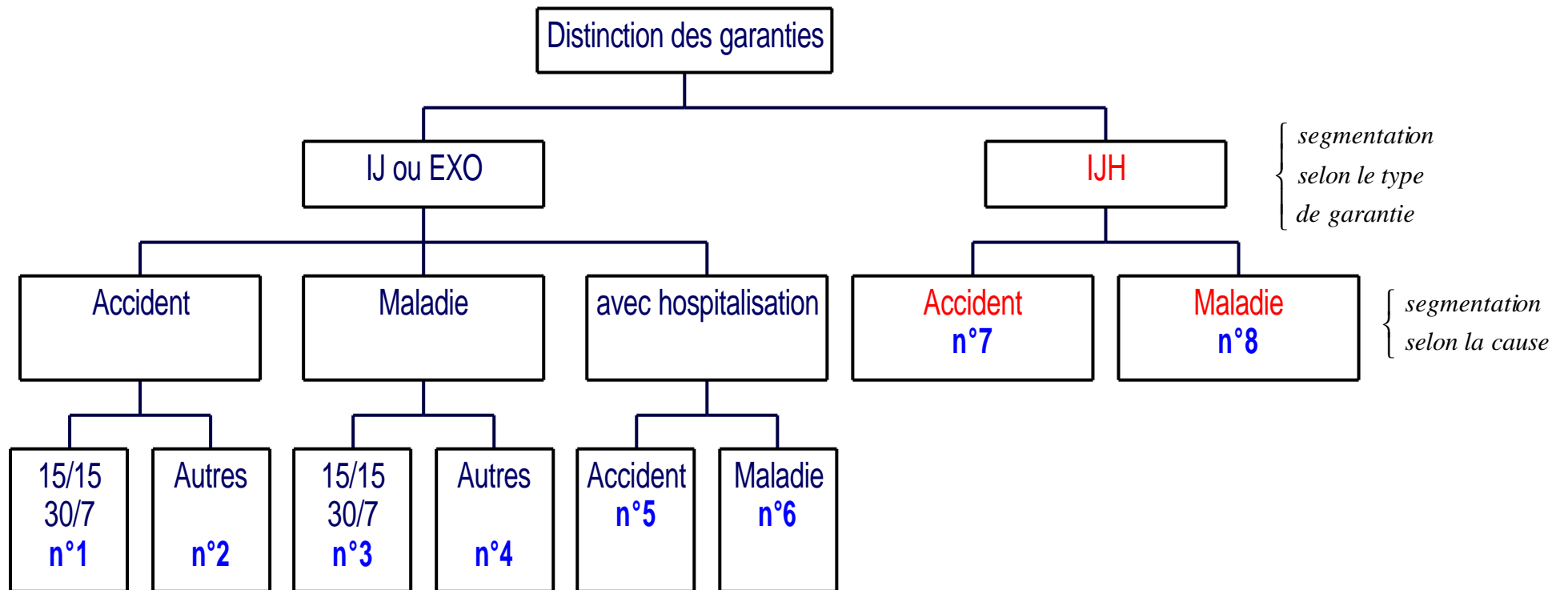
Au total, 8 tables de maintien ont été élaborées, en tenant compte de l'âge à l'entrée en incapacité.

Dans les chapitres suivants, pour alléger l'écriture, nous appellerons les 4 tables IJ précédentes de la manière suivante :

- « accident1 » pour les accidents avec les franchises 15/15 et 30/7
- « accident2 » pour les accidents avec les autres franchises
- « maladie1 » pour les maladies avec les franchises 15/15 et 30/7
- « maladie2 » pour les maladies avec les autres franchises

L'arbre de segmentation ci-après permet une meilleure visualisation des segmentations conduisant aux 8 tables de maintien en incapacité.

les tables de maintien



segmentation selon le couple de franchise *table spéciale hospitalisation*

Chapitre 5 : Les contraintes de volume

Nous venons de former des groupes de sinistralité homogène suivant certaines variables discriminantes (sexe, cause et franchise). Maintenant, nous désirons obtenir pour chacune de ces segmentations une table de maintien déclinée sur tous les âges.

Or, pour avoir des résultats fiables et robustes, il faut faire en sorte d'avoir pour chaque âge un volume de personnes exposées au risque de sortie d'incapacité suffisant.

Comme précédemment dans le cas de l'entrée en incapacité, ceci implique de regrouper certains âges en classes d'âges, et d'obtenir ainsi des estimations brutes par «paliers». Mais pour un meilleur lissage, nous préférons calculer préalablement des estimations brutes par âge.

Pour ce faire nous disposons des deux méthodes exposées dans le chapitre 5 de la 4^{ème} partie :

- l'interpolation à jonctions souples
- le glissement

Pour homogénéiser les méthodes et faciliter la mise en œuvre pratique, nous avons choisi d'effectuer ici aussi un glissement sur les âges pour atteindre le volume souhaité.

Il reste maintenant à déterminer ce volume minimum. Il n'existe pas vraiment de théorie donnant des chiffres de référence. Après avoir interrogé plusieurs personnes expérimentées dans le domaine des certifications de tables d'incapacité, nous avons fixé le volume minimum à 1000 sinistrés par tranche d'âge pour chaque segmentation.

Chapitre 6 : les tables brutes

Après avoir vérifié la stabilité de la loi de maintien dans le temps, ce chapitre exposera les principaux résultats concernant les tables brutes de maintien et leurs intervalles de confiance. De plus, pour mettre en perspective nos résultats, nous les comparerons à la table réglementaire.

Section 6.1 : La stabilité dans le temps

Allant de septembre 1994 à février 2000, notre période d'observation dure 5 ans et demi. Lors du choix de cette période, nous avons signalé que cette taille était raisonnable pour estimer le maintien (environ 2 cycles entiers d'incapacité). Pour la fiabilité de nos tables, il est cependant nécessaire de vérifier que le maintien en incapacité de nos assurés est resté stable sur cette période.

Pour ce faire, nous avons comparé la sinistralité en début de période avec celle en fin de période. Nous avons donc scindé notre échantillon de sinistres en 2 échantillons par rapport à une date de référence telle que la répartition sur les 2 échantillons soit équilibrée. Cette date correspond à peu près au milieu de la période d'observation soit le 1^{er} janvier 1997.

Donc tous les sinistres survenus avant le 1^{er} janvier 1997 appartiennent au 1^{er} échantillon et tous les autres appartiennent au 2^{ème}.

Mais pour avoir des échantillons comparables, il faut les retraiter :

- d'abord, il faut enlever les sinistres censurés. En effet, ces derniers se trouvant en majorité dans le 2^{ème} échantillon, ils biaiserait le maintien en incapacité dans celui-ci.
- de plus, il faut supprimer l'impact du vieillissement de la population sur le maintien. Nous allons donc segmenter chacun des échantillons suivant l'âge à l'entrée classé par tranche de 5 ans.

Pour réaliser la comparaison des échantillons, nous avons effectué à classe d'âge égale un test d'homogénéité du χ^2 .

Comme on veut étudier l'évolution du maintien, on va dans chaque échantillon classer les sinistres suivant leur durée en mois. Mais pour avoir des effectifs satisfaisant le critère de Cochran, au-delà de 7 mois, on constituera des classes annuelles pour la durée des sinistres.

Par exemple pour la tranche d'âge 30-35 ans, on obtient le tableau suivant :

30-35 ans		critère cochran						
durée	éch1	éch2	pi1	pi2	n1*pi	n2*pi	pi	(pi1-pi2) ² / pi
moins de 1 mois	1903	2199	0.392	0.420	1971.529	2130.471	0.407	0.002
de 1 à 2 mois	1048	1205	0.216	0.230	1082.851	1170.149	0.223	0.001
de 2 à 3 mois	549	619	0.113	0.118	561.372	606.628	0.116	0.000
de 3 à 4 mois	315	368	0.065	0.070	328.268	354.732	0.068	0.000
de 4 à 5 mois	262	244	0.054	0.047	243.197	262.803	0.050	0.001
de 5 à 6 mois	169	171	0.035	0.033	163.413	176.587	0.034	0.000
de 6 à 7 mois	123	105	0.025	0.020	109.583	118.417	0.023	0.001
de 7 mois à 1 an	380	282	0.078	0.054	318.175	343.825	0.066	0.009
de 1 à 2 ans	58	42	0.012	0.008	48.063	51.937	0.010	0.002
de 2 à 3 ans	43	6	0.009	0.001	23.551	25.449	0.005	0.012
total	4850	5241						
		$d^2 = 0.029$		p-value > 99.5%				

où :

- n_1 (respectivement n_2) est l'effectif de sinistre dans l'échantillon 1 (respectivement 2)
- p_{i1} (respectivement p_{i2}) est la proportion de sinistre dans la i^{me} classe de durée de sinistre pour l'échantillon 1 (respectivement 2)
- p_i est la proportion de sinistre dans la i^{me} classe de durée de sinistre si on regroupe les 2 échantillons.

Le critère de Cochran est bien satisfait puisque tous les $n_1 p_i$ et $n_2 p_i$ sont largement supérieurs à 1.

Si on définit r comme le nombre de classes de durée de sinistre (ici $r = 10$) alors la valeur de d^2 est donnée par :

$$d^2 = \sum_{i=1}^r \frac{(p_{i1} - p_{i2})^2}{p_i}$$

d^2 étant la valeur observée de D^2 qui converge en loi vers un χ^2_{r-1} .

On cherche à tester l'hypothèse suivante :

H_0 : La loi de maintien est stable sur notre période d'observation.

On rejettera H_0 lorsque la valeur de d^2 est trop grande. En fait on calcule la p-value $\hat{\alpha}$ (ou seuil critique) définie par :

$$\hat{\alpha} = P(\chi^2_{r-1} > d^2)$$

Si cette p-value est trop petite (<1%), on rejette H_0 .

Par exemple, le tableau précédent indique que $d^2 = 0,029$ et comme $r = 10$ notre p-value est donnée par : $\hat{\alpha} = P(\chi^2_9 > 0.029)$.

Or d'après les tables statistiques cette probabilité est strictement supérieure à 99,5%.

Ceci concerne la p-value du test sur la tranche d'âge 30-35 ans mais il en est de même sur les autres tranches d'âges.

Nous pouvons ainsi conclure à la stabilité du maintien en incapacité sur la période observée.

Cette condition essentielle étant remplie, nous allons maintenant exposer nos tables brutes.

Section 6.2 : les résultats bruts

6.2.1 : les ordres de grandeur des résultats

Nous présenterons ici les I_t , c'est à dire, pour un âge fixé, la loi de maintien en fonction du temps pour 10 000 entrées en incapacité à cet âge là.

En raison de leur double dimension et de leur caractère journalier, les tables obtenues sont difficilement visualisables dans leur totalité.

