

ISUP

PROMOTION 2010

Mémoire présenté devant

**L ' I n s t i t u t d e S t a t i s t i q u e
De l'Université Pierre et Marie Curie**

Pour l'obtention du

**Diplôme de Statisticien
M e n t i o n A c t u a r i a t**

Assurance ☒

Finance ☐

Par Mlle Emilie Despierres

Sujet : Construction et validation de tables d'expérience pour le provisionnement d'un portefeuille de rentes éducation.

Lieu du stage : Optimind

Responsable du stage : M. Thibaud Hager

Invité(s) :

CONFIDENTIEL ☒

Résumé :

Mots clés : table d'expérience ; rente éducation ; enfant à charge ; fonction de survie ; Kaplan-Meier ; lissage ; segmentation ; modèle à hasard proportionnel de Cox ; provisions ; backtesting.

Ce mémoire a pour objectif de répondre à la problématique réelle suivante : améliorer le provisionnement de la garantie rente éducation d'un régime national de prévoyance d'une branche professionnelle. Dans notre cas l'amélioration du provisionnement passe par la prise en compte de l'expérience du portefeuille pour déterminer les probabilités de payer les flux futurs, autrement dit par la construction de tables d'expérience à partir des données du portefeuille.

Pour bénéficier d'une rente éducation au décès d'un parent participant, l'orphelin doit, en plus d'être en vie, satisfaire à d'autres conditions (d'âge, de scolarité, de chômage) ; toutes ces conditions étant résumées par le statut « enfant à charge ».

Ainsi de la même manière que l'on utilise des tables décrivant la durée de la vie humaine pour provisionner des rentes en cas de vie, nous aurons besoin ici de tables d'expérience décrivant la survie de l'enfant dans l'état « enfant à charge » afin de provisionner au plus juste ces garanties.

Avant de se lancer dans la construction de la table, nous avons créé une première loi, dite de référence à l'aide de données de marché représentatives de la population française. Cette démarche nous a permis d'avoir une idée de la forme a priori que peut avoir une loi de survie dans l'état « enfant à charge ».

Pour obtenir la loi de survie d'expérience, nous avons ensuite réalisé toutes les étapes nécessaires à l'obtention d'une table fiable. L'analyse du portefeuille et la validation des données initiales ont été des étapes primordiales donnant du crédit aux résultats.

Ces derniers ont été acquis en estimant les taux bruts de sortie par la méthode de Kaplan-Meier et en les lissant avec la méthode non paramétrique de Whittaker-Henderson.

Nous avons ensuite souhaité segmenter le portefeuille en groupes de risques plus homogènes avec la variable « âge de l'enfant au décès du parent ».

La construction d'une loi d'expérience propre à chaque groupe avec les mêmes méthodes que pour la loi unique et dans un deuxième temps avec le modèle semi-paramétrique de Cox ne s'est pas révélée concluante.

En effet en étudiant les risques liés aux tables d'expérience, la prise en compte du risque d'estimation résultant d'un volume de données insuffisant pour la création des lois nous a conduit à préférer l'utilisation d'une loi unique représentant le portefeuille dans sa globalité.

Les derniers tests de validation sur cette loi unique basés sur des principes « backtesting » ayant été probants, l'utilisation de la table d'expérience permettra à notre client de provisionner de manière plus réaliste cette garantie.

Abstract :

Key words : experience table; education annuity ; survival function ; Kaplan-Meier ; smoothing ; segmentation ; Cox proportional hazards model ; reserves ; backtesting.

This report aims to resolve the following issue : to improve the reserving for education annuity guarantees.

In our case the improvement of reserving involves considering the experience of the portfolio to calculate the probability of future cash-flows. In other words, we have to determinate experience tables with portfolio data.

To enjoy education annuity at the death of a parent, the orphan must be alive and also meet other requirements (age, schooling, unemployment). All requirements could be summed up by « dependent child » status.

In the same way as we need life tables to work out annuity reserves, we need here experience tables to describe child survival time in « dependent child » status.

Before constructing the table, we created a first law with market data. This approach enabled us to have an idea about form for the distribution of survival time in « dependent child » status.

Next, we have carried out all necessary steps to have a reliable table. Preliminary steps such as portfolio analysis and data validation were essential to have good results.

Then, we have built gross exit rates with the Kaplan-Meier method and smoothed rates with the Whittaker Henderson method of graduation.

Next, we wanted to segment the portfolio into homogeneous risk groups with « age of the child at the date of parent's death » variable. To build an experience table for each group, we have tested the same method seen above and Cox model.

But by studying the estimation risk resulting from low volume data used for construction, prompted us to choose sole table representing the overall portfolio.

Final validation tests based on the concept of backtesting have been positive. Thus it is suggested using experience table. Indeed it allows our client to estimate fair and reasonable reserves based on his own experience.

Remerciements :

Je souhaite tout d'abord remercier Christophe Eberlé, président de la société de conseil Optimind pour son accueil au sein de sa société.

Je souhaite remercier tout particulièrement Thibaud Hager pour son suivi et ses conseils tout au long de mon stage. Il a su se rendre disponible malgré un emploi du temps déjà chargé. Je le remercie également de m'avoir donné l'opportunité de travailler avec un acteur du marché sur une problématique concrète.

Je remercie également l'entreprise partenaire avec laquelle les échanges ont été efficaces et cordiaux.

Mes remerciements s'adressent aussi à Olivier Lopez pour ses conseils techniques et sa grande réactivité ainsi qu'à Arnaud Cohen pour ses conseils expérimentés.

Enfin je remercie toutes les personnes qui ont contribué à la réalisation de cette étude.

Sommaire :

Introduction :	11
1 Contexte et objectif de l'étude	13
1.1 Contexte de l'étude : Etude d'un portefeuille de rentes éducation d'un régime de prévoyance collective	13
1.1.1 L'étude : Au cœur d'un partenariat « Recherche & Développement »	13
1.1.2 Présentation de la prévoyance collective complémentaire	14
1.1.3 Le marché de la prévoyance collective	18
1.1.4 La couverture du risque décès en prévoyance	20
1.2 Objectif de l'étude : Construction de tables d'expérience	23
1.2.1 Des modèles de durées aux tables d'expérience	23
1.2.2 Les tables d'expérience	25
1.2.3 Les tables d'expérience dans le contexte Solvabilité II	26
1.2.4 Les tables d'expérience pour les rentes éducation	27
1.2.5 Les grandes étapes de la construction d'une table d'expérience	29
2 Analyse et réflexion autour du portefeuille « rentes éducation »	31
2.1 Le produit étudié	31
2.2 Construction d'une loi dite de référence	33
2.2.1 Indicateurs synthétiques de la durée de scolarisation	33
2.2.2 Construction d'une loi de référence à partir de données de marché	35
2.3 Les données	43
2.3.1 Présentation	43
2.3.2 Création d'une base sinistre exploitable	44
2.4 Etudes statistiques	47
2.4.1 Statistiques descriptives du portefeuille	47
2.4.2 Facteurs discriminants a priori	49
2.5 Choix déterminants pour la construction des tables	50
2.5.1 Choix de la période d'observation	50
2.5.2 Format des tables	50
Conclusion de la partie 2	51
3 Construction de lois de survie dans l'état « enfant à charge » à partir des données d'expérience	53
3.1 L'estimation brute des taux de sortie	53
3.1.1 Généralités	53
3.1.2 Données incomplètes	53
3.1.3 La méthode de Kaplan-Meier	54
3.2 Le lissage des taux bruts de sortie	58
3.2.1 Généralités	58
3.2.2 La méthode de Whittaker-Henderson	59
3.2.3 Positionnement par rapport à la loi de référence	60
3.2.4 Critère de validation des courbes lissées	61
3.3 Estimation et lissage des taux bruts du portefeuille global	63

3.3.1	Taux bruts	63
3.3.2	Lissage	64
3.3.3	Adéquation de l'ajustement.....	67
3.4	Segmentation du portefeuille en sous-populations : estimation et lissage	70
3.4.1	Taux bruts	72
3.4.2	Lissage et adéquation.....	73
3.4.3	Les 5 courbes	75
3.5	Comparaison des courbes de survie et analyse de l'impact sur le provisionnement.....	76
3.5.1	Comparaison des courbes de survie.....	76
3.5.2	Impact sur le provisionnement.....	79
	Conclusion de la partie 3	80
4	Une autre approche, la modélisation semi-paramétrique : le modèle de Cox..	81
4.1	Présentation du modèle	81
4.1.1	Généralités.....	81
4.1.2	De la vraisemblance du modèle de Cox à la vraisemblance partielle de Cox	82
4.1.3	Estimation des paramètres du modèle	85
4.1.4	Tests statistiques	85
4.2	Mise en œuvre	86
4.2.1	Le modèle avec nos données	86
4.2.2	L'estimation des paramètres	87
4.3	Validité du modèle.....	88
	Conclusion de la partie 4 :.....	91
5	Validation et utilisation des tables	93
5.1	Les risques liés aux tables d'expérience	93
5.1.1	Risque de déformation	93
5.1.2	Risque d'estimation.....	94
5.2	Une méthode de validation : Le backtesting	96
5.2.1	Vérification de l'équilibre « sorties observées / sorties attendues ».....	96
5.2.2	Vérification de l'équilibre « charge constatée / charge provisionnelle »	96
5.3	Comparaison des résultats du backtesting sur la loi d'expérience construite et sur la loi utilisée actuellement.....	97
5.3.1	Vérification de l'équilibre « sorties observées / sorties attendues ».....	97
5.3.2	Vérification de l'équilibre « charge constatée / charge provisionnelle »	98
	Conclusion de la partie 5	100
	Conclusion.....	101
	Annexe 1 : Données de marché utilisées pour construire la loi de référence.....	102
	Annexe 2 : Traitement des données	103
	Annexe 3 : Estimateur de Greenwood.....	104
	Bibliographie	105

Introduction :

Les prochaines évolutions guidées notamment par la directive Solvabilité II conduisent les organismes d'assurance à identifier et évaluer leurs risques avec une vision de plus en plus économique.

C'est dans ce contexte que ces derniers basculent peu à peu vers une logique de provisionnement basée non plus sur des hypothèses prudentes mais sur des hypothèses réalistes.

Ce mémoire effectué dans le cadre d'un partenariat « recherche et développement » avec un acteur incontournable du marché français de la prévoyance complémentaire a pour vocation de répondre à un besoin réel, à savoir mieux connaître le risque lié aux garanties rentes éducation.

En réponse à cette problématique, nous allons prendre en compte l'expérience du portefeuille pour évaluer la probabilité de versements de flux futurs, autrement dit construire des tables d'expérience décrivant la durée passée dans l'état enfant à charge, état donnant lieu à une prestation en cas de décès du parent participant.

La première partie a pour objectif de définir le contexte de l'étude, nous y décrivons le marché de la prévoyance complémentaire collective et les mécanismes liés aux rentes éducation. Nous présentons également les outils mathématiques et les grandes lignes du processus de construction d'une table d'expérience.

La deuxième partie présente les travaux effectués en amont de la conception de la table.

- Recherche de données de marché concernant les durées en jeu,
- Vérification de la qualité des données,
- Analyse du portefeuille.

Toutes ces étapes sont importantes pour s'assurer que l'on construit des tables sur des informations fiables et maîtrisées.

La troisième partie, cœur de ce mémoire présente aussi bien les aspects théoriques que pratiques liés à l'élaboration des tables d'expérience.

En plus de la construction d'une loi unique représentative de tout le portefeuille, nous tenterons de segmenter le portefeuille en sous-groupes de risques plus homogènes vis-à-vis de la durée étudiée et de construire une loi d'expérience pour chaque groupe.

D'un point de vue technique, nous estimerons les taux bruts de sortie par la méthode de Kaplan-Meier et deux méthodes de lissages seront traitées en pratique, à savoir le positionnement par rapport à une référence externe et le lissage non paramétrique de Whittaker-Henderson.

Dans une quatrième partie, nous testerons le modèle à hasard proportionnel de Cox sur le portefeuille segmenté pour trouver un rapport de risque entre les différents groupes.

Enfin dans une cinquième partie, nous nous pencherons sur les différents risques inhérents à l'élaboration et à l'utilisation de tables, plus particulièrement sur les risques de déformation et d'estimation. Puis nous proposerons une loi candidate à la validation, sur laquelle sera effectuée deux test basés sur des principes « backtesting » afin de d'homologuer définitivement la loi pour le provisionnement.

1 Contexte et objectif de l'étude

1.1 Contexte de l'étude : Etude d'un portefeuille de rentes éducation d'un régime de prévoyance collective

1.1.1 L'étude : Au cœur d'un partenariat « Recherche & Développement »

Optimind est une société de conseil spécialisée dans le domaine de l'actuariat et des systèmes d'informations. Ses clients sont des organismes d'assurance, des banques et des grandes entreprises. La société attache une réelle importance à la recherche et au développement et a sans cesse l'objectif de relier la théorie actuarielle aux problématiques réelles de ses clients. C'est dans ce sens qu'officialie la cellule « Recherche & Développement ». Elle relie les travaux de recherche scientifique avec des travaux de recherche organisés par la profession. Des Pôles de Compétences (PCO) ont ainsi été mis en place dans le but de permettre une transmission du savoir-faire et de poursuivre certains travaux de recherche actuarielle.

C'est donc dans l'idée de relier la théorie à une problématique réelle que nous avons conclu un accord de partenariat « Recherche & Développement » avec un grand groupe de protection sociale dans le cadre de ce mémoire.

L'objet du partenariat est la construction de tables d'expérience propres aux enfants valides d'une branche professionnelle du portefeuille rentes éducation de notre client. En plus de traiter les aspects théoriques liés au sujet, l'objectif est également de répondre aux attentes d'un client de façon pratique.

Voici en quelques mots les principaux engagements des deux parties dans l'accord de partenariat :

Le groupe de protection sociale s'est engagé à délivrer :

- Les données du portefeuille rentes éducation permettant de construire une ou plusieurs tables d'expérience.
- La documentation commerciale et technique associée à ces données.
- Toutes les informations nécessaires au bon déroulement de l'étude.

Optimind s'est engagé en contrepartie de la mise à disposition des informations à fournir :

- Les résultats des travaux de recherche.
- Les principales sources électroniques, synthèses, notes, programmes informatique et mémoire de stage.

Il est convenu également que toutes les informations échangées dans le cadre de ce partenariat sont confidentielles, notamment les données du portefeuille de rentes éducation et les résultats obtenus. Une attention particulière a été apportée pour respecter cette exigence. Comme le mentionne un article du code de déontologie du syndicat des Actuaire Conseils et Actuaires Experts indépendants, « l'actuaire-conseil traite avec totale confidentialité les informations propres à son client ».

Afin de cerner au mieux le périmètre d'action de notre client, nous présenterons dans un premier temps la prévoyance collective complémentaire de façon générale. Puis nous étudierons les différents acteurs ainsi que leurs marchés, nous pourrons alors positionner notre client dans cet environnement.

Enfin nous verrons quelles sont les garanties usuelles en prévoyance, en portant une attention particulière aux garanties en cas de décès et notamment aux rentes éducation.

1.1.2 Présentation de la prévoyance collective complémentaire

a. Généralités

La loi Evin (décembre 1989) définit la prévoyance comme étant « *une opération ayant pour objet la prévention et la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque chômage* ».

Ces risques sont couverts par les régimes de base de la sécurité sociale (pilier I), auxquels viennent s'ajouter les régimes complémentaires négociés dans l'entreprise ou par branches professionnelles (pilier II), et les assurances individuelles (pilier III).

Nous appellerons prévoyance complémentaire les régimes relatifs aux piliers II et III.

Quelle que soit la forme que prend cette prévoyance, elle constitue désormais un relais nécessaire aux prestations de base de la sécurité sociale, en particulier en raison du recul progressif des couvertures qu'elle propose.

La prévoyance complémentaire collective à adhésion obligatoire offre des conditions tarifaires, fiscales et sociales plus avantageuses que la prévoyance individuelle ou que la prévoyance collective à adhésion facultative.

Ces garanties bénéficient à tous les salariés d'une même entreprise ou d'une même branche professionnelle, ou à une ou plusieurs catégories d'entre eux, définies de manière objective, générale et impersonnelle. Cela signifie qu'une catégorie objective de personnel ne peut pas être définie à partir de critères relatifs à la durée de travail, l'âge du salarié, son ancienneté ou son contrat de travail.

De plus à partir du moment où l'employeur contribue aux cotisations, tous les salariés concernés doivent adhérer au contrat de prévoyance mis en place dans l'entreprise ou dans la branche professionnelle. Sans ce caractère obligatoire, l'équité dans l'entreprise ne serait pas respectée et la participation de l'employeur pourrait être assimilée à des avantages en nature.

b. Une protection sociale à l'initiative des partenaires sociaux

En matière de prévoyance, la seule obligation qui s'impose à toute entreprise est la garantie décès pour les cadres prévue dans la convention collective nationale du 14 mars 1947. En matière d'incapacité de travail, la loi de mensualisation de 1973 oblige également les employeurs à garantir un certain niveau de salaire en cas d'arrêt de travail.

En dehors de ces deux obligations, la mise en place de garanties de prévoyance relève entièrement de l'initiative des partenaires sociaux des branches professionnelles et des entreprises.

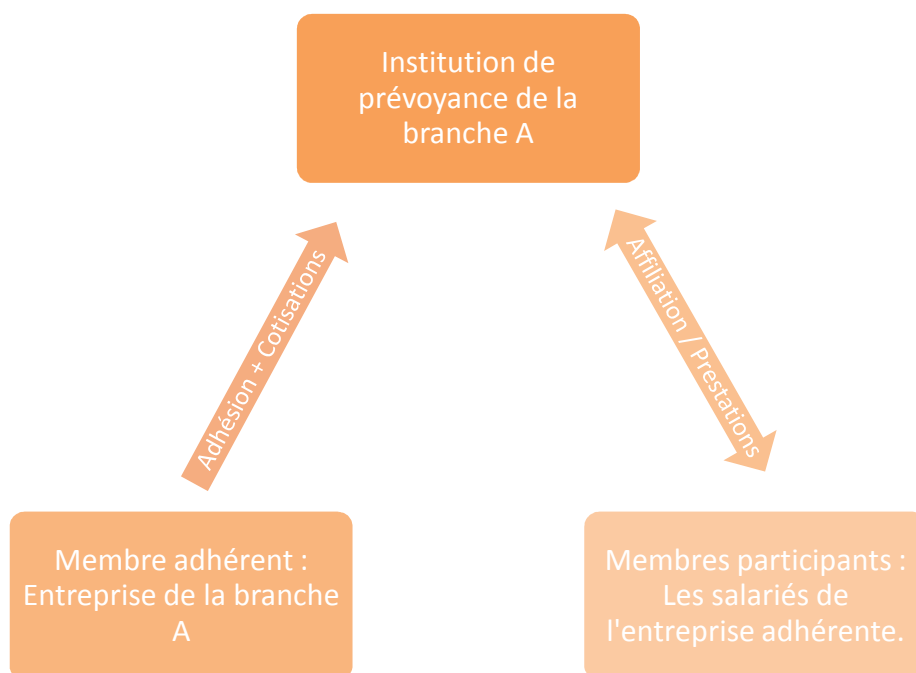
Lorsqu'un régime de prévoyance est négocié au niveau de la profession puis étendu au plan national, toutes les entreprises relevant de la convention collective doivent s'y conformer.

A défaut d'accord de branche, la prévoyance complémentaire n'a pas de caractère obligatoire pour l'entreprise. L'exonération de cotisations sociales de la part employeur constitue alors la condition de son développement. Elle est aussi la contrepartie du caractère collectif et obligatoire de l'adhésion des salariés, sans lequel il ne peut y avoir une mutualisation solidaire au sein de l'entreprise ou de la branche.

Quatre étapes sont indispensables à la mise en place de la prévoyance dans l'entreprise :



Le diagramme ci-dessous présente le mécanisme d'un contrat de prévoyance d'une entreprise relevant d'une branche professionnelle particulière notée branche A.



c. Un cadre législatif et conventionnel

Au fil du temps les pouvoirs publics ont permis de renforcer les garanties des assurés et d'améliorer la protection de leurs droits. Les principales lois ou décrets ayant marqué la protection sociale des salariés sont décrits ci-dessous.

La convention collective nationale pour les cadres de mars 1947

Cet accord crée une obligation pour l'employeur de souscrire une garantie décès pour les cadres. La contribution minimum de l'employeur est de 1,50% du salaire de la tranche A.

La mensualisation du 19 janvier 1978

Cette loi impose aux employeurs d'assurer aux salariés, sous certaines conditions, un certain niveau de salaire en cas d'arrêt de travail.

La loi Evin du 31 décembre 1989

Evènement majeur pour la prévoyance, cette loi renforce les garanties et améliore la protection des salariés. Cette loi précise les conditions de la couverture collective et les conditions de mise en place d'un régime de prévoyance.

La loi du 8 août 1994

Cette loi marque une nouvelle étape importante pour la prévoyance, qui devient un secteur d'activité autonome. La séparation juridique des activités retraite et prévoyance redessine le paysage de la protection sociale complémentaire. Les institutions de prévoyance sont alors considérées comme des entreprises d'assurance à part entière au sens des directives européennes.

De plus en cas de changement d'assureur, l'entreprise souscriptrice est chargée d'organiser :

- La revalorisation des prestations en cours de service lors du changement.
- Le maintien des garanties décès aux personnes en arrêt de travail au moment du changement d'assureur.

Ce dernier point est précisé par la loi du 17 juillet 2001.

La loi du 17 juillet 2001

L'article 7 de la loi Evin précise que « *la résiliation ou le non renouvellement du contrat ou de la convention est sans effet sur le versement des prestations immédiates ou différées* ». La nature des prestations visées dans cet article n'étant pas mentionnée, la loi du 17 juillet 2001 rend obligatoire dans les contrats collectifs, le maintien des garanties décès aux personnes en arrêt de travail.

Ainsi, depuis le 1^{er} janvier 2002, l'assureur est dans l'obligation de maintenir les garanties décès des contrats collectifs aux personnes en arrêt de travail même en cas de résiliation ou de non renouvellement du contrat. Cet engagement doit faire l'objet d'un provisionnement.

En cas de changement d'assureur suite à une résiliation ou un non-renouvellement, deux possibilités existent :

- Soit les garanties sont maintenues par l'ancien assureur contre le versement par le souscripteur d'une prime de résiliation, celle-ci est destinée à couvrir la charge de provisionnement restant à constituer.
- Soit les engagements sont repris par le nouvel assureur avec transfert des provisions déjà constituées par l'ancien.

La loi Fillon du 21 août 2003

Cette loi porte sur la réforme des retraites, mais elle a également modifié les règles d'exonération sociale et de déductibilité fiscale des cotisations au titre de la prévoyance. Ces nouvelles règles s'appliquent à tout contrat collectif, existant ou nouveau, à compter du 1^{er} janvier 2009.

Exonération sociale :

La contribution de l'employeur pour chaque salarié est exonérée de cotisations de sécurité sociale jusqu'à 6% du PASS + 1,5% de la rémunération annuelle brute du salarié, le total ne pouvant dépasser 12% du PASS.

La contribution de l'employeur est néanmoins soumise à la CSG/CRDS et à une taxe de 8% pour les entreprises de plus de 9 salariés.

Déduction fiscale :

La totalité de la cotisation est déductible du revenu imposable du salarié dans la limite de 7% du PASS + 3% de la rémunération brute annuelle, le total ne pouvant dépasser 24% du PASS.

En 2010, le PASS s'élève à 34 620 euros (soit 2 885 euros par mois).

L'article 14 de l'accord national interprofessionnel (ANI) du 11 janvier 2008

Cet article a mis en place un mécanisme de « maintien des garanties santé et prévoyance d'entreprise ». Depuis le 1^{er} juillet 2009, ce dispositif permet aux salariés licenciés qui le souhaitent de continuer à bénéficier des garanties santé et prévoyance mises en place dans leur entreprise pendant une durée de 9 mois maximum.

Cette « portabilité » des droits est souvent financée par un système de mutualisation où les cotisations des salariés en activité financent la portabilité des salariés licenciés.

Un autre article a également assoupli deux critères ouvrant droits au versement de l'indemnité conventionnelle de la maladie en cas d'arrêt de travail : le délai d'ancienneté passe de trois à une année et le délai de carence de onze à sept jours. (26 juin 2008).

1.1.3 Le marché de la prévoyance collective

a. Les différents acteurs

Les institutions de prévoyance

Ce sont des personnes morales de droit privé à but non lucratif, administrées paritairement par des membres adhérents et des membres participants (Art. L931-1). Elles sont régies par le livre IX du code de la sécurité sociale et relèvent des directives européennes « Assurance ».

Elles n'ont pas d'actionnaires à rémunérer, ainsi les résultats servent à renforcer la sécurité des engagements, à améliorer le niveau des garanties et la qualité des services. Relevant de la directive européenne « Assurance », elles sont soumises aux mêmes règles techniques que toutes entreprises d'assurance. Elles offrent ainsi un très bon niveau de sécurité dans les gestions des engagements.

Ils existent trois types d'institutions de prévoyance :

- Les institutions professionnelles : leur périmètre d'intervention s'étend à une branche professionnelle, à une profession ou à une catégorie professionnelle.
- **Les instituts interprofessionnels** : leur périmètre d'intervention s'étend à plusieurs branches professionnelles, professions ou entreprises.
- Les institutions d'entreprise qui sont issues de la conclusion d'une convention ou d'un accord collectif, ou de la ratification d'un projet de l'employeur par les salariés.

La plupart des institutions de prévoyance appartiennent à un groupe de protection sociale. Celui-ci regroupe en plus des institutions de prévoyance, des institutions de retraite complémentaires, des mutuelles, parfois même des sociétés d'assurance ou de gestion d'épargne salariale. Pour l'entreprise c'est l'avantage d'un même interlocuteur pour tous ses besoins en termes de protection sociale complémentaire de ses salariés.

Les sociétés d'assurances

Elles sont régies par le code des assurances et peuvent prendre différentes formes juridiques : société anonyme, société d'assurance mutuelle ou société mutuelle d'assurance. Contrairement aux autres intervenants, elles ont un but lucratif sur le marché de la prévoyance. Elles doivent obéir à des règles très strictes en matière de solvabilité et respecter des normes visant à protéger les assurés.

Les mutuelles

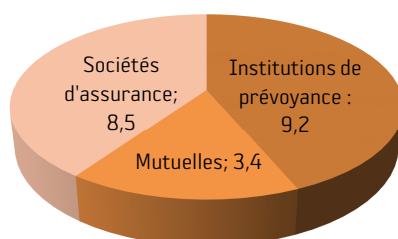
Elles sont régies par le code de la mutualité. Celui-ci les définit comme « *des groupements à but non lucratif qui, essentiellement au moyen des cotisations de leurs membres, se proposent de mener, dans l'intérêt de ceux-ci ou de leur famille, une action de prévoyance,...* ». Ce sont des associations sur le plan juridique.

b. Leurs marchés

Les garanties de prévoyance collective se concentrent aujourd'hui essentiellement vers la protection du salarié et de sa famille. Elles permettent le plus souvent le maintien du revenu en cas d'arrêt de travail temporaire (incapacité) ou permanent (invalidité) et la mise en place de temporaires décès, de rentes de conjoint et d'éducation en cas de décès du salarié.

Ces garanties, qui constituent le socle de la prévoyance complémentaire doivent s'adapter aux évolutions socio-économiques de la société et aux politiques de l'emploi. Un des enjeux futur pour le marché sera d'intégrer la prise en charge de la dépendance notamment à travers un partenariat entre le secteur public et le secteur privé.

En 2009, les cotisations totales des contrats collectifs d'entreprise représentaient 21 milliards d'euros (ce chiffre comprend le périmètre de la prévoyance et de la santé).



Répartition des cotisations totales des contrats collectifs en 2008 en Mds d'€.

Il faut tout de même mentionner que la crise économique de 2009 a pesé sur le marché de la prévoyance. Les cotisations étant assises sur la masse salariale et celle-ci ayant baissé ; le chiffre d'affaire de la prévoyance a également baissé.

En 2009, les institutions de prévoyance sont les premiers opérateurs en prévoyance collective, ils enregistrent une hausse de 3,2% de leur activité avec 11,4 milliards d'euros de cotisations au titre des opérations directes (ce chiffre prend en compte la prévoyance au sens large, il englobe les cotisations des complémentaires santé, prévoyance et retraite supplémentaire).

Ce taux de progression masque en réalité un faible développement de l'activité de prévoyance. S'élevant à 5,1 milliards d'euros, les cotisations liées aux couvertures de prévoyance classique (décès, incapacité

de travail et invalidité, dépendance) augmentent de seulement 1,5% contre 4% en 2008. La diminution de la masse salariale due à la crise de l'emploi a fortement pesée sur le niveau de cotisations.

Cotisations (Opérations directes, hors réassurance et acceptations) :			
	2008	2009	Variation 2008/2009
Décès, rentes éducation et de conjoint	2250	2269	0.8%
Incapacité, invalidité, mensualisation	2811	2865	1.9%
Dépendance	18	19	5.6%
Total prévoyance	5079	5158	1.5%

Evolution des cotisations en M d'€, source : CTIP.

La charge des prestations, qui intègrent les prestations versées sous forme de rentes, de capitaux ou de remboursements de frais médicaux ainsi que la variation des provisions techniques, enregistre une augmentation de 2.6%. Le tableau ci-dessous montre l'évolution des garanties liées aux couvertures de prévoyance classique (décès, incapacité de travail et invalidité, dépendance).

Prestations (versées au cours de l'exercice + variation de provisions mathématiques et des prestations à payer, hors frais de gestion) :			
	2008	2009	Variation 2008/2009
Décès, rentes éducation et de conjoint	1627	1553	-4.5%
Incapacité, invalidité, mensualisation	2493	2633	5.6%
Dépendance	11	15	30.17%
Total prévoyance	4131	4201	1.7%

Evolution des prestations par garanties en 2009 en M d'€

1.1.4 La couverture du risque décès en prévoyance

a. Le régime de base

Les ayants droits du défunt vont percevoir des prestations de la part de la sécurité sociale. Ces prestations sont fonction du régime auquel l'assuré est rattaché et des causes du décès.

Par exemple, les ayants droits des salariés du secteur privé peuvent prétendre au versement d'un capital décès sous certaines conditions. Cette indemnité, égale aux trois derniers mois de salaire de l'assuré permet de faire face aux frais immédiats entraînés par le décès (les frais d'obsèques par exemple).

Le régime général prévoit également de protéger les veufs et veuves précoces au moyen d'une allocation veuvage s'élevant à 565.13€ par mois (pour une durée maximale de 2 ans). Cependant la loi du 21 août 2003 a prévu une suppression de l'allocation veuvage en 2011, date à laquelle toute personne dans cette situation pourra demander, sous conditions de ressources, une pension de réversion, auparavant réservée aux veuves ou veufs de 55 ans ou plus.

Parallèlement à ces dispositifs les conjoints survivants peuvent percevoir des prestations familiales du fait de leur nouvelle situation :

- L'allocation de parent isolé (API) qui concerne toute personne vivant seule ayant un ou des enfants à charge est versée sous conditions de ressources pour une durée d'un an ou jusqu'aux trois ans de l'enfant.
- L'allocation de soutien familiale (ASF) est versée sans conditions de ressources aux orphelins de père ou de mère jusqu'aux 20 ans de l'enfant. Elle s'élève à 87.14€ par mois (orphelin d'un parent) ou à 116.18€ par mois (orphelin des deux parents).

b. Une couverture complémentaire nécessaire

Au regard de la faiblesse des prestations des régimes obligatoires en matière de décès, une couverture complémentaire est particulièrement utile. Elle permettra d'assurer la sécurité financière des proches en cas de décès par le versement d'un capital ou d'une rente au bénéfice du conjoint et/ou des enfants. Ces prestations s'ajouteront à celles versées par les régimes obligatoires.

Le versement d'un capital décès

La plupart des contrats de prévoyance collective prévoient des garanties liées au décès, souvent ces garanties prennent la forme d'un capital versé à la survenance du décès. On rappelle d'ailleurs que les employeurs sont tenus de verser une cotisation minimale pour la couverture décès des salariés cadres. Le capital décès est généralement calculé en fonction de la rémunération annuelle brute du salarié assuré. Il est le plus souvent exprimé en pourcentage de la totalité de cette rémunération ou d'une ou plusieurs tranches de celle-ci.

Certaines garanties annexes peuvent s'ajouter au capital décès :

- Décès par accident : doublement du montant du capital
- Garantie « double effet » : en cas de décès du conjoint survivant, versement d'un capital aux enfants à charge
- Garantie décès d'un membre de la famille : versement d'un capital à l'assuré.
- Garantie « obsèques » : versement d'une somme forfaitaire destinée à couvrir les frais d'obsèques.

La rente de conjoint et d'orphelin

Le décès de l'assuré peut également donner lieu au versement d'une rente au conjoint survivant. La rente de conjoint peut être déterminée en fonction des droits acquis auprès des régimes de retraites complémentaires pendant la période d'activité de l'assuré, ou être calculée sur la base de son dernier salaire.

La rente de conjoint est constituée :

- Soit d'une rente viagère seule,
- Soit de cette rente viagère, à laquelle s'ajoute une rente temporaire.

Si le conjoint survivant remplit les conditions d'ouverture des droits de réversion des régimes complémentaires de retraite dès le décès, il perçoit la rente viagère en complément de sa pension de réversion. Si le conjoint survivant ne remplit pas les conditions d'ouverture des droits de réversion des régimes complémentaires de retraite au moment du décès, il perçoit la rente temporaire en complément de la rente viagère. Une fois que ses droits à la pension de réversion sont ouverts, l'organisme assureur ne lui versera plus que la rente viagère.