Par conséquent, pour évaluer les ordres de grandeur de nos résultats, nous allons considérer simplement deux âges d'entrée : 30 et 50 ans, en nous limitant aux tables IJ. De plus les anciennetés significatives retenues sont les suivantes : 0, 1 mois, 3mois, 6 mois, 1 an et 2 ans.

voici les ordres de grandeur obtenus :

	accident 1		accident 2		maladie 1		maladie 2	
	30 ans	50 ans	30 ans	50 ans	30 ans	50 ans	30 ans	50 ans
I_0	10 000	10 000	10 000	10 000	10 000	10 000	10 000	10 000
$I_{1 \text{ mois}}$	4 823	5 747	6 237	7 610	6 071	6 962	7 143	7 817
$I_{3 \text{ mois}}$	1 359	2 208	2 604	4 019	2 688	3 452	3 226	5 228
$I_{6 \text{ mois}}$	482	946	834	2 032	738	1 850	918	3 167
$I_{1 \text{ an}}$	122	361	156	479	161	750	152	840
$I_{13 \text{ mois}}$	32	168	141	425	63	312	129	718
$I_{2 \text{ ans}}$	14	60	34	80	16	71	4	258

Nous constatons que la loi de maintien diminue très rapidement dans les 6 premiers mois :

- Sur 10 000 sinistrés à 30 ans, il n'en reste plus que 500 à 1 000 au bout de 6 mois (selon la cause du sinistre et le type de franchise).
- Sur 10 000 sinistrés à 50 ans, il n'en reste plus que 1 000 à 3 000 au bout de 6 mois (selon la cause du sinistre et le type de franchise).

De plus, très peu de sinistres se prolongent au-delà d'1 an d'incapacité :

- de 30 à 140 à 30 ans
- de 170 à 700 à 50 ans

Bien sûr, le maintien croit avec la franchise. En effet, à ancienneté identique, il reste toujours plus de sinistrés sur les tables de type 2 où la franchise est plus importante.

Enfin, à valeurs comparables, le maintien en maladie est toujours plus fort que le maintien pour cause d'accident.

6.2.2 : Etude des taux conditionnels de maintien mensuel

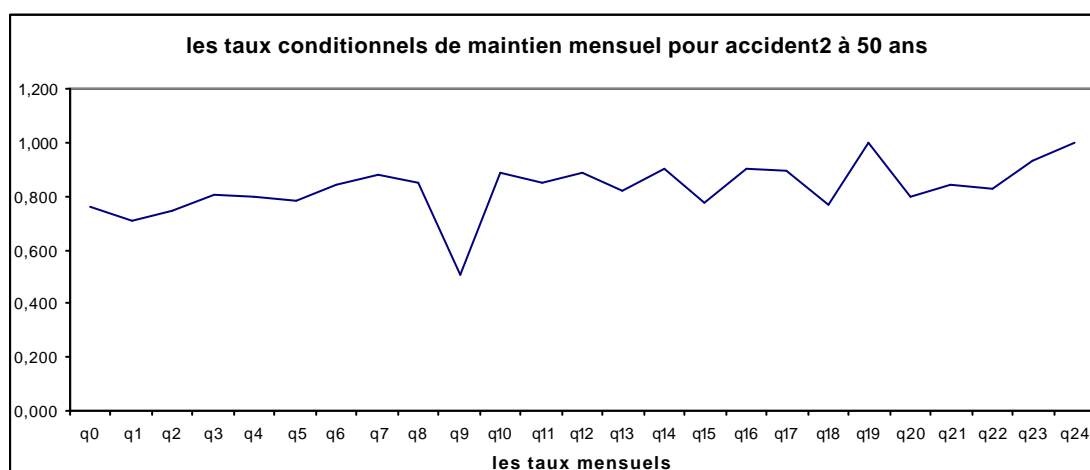
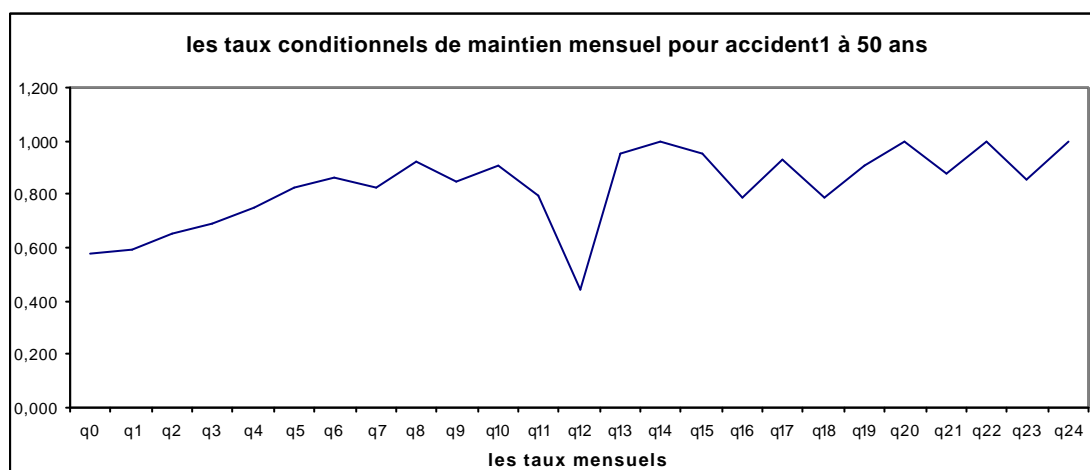
Nous nous intéressons ici aux taux conditionnels de maintien mensuel q_x avec :

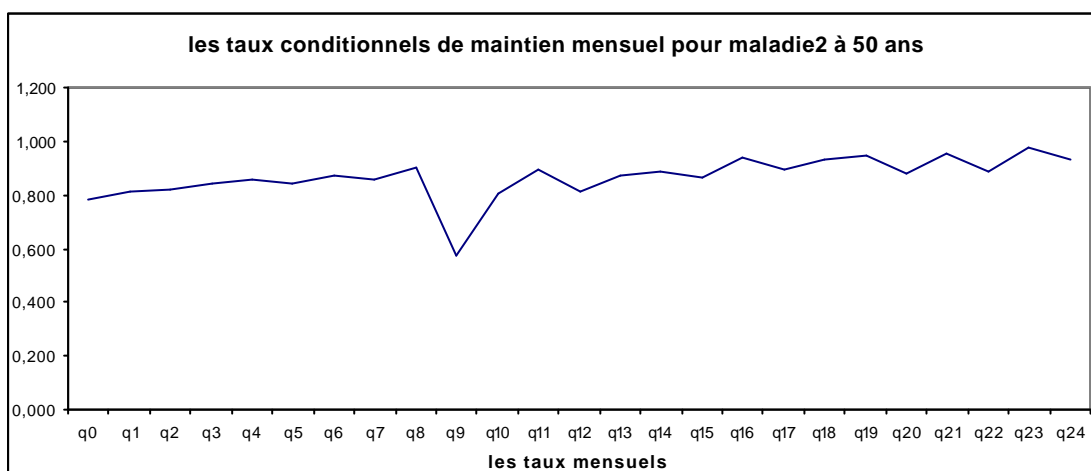
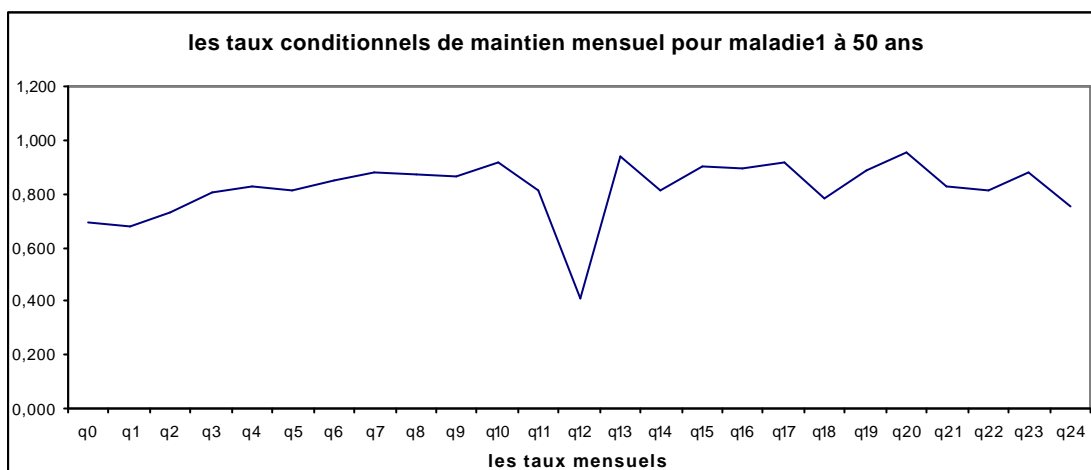
- x qui représente l'âge à l'entrée en incapacité.
- les q_x qui représentent la probabilité d'être encore en incapacité le mois suivant sachant qu'on est déjà en incapacité, par exemple le taux mensuel à l'ancienneté 3 mois est défini par :

$$q_x^{3\text{mois}} = \frac{l_x^{4\text{mois}}}{l_x^{3\text{mois}}}$$

Pour des raisons de visibilité, nous limiterons encore notre étude aux tables IJ pour un âge de 50 ans à l'entrée en incapacité.

Les graphes suivants présentent pour chaque table l'évolution de $q_{50\text{ans}}$ en fonction de l'ancienneté atteinte :





On constate globalement que les taux de maintien mensuels sont croissants jusqu'à atteindre quasiment la valeur 1 à 2 ans d'ancienneté. Ceci signifie que plus on a d'ancienneté en sinistre plus on a de chance de rester en incapacité le mois suivant.

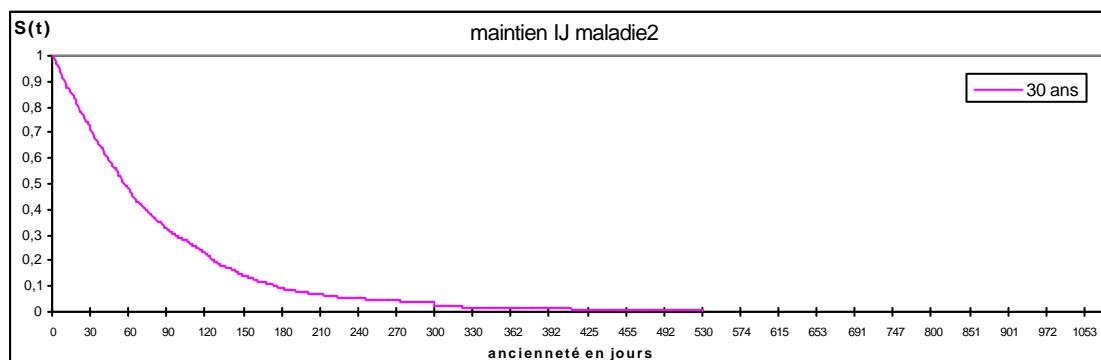
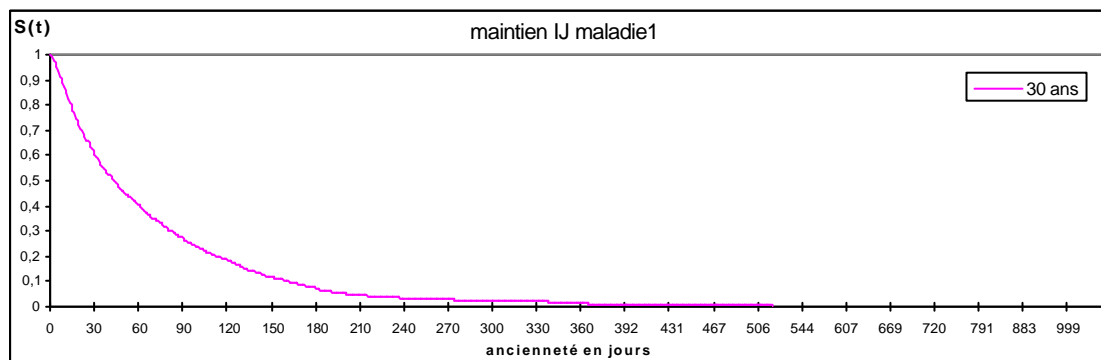
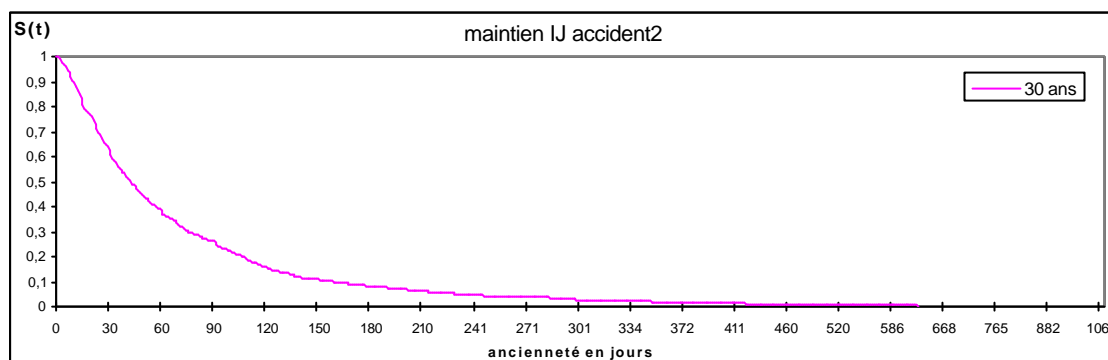
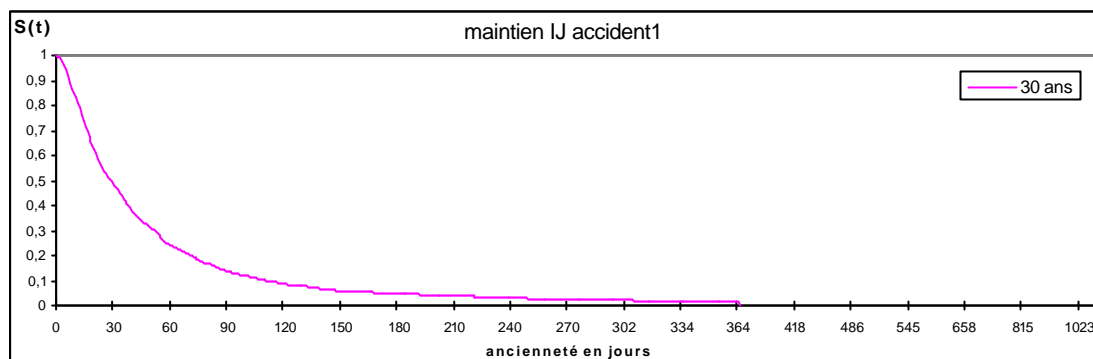
Selon les tables, on a entre 60 et 80% de chance de rester au moins un mois en incapacité en entrant à 50 ans (c'est le taux q_0).

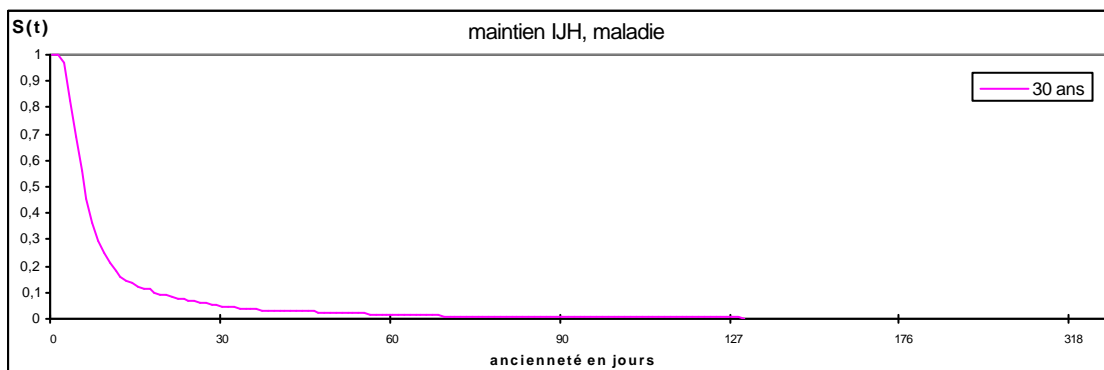
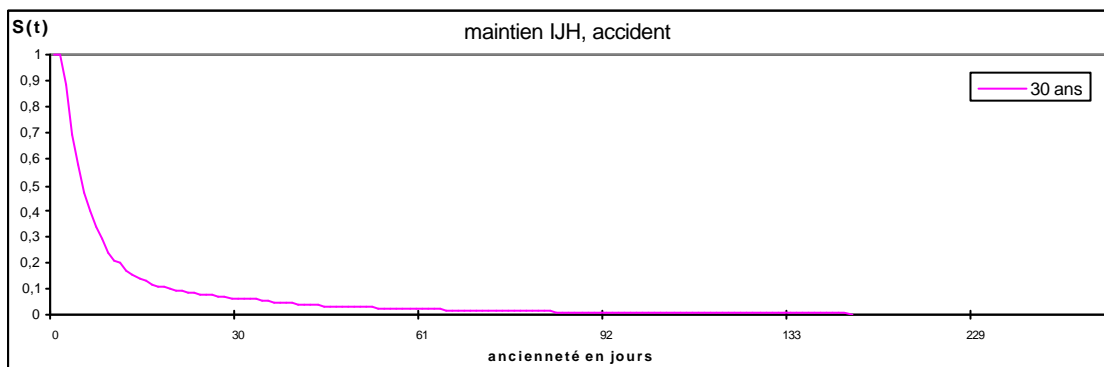
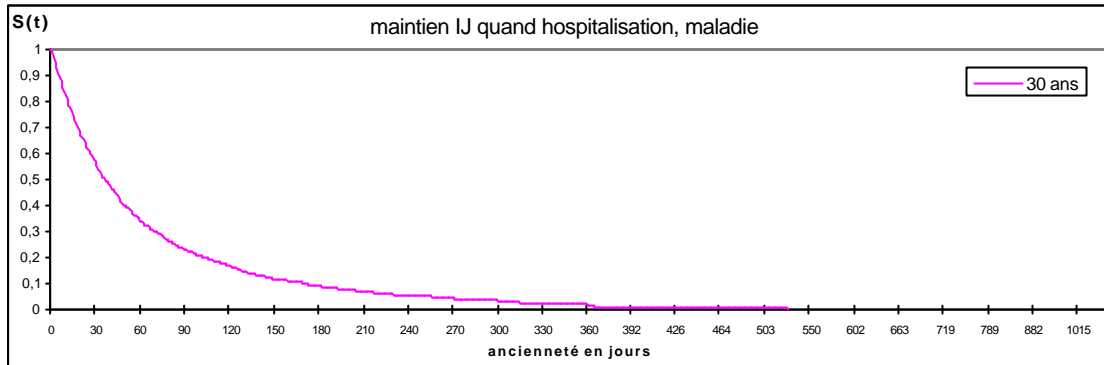
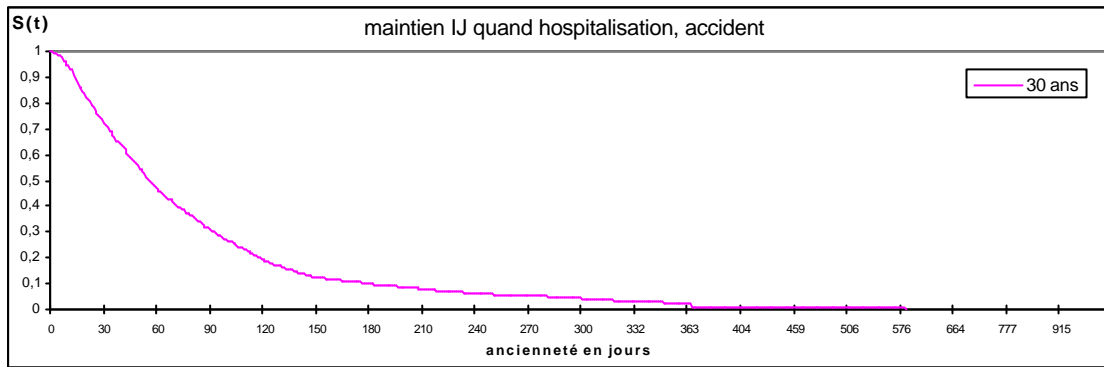
Par ailleurs, on constate une chute importante du taux à 1 an pour les tables accident1 et maladie1 et du taux à 9 mois pour les autres. Ces creux traduisent une sortie massive d'incapacité à ces dates là (vraies sorties ou censures). Il s'ensuit un certain chaos dans les taux supérieurs (de 1 à 2 ans). En effet, un faible volume de sinistrés restant après 1 an peut compromettre la régularité et la fiabilité de la loi au-delà d'1 an. Des mesures appropriées ont été prises pour palier à ce problème, elles seront exposées dans la section 6.4.

Nous allons maintenant donner l'allure générale des courbes de maintien suivant l'ancienneté.

6.2.3 : allure générale des courbes de maintien

Etant donné nos observations journalières, la loi de maintien estimée par KM est très précise. Nous ne présenterons ici que les tables de maintien pour un âge à l'entrée de 30 ans puisque d'après les volumes, c'est l'âge le plus représentatif en nombre de sinistres.





Commentaires :

Sur toutes les tables, la loi de maintien décroît rapidement au début puis se stabilise un peu jusqu'à une chute vers 1 an après laquelle elle devient pratiquement nulle.

Le redressement des courbes traduit le fait que plus on a d'ancienneté en incapacité plus on a de chance d'y rester.

Il se produit plus ou moins tôt selon les tables :

- au bout de 3 mois d'ancienneté pour les tables IJ accident
- au bout de 4 ou 5 mois d'ancienneté pour les tables IJ maladie ou IJ quand il y a hospitalisation.
- Au bout de 2 semaines seulement pour les tables IJH.

Le maintien des tables IJH est beaucoup plus court ce qui paraît logique puisqu'il représente uniquement le maintien en hospitalisation et non le maintien total en incapacité.