Dans le cas du décès simultané ou postérieur du conjoint survivant, certaines formules prévoient le versement aux orphelins d'une rente temporaire correspondant à un pourcentage (le plus souvent 30%, 40% ou 50%) de la rente de conjoint survivant.

La rente d'éducation

Généralités

La rente d'éducation a pour objet d'assurer le versement d'une rente **aux enfants à charge** d'un assuré si celui-ci décède ou, dans certains cas, s'il est reconnu en état d'invalidité absolue et définitive. Cette rente a pour but d'assurer l'avenir financier des enfants. Elle est versée temporairement aux enfants à charge survivants ayants-droits de l'assuré sous certaines conditions.

Versement de la rente d'éducation

Elle est généralement calculée en fonction du dernier salaire d'activité de l'assuré ou du plafond annuel de la sécurité sociale, auquel on applique un pourcentage qui peut être fixe ou variable en fonction de l'âge. Certains organismes prévoient un doublement de la rente si l'enfant est orphelin de ses deux parents.

Le maintien en rente

La rente d'éducation est versée aux enfants à charge. Chaque organisme définit sa notion « d'enfant à charge ». On trouve quasi systématiquement une condition sur l'âge limite de l'enfant et très souvent une condition sur la poursuite des études.

La rente éducation est servie à vie aux enfants reconnus invalides ou handicapés.

1.2 Objectif de l'étude : Construction de tables d'expérience

L'objectif de l'étude est de modéliser la durée passée dans l'état « enfant à charge » pour une population donnée.

Dans un premier temps, faisons un tour d'horizon théorique sur les modèles de durée à disposition, nous pourrons alors choisir lequel d'entre eux est le mieux adapté à notre étude. Ensuite nous verrons comment passer d'un modèle de durée à une table d'expérience directement exploitable par l'assureur ainsi que les grandes étapes d'une construction de table.

1.2.1 Des modèles de durées aux tables d'expérience

a. Généralités

Les modèles de durées sont utilisés dans de nombreux domaines de l'assurance. Que ce soit pour provisionner ses engagements ou tarifier ses contrats, l'assureur a besoin de modéliser de nombreuses durées, en voici une liste non exhaustive :

- la durée de la vie humaine pour tous les contrats vie/décès,
- la durée d'un arrêt de travail pour les contrats incapacité/invalidité,
- la durée de chômage pour certains contrats de prévoyance,
- la durée de scolarisation pour les contrats de rentes éducation.

Les modèles de durée présentent certaines particularités :

- ✓ Les données de durée sont engendrées par des variables aléatoires positives.
- ✓ La fonction de survie et la fonction de hasard ont une interprétation physique naturelle dans le cadre des modèles de durée et nous détaillerons celle-ci plus bas.
- ✓ Les données de durée utilisent en général des variables explicatives, ce qui implique une segmentation de la population en sous-populations homogènes.
- ✓ Les données de durée sont souvent incomplètes et font appel à des modèles dits tronqués et/ou censurés. L'expérience ayant souvent une durée limitée, les individus ne sont que partiellement observables sur cette durée.

Avant de présenter les différents modèles statistiques utilisés pour appréhender des durées, il est nécessaire de faire quelques rappels théoriques.

b. Rappels et notations

Considérons la variable aléatoire T représentant la durée passée dans l'état « enfant à charge ». Voici une liste de fonctions mathématiques qui permettent d'apprécier cette variable aléatoire sous différents angles.

- ❖ La fonction de répartition $F(t) = P(T \leq t)$ est la probabilité que la sortie de l'état « enfant à charge » de l'individu se produise entre 0 et t .

- ❖ La densité de F : $f(t) = \frac{dF(t)}{dt} = \lim_{h \rightarrow 0} \frac{P(t \leq T \leq t+h)}{h}$

- ❖ La fonction de survie est par définition, le complément à 1 de la fonction de répartition : $S(t) = P(T > t)$, elle est décroissante et telle que $S(0) = 1$ et $\lim_{t \rightarrow \infty} S(t) = 0$. Elle est la probabilité que la sortie de l'état enfant à charge de l'individu se produise après l'instant t .

- ❖ La fonction de survie conditionnelle représente la probabilité que le temps passé dans l'état « enfant à charge » d'un individu dépasse l'instant $t + u$, sachant qu'il était dans l'état à l'instant u .

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

- ❖ La fonction de hasard est souvent utilisée pour spécifier un modèle de durée car elle a une interprétation physique naturelle, on peut l'assimiler à la probabilité que l'enfant sorte de l'état à l'instant t sachant qu'il y a survécu jusqu'à cet instant. De grandes valeurs de cette fonction indiquent les âges pour lesquels le risque de sortie est important. Elle se définit comme ceci :

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

- ❖ La fonction de risque cumulé est donnée par :

$$\Lambda(t) = \int_0^t h(s) ds$$

La distribution de la variable aléatoire T repose sur l'une des quatre données suivantes, qui sont équivalentes : $S(t)$, $f(t)$, $h(t)$ et $\Lambda(t)$.

Il existe une relation qui lie le taux de hasard, le taux de hasard cumulé et la fonction de survie :

$$S(t) = \exp\left(-\int_0^t h(s) ds\right) = \exp(-\Lambda(t)).$$

c. Les familles de modèle de durée

Les différents modèles disponibles peuvent être classés en trois catégories distinctes :

- Les modèles paramétriques

Ils modélisent le plus souvent la fonction de hasard par des distributions connues telles que la loi exponentielle, la loi Log-normale, la distribution de Weibull. Outre ces lois de distributions usuelles,

le modèle de Gompertz-Makeham est un modèle de référence pour construire des tables de mortalité.

Avec ces modèles une hypothèse sur la forme du taux de hasard et donc sur la forme de la fonction de survie est posée. Les paramètres sont ensuite estimés sur la base des statistiques disponibles par des méthodes connues comme la méthode du maximum de vraisemblance ou la méthode des moindres carrés.

Ces modèles possèdent de nombreux avantages :

- ils permettent une extension en dehors de la plage d'observation,
- les paramètres sont souvent facilement interprétables,
- ces modèles sont très utiles si on a peu de données et que l'on connaît la forme a priori de la courbe à modéliser.

- Les modèles non paramétriques

Le premier réflexe est de calculer un estimateur empirique de la fonction de répartition avec la formule $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{T_i \leq t}$ cependant la variable aléatoire T_i n'est pas toujours observable car notre modèle est censuré. Les modèles non paramétriques font alors appel à d'autres estimateurs empiriques qui vérifient de « bonnes propriétés » en présence de censure et/ou de troncature. Les plus connus sont l'estimateur de Kaplan-Meier de la fonction de survie et l'estimateur de Nelson-Aalen du taux de hasard cumulé.

- Les modèles semi-paramétriques

Les modèles semi-paramétriques proviennent de la volonté de faire un compromis entre les forces et les faiblesses des deux approches précédentes.

Ainsi pour modéliser la fonction de survie, on souhaite faire appel à l'estimation de paramètres mais aussi à une estimation dite non paramétrique, ce qui permet un ajustement propre aux données.

Les modèles « à hasard proportionnel » sont des modèles semi-paramétriques, ils sont appelés ainsi car la fonction de hasard estimée est proportionnelle à une fonction de hasard « de référence ».

Il y a deux types de modèles à hasard proportionnel :

- Ceux dont la fonction de hasard dite « de référence » est connue
- Ceux dont la fonction de hasard dite « de référence » est inconnue. Ce dernier groupe forme le modèle de Cox.

Le modèle de Cox apparaît comme l'un des modèles fréquemment utilisés en présence de données censurées et d'une population hétérogène.

1.2.2 Les tables d'expérience

Une table qu'elle soit d'expérience ou réglementaire retrace la survie d'une population dans un état particulier au fil du temps. Elle peut se présenter sous forme d'un tableau à simple ou double entrée et donne le nombre de survivants dans l'état pour chaque unité de temps.

Ainsi, lors de la construction d'une table d'expérience il faut compter l'occurrence d'un événement (la sortie de l'état) sur une population **donnée** en fonction de variables explicatives.

Les événements peuvent être par exemple :

- le décès de l'assuré, i.e. la sortie de l'état « être en vie »
- la survenance d'un arrêt de travail (incapacité, invalidité, chômage), i.e. la sortie de l'état « être en vie et ne pas être en arrêt de travail ».
- la reprise du travail, i.e. la sortie de l'état « être en arrêt de travail ».
- la fin des études, i.e. la sortie de l'état « être scolarisé ».

L'obtention d'une table fiable requiert un certain nombre de précautions : en effet, il faut prendre en compte les évolutions possibles de la population, les contraintes d'observation (franchises, résiliations, qualité et quantité des données disponibles) et l'impact des couvertures sur le risque (sélection médicale, niveau et étendue des garanties).

L'utilisation la plus courante des tables d'expérience est de se substituer aux tables réglementaires pour la tarification et le provisionnement des risques « vie/décès » et « arrêt de travail ». La tarification et le provisionnement sont strictement encadrés par la réglementation qui impose des contraintes sur le taux technique utilisé et sur la table de survenance ou de maintien utilisée.

Ce sont les articles A335-1 et A331-22 du code des assurances qui fixent ce cadre réglementaire. Ils permettent aux organismes assureurs d'utiliser, sous certaines conditions, des tables d'expérience (à la place des tables réglementaires) construites sur la base des données propres au portefeuille, ou de toute autre source, pour autant qu'elles permettent de mieux évaluer les engagements contractés.

En effet il est utile de cerner au mieux tout comportement de la population assurée qui serait significativement différent de la population ayant servi à construire les tables réglementaires. De manière générale la construction d'une table d'expérience se fera toujours dans le but d'exprimer au mieux le comportement de la population assurée.

Outre les utilisations autorisées par la réglementation, la nécessité de l'utilisation de tables d'expérience augmente avec :

- l'application de la directive Solvabilité II
- les calculs d'Embedded value et de capitaux économique.
- les études de rentabilité internes

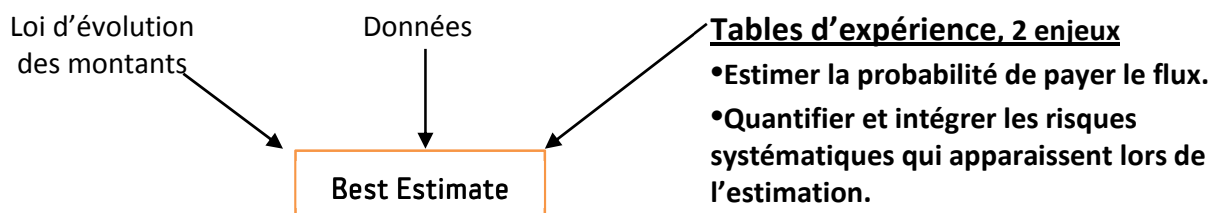
1.2.3 Les tables d'expérience dans le contexte Solvabilité II

L'intérêt porté par la directive Solvabilité II aux hypothèses réalistes conduit les assureurs à une prise en compte aussi large que possible de l'expérience du portefeuille. Dans ce contexte les tables d'expérience trouvent tout à fait leur place.

Sous Solvabilité II, la valorisation économique est basée sur le calcul d'un best estimate.

Le CEIOPS définit le best estimate comme étant :

« ...la moyenne pondérée en fonction de leur probabilité des flux futurs de trésorerie compte tenu de la valeur temporelle de l'argent, laquelle est estimée sur la base de la courbe des taux sans risque pertinente ».



Les tables d'expérience peuvent être vues comme un des paramètres nécessaires au calcul du best estimate.

Concernant celui-ci, il faut également poser une réflexion sur le niveau de segmentation à retenir pour les tables. Si on considère une table unique pour l'ensemble du portefeuille, on court le risque de n'être « best estimate » sur aucune des sous-populations homogènes composant le portefeuille et la table devient obsolète dès que le portefeuille se déforme dans le temps.

Il convient alors de déterminer une segmentation optimale de manière à capter l'hétérogénéité du portefeuille, elle est déterminée au cas par cas après une analyse du portefeuille. Cette segmentation implique de nouveaux risques qui doivent être quantifiés et intégrés dans la marge pour risque.

Dans une réglementation dite « prudentielle », les risques statistiques liés à la segmentation ne sont pas explicités, mais ils sont largement intégrés dans la marge de prudence imposée lors de la construction d'une table. Dès lors que l'on passe d'un univers « prudentiel » à un univers « réaliste », l'ensemble des risques impactant le niveau des engagements doit être explicités et pris en compte, on doit donc constater l'apparition de nouveaux risques appelés risque de modèle et risque d'estimation. Ces risques sont systématiques et potentiellement dangereux et de manière générale ils augmentent avec le niveau de segmentation retenu.

Le risque d'estimation augmente si le volume des données diminue, car la volatilité des estimateurs augmente.

Pour le risque de modèle, il faut poser un certain nombre de contraintes afin de structurer l'impact de l'hétérogénéité.

1.2.4 Les tables d'expérience pour les rentes éducation

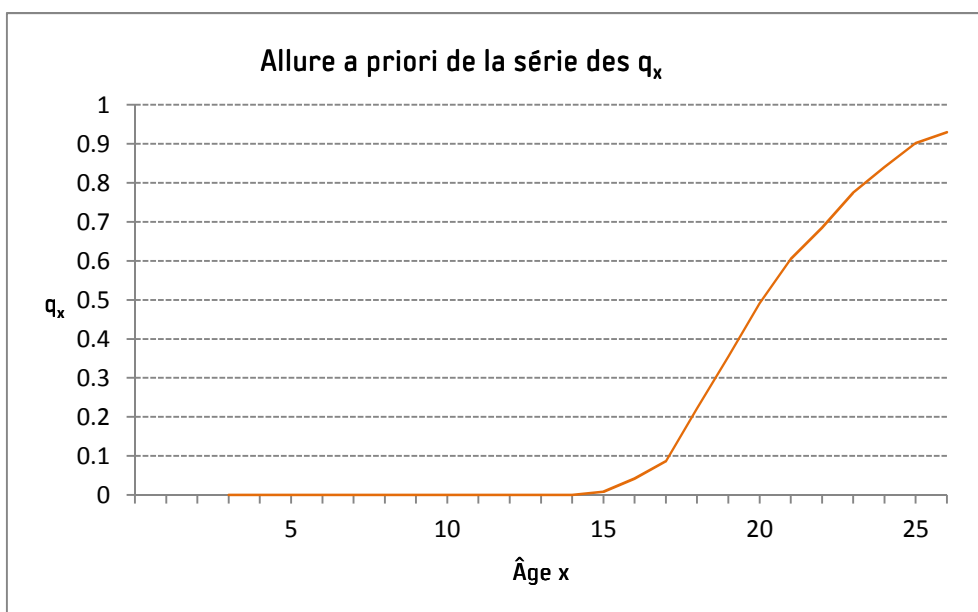
Les tables d'expérience qui vont être construites vont suivre le cheminement d'une génération fictive de 10 000 nouveau-nés à qui l'on fera subir aux différents âges les conditions de sortie observées sur la population réelle du portefeuille de rentes éducation. Les tables nous donneront pour la suite des anniversaires x (x variant de 0 à âge limite de la rente) :

- Le nombre théorique d'enfants dans l'état « enfant à charge » à l'âge entier x noté l_x
- Le nombre théorique de sorties entre x et $x + 1$ noté $d_x = l_x - l_{x+1}$

L'élaboration de nos tables fera appel à différentes fonctions définies précédemment :

- Fonction de survie : $S(x) = \frac{l_x}{l_0}$
- Probabilité de survie à l'âge x pendant t années : ${}_t p_x = \frac{l_{x+t}}{l_x}$
- Probabilité de sortir de l'état « enfant à charge » dans la plage d'âge $[x, x + 1[$:
 $q_x = \frac{d_x}{l_x} = 1 - p_x$

La série des q_x nous servira de base pour la construction de nos tables d'expérience, nous pouvons dès à présent dire que le taux de sortie sera très faible jusqu'à 16 ans, l'école étant obligatoire jusqu'à cet âge-là et que la pente de la courbe sera de plus en plus rude. Intuitivement la courbe des q_x aura l'allure suivante :



Une utilisation directe des tables d'expérience est de calculer les provisions en représentation des engagements de l'assureur vis-à-vis des rentiers dans le futur. De façon générale, pour chaque rentier d'âge x la formule utilisée est la suivante :

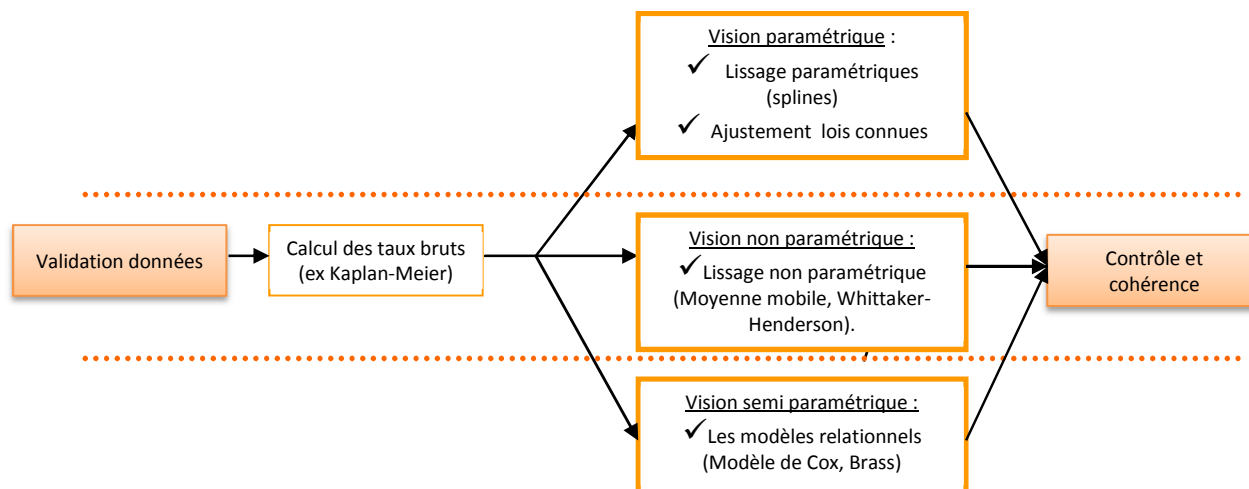
$$P = R \times \sum_{k=1}^{\text{âge limite}-x} \frac{l_{x+k}}{l_x} \times \frac{1}{(1+i)^k}$$

En notant :

- R : montant de la rente annuelle
- i : taux technique annuel
- l_{x+k} : nombre de personnes dans l'état enfant à charge d'âge $x + k$ extrait de la table construite dite « d'expérience »

1.2.5 Les grandes étapes de la construction d'une table d'expérience

Les grandes étapes de la construction d'une table peuvent être schématisées comme ceci :



La justesse des tables d'expérience est avant tout conditionnée par la pertinence et la qualité des données utilisées pour sa construction. A ce titre la validation des données est une étape incontournable du travail.