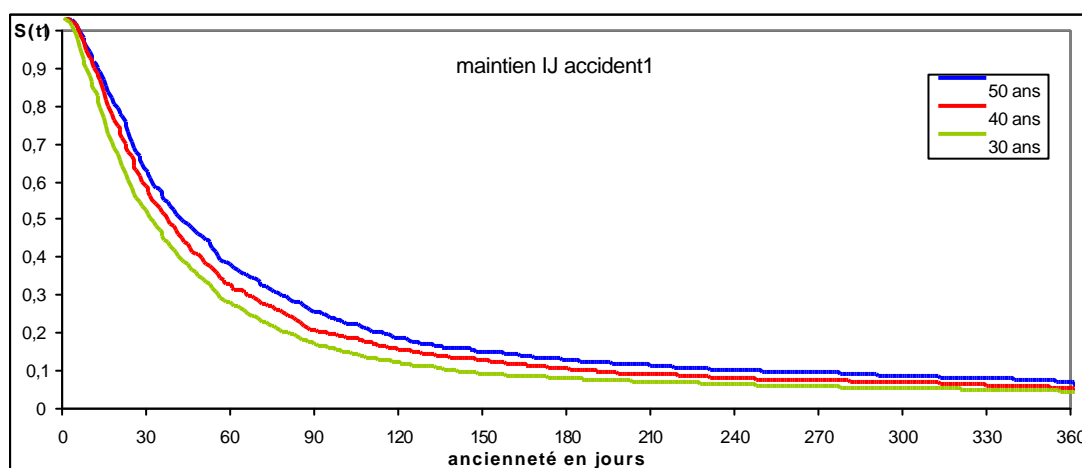
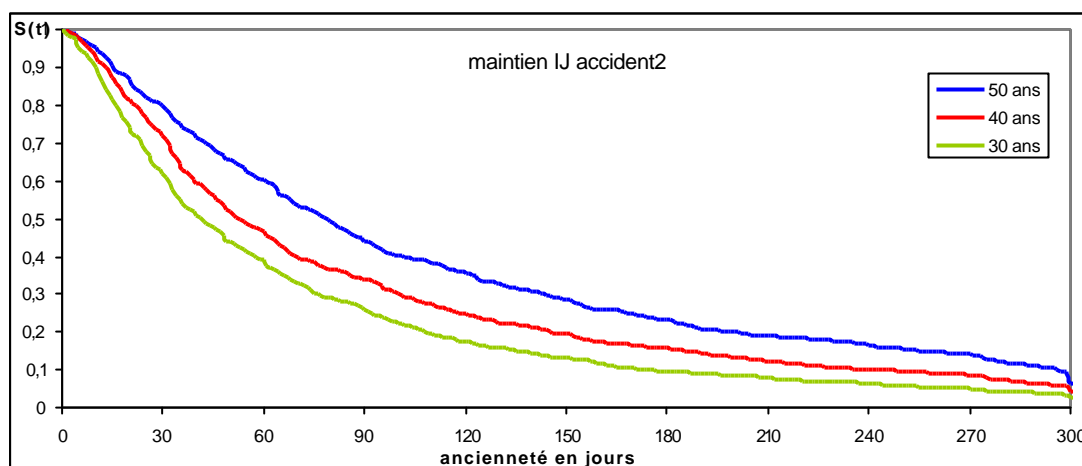
La chute brutale du maintien au bout d'1 an d'ancienneté s'explique par la présence dans notre portefeuille de garanties où le nombre de jours indemnisable par sinistre est limité à 300 ou 365 jours selon le type de garantie. Ces garanties représentent à peu près 30% de notre portefeuille, de ce fait, environ 30% de nos sinistrés qui ont atteint 1 an d'ancienneté sortent brutalement d'incapacité à cause de la limite de leur garantie. Cette sortie en masse explique donc la chute de la loi de maintien ainsi que celle des taux conditionnels de maintien mensuel. Nous verrons dans la section 6.4 comment remédier à ce problème.

Nous allons maintenant nous intéresser à l'influence de l'âge sur la loi de maintien.

6.2.4 : Influence de l'âge sur le maintien

Pour éviter de surcharger notre étude, nous limiterons notre présentation aux résultats des tables IJ accident.

Sur ces tables, nous allons mettre en évidence l'influence de l'âge sur la structure et le niveau de la loi de maintien. Les graphes suivant présentent les courbes obtenues pour une évolution de l'âge de 30 à 50 ans par tranches de 10 ans. Ils sont volontairement arrêtés à 1 an (ou 300 jours) puisque ensuite la loi est peu lisible.



Naturellement le niveau des courbes de maintien croit avec l'âge, ce qui confirme bien que l'âge est un facteur aggravant du maintien en incapacité.

Cependant, on constate que l'écart entre les courbes n'est pas constant :

- il augmente avec l'ancienneté jusqu'à trois mois environ. En effet, pour des faibles anciennetés en sinistre, le fait d'être jeune permet un rétablissement plus rapide.
- ensuite l'écart diminue jusqu'à devenir presque insignifiant vers 1 an. En fait, la courbe se redresse plus sur les jeunes âges. Ceci traduit le fait qu'à partir d'une certaine ancienneté (3 ou 4 mois), les taux conditionnels de maintien mensuels sont plus importants sur les jeunes âges, comme si les incapacités de plus de 3 ou 4 mois été plus graves chez les jeunes que chez les personnes plus âgées.
- Après plus d'un an d'ancienneté, les courbes se confondent presque car c'est vraiment la gravité du sinistre qui détermine le maintien indépendamment de l'âge de l'assuré à l'entrée en incapacité.

La section suivante va comparer nos tables brutes avec la table réglementaire.

Section 6.3 : Comparaison avec la table réglementaire

Ici, nous allons comparer uniquement avec les tables IJ car ce sont les seules qui reposent sur la même définition d'incapacité que la table réglementaire.

Cependant, avant toute chose, pour permettre la comparaison, il faut d'abord ajuster la table réglementaire que par abus de langage nous appellerons «table BCAC ». En effet, il ne faut pas oublier que nos tables représentent le maintien au-delà de la franchise alors que la table BCAC traduit le maintien sans franchise. Pour être cohérent, il est donc nécessaire de décaler cette table de manière à ce que la population sinistrée au premier jour après la franchise représente le nouveau h initial. En fait pour chaque âge, on divise par le nouveau h puis on remultiplie par 10 000 et on décale la numérotation de l'ancienneté pour que l'ancienneté 0 corresponde au nouveau jour de référence.

Sur nos 4 tables IJ, la table accident1 correspond principalement à une franchise de 7 jours alors que les 3 autres correspondent à une franchise d'un mois.

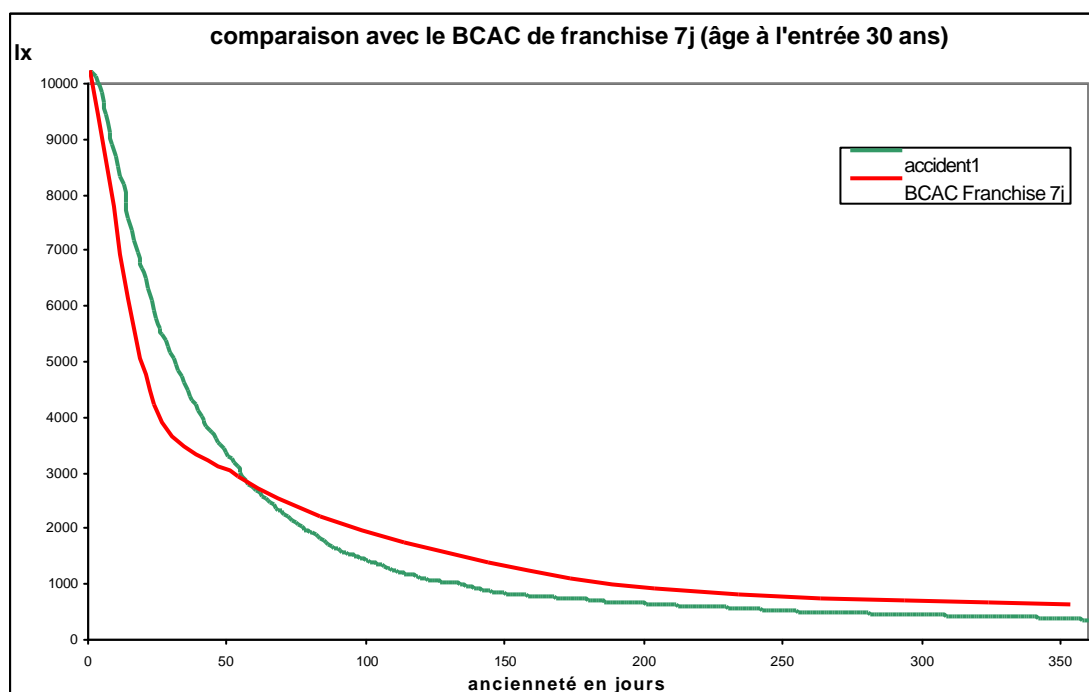
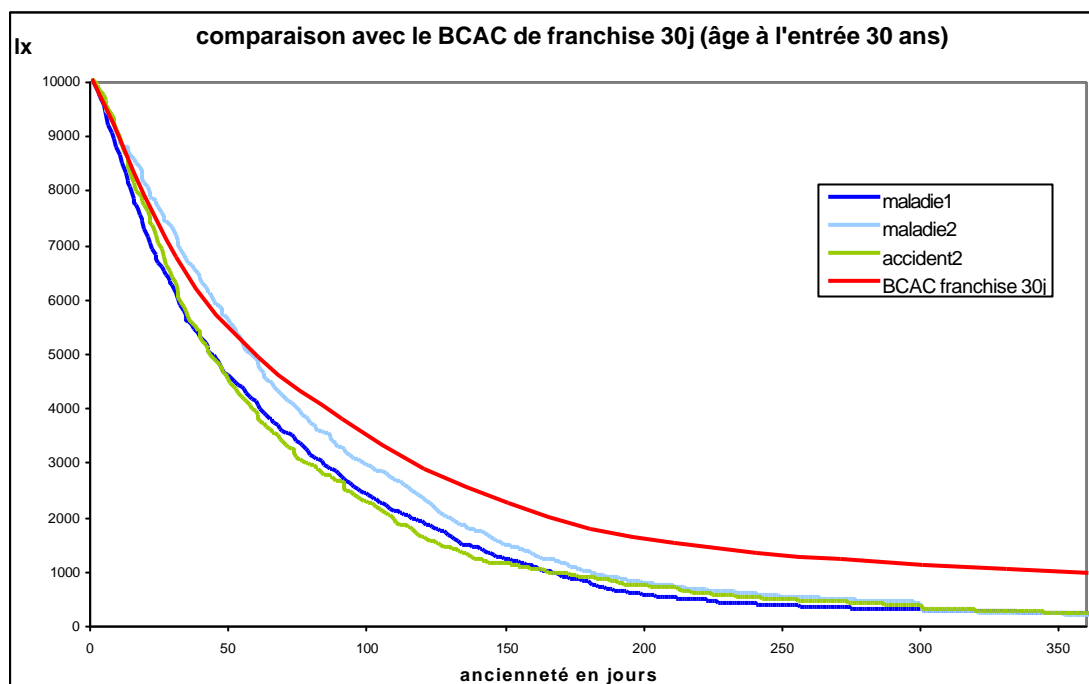
Il faut donc créer artificiellement deux tables BCAC, une pour chaque franchise.

Pour créer la table BCAC avec une franchise de 30 jours, nous disposons bien de la valeur de l_x à un mois, mais pour la franchise 7 jours, il est nécessaire de calculer le l_x à 7 jours car le pas de la table est mensuel. Pour cela, nous avons procédé à une interpolation linéaire entre 0 et 1 mois.

Remarque :

Pour une meilleure lisibilité, nos comparaisons se limiteront à un an d'ancienneté.

Les deux graphes suivants comparent, pour l'âge à l'entrée égal à 30 ans, la table BCAC franchise 7 jours avec notre table accident1 et la table BCAC franchise 30 jours avec nos autres tables.