Il est important de se poser des questions sur le système d'information afin de bien comprendre d'où viennent les données, ce qu'elles signifient et comment elles ont été traitées (mode de stockage des données, durée de l'historique conservé, y-a-t-il eu un biais dans l'extraction). De plus il est nécessaire de vérifier la cohérence des données avant de pouvoir les exploiter.

On ne peut parler de qualité des données sans penser au piler II de la réforme Solvabilité II, celui-ci insiste en effet sur les exigences qualitatives que devront remplir les organismes assureurs.

Il faudra être en mesure de démontrer que les données utilisées pour les calculs sont exactes, exhaustives et appropriées. La directive va même plus loin en demandant de mettre en place des dispositifs d'évaluation de la qualité des données, de pilotage de celle-ci et d'intervention en cas de détection d'anomalie. La gouvernance des données est un enjeu au cœur d'une multitude de problématique comme la construction de tables d'expérience.

2 Analyse et réflexion autour du portefeuille « rentes éducation »

2.1 Le produit étudié

Le produit a pour objet le versement d'une rente éducation en cas de décès du participant aux enfants qui se trouvent dans l'état « enfant à charge ». Cette garantie fait partie d'un régime de prévoyance nationale obligatoire d'une branche professionnelle particulière. C'est un accord collectif national qui institue ce régime national de prévoyance et il prévoit que toutes les entreprises qui relèvent de la branche professionnelle en question sont tenues d'adhérer à ce régime auprès d'une institution de prévoyance désignée. Le partenariat « Recherche & Développement » a été conclu avec cette institution de prévoyance.

Il est donc possible de dire que :

- Le portefeuille est homogène vis-à-vis de la catégorie professionnelle.
- Le portefeuille contient l'ensemble des prestations rentes éducation déclenchées par le décès d'un participant appartenant à la branche professionnelle concernée.

Deux types de rentes sont proposés par le régime :

- Rente R1 : Rente à l'orphelin d'un seul parent : En cas de décès du participant, non provoqué par un accident du travail ou une maladie professionnelle, il est versé une rente à chaque enfant du participant :
 - Qui est **enfant à charge**,
 - Qui n'est pas orphelin de père et de mère.
 Le montant annuel de la rente est égal à 10 % de S.
- Rente R2 : Rente à l'orphelin des deux parents : En cas de décès d'un participant quelle qu'en soit la cause, il est versé une rente à chaque enfant du participant :
 - Qui remplit les conditions d'âge et de situation prévues **pour les enfants à charge**, et dans lesquelles la condition d'âge « moins de 21 ans » se substitue à la condition d'âge « moins de 18 ans »
 - Qui est orphelin de père et de mère.

Le montant annuel de la rente est égal à 10 % de S pour les cas où le décès est provoqué par un accident du travail ou une maladie professionnelle, 20 % dans les autres cas.

Branche professionnelle A	
Si orphelin d'un parent participant : R1	
- décès non suite à AT ou MP :	10%S
- décès suite à AT ou MP :	-
Si orphelin de père et de mère : R2	
- décès non suite à AT ou MP :	20%S
- décès suite à AT ou MP :	10%S

S : Salaire annuel du participant soumis aux cotisations de l'exercice précédent.

Si l'enfant est orphelin d'un parent, il n'y a pas de prestation dans le cas où le décès du participant est provoqué par un accident du travail ou une maladie professionnelle car dans ce cas-là une rente de 80% ou 100% du salaire du participant est versée au conjoint survivant pour le foyer.

La principale condition pour percevoir une rente éducation est qu'il faut être dans l'état « **enfant à charge** ».

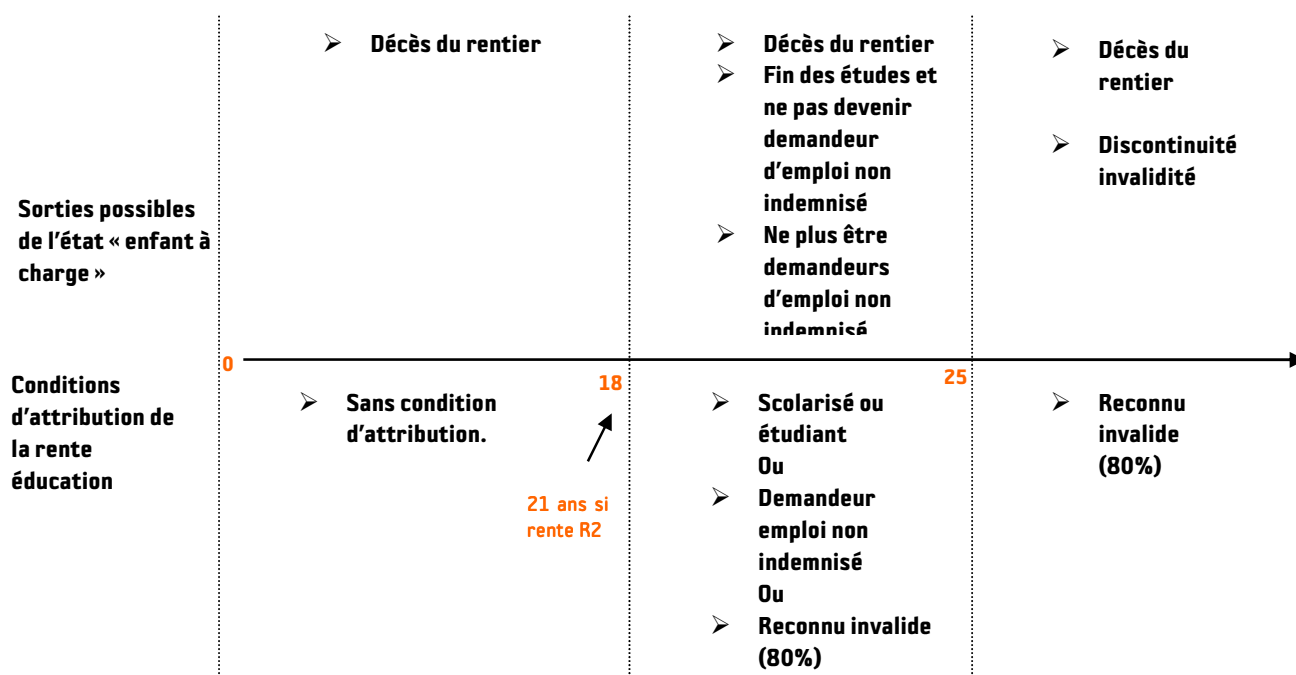
Sont considérés comme à charge, les enfants nés du participant ou adoptés par le participant :

- âgés de moins de 18 ans,
- âgés de 18 à 25 ans, célibataires, s'ils sont dans l'une des situations suivantes :
 - Apprentis,
 - Scolarisés dans un établissement du second degré ou étudiants, sans être rémunéré au titre de leur activité principale,
 - en contrat de professionnalisation ou en formation en alternance,
 - demandeurs d'emplois inscrits au pôle emploi et non indemnisés par le régime d'assurance chômage.
- Sans limite d'âge, s'ils sont reconnus atteints, avant 21 ans et sans discontinuité depuis cet âge, d'une invalidité au taux de 80% ou plus au sens de la législation sociale. Dans ce cas, l'enfant doit être à charge fiscale du participant.

Sont également considérés comme enfants à la charge du participant :

- Les enfants du conjoint, répondant aux critères ci-avant et à la charge fiscale du participant,
- les enfants du participant nés viables, moins de 300 jours après le décès de ce dernier.

Ce sont ces conditions qui vont déterminer les causes de sorties de l'état « enfant à charge » et par la suite les taux de sortie. Voici un schéma récapitulatif des sorties possibles de l'état enfant à charge.



Conditions de versement

Le point de départ des rentes est le premier jour du mois civil qui suit le fait générateur (décès du participant), dès lors que les conditions d'attribution des droits auront été réunies.

Elles sont versées trimestriellement et d'avance aux dates déterminées par le mois de naissance du participant et sont supprimées quand l'enfant ne remplit plus les conditions pour avoir le statut « enfant à charge ».

Le besoin de notre client se situe sur le périmètre des **enfants « valides »**. Les cas des enfants « invalides » ne seront pas traités car ils touchent une rente éducation à vie et ne sont donc pas provisionnés de la même manière que les enfants valides touchant une rente temporaire. L'âge limite concernant notre portefeuille pour toucher une rente sera alors 25 ans.

2.2 Construction d'une loi dite de référence

Selon la description du produit, les principales conditions pour bénéficier de la rente éducation sont d'être en vie et d'être scolarisé ou demandeur d'emploi non indemnisé. La durée passée dans l'état enfant à charge peut être approchée en tenant compte des trois éléments ci-dessous :

- le décès,
- la durée de scolarisation d'un enfant,
- la prise en compte du chômage post scolarisation.

La prise en compte de ces 3 facteurs permet de créer une table de référence à partir de données de marché.

2.2.1 Indicateurs synthétiques de la durée de scolarisation

a. Taux de scolarisation et taux de déscolarisation :

Sur le document « Repères et références statistiques sur les enseignements, la formation et la recherche » publié sur le site internet de l'éducation nationale, nous obtenons pour chaque âge les effectifs de la population et les effectifs de la population scolarisée pour l'année 2007/2008.

Ces données vont permettre de calculer des indicateurs synthétiques du risque « poursuite des études » dans un premier temps, puis dans un second temps de construire une loi de survie dans le système scolaire représentative de la population française.

Les taux de scolarisation de la population française sont calculés pour chaque âge avec la formule suivante :

$$TS_x = \frac{ES_x}{POP_x}$$

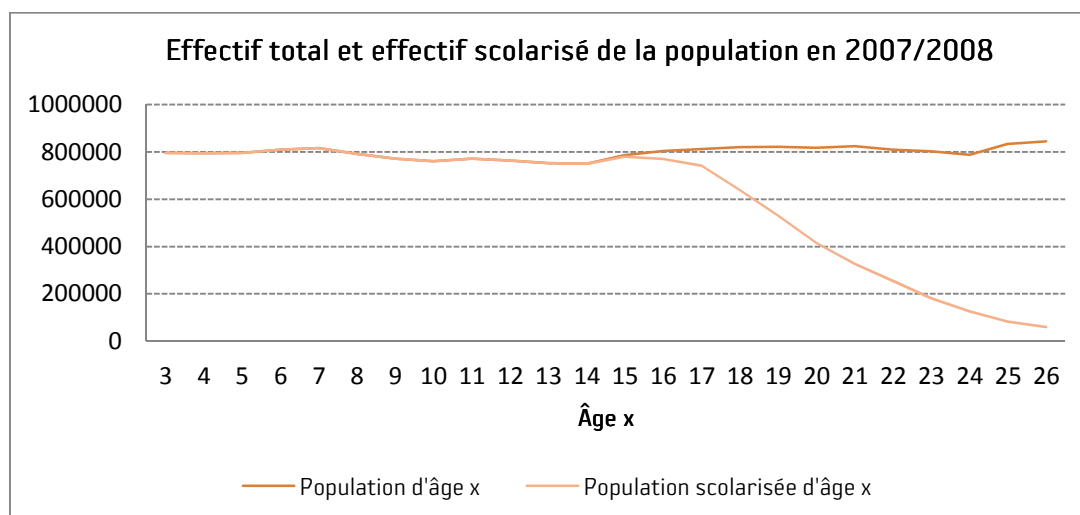
TS_x : Taux de scolarisation pour les personnes d'âge x .

ES_x : Nombre de personnes scolarisées d'âge x , tous niveaux d'enseignement confondus.

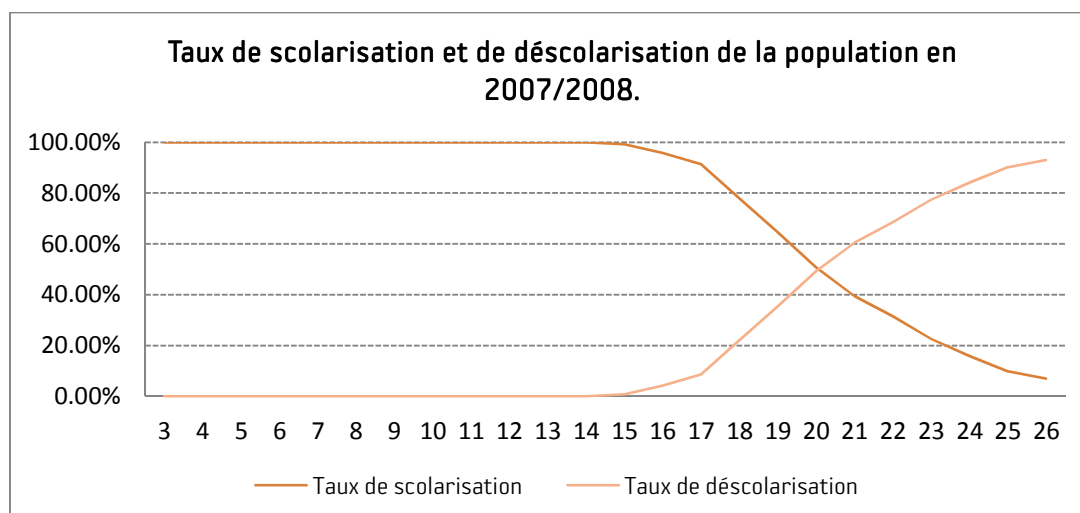
POP_x : Effectif de la population d'âge x .

Le taux de déscolarisation pour chaque âge s'obtient avec la formule : $TDS_x = 1 - TS_x$

Ci-dessous l'évolution des effectifs totaux et scolarisés de la population en 2007/2008.



Ci-dessous les taux de scolarisation et de déscolarisation par âge en 2007/2008.

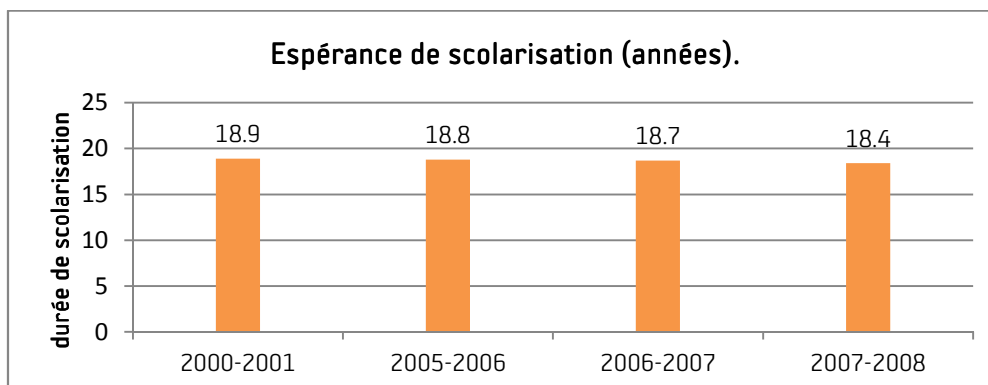


Nous allons considérer que les taux de déscolarisation peuvent représenter la probabilité de sortir du système scolaire à l'âge x , sachant que l'enfant a été scolarisé pendant ces x premières années. Il convient de remarquer que si les conditions de sortie évoluent au fil du temps, cette approche est biaisée, car on considère des individus de générations différentes pour calculer les taux de sortie.

b. Espérance de scolarisation et évolution de la durée de scolarisation

L'espérance de scolarisation est une estimation de la durée de scolarité d'un enfant entrant en maternelle. Autrement dit c'est la durée moyenne des études d'une génération fictive soumise aux conditions de sortie observées au cours de l'année 2007/2008. Comme l'espérance de vie, cet indicateur exprime une situation ponctuelle, reflet de la scolarisation de l'année étudiée. On peut l'estimer en sommant les taux de scolarisation observés aux différents âges.