Pour les 3 premières tables, on peut légitimement affirmer que globalement le maintien chez les sinistrés de notre portefeuille est inférieur à celui de la table réglementaire.

Par contre, il est plus difficile de trancher en ce qui concerne la table accident1 puisque jusqu'à environ 2 mois, la loi de maintien est plus élevée sur notre table d'expérience par rapport à la table réglementaire, alors qu'ensuite elle est plus faible.

En cas de certification, on peut tout de même espérer obtenir avec ces tables un provisionnement en général inférieur à celui préconisé par la réglementation.

Section 6.4 : les intervalles de confiance

Construction :

Comme signalé dans les propriétés de l'estimateur de Kaplan Meier :

$$\frac{\sqrt{n}[\hat{S}(t) - S(t)]}{\sqrt{\text{Var}(\hat{S}(t))}} \xrightarrow{L} N(0,1) \quad \text{quand } n \rightarrow +\infty$$

d'où un intervalle de confiance pour S(t) de niveau asymptotique (1-η) :

$$\text{IdC}_{(1-\eta)}\{ S(t) \} = \left[\hat{S}(t) - q_{1-\frac{\eta}{2}} * \sqrt{\frac{\text{Var}[\hat{S}(t)]}{n}} ; \hat{S}(t) + q_{1-\frac{\eta}{2}} * \sqrt{\frac{\text{Var}[\hat{S}(t)]}{n}} \right]$$

Comme pour l'entrée en incapacité, nous avons choisi de calculer des intervalles avec un niveau de confiance de 95% donc $q_{1-\frac{\eta}{2}} = 1,96$.

Etant donné la forme de la variance de l'estimateur de KM, la largeur de l'intervalle de confiance pour S(t) est directement proportionnelle à l'estimation $\hat{S}(t)$.

En effet :

$$\text{Var}\hat{S}(t) = \hat{S}^2_{KM}(t) \sum_{T'_i \leq t} \frac{d(T'_i)}{n(T'_i)[n(T'_i) - d(T'_i)]}$$

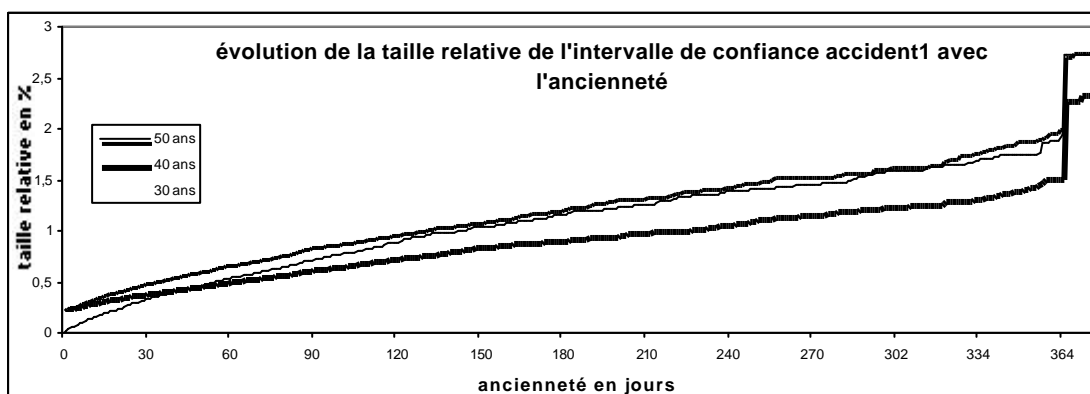
d'où la largeur de l'intervalle:

$$L\{ \text{IdC}_{(1-\eta)} \} = 2 * \left[1,96 * \frac{\hat{S}(t)}{\sqrt{n}} \sqrt{\sum_{T'_i \leq t} \frac{d(T'_i)}{n(T'_i)[n(T'_i) - d(T'_i)]}} \right]$$

A cause de cette proportionnalité, nous allons plutôt étudier la largeur **relative** de l'intervalle pour déterminer les endroits où nos résultats ne sont pas assez fiables.

Résultats :

Les courbes suivantes montrent pour la table « IJ accident1 » l'évolution de cette largeur relative en fonction de l'ancienneté.



Remarque :

Des courbes similaires ont été obtenues avec les autres tables.

Au vu de ce graphe (il en est de même pour les autres tables), on ne peut pas raisonnablement utiliser nos estimations brutes au-delà de 365 jours d'ancienneté pour les tables IJ accident1 et maladie1 ainsi que pour les tables IJ lorsqu'il y a hospitalisation, ni au-delà de 300 jours pour les tables IJ accident2 et maladie2. En effet, on remarque une augmentation brutale de la largeur relative de l'intervalle de confiance pour ces anciennetés là, ce qui traduit en fait une chute de volume trop importante. D'ailleurs, les courbes brutes et les taux conditionnels mensuels montraient déjà ce saut à 365 ou 300 jours.

Pour ce qui est des intervalles sur les tables IJH, c'est un peu moins net, mais par cohérence avec les intervalles précédents, nous choisirons de limiter les tables IJH à 3 mois. En effet, au-delà de cette ancienneté, la taille de l'intervalle relatif dépasse 2% et cette taille correspond à la taille maximum acceptée sur les tables précédentes.

Cependant le but de ce mémoire était d'obtenir des tables sur 3 ans, il n'est donc pas satisfaisant de garder simplement nos tables limitées à 1 an. Pour palier à cette limitation, nous avons décidé de raccorder nos tables d'expérience avec la table réglementaire préalablement interpolée en jours. Voici la formule utilisée pour calculer le $l_x(t)$ après un raccordement à 365 jours par exemple :

$$l_x(t) = {}_{BCAC}l_x(t) * \frac{\exp l_x(364)}{{}_{BCAC}l_x(364)}, \text{ pour } t \in [365j, 1095j]$$

avec :

- ${}_{BCAC}l_x(t)$ issu de la table réglementaire
- $\exp l_x(364)$ issu de nos tables d'expérience

De cette manière, on récupère des tables de maintien de 0 à 3 ans, tout en restant prudent vis à vis des données.

NB : attention, comme nos tables représentent le maintien au-delà de la franchise alors que la table réglementaire traduit le maintien sans franchise, pour être cohérent, il faudra raccorder chaque table avec la table BCAC créée avec la franchise correspondante.

Maintenant que nous disposons de tables brutes sur 3 ans, nous allons les lisser.

Chapitre 7 : lissage des lois de maintien

Le modèle d'estimation utilisé dans le cas du maintien en incapacité permet de calculer une loi brute de survie qui se présente sous la forme d'une fonction en escalier. Ce résultat n'est pas toujours satisfaisant puisque les paliers obtenus peuvent être trop grossiers et donner une loi brute de maintien irrégulière.

Tant pour le tarif que pour le provisionnement, les irrégularités doivent être lissées pour éviter d'avoir des écarts trop importants entre deux anciennetés ou deux âges consécutifs. Il est bien sûr envisageable de garder des tables brutes et de lisser le résultat final, c'est à dire les provisions ou le tarif.

Cependant, ici nous souhaitons lisser directement les tables et non les provisions ou les tarifs calculés à partir des lois brutes. En effet, les tarifs et les provisions sont fonction d'un taux technique qui évolue avec le TME (Taux moyen des emprunts d'Etat) donc si on lisser le résultat et non les tables, il faudra relisser à chaque changement de taux.

Les tables de maintien étant à double entrée (l'âge et l'ancienneté), il faudra effectuer un **double lissage**, c'est à dire un lissage qui gommara les irrégularités de la loi de maintien aussi bien en fonction de l'âge qu'en fonction de l'ancienneté.

Après avoir présenté la problématique de ce double lissage, nous étudierons quelques méthodes de lissage traditionnellement utilisées avant de choisir la mieux adaptée à nos tables brutes.

Section 7.1 : La problématique du double lissage

La double dimension de nos tables implique de réfléchir à la méthode de lissage la mieux adaptée dans chacune des dimensions.

En effet, si l'on lisser dans un sens puis ensuite dans l'autre, on risque de détériorer le premier lissage. Le mieux serait de pouvoir lisser dans les deux sens en une seule fois.

La méthode des moyennes mobiles semble a priori la mieux adaptée. Elle est d'ailleurs communément utilisée pour lisser les provisions calculées à partir des tables réglementaires brutes.

En effet, cette méthode non paramétrique permet de calculer une valeur lissée de la loi à partir de son estimation brute mais aussi des estimations qui l'entourent. Voici selon cette méthode la formule utilisée pour lisser l_x pour un âge d'entrée x et pour une ancienneté atteinte anc :

$$l_{lissé}(anc, x) = \frac{1}{2} * l(anc, x) + \frac{1}{16} * l(anc - 1, x) + \frac{1}{16} * l(anc - 1, x - 1) + \frac{1}{16} * l(anc, x - 1) + \frac{1}{16} * l(anc + 1, x) \\ + \frac{1}{16} * l(anc + 1, x + 1) + \frac{1}{16} * l(anc, x + 1) + \frac{1}{16} * l(anc - 1, x + 1) + \frac{1}{16} * l(anc + 1, x - 1)$$

Cette formule permet donc d'effectuer un double lissage en un temps.

Elle pose cependant un problème pour lisser les valeurs extrêmes, c'est à dire le premier et le dernier âge ainsi que la dernière ancienneté.

De plus, le lissage doit conserver certaines propriétés de la loi de maintien, comme par exemple la décroissance de la loi suivant l'ancienneté; or la moyenne mobile ne permet pas de maîtriser cette contrainte.

Enfin, si l'on applique cette formule à nos estimations brutes, le résultat n'est pas très satisfaisant du point de vue de la régularité.

Pour toutes ces raisons, nous ne retiendrons pas cette méthode de lissage.

Nous devons maintenant envisager un lissage en 2 temps et choisir l'ordre à respecter.

Pour assurer la décroissance de la loi de maintien en fonction de l'ancienneté, il est préférable de lisser en fonction de l'ancienneté en dernier et donc de lisser d'abord en fonction de l'âge et à ancienneté constante. Nous appellerons ce premier lissage le «lissage suivant l'âge », et le second le «lissage suivant l'ancienneté ».

Etant donné que nous allons effectuer deux lissages distincts, il est possible de choisir deux méthodes de lissage différentes. Cependant, il serait illogique de lisser dans un sens par une méthode non paramétrique puis dans l'autre sens par une méthode paramétrique.

Nous allons maintenant présenter quelques méthodes en commençant par le lissage suivant l'âge.

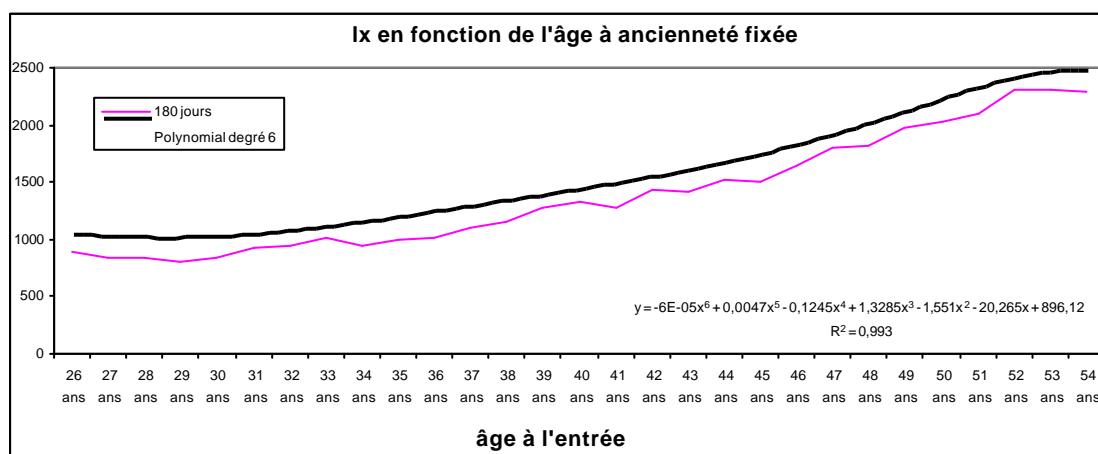
Section 7.2 : Le lissage suivant l'âge

Lorsqu'on visualise les courbes brutes de la loi de maintien suivant l'âge, globalement, on remarque une allure assez similaire aux courbes obtenues pour l'entrée en incapacité, avec cependant une caractéristique récurrente : la croissance avec l'âge. Cette forme croissante paraît évidente puisque qu'elle traduit le facteur aggravant de l'âge.

De même que pour le lissage de l'entrée en incapacité, des comparaisons graphiques montrent qu'une régression polynomiale est tout à fait adaptée pour lisser les courbes de maintien suivant l'âge. Pour des raisons pratiques, nous choisirons un degré de polynôme constant et égal à 6 pour toutes les anciennetés. En effet, sur chaque table on effectue un lissage pour chaque ancienneté (journalière), soit en fait 1095 lissages pour chaque table. C'est pourquoi nous avons renoncé à effectuer un lissage avec un degré de polynôme modulable.

Nous avons fixé un degré de 6 qui permettait généralement d'obtenir un R^2 très satisfaisant de l'ordre de 95%.

Voici un exemple du lissage de la table accident2 pour l'ancienneté 6 mois :



Nous allons maintenant choisir la méthode de lissage suivant l'ancienneté.

Section 7.3 : Le lissage en fonction de l'ancienneté

Dans cette section, nous exposerons 3 méthodes de lissage différentes :

- le lissage par Whittaker et la variante de Lowrie
- l'ajustement à une loi de Makeham
- l'ajustement à des fonctions usuelles (polynomiale, logarithmique, exponentiel, puissance, etc ...) par la méthode des moindres carrés.

7.3.1 : lissage par Whittaker et la variante de lowrie

Principe d'une méthode de lissage:

Tout lissage est un compromis entre le respect des données (résultat brut) et la régularité du lissage obtenu (résultat lissé).

La fidélité à la courbe initiale se mesure en calculant l'erreur F suivante:

$$F = \sum_{i=1}^n w_i (v_i - u_i)^2$$

où: - i indice les données à lisser;

- n est le nombre de données à lisser;

- w_i est le poids donné au respect de la donnée i (en général $w_i = \frac{1}{\text{var}[u_i]}$);

- v_i est la donnée i lissée;

- u_i est la donnée brute à lisser.

La régularité du résultat se mesure par la valeur S suivante: $S = \sum_{i=1}^{n-Z} (\Delta^z v_i)^2$

où: $z-1$ est le degré du polynôme que l'on considère comme le standard de la régularité par rapport aux données.

Si les données lissées sont égales aux données brutes, F est nulle. Par contre si les données lissées suivent exactement un polynôme de degré $z-1$, S est nulle car alors $\Delta^z v_i = 0$.

Toute méthode de lissage tente d'obtenir F et S les plus petits possibles, sur la base d'un compromis entre ces éléments "antagonistes".

La méthode de Whittaker:

La méthode de lissage de Whittaker consiste à trouver les valeurs lissées v_i qui minimise M avec $M = F + h S$ pour une valeur de h , poids relatif de S , les données u_i fixés.

$$M = \sum_{i=1}^n w_i (v_i - u_i)^2 + h \sum_{i=1}^{n-Z} (\Delta^z v_i)^2 \quad \text{est minimale si pour tout } i \quad \frac{\partial M}{\partial v_i} = 0$$

Une approche matricielle permet de synthétiser le système d'équations et de le résoudre, sachant qu'il faut procéder à l'inversion de certaines matrices.

La variante de Lowrie:

Lowrie cherche plutôt à minimiser la valeur suivante:

$$M=(1-h_1) \sum_{i=1}^n w_i (v_i - u_i)^2 + h_1 \sum_{i=1}^n w_i^s (v_i - s_i)^2 + h_2 \sum_{i=1}^{n-Z} (\Delta^z v_i - r \Delta^{z-1} v_i)^2$$

Interprétation des différents termes :

1. Le terme $\sum_{i=1}^{n-Z} (\Delta^z v_i - r \Delta^{z-1} v_i)^2$ correspond au fait que le critère de régularité n'est plus l'adéquation à un polynôme de degré $z-1$ mais à un polynôme de degré $z-2$ plus un terme exponentiel $(r+1)^i$.

Par exemple: $v_i = A + H i + B c^i$

2. $r = c-1$ en fonction de la distribution exponentielle supposée sur certaines anciennetés.

3. $z-2$ est le degré du polynôme correspondant à la distribution des données.

4. h_2 est le poids que l'on veut donner au lissage par rapport au respect des données.

5. Le terme d'ajustement: $(1-h_1) \sum_{i=1}^n w_i (v_i - u_i)^2 + h_1 \sum_{i=1}^n w_i^s (v_i - s_i)^2$ est particulier.

En effet, la seconde partie de l'addition permet sur certaines anciennetés de lisser en respectant non pas les données brutes mais une pondération entre les données brutes u et une loi de référence s . Ainsi, sur les anciennetés où, pour des problèmes de volumes, on considère qu'il serait plus prudent de respecter la table de référence plutôt que nos estimations brutes, w_i sera faible et w_i^s sera plus élevé.

6. $(1-h_1)$ est le poids que l'on veut donner au respect des données brutes par rapport à la table de référence.

7. w_i (resp w_i^s) est le poids attribué à la valeur u_i (resp à la valeur s_i).

Application aux données :

Sous tableur, l'inversion d'une matrice est possible mais la dimension de la matrice est limitée (57; 57). Pour lisser la loi sur les 1095 anciennetés il faudrait effectuer le lissage en plusieurs temps (une vingtaine de fois) et procéder à des raccordements (par interpolation linéaire). La mise en œuvre de ce lissage semble donc assez contraignante et longue.

Ensuite, l'intérêt de cette méthode est de pouvoir s'appuyer sur une table de référence lorsque les données brutes sont jugées insuffisantes, or nous avons déjà paré à cette éventualité en raccordant (à 100%) nos tables brutes à la table réglementaire dès qu'une sortie trop importante d'individus compromettrait la fiabilité de nos résultats.

Enfin, nous avons choisi un lissage paramétrique suivant l'âge il serait donc malvenu d'utiliser une méthode non paramétrique pour lisser suivant l'ancienneté.

Nous allons donc plutôt nous pencher sur des lissages paramétriques, c'est à dire en fait des ajustements à une loi définie.

7.3.2 : Modèle de Makeham

Notations :

- Les données sont classées par intervalle $[t, t + \Delta t]$
- n le nombre d'intervalles observés
- $SORT_t$: le nombre de sorties d'incapacité pendant $[t, t + \Delta t]$
- $SRIS_t$: la population sous risque pendant $[t, t + \Delta t]$
- $q_t = P(t - \Delta t < T \leq t / T > t - \Delta t)$ la probabilité de sortir sur $[t - \Delta t, t]$ sachant qu'il est en incapacité en $t - \Delta t$ (T : variable aléatoire « durée de maintien de l'individu en incapacité »)

a) principe

Ce modèle suppose que les taux instantanés de sortie d'incapacité sont fonction du temps par la fonction suivante appelée fonction de Makeham : $\mu_t = a + b * e^{-ct}$

Donc la loi de maintien théorique l_t s'écrit :

$$l_t = l_0 e^{-\int_0^t m_h da} = l_0 \exp \left[- \left\{ at + \frac{b}{\log(c)} (c^t - 1) \right\} \right]$$

Les estimations des paramètres a , b , c doivent maximiser la vraisemblance ou la log vraisemblance du modèle statistique :

$$V = \prod_t (q_t)^{SORT_t} * (p_t)^{(SRIS_t - SORT_t)} \quad \text{et} \quad L = \log V = \sum_t (SORT_t \log \frac{q_t}{p_t} + SRIS_t \log p_t)$$

$$\text{avec } q_t = 1 - \frac{l_{t+\Delta t}}{l_t} = 1 - \exp \left[- \left\{ a\Delta t + \frac{b}{\log c} c^t (c^{\Delta t} - 1) \right\} \right] \quad \text{et } p_t = 1 - q_t$$

l'ajustement par la loi de Makeham se fait en deux étapes :

celle qui consiste à fournir une première estimation de a , b et c , appelée méthode de King et Hardy, puis celle qui affine cette estimation par l'algorithme itératif de Newton et qui est basée sur le maximum de vraisemblance.

Mais auparavant, il serait intéressant de mesurer la validité de l'hypothèse de Makeham à l'aide d'une régression linéaire.

b) Validité de Makeham

Si l'on accepte l'hypothèse de Makeham, on peut écrire :

$$\log(p_t) = -a - \frac{b}{\log c} c^t (c^{\Delta t} - 1), \quad \text{ou encore :} \quad \log \left(\frac{p_{t+\Delta t}}{p_t} \right) = - \frac{b}{\log c} c^t (c^{\Delta t} - 1)^2$$

$$\text{On a finalement : } \log \left[\log \left(\frac{1 - q_{t+\Delta t}}{1 - q_t} \right) \right] = \text{constante} + t \log c$$

Par conséquent, si l'hypothèse de Makeham est vérifiée, les points de coordonnées :

$$\left(t ; \log \left[\log \left(\frac{1 - q_{t+\Delta t}}{1 - q_t} \right) \right] \right) \text{ devraient être sensiblement alignés sur une droite de pente } \log c.$$

Malheureusement, nos données ne permettant pas de valider (graphiquement) l'adéquation à une loi de Makeham, nous devons rejeter cette loi pour l'ajustement et en considérer d'autres.

7.3.3 : Régressions paramétriques par les moindres carrés

Par cette méthode, on considère que la loi de maintien en incapacité suit une fonction particulière. La régularité des estimations lissées est alors incluse dans la fonction. Les paramètres de cette fonction sont calculés en fonction des estimations brutes par la méthode des moindres carrés.

Le fait de pouvoir moduler la fonction et ses paramètres permet d'adapter cette méthode de lissage à toutes nos lois. Elle permet aussi de s'assurer de la décroissance de la loi de maintien. Nous avons donc choisi de retenir cette démarche pour lisser nos tables en fonction de l'ancienneté.

Il faut maintenant déterminer les fonctions à appliquer et leurs paramètres.

Pour les tables IJ :

Si l'on regarde les différents tracés des courbes de maintien suivant l'ancienneté (cf « l'allure des courbes de maintien » dans le chapitre précédent), on remarque deux choses :

- un changement d'allure entre 3 et 6 mois d'ancienneté.
- une cassure de la courbe vers 300 ou 365 jours.