Voici un tableau qui récapitule les espérances de scolarisation évaluées à différentes années :



Malgré une légère baisse au fil du temps, les espérances de scolarisation sont équivalentes entre les années 2000 et 2008. Ce graphique permet de poser l'hypothèse que pendant cette période les conditions de scolarité et donc les durées de scolarisation sont considérées comme identiques.

2.2.2 Construction d'une loi de référence à partir de données de marché

Nous allons maintenant construire une loi de survie dans l'état « enfant à charge » de la population française à l'aide de données de marché fournis par :

- les tables de mortalités réglementaires,
- le ministère de l'éducation nationale pour la condition « être scolarisé »,
- l'institut national de la statistique et des études économiques (INSEE) pour la condition « être demandeur emploi non indemnisé ».

L'objectif est de construire une loi dite de « référence » qui nous renseignera sur la forme a priori que devrait avoir la loi de survie dans l'état « enfant à charge » si notre portefeuille était composé de la population française.

Nous rappelons que pour être dans l'état enfant à charge, il faut :

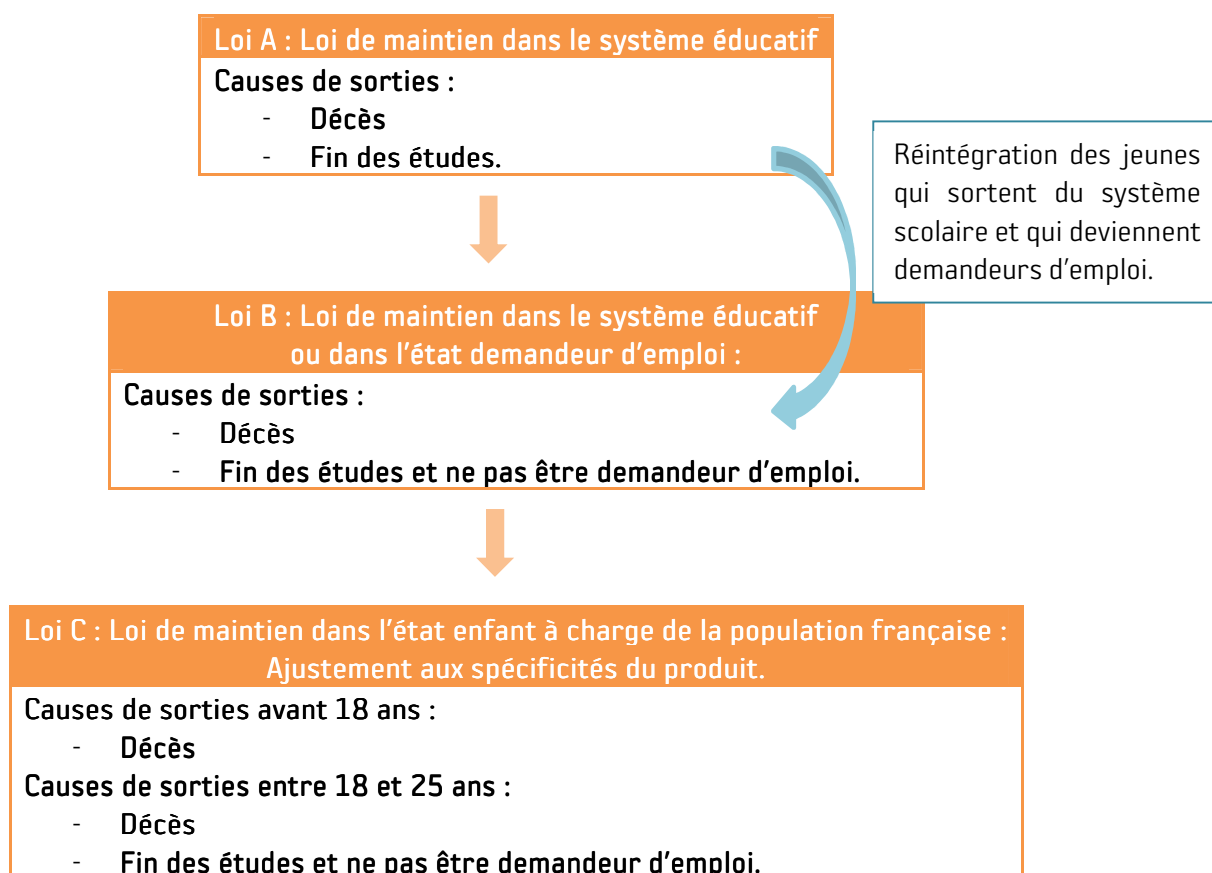
- être en vie,
- être scolarisé ou demandeur d'emploi non indemnisé,
- respecter les limites d'âges contractuelles.

Statistiquement parlant, il convient de s'intéresser à la variable aléatoire T représentant la durée passée dans l'état « enfant à charge » ou de manière équivalente de considérer les variables T_x représentant la durée résiduelle passée dans l'état « enfant à charge » d'un individu conditionnellement au fait qu'il était dans l'état à l'âge x , i.e. :

$$T_x = [T - x/T > x].$$

Nous posons l'hypothèse que lorsque l'individu sort de l'état « enfant à charge », c'est de manière définitive et qu'un enfant se trouve automatiquement dans l'état « enfant à charge » s'il a moins de 18 ans et s'il est en vie.

La construction de la loi de référence s'effectue en trois étapes :



Phénomène à modéliser pour apprécier le statut « enfant à charge »	LOI A	LOI B	LOI C
Mortalité	OUI	OUI	OUI
Scolarisation	OUI	OUI	OUI
Chômage (demandeur d'emploi)	NON	OUI	OUI
Limites d'âges contractuelles.	NON	NON	OUI

D'après ce tableau, la loi C représentera le comportement de la population française dans l'état « enfant à charge » défini par le contrat étudié.

a. Loi de maintien dans le système éducatif

La probabilité pour un individu d'être dans le système éducatif au-delà de l'âge $t > s$ peut s'écrire :

$$S(t) = P(T > t/T > s) \times P(T > s) = P(T > t/T > s) \times S(s)$$

En effet la probabilité conditionnelle de l'évènement $\{T > t\}$ sachant l'évènement $\{T > s\}$ s'écrit :

$$P(T > t/T > s) = \frac{P(T > t, T > s)}{P(T > s)} = \frac{P(T > t)}{P(T > s)}$$

Nous pouvons renouveler l'opération et cela fait apparaître des produits de termes en $P(T > t/T > s)$.

Si nous choisissons comme instants de conditionnement les différents âges où sont réévaluées les sorties, le problème se ramène à estimer des probabilités de la forme :

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

où p_i est la probabilité que l'enfant soit dans le système éducatif à l'âge T_i sachant qu'il y était à l'âge T_{i-1} .

Il est alors possible d'estimer $q_i = 1 - p_i$ par $\hat{q}_x$ taux de sortie estimé dans la plage d'âge $[x ; x + 1[$.

Pour un individu d'âge x , notons :

- ✓ $q_x^{\text{décès}}$: la probabilité que l'enfant décède dans la plage d'âge $[x ; x + 1[$.
- ✓ $q_x^{\text{fin de scolarité}}$: la probabilité que l'enfant sorte du système scolaire à l'âge x .

Les deux causes de sorties sont indépendantes. La somme des taux de sortie exprimés cause par cause est donc égale au taux de sortie de la table de maintien toutes causes.

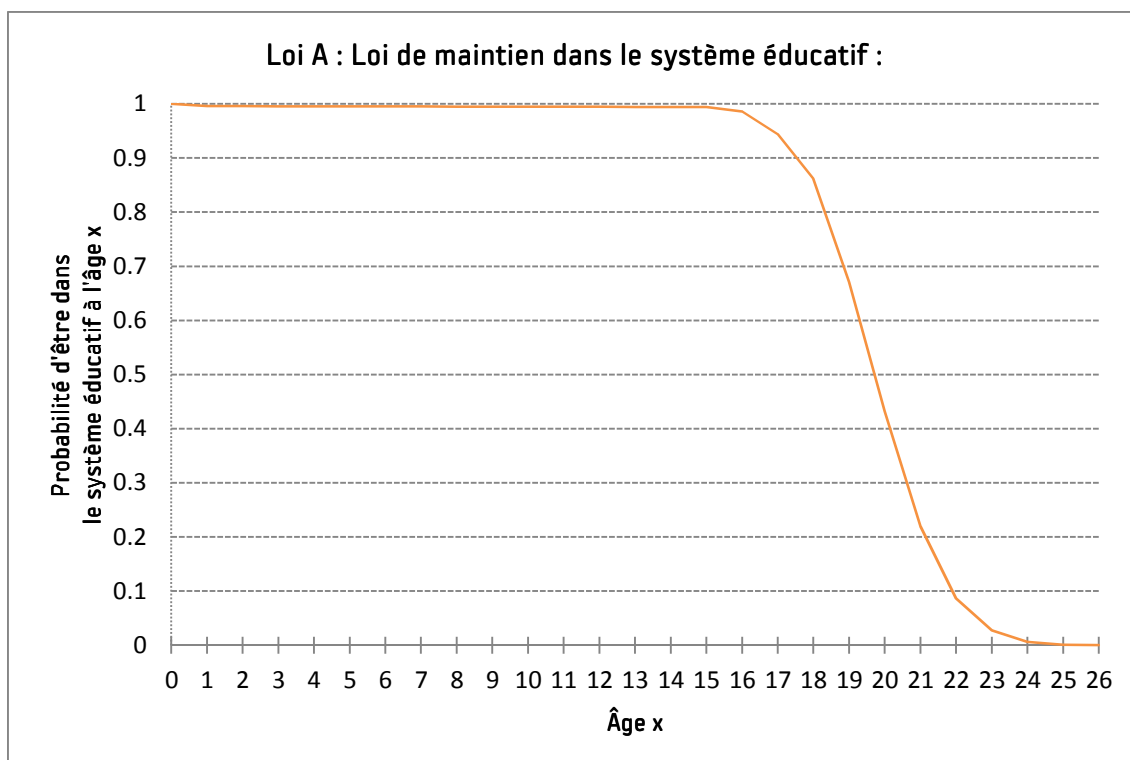
Ainsi $q_x = q_x^{\text{décès}} + q_x^{\text{fin de scolarité}}$ s'estime par $\hat{q}_x = \hat{q}_x^{\text{décès}} + \hat{q}_x^{\text{fin de scolarité}}$ avec

$\hat{q}_x^{\text{fin de scolarité}} = TDS_x$: taux de déscolarisation à l'âge x TDS_x

$\hat{q}_x^{\text{décès}} = \hat{q}_x^{\text{TGF05}}$: quotient de mortalité de la table réglementaire TGF05.

On choisit la table de génération féminine, car c'est la table la plus prudente pour le risque vie.

La formule $S(t) = \prod_{i < t} (1 - \hat{q}_i)$ nous permet alors de tracer ci-dessous la loi de maintien dans le système éducatif.



b. Loi de maintien dans le système éducatif ou dans l'état demandeur d'emploi

Un jeune qui sort des études a une probabilité non négligeable de devenir un demandeur d'emploi et de le rester quelques années. Nous souhaitons intégrer à la loi précédente une condition supplémentaire sur la sortie afin d'intégrer ce phénomène de chômage.

Pour sortir de l'état « enfant à charge », il doit s'il est vivant, être non scolarisé et non demandeur d'emploi. Il convient donc de réintégrer tous les jeunes qui sortent du système scolaire et qui deviennent demandeurs d'emploi.

Pour modéliser ce phénomène, nous nous appuyons sur des statistiques publiées par l'INSEE qui nous indiquent le taux de chômage selon la durée écoulée depuis la sortie de formation initiale.

Taux de chômage selon la durée depuis la sortie de formation initiale	
Sortis depuis 1 à 4 ans de formation initiale	20.2%
Sortis depuis 5 à 10 ans de formation initiale	10.5%

$$\frac{POP \text{ sortie de formation initiale depuis 1 à 4 ans au chômage}}{POP \text{ sortie de formation initiale depuis 1 à 4 ans}} = 20.2\%$$

Et

$$\frac{POP \text{ sortie de formation initiale depuis 5 à 10 ans au chômage}}{POP \text{ sortie de formation initiale depuis 5 à 10 ans}} = 10.5\%$$

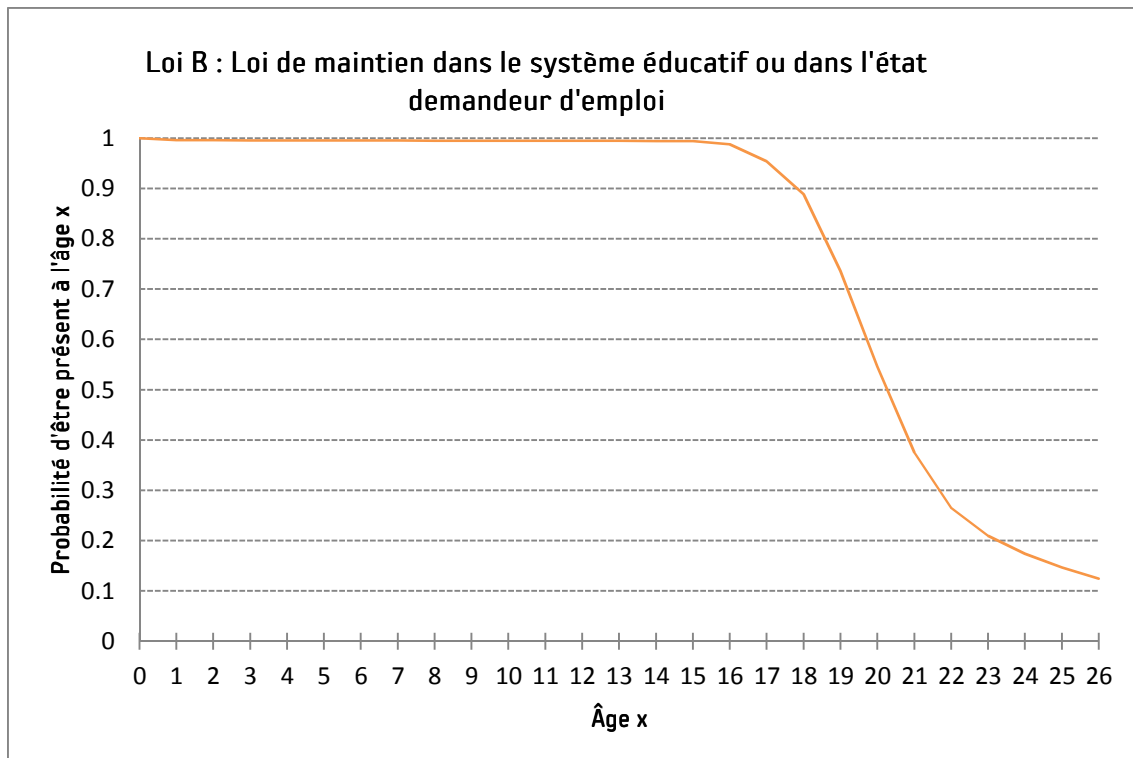
Ainsi pour chaque âge x , nous allons recenser le nombre de jeunes d'âge x sortis de la formation initiale depuis au maximum 4 ans puis multiplier ce nombre par le pourcentage fourni par l'INSEE. Nous obtenons alors le nombre de jeunes au chômage à réintégrer pour chaque âge.

Par exemple pour l'âge $x = 20$ ans :

Âge	Loi A	Nombre de jeunes sortis du système scolaire à l'âge x	Loi B
16	l_{16}	d_{16}	
17	l_{17}	d_{17}	
18	l_{18}	d_{18}	
19	l_{19}	d_{19}	
20	l_{20}	d_{20}	$l_{20}^{loi B} = l_{20} + 20.2\% \times \sum_{i=16}^{i=20} d_i$

Nous faisons cependant quelques approximations :

- En utilisant le nombre de jeunes sortis du système scolaire à chaque âge pour calculer le nombre de jeunes à réintégrer, les jeunes sortis pour cause de décès sont inclus dans l'assiette de calcul. Au regard des probabilités de décès à ces âges, cette approximation reste acceptable.
- La statistique de l'INSEE ne distingue pas les chômeurs indemnisés des chômeurs non indemnisés alors que la définition de l'état « enfant à charge » s'applique uniquement aux chômeurs non indemnisés. L'approche la plus prudente consiste à les réintégrer tous.
- Nous appliquons la statistique sur les jeunes sortants immédiatement du système scolaire, alors qu'elle représente le taux de chômage pour les jeunes sortis de 1 à 4 ans. Il faudrait sans doute appliquer un taux de chômage plus fort pour les jeunes sortis du système scolaire depuis moins d'un an. Nous sous-estimons peut-être ainsi le nombre de jeunes à réintégrer.



c. Loi de maintien dans l'état « enfant à charge » de la population française

Il nous reste à adapter la loi précédente de façon à respecter les limites d'âges contractuelles du contrat. D'après les dispositions liées aux versements d'une rente éducation :

Aucune sortie hormis le décès ne peut être observée avant l'âge de 18 ans, la rente étant obligatoirement versée jusqu'à cet âge-là. Ainsi avant 18 ans, seuls les taux de mortalité s'appliquent sur la population. Néanmoins un jeune sortant avant 18 ans du système scolaire et n'ayant pas le statut demandeur d'emploi va sortir automatiquement de l'état enfant à charge à 18 ans, il nous faut donc concentrer tous les taux de sortie non liés au décès avant 18 ans sur le q_{18} .

Concernant l'âge limite, le versement de la rente est interrompu lorsque le rentier atteint 25 ans.

A l'aide de la loi B, nous recalculons les taux de sortie q_x toutes causes confondues par la formule

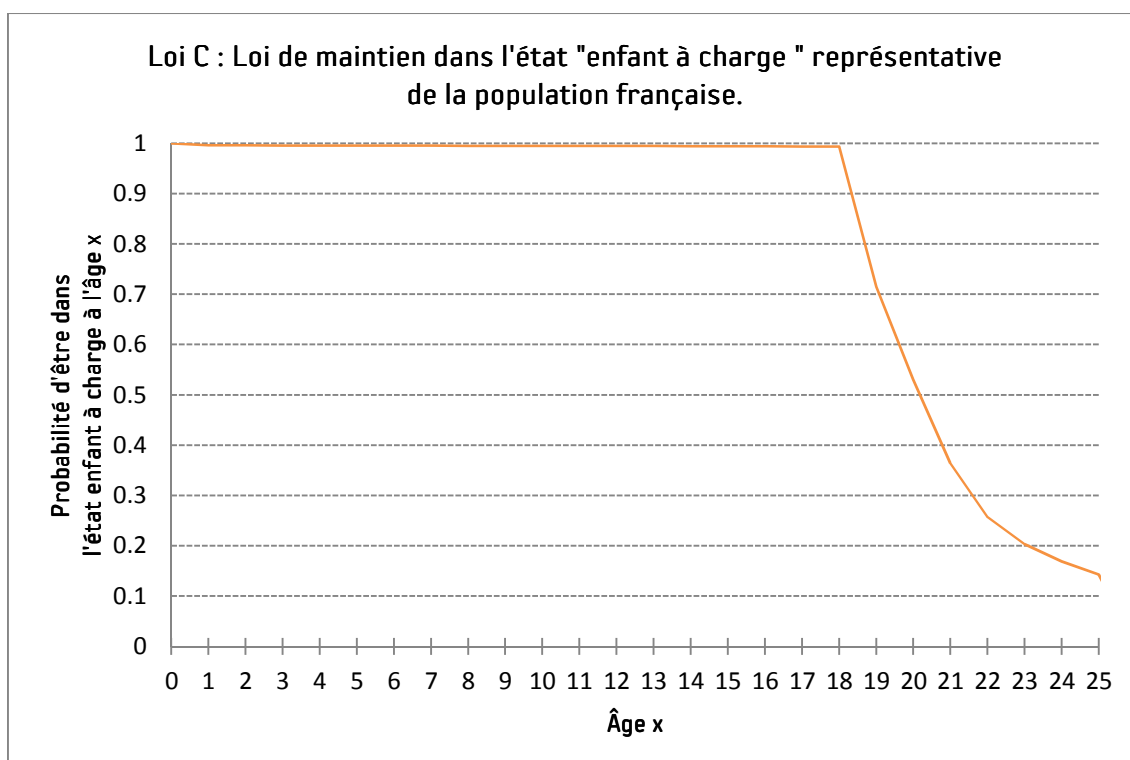
$$q_x = 1 - \frac{l_{x+1}}{l_x}$$

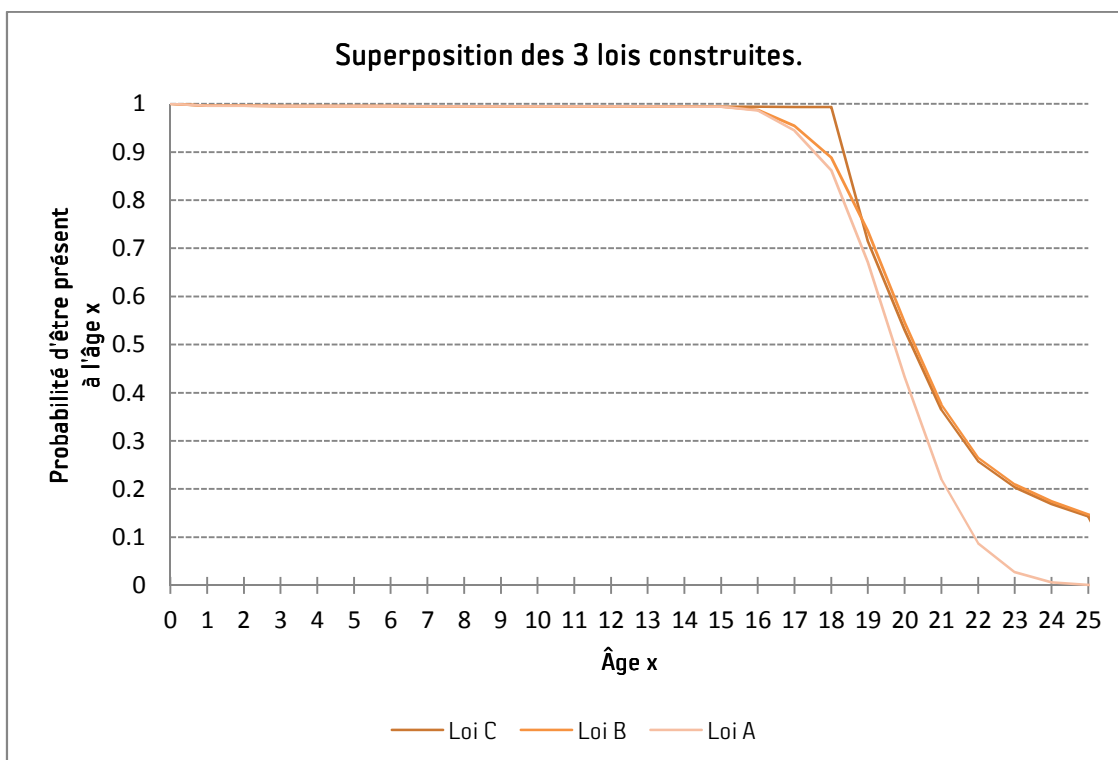
Puis nous décomposons q_x de manière à distinguer le taux de mortalité et le taux de sortie pour les autres causes.

$$q_x^{loi B} = q_x^{décès} + q_x^{fin\ sco\ et\ non\ chômeur}$$

La loi C s'obtient en appliquant les taux de sortie suivants pour chaque âge :

$$\left\{ \begin{array}{l} q_x^{loi C} = q_x^{décès} \text{ si } x \leq 17 \\ q_{18}^{loi C} = q_{18}^{décès} + \sum_{i \leq 18} q_i^{fin\ sco\ et\ non\ chômeur} \\ q_x^{loi C} = q_x^{loi B} \text{ si } 19 \leq x \leq 24 \\ q_{25}^{loi C} = 1 \end{array} \right.$$





Nous obtenons une loi référence pour la loi de maintien dans l'état enfant à charge. Celle-ci a deux utilités :

- ✓ Elle permet d'avoir une idée de la forme que pourrait avoir notre courbe.
- ✓ Elle nous permettra de positionner le comportement de nos rentiers par rapport à la population française.

2.3 Les données

Les données fournies par notre client sont des données brutes issues de son système d'information. Avant de commencer tous calculs, il convient de présenter les données telles que nous les avons reçues et de décrire le cheminement effectué pour arriver à une base sinistre exploitable.

2.3.1 **Présentation**

Le groupe a mis à notre disposition 5 fichiers correspondant à 5 arrêtés comptables aux dates suivantes :

- 31 décembre 2008
- 31 mars 2009
- 30 juin 2009
- 30 septembre 2009
- 31 décembre 2009.

Chaque arrêté nous donne une vision instantanée du portefeuille rente éducation à la date correspondante. Ils ont été convertis en tables SAS où chaque colonne est une variable informative et chaque ligne un enregistrement traduisant une entrée ou une modification de gestion. Ces tables brutes comportent chacune 80 colonnes et plus de 20000 lignes. Elles contiennent donc beaucoup d'information qu'il a fallu structurer et affiner.

Le premier travail fut de sélectionner les colonnes utiles à l'étude. Nous avons retenu les variables suivantes :

Caractéristiques individuelles des bénéficiaires :

- Muna Personne : Identifiant de l'enfant
- **Date de naissance de l'enfant.**

Caractéristiques individuelles du participant :

- P0032005K : Identifiant du participant.
- Date de naissance
- Date fait générateur (date du décès du participant)
- Salaire retenu
- Catégorie professionnelle : Permet d'identifier la branche professionnelle.

Données relatives à la garantie rente éducation :

- Date de début de versement de la rente
- **Date de fin de versement de la rente**
- Prestation : type de la rente R1 ou R2
- Montant de rente versé.

Données concernant le contrat rente éducation :

- Numéro de dossier.
- **Situation du dossier** : fermé, liquidé, ouvert, en cours de révision, annulé, bloqué ou en rejet d'instruction.

- Date situation : Dernière date d'évaluation de la situation du dossier.

Les variables en gras sont les trois variables fondamentales pour l'étude.

La date de fin de versement de la rente est interprétée comme la date de sortie de l'état enfant à charge. Cette sortie a lieu soit parce que l'enfant ne remplit plus les conditions d'âges, de scolarité, de chômage (au sens demandeur d'emploi) ou soit parce qu'il est décédé.

La variable date de naissance nous permet de calculer l'âge exact à la sortie.

La variable situation nous indique si la date de fin de versement est la date de fin définitive (si le dossier est fermé) ou si c'est une date de censure i.e. si le dossier est encore en cours de prestations.

Comme nous l'avons évoqué plus haut, la construction de tables d'expérience est fondée sur l'exploitation de données réelles et fiables.

Nous nous sommes employés à créer une seule base sinistre exploitable, cette base a été obtenue au travers de plusieurs manipulations, traitements, vérifications et tests que nous décrivons ci-dessous.

2.3.2 Création d'une base sinistre exploitable

Nous avons le choix de nous placer pour la construction des tables à la vision au 31 décembre 2008 ou au 31 décembre 2009 puisque nous connaissons l'état du portefeuille à ces deux dates. Nous décidons de nous placer au 31 décembre 2008 pour deux raisons :

- Nous pourrions récupérer les déclarations tardives de sinistres relatives à l'année 2008, ainsi que les dossiers fermés ou annulés pour une raison puis réouverts sur l'arrêté suivant.
- Les nouvelles informations disponibles sur l'arrêté au 31.12.2009 pourront être vues comme un échantillon de validation. Il nous servira pour homologuer la table d'expérience construite avec l'échantillon d'apprentissage.

L'arrêté au 31 décembre 2008 nous sert donc de socle pour créer notre base sinistre, c'est sur les dossiers présents dans cet arrêté que nous allons construire les tables d'expérience.

Le retraitement des données se découpe en plusieurs étapes.

Etape n°1 : Traitement des arrêtés de façon indépendante :

Nous allons décrire le traitement que nous avons effectué sur chaque arrêté à l'aide du logiciel SAS.

1. Sélection des enregistrements concernant :
 - la catégorie socio-professionnelle étudiée.
 - le type de prestation R1/R2.
 - Les enfants « valides ».
2. Fusion des lignes concernant le même rentier et le même type de prestation en une seule ligne cohérente :

	Prestation	Date situation	Situation	Date début rente	Date fin de rente
Rentier x	R1	06/12/2008	liquidé	01/11/2001	30/09/2017
Rentier x	R1	19/08/2003	annulé	01/11/2001	30/09/2017



	Prestation	Date situation	Situation	Date début rente	Date fin de rente
Rentier x	R1	06/12/2008	liquidé	01/11/2001	30/09/2017

3. Fusion des lignes concernant le même rentier avec des prestations différentes (R1 puis R2) en une seule ligne.

	Prestation	Date situation	Situation	Date début rente	Date fin de rente
Rentier x	R1	06/12/2008	Fermé	01/11/2001	30/09/2005
Rentier x	R2	19/08/2009	liquidé	01/11/2005	30/09/20017



	Prestation	Date situation	Situation	Date début rente	Date fin de rente
Rentier x	R1	19/08/2009	liquidé	01/11/2001	30/09/2017

La durée passée dans l'état « enfant à charge » est indépendante du type de prestation touchée, la fusion nous permet de récupérer la durée totale passée dans l'état « enfant à charge ».

4. Contrôles de cohérence des données :

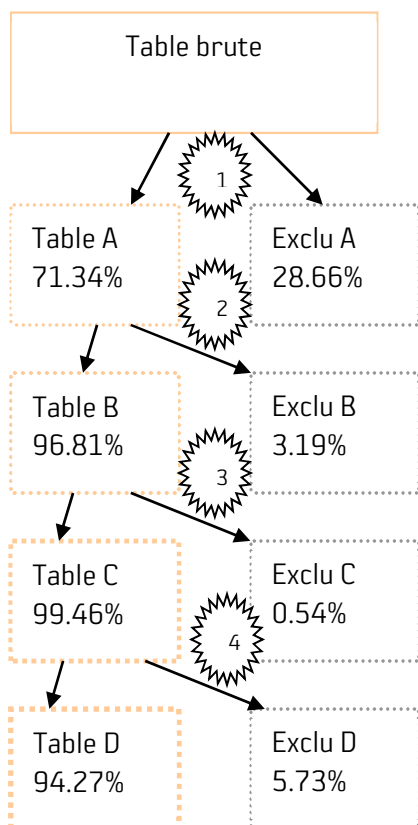
Voici les situations anormales qui doivent être supprimées :

- Âge du l'enfant en rente éducation > 26 ans au cours de la période, l'enfant a pu être noté « valide » par erreur, car les rentiers invalides n'ont pas de conditions d'âge limite.
- Date du fait générateur > 31 décembre 2008
- Date d'entrée en rente éducation < Date du fait générateur.
- Date d'entrée en rente éducation > Date de sortie de la rente
- Date situation > 31 décembre 2008

Etape n°2 : Utilisation des arrêtés qui nous offrent une vision prospective :

1. Fusion des 5 tables propres : Nous avons la chance grâce aux arrêtés qui suivent le 31 décembre 2008 de posséder une information prospective. Ainsi il est possible d'identifier les dossiers où la situation (par exemple : la variable situation ou la variable date de fin de rente) a évolué et rectifier celle-ci suivant l'évolution.
Par exemple si un dossier est annulé ou fermé puis réouvert par la suite, nous considérons qu'il a toujours été ouvert.
2. Suppression des dossiers annulés restants.
3. Récupération des dossiers ayant une date de fait générateur d'année 2008 et qui sont non présents dans l'arrêté 1 mais que l'on peut récupérer dans les arrêtés suivants. Ainsi on ne sous-estime pas le nombre de rentiers exposés au risque de sortir.

Voici un schéma récapitulatif du traitement des données effectué :



Etape n°1 :

On sélectionne les enregistrements concernant la branche professionnelle et les types de prestations

Etape n°2:

Fusion des lignes concernant le même rentier pour un même type de prestation en seule ligne.

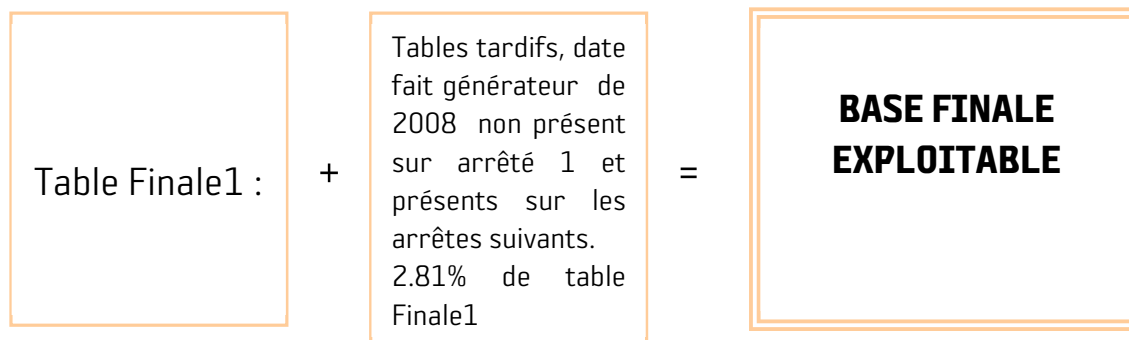
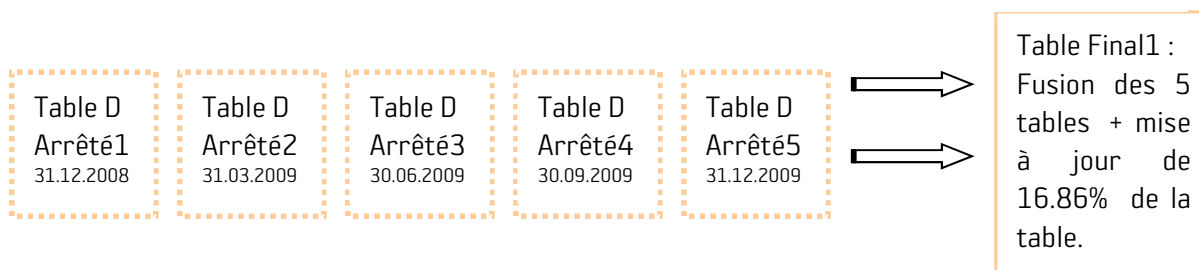
Etape n°3 :

Fusion des lignes concernant le même individu pour des prestations différentes en une seule ligne

Etape n°4 :

Contrôle de cohérence et suppression des lignes aberrantes.