Le premier changement paraît logique. En effet, la majorité des sinistres sont de courte durée ce qui explique la forte pente de 0 à 5, 6 mois. D'autre part, après 6 mois d'ancienneté, on entre dans la catégorie des sinistres « graves » où les sorties sont plutôt marginales. De ce fait, la pente de la courbe se redresse fortement.

Quant à la cassure, nous l'avons créée artificiellement en raccordant à nos tables la table réglementaire au bout de 300 ou 365 jours.

Ces remarques nous amènent donc à un lissage en trois parties :

- avant à peu près 4 mois d'ancienneté, le maintien évolue suivant un polynôme de degré 6. Nous avons opté pour un polynôme de degré 6 pour la première partie à lisser, car sa forme permet de s'approcher au mieux de la courbe brute et notamment sur les faibles anciennetés. En effet, avec d'autres fonctions d'ajustement, la loi de maintien brute est généralement sous estimée au début de l'incapacité, alors qu'un polynôme de degré 6 respecte la loi brute dès le début, ce qui compte beaucoup puisque les sinistres les plus nombreux sont les plus courts. Ainsi une sous estimation du maintien au début aurait un impact désastreux.
- ensuite, de 4 mois à 1 an (ou 300 jours suivant les cas), la loi brute s'ajuste à un autre polynôme de degré 6.
- Enfin un lissage de la dernière partie de la courbe (de 1 à 3 ans) qui n'est autre que la représentation de la table réglementaire. Ce lissage se fait principalement par un ajustement à une fonction logarithmique de la forme $f(x)=A*\ln(x)+B$.

Pour les tables IJH :

on remarque la même chose que pour les tables IJ mais à des anciennetés différentes :

- un changement d'allure vers 15 jours d'ancienneté.
- une cassure de la courbe vers 3 mois.

Le lissage qui en découle suit le même principe que précédemment mais avec une variante sur la dernière partie de la courbe qui est ici plus longue (de 3 mois à un an). Au lieu d'un ajustement à une fonction logarithmique nous avons utilisé un ajustement à un polynôme de degré 6 pour l'accident et à une fonction puissance pour la maladie.

Les raccordements :

Le 1^{er} raccordement entre les fonctions se fera de la manière suivante :

à l'endroit du raccordement, les deux lissages se chevauchent sur une période de dix jours [a,b], sur cette période on calcule alors la loi lissée $l(t)$ par la formule suivante :

$$l(t)_{\text{lissée}} = l_1(t) * \left(\frac{b-t}{b-a} \right) + l_2(t) * \left(\frac{t-a}{b-a} \right)$$

L'endroit de raccordement de chaque courbe est bien sûr déterminé pour le mieux, c'est à dire pour que les deux lissages soient proches l'un de l'autre à cet endroit, que la décroissance de la loi lissée soit assurée et enfin, que le coefficient R^2 (propre à la méthode des moindres carrés) soit optimum.

Pour le 2^{ème} raccordement, nous avons choisi de passer directement d'un lissage à l'autre à l'endroit de l'intersection et donc de cette manière la cassure sous jacente est conservée.

Nous avons fait ce choix pour deux raisons.

La première, c'est qu'un lissage de la cassure modifierait l'allure de notre loi brute avant le raccordement, or celle-ci a été jugée fiable puisque la chute de volume ne se produit qu'un jour bien précis (365 ou 300 jours). Avant ce jour là, notre loi représente bien la sinistralité et elle n'est pas encore faussée par la fin contractuelle des garanties.

La deuxième raison, c'est que l'atténuation de la cassure implique une surestimation du maintien sur la période proche du raccordement et donc sur la partie de la courbe qui correspond à la table du BCAC. Or être encore plus prudent que la table réglementaire ne semble pas nécessaire.

Etant donné la double dimension des tables et leur taille importante (plus de 1 000 lignes et autant de colonnes que d'âges estimés soit une trentaine), nous ne pouvons pas produire ici les tables finales dans leur totalité.

CONCLUSION

I. Rappel des principaux résultats :

L'étude réalisée s'inscrit dans le cadre de la maîtrise des risques de la prévoyance. Ce projet nécessite une meilleure connaissance de la sinistralité du portefeuille d'assurés, à commencer par le risque incapacité. Dans ce but, nous avons effectué des statistiques descriptives puis élaboré des tables d'entrée et de maintien en incapacité.

A. Le descriptif du portefeuille

La description du portefeuille sur lequel ont été réalisées les tables d'expérience est essentielle dans l'utilisation future des tables. En effet, la connaissance de cette base permet de dire si les résultats, obtenus en utilisant nos tables sur un nouveau portefeuille, seront pertinents. De plus cette connaissance du portefeuille peut influencer les choix de politique commerciale.

Nous avons tenté de décrire le portefeuille en terme de contrat et de population assurée. Voici le portrait de l'assuré type sur ce portefeuille (assuré qui possède les caractéristiques majoritaires) :

« Un homme, ayant souscrit un contrat de prévoyance avec au moins une garantie exonération de primes dont la franchise est d'un mois en cas de maladie et d'une semaine en cas d'accident, et le cas échéant avec un montant de 50 à 150F d'indemnités journalières. Cet homme aurait aujourd'hui entre 30 et 55 ans, il aurait souscrit son contrat entre l'âge de 20 ans et de 45 ans soit en fait il y a 10 ans et il serait salarié. »

Pour avoir un aperçu du risque incapacité et pour identifier les variables qui n'étaient pas utilisables par la suite car incomplètes, nous avons aussi étudié les sinistres du portefeuille.

Globalement, on obtient les résultats suivants :

- presque autant d'accidents que de maladies
- la principale franchise sur nos sinistres dure 1 mois
- $\frac{3}{4}$ des sinistres durent moins de 3 mois
- ils sont plus fréquents :
 - en hiver
 - chez les femmes
 - chez les agriculteurs, les artisans ou les salariés
 - dans le Sud-Ouest de la France et en Corse
 - sur les gros montants d'IJ

Après ce descriptif, nous avons élaboré les tables d'entrée en incapacité.

B. Les tables d'entrée en incapacité

Pour l'élaboration des tables d'entrée en incapacité, la principale difficulté a été d'identifier réellement l'élément que nos tables devaient traduire et qui dépend en fait de l'utilisation future des tables. Il a donc fallu anticiper une éventuelle retarification pour déterminer ce que l'on avait besoin d'estimer.

Dans le cadre d'un modèle de tarif de type IARD, nous avons choisi d'estimer le nombre espéré annuel de sinistre par assuré sous l'hypothèse d'une loi de Poisson. Pour traduire exactement les risques garantis et pour immuniser au maximum notre étude contre le risque d'évolution du portefeuille, nous avons segmenté nos tables de manière à obtenir des groupes de sinistralité homogène. De ce fait, nous avons réalisé 19 tables d'entrée toutes déclinées par âge de 20 à 60 ans et segmentée suivant le type de garantie souscrit, IJ (ou EXO) et IJH,

suivant le sexe de l'assuré, suivant la cause du sinistre (maladie ou accident), et suivant la durée de franchise souscrite.

Le tarif dépendant généralement de l'âge de l'assuré, on comprend bien que des écarts trop importants dans les résultats d'un âge à l'autre soient difficilement acceptables commercialement. Pour rendre nos tables brutes régulières, nous les avons donc lissées par une régression polynomiale en fonction de l'âge de l'assuré.

Contrairement aux idées reçues, nos tables ne sont pas toujours croissantes avec l'âge.

Elles présentent souvent un pic de sinistralité pour les femmes de 25 à 30 ans et pour les hommes de 50 à 55 ans.

Le nombre espéré annuel de sinistre par assuré est toujours inférieur à 18% il est même inférieur à 5% pour 15 tables sur 19.

Parallèlement à l'élaboration des tables d'entrée en incapacité, nous nous sommes aussi consacrés à la construction des tables de maintien. Les deux études ont été menées de front car elles sont extrêmement liées, notamment lorsqu'on envisage une nouvelle tarification.

C. Les tables de maintien en incapacité

Si l'on veut à terme utiliser ces tables pour le provisionnement, leur forme et leur contenu est défini par la table réglementaire fournie par le BCAC. En effet, il faut estimer la loi de maintien suivant l'ancienneté en sinistre, c'est à dire pour chaque valeur de t comprise entre 0 et 3 ans (durée maximum de l'incapacité) il faut calculer la probabilité que la variable durée de l'incapacité dépasse t . Pour ce faire, nous avons utilisé l'estimateur de Kaplan-Meier, car certaines de nos données sont censurées et c'est l'estimateur qui traite le mieux les censures.

Etant donnée la précision de nos données, nous avons pu réaliser des tables journalières (la loi est donnée avec un pas journalier). Cependant pour des raisons de volume, nous avons arrêté nos tables à 1 an d'ancienneté en sinistre et pour avoir des tables sur 3 ans nous les avons ensuite raccrochées à la table réglementaire.

En plus du provisionnement, les tables de maintien peuvent servir dans la tarification et doivent donc être cohérentes avec les tables d'entrée en incapacité. C'est pourquoi nous avons utilisé les mêmes principes de segmentation. Cependant, certaines variables discriminantes au niveau de l'entrée en incapacité ne le sont pas au niveau du maintien, c'est le cas par exemple du sexe de l'assuré qui apparemment n'influence pas la durée des sinistres.

Au final, nous avons construit 8 tables de maintien déclinées sur chaque âge, entre deux âges extrêmes qui dépendent du volume de données sur chaque table. De manière générale, ces extrémités se situent aux alentours de 25 ans et de 55 ans. Au-delà, nous n'avons pas le détail de la loi de maintien par âge.

Etant donnée la double dimension des tables de maintien (âge et ancienneté), nous avons effectué un double lissage des tables brutes en commençant par le lissage suivant l'âge de l'assuré et à ancienneté fixée. Ce premier lissage a été effectué par une régression polynomiale, le deuxième lissage (en fonction de l'ancienneté et à âge fixé) s'appuie sur les résultats du premier et consiste en un ajustement de la loi avec des fonctions usuelles (exponentielle, puissance, etc...).

Globalement, nos résultats montrent que la loi de maintien des assurés d'AXA se positionne en dessous de la table réglementaire fournie par le BCAC.

L'âge est un facteur aggravant du maintien mais de moins en moins déterminant lorsque l'ancienneté du sinistre augmente.

Pour les assurés (IJ ou EXO), de $\frac{1}{4}$ à la moitié sortent d'incapacité avant un mois et pour les assurés (IJH) 90% sortent dès le 1^{er} mois.

II. L'utilisation des tables :

La description de la sinistralité du portefeuille à la fois par les tables d'expérience pour l'entrée et pour le maintien en incapacité n'est pas une fin en soi. En effet, l'objectif du mémoire est de fournir un outil de tarification et de provisionnement à AXA. Ce chapitre termine le mémoire en traitant des aspects pratiques de l'utilisation de la table en deux points:

- analyse des conditions de validité de la table;
- présentation de l'utilisation de la table par AXA :
 - les tarifs
 - les provisions

A. Les conditions de validité de la table

1. Les limites des données et de l'étude

Les fichiers utilisés contenaient des données aberrantes et ne décrivaient pas toute l'information nécessaire à la description exhaustive du risque étudié. Les résultats de l'étude sont donc limités par la validité des données qui lui ont servi de base.