Le processus ci-dessus est effectué pour les 5 arrêtés bruts afin d'obtenir 5 tables propres. Les chiffres inscrits sur le schéma sont ceux concernant l'arrêté au 31 décembre 2008 (les autres statistiques sont disponibles en annexe). Il reste à fusionner ces cinq tables pour récupérer la dernière situation de chaque dossier et pour récupérer les déclarations tardives : les faits générateurs de 2008 non déclarés au 31 décembre 2008.



2.4 Etudes statistiques

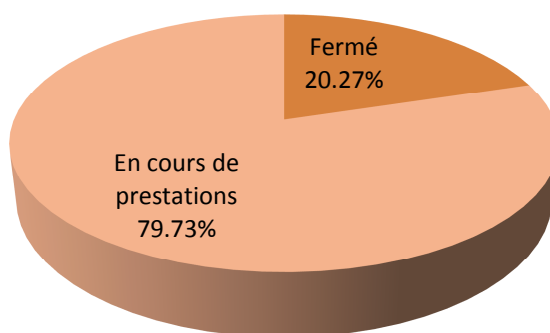
2.4.1 Statistiques descriptives du portefeuille

Les retraitements décrits dans la partie précédente conduisent à obtenir une base de rentiers comportant 14554 lignes, i.e. nous allons baser notre étude en observant le comportement de 14554 rentiers.

On rencontre deux situations de dossiers :

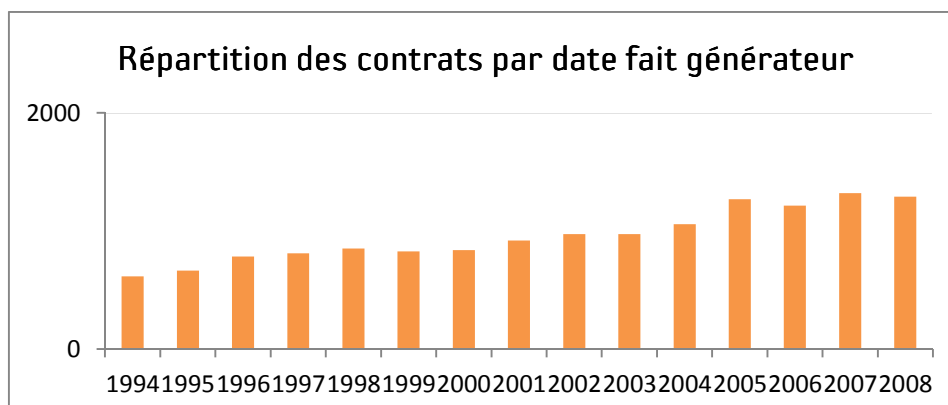
- Soit le dossier est fermé : la sortie du rentier est observée et il est possible de déterminer précisément l'âge de sortie du rentier.
- Soit le dossier est en cours de prestation : la sortie du rentier n'est pas observée, mais cela apporte quand même de l'information, à savoir que le rentier reste dans l'état enfant à charge au moins jusqu'à l'âge qu'il a à la date du 31 décembre 2008.

Situation du dossier :



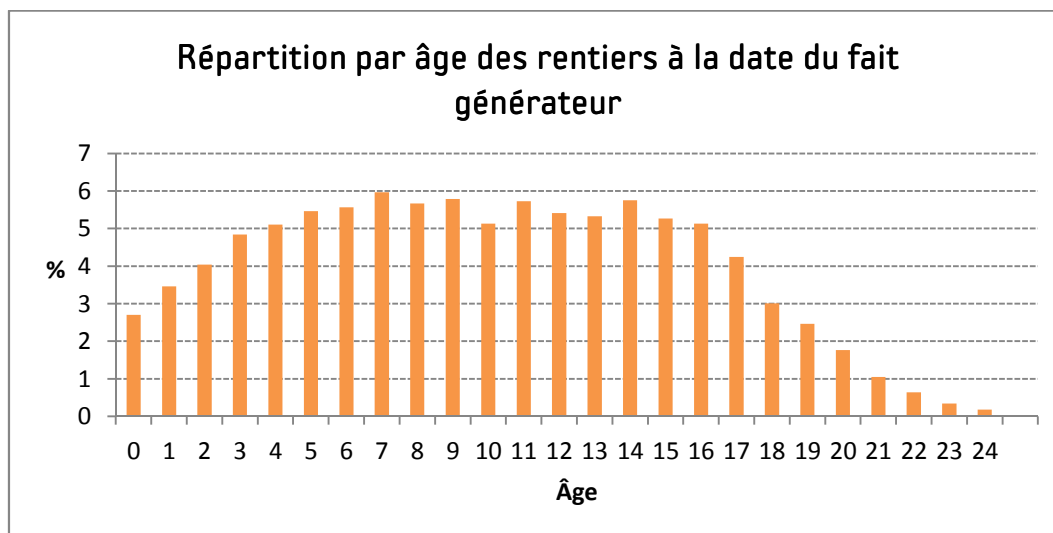
Seulement 20.27% des contrats du portefeuille permettent d'observer une sortie réelle de l'état « enfant à charge ».

Le graphique suivant permet de savoir s'il ne manque pas de données par année de survenance.



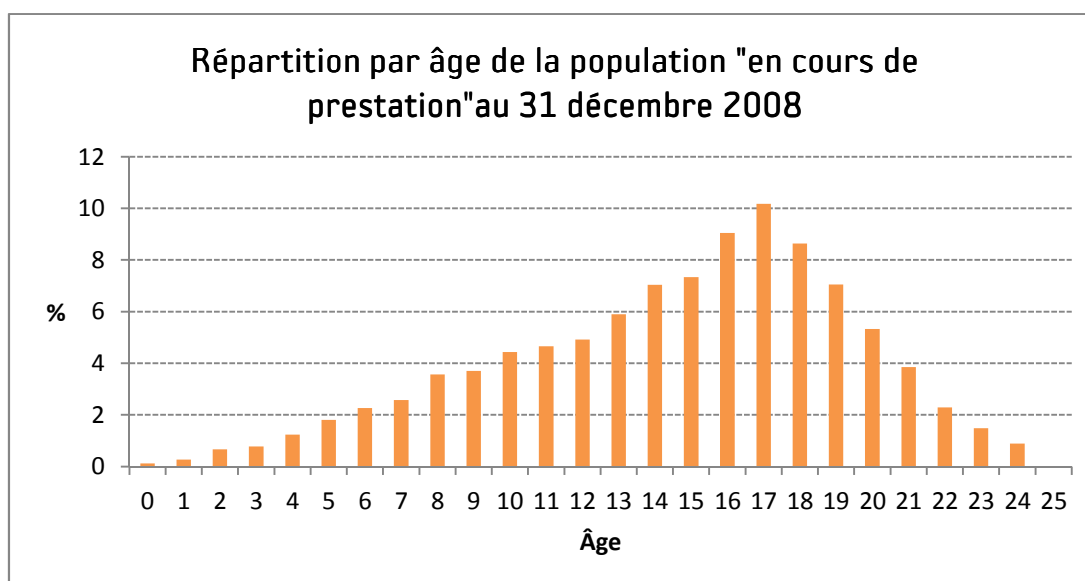
L'évolution de la courbe n'est pas anormale. Nous n'observons pas « de trou » à une année particulière. Bien qu'il faille la comparer avec l'évolution du nombre d'assurés, il semble raisonnable de conclure que la base sinistre est complète.

Le graphique suivant représente la répartition par âge des rentiers à la date du décès du participant.



La proportion de rentiers à l'entrée pour les âges inférieurs à 18 ans est homogène, elle décroît à partir de 18 ans. La baisse significative à partir de 18 ans s'explique par le fait que lorsqu'un participant décède, si l'enfant a plus de 18 ans, il a potentiellement déjà terminé ses études et ne bénéficie donc pas de la rente éducation.

Afin de bien comprendre la composition du portefeuille « en cours de prestations », intéressons-nous à la répartition de la population par âge au 31 décembre 2008. Ici on ne considère que la population « en cours de prestation ». C'est-à-dire la population qui perçoit encore des versements.



2.4.2 Facteurs discriminants a priori

Dans un premier temps nous aimerions savoir si la perte d'un parent pendant l'enfance peut avoir un effet sur le parcours scolaire et donc sur la durée des études. Pour cela nous nous appuyons sur une étude menée par la DREES qui répond à cette question.

Elle indique que pour connaître l'effet du décès d'un des parents sur la scolarité des enfants, il faut aussi prendre en compte d'autres facteurs de réussite ou d'échec comme la catégorie socio-professionnelle des parents ou la taille de la fratrie.

Nous partons de l'hypothèse que vivre le décès d'un parent pendant l'enfance joue un rôle pénalisant sur la poursuite des études. Nous aimerions aller un peu plus loin en mettant en évidence les périodes de l'enfance les plus critiques pour vivre le décès d'un parent vis-à-vis de sa future scolarité. Par exemple on peut penser que si l'enfant vit le décès d'un parent pendant son adolescence, les conséquences sur sa durée de scolarité peuvent être plus marquées que s'il vit le décès d'un parent alors qu'il a plus de 20 ans et qu'il est déjà lancé dans un cycle d'études supérieures.

Un des enjeux de cette étude est d'étudier l'impact de la variable âge de l'enfant à la date du fait générateur sur l'âge à la sortie i.e. sur la durée passée dans l'état « enfant à charge ». Nous pourrions voir si le décès d'un parent suivant l'âge où il est vécu impacte de façon significative sur la durée de scolarisation. Tout cela nous permettra de modéliser au mieux la durée dans l'état enfant à charge et donc de provisionner au plus juste les engagements de l'organisme de prévoyance.

Pour cela nous découpons notre population en quatre sous-populations.

- **Population 1 « Non scolarisé / Maternelle »** : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]0 ; 6]. On peut supposer que ces enfants ne sont pas encore scolarisés ou en classe de maternelle.
- **Population 2 « Primaire / Collège »** : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]6 ; 14]. On peut supposer que ces enfants sont scolarisés à l'école primaire ou au collège.
- **Population 3 « Fin de collège/ Lycée »** : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]14 ; 18]. On peut supposer que ces enfants sont en fin de collège ou au lycée. Pendant cette période ils auront à faire leur premier choix d'orientation.
- **Population 4 « Fin de lycée / Etudes supérieures »** : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]18 ; 25]. On peut supposer que ces enfants sont en fin de lycée ou déjà engagé dans un cycle d'études supérieures.

Il n'est pas possible de prendre en compte l'éventuel redoublement de certains élèves.

Nous pouvons dès à présent calculer une première statistique sur les dossiers où l'on a pu observer la sortie : ci-dessous un tableau qui récapitule l'âge moyen de sortie observé pour chaque sous-population.

Variable âge à la sortie	
Sous-population	Moyenne
Population 1	18.11
Population 2	19.86
Population 3	20.31
Population 4	22.64

Suivant la classe d'appartenance, la variable « âge à la sortie de l'état d'enfant à charge » a une valeur moyenne différente.

Ces premiers chiffres nous indiquent qu'il y a peut-être un comportement différent en termes de survie dans l'état enfant à charge suivant l'âge auquel l'enfant vit le décès d'un parent. Attention ces chiffres sont à prendre avec précautions car ici nous n'avons pas pris en compte l'information censurée. Nous verrons par la suite comment valider ce facteur a priori.

2.5 Choix déterminants pour la construction des tables

2.5.1 Choix de la période d'observation

Le choix de la période de d'observation présente un enjeu double : obtenir suffisamment de données pour que le volume soit suffisamment important et que les résultats soient pertinents, sans toutefois mélanger des générations qui n'auraient pas le même comportement dans le temps car l'objectif est d'être représentatif du portefeuille dans le présent.

L'Institut des Actuaire préconise plusieurs vérifications à effectuer afin de choisir une période d'observation adéquate :

- Les déclarations tardives de sinistres :
Comme nous l'avons vu lors du traitement des données, notre table intègre les déclarations tardives de sinistre grâce aux arrêts suivants. On peut donc choisir comme date de fin d'observation la date du 31 décembre 2008 sans perdre de dossiers.
- La longueur de la période d'observation :
La contrainte sur le volume des données nous invite à choisir une longueur d'observation assez importante, de plus nous nous sommes assurés auparavant que les conditions de scolarité pouvaient être considérées comme étant homogènes depuis 2000.

La période d'observation retenue s'étend du 1^{er} janvier 2000 au 31 décembre 2008.

2.5.2 Format des tables

Si nous construisons nos lois en considérant le portefeuille comme une seule population homogène, nous obtiendrons une table d'expérience « classique » sous le format d'une table de mortalité à une dimension. On appliquera à une population fictive les taux de sortie estimés q_x de l'état « enfant à charge ». q_x étant la probabilité de sortir entre l'âge x et $x + 1$.

Âge x	L_x
0	$L_0 = 10000$
1	$L_1 = L_0 \times (1-q_0)$
2	$L_2 = L_1 \times (1-q_1)$
...	...
...	...
...	...
18	$L_{18} = L_{17} \times (1-q_{17})$
19	$L_{19} = L_{18} \times (1-q_{18})$
20	$L_{20} = L_{19} \times (1-q_{19})$
21	$L_{21} = L_{20} \times (1-q_{20})$
22	$L_{22} = L_{21} \times (1-q_{21})$
23	$L_{23} = L_{22} \times (1-q_{22})$
24	$L_{24} = L_{23} \times (1-q_{23})$
25	$L_{25} = L_{24} \times (1-q_{24})$

Pour calculer la provision mathématique d'un individu d'âge 18 dans le cas d'une rente annuelle on utilise la formule :

$$PM_{exemple} = R \times \sum_{k=1}^{25-18} \frac{l_{18+k}}{l_{18}} \cdot \frac{1}{(1+i)^k}$$

Conclusion de la partie 2

L'objectif de cette partie a été de réaliser des travaux préparatoires afin de construire des tables d'expérience pertinentes.

Les premiers travaux d'analyse ont eu pour objectif de nous faire réfléchir sur la problématique, sur les durées en jeu et sur les mécanismes des rentes éducation.

L'étude contractuelle du produit et la construction d'une loi de référence nous ont permis d'avoir une première idée de la forme d'une loi de survie dans l'état « enfant à charge » représentative de la population française.

Nous nous sommes ensuite concentrées sur les données du portefeuille. Un processus de vérification et de traitement des données a été mis en place pour obtenir une base sinistre exploitable sur laquelle nous baserons nos calculs détaillés en partie 3.

L'étude statistique du portefeuille nous a permis d'apprécier les caractéristiques des rentiers et de poser l'hypothèse d'un facteur discriminant vis-à-vis de la durée, à savoir l'âge au décès du parent.

Enfin la prise en compte des conditions de scolarité et de la nécessité d'avoir un volume de données suffisant nous ont permis de choisir une période d'observation s'étalant du 1^{er} janvier 2000 au 31 décembre 2008.

3 Construction de lois de survie dans l'état « enfant à charge » à partir des données d'expérience

3.1 L'estimation brute des taux de sortie

3.1.1 Généralités

Pour estimer le taux annuel de sortie à l'âge x , noté q_x nous utiliserons uniquement les individus observés dans l'état « enfant à charge » sur une partie ou sur la totalité de leur $x + 1^{\text{ème}}$ année. On ne tiendra pas compte des observations relatives aux âges voisins, qu'elles proviennent des mêmes individus ou d'individus différents. C'est pourquoi on parle d'estimation brute des taux q_x par opposition aux estimations dites lissées qui, elles, prennent en compte un ensemble d'informations relatives à une plage d'âge.

Nous posons dès à présent une première hypothèse : supposons que sur l'intervalle d'âge $[x ; x + 1]$, toutes les sorties éventuelles des rentiers sont des événements indépendants et que tous les individus utiles à l'estimation du taux q_x ont la même probabilité de sortir dans leur $x + 1^{\text{ème}}$ année.

Ces deux hypothèses font penser au modèle binomiale.

Dans ce modèle $\hat{q}_x = \frac{D_x}{N_x}$ est à la fois l'estimateur du maximum de vraisemblance et celui de la méthode des moments.