2. L'évolution future du portefeuille.

L'étude a été faite à partir de l'observation du portefeuille à un moment donné. Si la structure de ce portefeuille change, sa sinistralité peut évoluer et les résultats de notre étude ne seront plus valables.

Le nombre de dimensions de la table a été choisi de façon à immuniser le plus possible la table contre la plupart des changements de structure du portefeuille (proportion de femmes, type de garantie et de franchise...). En effet, les segmentations des résultats permettent de reconstituer la sinistralité globale en fonction des proportions d'hommes ou de femmes, des différents types de franchises et des âges des assurés dans le portefeuille.

Toutefois, il reste des évolutions contre lesquelles nous n'avons pas pu nous prémunir.

- Il est presque certain que la CSP de l'assuré intervient dans sa sinistralité. Cependant notre portefeuille est constitué à 70% de salariés, c'est pourquoi notre segmentation ne tient pas compte de la CSP. Pourtant, si cette proportion diminuait dans un futur proche, cela fausserait l'adéquation entre la sinistralité future des assurés du groupe et les tables d'expérience qui restent une description de la sinistralité du portefeuille étudié.
- Il est aussi fort possible que la sinistralité dépende des capitaux souscrits. Cependant, nous ne disposons pas d'une table en fonction du montant du capital assuré pour nous permettre de maîtriser ce paramètre. Une modification dans le portefeuille des proportions entre les contrats populaires (faibles capitaux souscrits) et les autres (cible commerciale haute gamme) biaiserait les résultats obtenus en utilisant nos tables d'expérience.
- Certains choix de la compagnie pourraient notamment rendre caduques nos résultats: modification des procédures de sélection médicale, changement du mode de distribution des contrats (utilisation d'Internet par exemple)...

3. Future évolution de la santé publique.

Des progrès dans la recherche médicale ou au contraire, une recrudescence ou l'apparition de nouvelles épidémies pourront influencer dans un sens ou dans l'autre la sinistralité de la population française et donc celle des assurés d'AXA en France.

De manière générale, la validité de la table est menacée par toute nouvelle épidémie ou évolution de la science médicale (notamment les tests génétiques).

B. Les domaines d'utilisation de la table.

1. Les tarifs

Les tarifs sont élaborés à partir de calculs d'études de rentabilité des produits qui tiennent compte des paramètres suivants:

- la sinistralité (primes pures et hypothèses de sinistralité)
- les frais
- les hypothèses de commercialisation...

Les résultats de notre étude fournissent l'hypothèse de sinistralité qui peut être prise en compte pour l'élaboration des tarifs. Les segmentations des tables d'expérience donneront une meilleure vision de la sinistralité de nos assurés qu'une référence à une table réglementaire sans rapport avec la réalité des affaires de la compagnie.

remarque pratique :

Certaines cotisations évoluent à date anniversaire du contrat en fonction de l'âge de l'assuré. Nos tables d'expérience utilisent une définition de l'âge en adéquation avec la convention de la plupart des tarifs (l'âge à la souscription égal à l'âge au plus proche anniversaire). Dans cette situation, les résultats peuvent être utilisés sans approximations ou interpolations sur les âges. Pour d'autres conventions, il est facile de développer une approximation adéquate.

2. Les provisions pour sinistres à payer (PSAP)

Ces provisions nécessitent une hypothèse de sinistralité. Nos résultats permettront après certification d'établir les PSAP sur les sinistres incapacité touchant le portefeuille AXA.

Notre étude donne la sinistralité globale de l'ensemble du portefeuille. Cependant, ce portefeuille regroupe différentes entités qui correspondent à des sociétés juridiques séparées, il est nécessaire de connaître la sinistralité du portefeuille de chaque société pour établir les PSAP qui influencent leurs bilans comptables.

Pour ce faire, il faudrait réaliser l'étude des sous populations correspondant à chaque réseau.

En fait plusieurs études complémentaires à ce mémoire sont envisageables pour palier aux limites de validité précédemment exposées.

III. Les études supplémentaires possibles :

A. Etude de sous populations particulières

Axa dispose d'un portefeuille à la fois d'une taille conséquente et représentatif de différents "types" d'assurés. Il serait intéressant d'utiliser ces atouts pour déterminer la sinistralité de certaines sous-populations. Il ne s'agit pas de refaire une table en ajoutant des variables de segmentations supplémentaires car le volume de données serait alors insuffisant mais de tester des différences de sinistralité par rapport à l'ensemble du portefeuille. L'étude de la sinistralité de certains segments d'assurés permettrait d'adapter éventuellement les cibles commerciales futures ou d'intégrer dans les tarifs les particularités détectées.

Pour comparer la sinistralité de deux populations, la méthode la plus utilisée par les actuaires et les bio statisticiens est le test du SMR (Ratio Standardisé de Mortalité). Il se rapproche du test du Chi-deux pour l'adéquation d'un échantillon (sous-population du portefeuille) à une loi (table de référence). Il présente l'intérêt par rapport au test d'adéquation de Kolmogorov de fournir directement un indice comparatif de l'incidence de la sinistralité dans le groupe (population ou échantillon) étudié par rapport à la population de référence. De plus, ce test est immunisé contre les différences de structure des populations concernées (par exemple si les assurés sont plus âgés dans une sous population).

L'expérience et les spécificités du portefeuille suggèrent par exemple de tester la sinistralité des sous populations définies par les variables telles que la catégorie socio-professionnelle ou le réseau de commercialisation du produit.

B. Elaboration de tables par capitaux

Dans ce mémoire, nous n'avons exploité que des estimateurs par tête, c'est à dire que tous les assurés avaient le même poids dans l'expérience. Cependant, il existe aussi des estimateurs pondérés utilisés par les compagnies d'assurance notamment en Grande-Bretagne et en Amérique du Nord.

L'intérêt de ce type d'estimateur est qu'il correspond réellement au risque de la compagnie d'assurance.

En effet, le risque encouru par AXA n'est pas vraiment l'entrée en incapacité de l'assuré mais le versement des indemnités journalières qu'il avait souscrit et/ou l'exonération de sa prime. L'assimilation du risque de l'assuré, qui est l'incapacité, avec le risque de l'assureur, qui est la propension à verser un franc, postule l'indépendance de la sinistralité et des montants garantis. Certains pensent au contraire que la sinistralité et le montant garanti sont fortement liés. En Amérique du Nord par exemple, il existe des tables de mortalité différentes pour les contrats à gros capitaux et à faibles capitaux car on constate que la mortalité des premiers est plus faible. Ceci s'explique notamment par le fait que les personnes souscrivant un contrat avec un capital important subissent une sélection plus stricte et ont souvent un statut social plus élevé.

Il serait donc judicieux de compléter nos estimateurs par tête par des estimateurs pondérés par capitaux qui estimeraient non plus la sinistralité mais la propension de l'assureur à verser un franc par arrêt de travail d'un assuré.

C. Construction de tables sur 2 têtes pour l'entrée et le maintien en incapacité

Nous savons qu'il existe dans notre portefeuille des contrats avec deux assurés qui garantissent l'exonération de la prime lorsqu'un des deux assurés entre en incapacité.

De ce fait, pour être adapté aux garanties souscrites, il faudrait, en plus des tables par tête, construire une table sur deux têtes basée sur l'échantillon de ces contrats particuliers. Cependant, cet échantillon est trop restreint pour permettre un niveau de segmentation aussi poussé que pour les tables par tête. De plus les caractéristiques client ne peuvent pas être prises en compte pour segmenter la population puisque sur une seule garantie, il existe deux assurés avec chacun ses propres caractéristiques.

BIBLIOGRAPHIE

MEMOIRES D'ACTUARIAT :

- 3) **KATKOFF et RANSAC** (CEA 1992) : *'Lois de maintien et provisions arrêt de travail dans les contrats collectifs d'emprunteurs'*
- 4) **FAIVRE** (CNAM 1978) : *'Contribution à l'étude du risque incapacité de travail dans les assurances de groupe'*
- 5) **FOUILLET** (ISFA 1997) : *'Etude d'un portefeuille d'arrêts de travail et élaboration de lois de maintien en incapacité et invalidité'*
- 6) **DONNART** (ISUP 1997) : *'Construction d'une table de mortalité' (confidentiel)*
- 7) **LHERBIERE** (ISUP 1997) : *'Estimation de taux d'entrée en arrêt de travail'*
- 8) **ALGARON, MAURENT et VIALAT** (CEA 1984) : *'Contribution à l'étude du risque 'Arrêt de Travail' dans les assurances de groupes, une approche des lois de maintien spécifiques par modélisation dynamique'*
- 9) **WOLFRUM** (ISUP 1992) : *'Une alternative non paramétrique au calcul des provisions techniques en Assurance Invalidité'*

COURS :

- 10) **SERANT** : Lois de maintien, ISFA
- 11) **PARTRAT** : Statistiques Inférentielles, ISFA

OUVRAGES :

- 12) **LAPLANCHE , COM-NOUGE et FLAMANT** : *'Méthodes statistiques appliquées à la recherche clinique'*, collection statistique en biologie et en médecine, Médecine-Sciences Flammarion
- 13) **HILL, COM-NOUGUE, FRAMAR, (...)** : *'Analyse statistique des données de survie'*, collection statistique en biologie et en médecine, Médecine-Sciences Flammarion
- 14) **BCAC** (note du 27/08/1993): *'Guide de l'assurance de groupe'*
- 15) **LONDON** : *'Graduation and the Revision of Estimates'*, Actex

ARTICLES :

- 16) **SERANT et RULLIERE** (1997) : *'Généralisation de l'estimateur de Kaplan-Meier d'une loi de durée de maintien en présence d'observations tronquées à gauche. Extension à l'étude conjointe de deux durées de maintien'*, Bulletin français d'actuariat ,Vol. 1, n°2.
- 17) **SERANT et RICHARD** (1999) : *'Un estimateur non paramétrique construit sur le concept d'exposition au risque de la loi conjointe de deux durées de maintien'*, Bulletin français d'actuariat, Vol. 3, n°5.
- 18) **WETZEL** : *'D'une loi générale de maintien en Incapacité de travail'*, Bulletin IAF n°245.
- 19) **FFSA** (sept. 2000): *'Annuaire des assurances de personnes 2000'*.

COMPTE RENDU DE CONFERENCE :

- 20) **GERBER** (1993) : *Un aperçu des méthodes actuarielles de lissage*, Université de Lausanne, ISFA.
- 21) **PICARD** (1993) : *L'estimateur de Kaplan-Meier et l'approche non paramétrique*, ISFA.
- 22) **SERANT** (1993) : *Modèles de durée de vie et logiciels statistiques*, ISFA.
- 23) **SERANT** (1993) : *Partition d'une Population en Groupes Homogènes*, ISFA.
- 24) **SERANT** (1993) : *Principes de Construction d'une Table de Mortalité avec ou sans Censures*, ISFA.

LOIS :

- 25) **ARRETE du 28 mars 1996** (J.O. 30/04/1996) qui a créé l'article A331-22 du Code des Assurances sur le provisionnement des prestations d'incapacité et d'invalidité.
- 26) **LOI EVIN du 31 décembre 1989** (J.O. 2/01/1990) renforçant les garanties offertes aux personnes assurées contre certains risques.