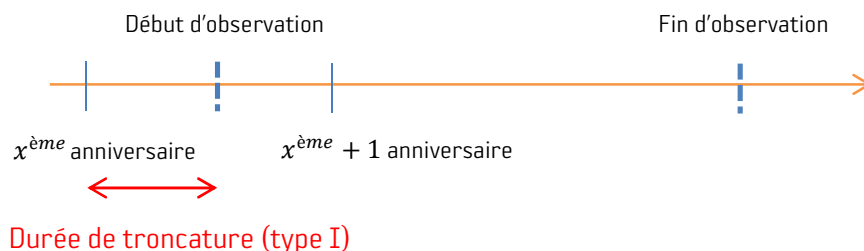
Avec N_x l'ensemble des individus utiles à l'estimation du taux q_x et D_x le nombre de décès dans la population N_x .

Les individus n'étant pas tous observables sur la totalité de la plage $[x ; x + 1]$, nous ne pouvons donc pas appliquer ce modèle. Il existe cependant de nombreuses extensions de cet estimateur en présence de données incomplètes dans la lecture actuarielle.

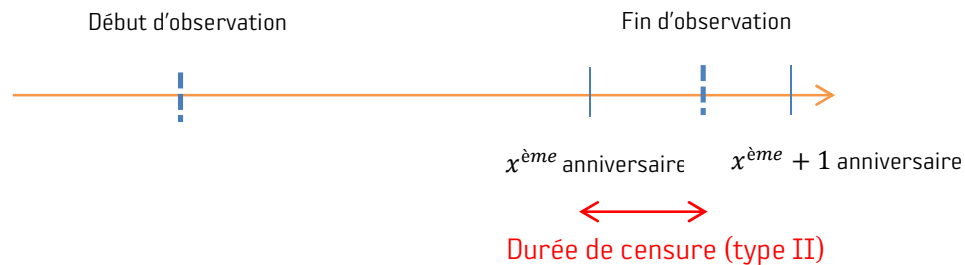
3.1.2 Données incomplètes

Un individu peut ne pas être observable durant la totalité de sa $x + 1^{\text{ème}}$ année principalement pour deux raisons :

- Soit il a fêté son $x^{\text{ème}}$ anniversaire avant le début de la période d'observation et il rentre donc sous observation entre les âges x et $x + 1$. On parle d'observation tronquée à gauche (type I).

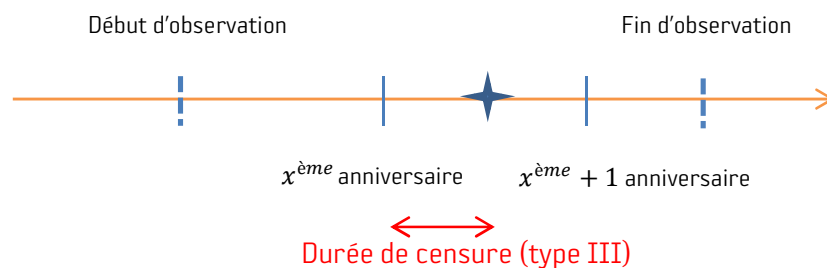


- Soit son $x^{\text{ème}} + 1$ anniversaire (qu'il soit sorti ou encore dans l'état « enfant à charge ») est postérieur à la date de fin d'observation. On parle d'observation censurée à droite (type II).



Dans ces deux situations les âges au début et à la fin d'observation sont connus et non aléatoires. Seule la sortie est aléatoire.

- A ces deux raisons peut s'ajouter un troisième motif d'observation incomplète : l'individu, pour une cause non calendaire (résiliation du contrat, perte d'informations), sort du champ d'observation dans l'état d'enfant à charge à un âge imprévisible (inférieur à $x + 1$). On parle alors de censure à droite à âge imprévisible a priori (type III).



Dans notre étude, nous rencontrerons des données incomplètes de type I ou II.

3.1.3 La méthode de Kaplan-Meier

La méthode de Kaplan-Meier fournit une estimation de la fonction de survie, et à partir de celle-ci on retrouve les taux bruts de sortie. Cette méthode connue dans le cas de données pouvant être censurées à droite a été prolongée aux cas de données pouvant être tronquées à gauche pour des variables discrètes.

Notations :

- S_x : loi discrète de la forme $(a_i, s_i)_{i \in \{1, \dots, m\}}$, les a_i étant les dates connues où il se produit un évènement entre x et $x + 1$, et les s_i les valeurs prises par S_x en a_i .
- T_x : durée de vie résiduelle d'un individu dans l'état enfant à charge conditionnellement au fait qu'il soit dans l'état à l'âge x , i.e. $T_x = [T - x/T > x]$.

- q_i : probabilité de sortir à l'instant a_i .
- n_i : nombre d'individus exposés au risque de sortir de l'état enfant à charge juste avant a_i .
- d_i : nombre d'individus qui sortent à l'instant a_i .
- c_i : nombre d'individus qui sortent par censure à l'instant a_i .
- t_i : nombre d'individus qui entrent par troncature à gauche à l'instant a_i .

Méthode :

On veut estimer q_x , probabilité de sortir dans la classe d'âge $[x, x + 1[$.

On a $q_x = 1 - p_x = 1 - \frac{S(x+1)}{S(x)}$, on cherche donc d'abord à estimer S_x , fonction de survie sur $[x, x + 1[$.

On a $\forall t \in [0; a_m] S_x(t) = \prod_{r/a_r < t} (1 - q_r)$.

On va donc estimer dans un premier temps les $q_r \forall r \in \{1, \dots, m\}$.

Calculs :

On va s'appuyer sur le même raisonnement énoncé plus haut, à savoir que la probabilité de survivre au-delà de $t > s$ peut s'écrire :

$$S(t) = P(T > t/T > s) \times P(T > s) = P(T > t/T > s) \times S(s)$$

Nous pouvons renouveler l'opération et cela fait apparaître des produits de termes en $P(T > t/T > s)$. Si nous choisissons alors comme instant de conditionnement les instants où se produit un évènement (sortie, censure ou troncature), on se ramène à estimer des probabilités de la forme :

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

Un estimateur naturel de $q_i = 1 - p_i$ est $\forall i \in [1, \dots, m], q_i = \frac{d_i}{n_i}$

On obtient donc une expression de $\hat{S}_x(t) = \prod_{i/a_i < t} \left(1 - \frac{d_i}{n_i}\right)$

$$\text{Ainsi } \hat{q}_x = 1 - \prod_{i=a_1}^{i=a_m} \left(1 - \frac{d_i}{n_i}\right)$$

Pour calculer n_i , le nombre de personnes exposées au risque de sortir, nous devons poser des hypothèses au cas où des évènements se produiraient simultanément.

Traitement des ex-æquo :

La probabilité que deux personnes sortent en même temps est nulle presque sûrement. En pratique, vu que le l'on discrétise le temps, on est confronté à la présence d'ex æquo.

Afin d'être en accord avec les lignes directrices de la commission d'agrément de l'Institut des Actuaire : « ... un individu entrant ou sortant par censure en t est conventionnellement considéré comme exposé au risque de décès... » (Ici risque de sortie).

On prend les hypothèses suivantes :

- A l'instant a_i , la sortie se produit avant une censure, c'est à dire un individu sortant par censure en a_i est conventionnellement exposé au risque de sortie à a_i .
- A l'instant a_i , la sortie se produit après une troncature, c'est à dire un individu entrant par troncature en a_i est conventionnellement exposé au risque de sortir en a_i .

Ainsi chronologiquement à l'instant t nous avons : troncature < sortie < censure.

Nous devons calculer pour chaque âge où il se produit un évènement, le nombre d'assurés exposés au risque de sortie.

Les formules suivantes nous permettent de calculer le nombre d'assurés sous risque à l'instant a_i .

$$n_1 = \text{nombre personnes exposées au risque à l'âge } x$$

$$\forall i \geq 2, n_i = n_{i-1} - d_{i-1} - c_{i-1} + t_i$$

Voici un exemple pratique :

Nous souhaitons par exemple estimer q_{19} , c'est-à-dire la probabilité de sortir dans la plage d'âge [19,20[, sachant qu'on est dans l'état enfant à charge à 19 ans. Voici un extrait du tableau qui permet de calculer q_{19} ,

Les deux cellules avec un fond de couleur orange sont utilisées pour calculer le premier terme et le dernier terme du produit de la formule suivante $\hat{q}_x = 1 - \prod_{i=a_1}^{i=a_m} \left(1 - \frac{d_i}{n_i}\right)$.

On obtient dans ce cas-là $q_{19} \cong 0.158$.

Âge a_i	Nb Sorties à l'instant a_i	Nb troncatures a_i	Nb censures a_i	Nb exposés au risque de sortir en a_i	<i>Nb sorties</i> <i>Nb exposés</i>
18.95	6	9	13	3225	0.001860465
18.96	5	4	7	3210	0.001557632
18.97	7	5	16	3203	0.002185451
18.98	3	4	10	3184	0.000942211
18.99	2	3	11	3174	0.00063012
19	5	7	6	3168	$\frac{d_{a_1}}{n_{a_1}} = 0.00157828$
19.01	7	8	7	3165	0.00221169
19.02	2	2	1	3153	0.000634317

19.03	9	7	13	3157	0.002850808
19.04	6	7	7	3142	0.001909612
19.05	3	4	8	3133	0.000957549
19.06	5	6	12	3128	0.001598466
...	...	...	...	...	...
19.88	7	5	6	2441	0.002867677
19.89	2	5	4	2433	0.00082203
19.9	4	3	6	2430	0.001646091
19.91	6	4	6	2424	0.002475248
19.92	6	4	9	2416	0.002483444
19.93	7	5	9	2406	0.002909393
19.94	4	4	1	2394	0.001670844
19.95	4	7	6	2396	0.001669449
19.96	3	6	6	2392	0.001254181
19.97	3	3	5	2386	0.001257335
19.98	6	5	10	2383	0.002517835
19.99	5	4	10	2371	$\frac{d_{am}}{n_{am}} = 0.00210881$
20	1	3	5	2359	0.000423908

3.2 Le lissage des taux bruts de sortie

3.2.1 Généralités

Les estimations âge par âge des taux annuels de sortie forment une courbe de sortie qui se révèle en général assez irrégulière. Au vue de l'étude que l'on a faite sur la durée de scolarisation en France, on peut raisonnablement penser que le taux de sortie augmente avec l'âge. Pour améliorer les estimations brutes, il existe de nombreuses méthodes de lissage qui permettent d'obtenir des estimations corrigées qui progressent avec l'âge.

De manière formelle, la procédure d'estimation initiale a conduit à une valeur $\hat{q}_x$ pour estimer q_x . Cette estimation a donc fait apparaître une erreur $e_x = \hat{q}_x - q_x$. L'objectif du lissage est de diminuer cette erreur, tout en construisant une courbe des taux de sortie en fonction de x plus lisse que la courbe des taux bruts.

Ainsi le choix d'une procédure de lissages fait intervenir deux contraintes qui doivent être prise en compte de manière conjointe :

- La fidélité, il faut que $e_x = \hat{q}_x - q_x$ soit la plus faible possible.
- La régularité.

Il faudra bien entendu valider les procédures de lissage par des tests.

Nous présentons quatre catégories de méthodes permettant d'obtenir des estimations lissées des taux de sortie, nous pourrons ensuite utiliser les taux de sortie pour construire les tables d'expérience.

La première méthode est d'utiliser un modèle paramétrique. **Dans notre étude, nous n'avons pas assez d'informations et d'expérience sur le risque lié aux rentes éducation pour poser une hypothèse forte sur la forme de la loi, nous n'utiliserons donc pas ces modèles.**

D'autres méthodes permettent de ne pas faire d'hypothèses sur la forme de la loi. Par exemple les méthodes de lissages non paramétriques permettent un ajustement assez fidèle aux données d'expérience sans chercher de représentation paramétrique, mais simplement en appliquant aux taux bruts un certain nombre de traitement pour les rendre plus lisses. Nous étudierons l'une d'entre elle, à savoir la méthode de Whittaker-Henderson.

Les méthodes de lissages paramétriques consistent à trouver une courbe paramétrique qui représente bien l'évolution des taux de sortie. Aucune hypothèse stricte ne sera posée sur la forme de la courbe (sinon nous serions dans le cas de la méthode 1), seulement une hypothèse d'appartenance à une famille de fonctions mathématiques (par exemple les fonctions splines). Ces fonctions seront assez flexibles pour obtenir une courbe lissée proche des taux bruts.

Enfin les modèles relationnels peuvent être utilisés, ils partent du même principe que la modélisation paramétrique, à la seule différence que le taux de sortie est désormais exprimé en fonction d'un taux de sortie qui représente une population dite « de référence ». Nous tenterons ainsi de positionner nos lois par rapport à la loi dite de référence et nous étudierons le modèle de Cox dans un chapitre suivant.

3.2.2 La méthode de Whittaker-Henderson

Comme nous l'avons dit en introduction, un lissage fait intervenir deux contraintes qui doivent être prises en compte de manière conjointe : la fidélité et la régularité.

- On peut mesurer la fidélité des taux lissés $\tilde{q}_x$ par rapport aux taux bruts $\hat{q}_x$ à l'aide d'un critère de fidélité noté F . F est la somme des distances euclidiennes pondérées entre $\tilde{q}_x$ et $\hat{q}_x$ pour chaque âge. Il est défini par la formule suivante :

$$F = \sum_{x=0}^{x=25} w_x \times (\tilde{q}_x - \hat{q}_x)^2$$

- On peut mesurer la régularité de la courbe à l'aide d'un critère de régularité noté R qui mesure une distance appelée « distance par différences avant d'ordre z » entre les taux lissés $\tilde{q}_x$. R est défini par la formule suivante :

$$R = \sum_{x=0}^{25-z} (\Delta^z \tilde{q}_x)^2$$

Avec

$$\begin{aligned} \Delta q_x &= q_{x+1} - q_x \\ \Delta^2 q_x &= \Delta(\Delta q_x) = \Delta(q_{x+1} - q_x) = q_{x+2} - 2q_{x+1} + q_x \\ \Delta^z q_x &= \Delta(\Delta^{z-1} q_x) = \Delta(\Delta(\Delta^{z-2} q_x)) = \dots \end{aligned}$$

z fixe le degré du polynôme utilisé pour le critère de régularité. Si l'on fixe $z = 1$, on retrouve la distance euclidienne entre deux valeurs lissées consécutives. Pour tenir compte de suffisamment d'informations consécutives, z est généralement compris entre 2 et 5.

La méthode dite de Whittaker-Henderson combine ces deux critères en minimisant M qui est une combinaison linéaire de F et de R . Elle permet de mettre plus ou moins d'accent sur la régularité grâce au paramètre h .

M se définit donc par la formule suivante : $M = F + h \times R$

Pour trouver les $\tilde{q}_x$, il faut que les n -équations provenant des dérivées partielles de M par rapport à chacun des $\tilde{q}_x$ soient nulles. Il faut donc résoudre le système : $\forall x, \frac{\partial M}{\partial \tilde{q}_x} = 0$

La résolution de ce système d'équations peut être effectuée au moyen de manipulation matricielle. A cet effet, on pose

$$\hat{q}_x = \begin{bmatrix} \hat{q}_0 \\ \hat{q}_1 \\ \dots \\ \dots \\ \hat{q}_{25} \end{bmatrix}; \tilde{q}_x = \begin{bmatrix} \tilde{q}_0 \\ \tilde{q}_1 \\ \dots \\ \dots \\ \tilde{q}_{25} \end{bmatrix}; W = \begin{pmatrix} w_1 & 0 & \dots & \dots & 0 \\ 0 & w_2 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & w_n \end{pmatrix}; \Delta^z \tilde{q}_x = \begin{bmatrix} \Delta \tilde{q}_0 \\ \Delta \tilde{q}_1 \\ \dots \\ \dots \\ \Delta \tilde{q}_{25-z} \end{bmatrix}$$

Avec W matrice diagonale des poids accordés à chaque âge et $\Delta^z \tilde{q}_x$ vecteur de $n - z$ lignes représentant les différences avant d'ordre z de $\tilde{q}_x$.

On peut réécrire en notation matricielle le critère F et le critère R :

$$F = (\tilde{q}_x - \hat{q}_x)^T \times W \times (\tilde{q}_x - \hat{q}_x)$$

$$R = (\Delta^z \tilde{q}_x)^T \times (\Delta^z \tilde{q}_x)$$

Pour détailler l'écriture de R , on introduit la matrice K_z de taille $(25 - z, 25)$ dont les termes sont les coefficients binomiaux d'ordre z et dont le signe alterne et commence positivement pour z pair. Par exemple pour $z = 2$ et si le nombre de taux à lissés est $p = 5$ (dans notre étude $p = 25$).

$$K_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 \end{pmatrix} \text{ et } (\Delta^z \tilde{q}_x) = K_z \tilde{q}_x$$

M s'écrit finalement sous la forme :

$$M = (\tilde{q}_x - \hat{q}_x)^T \times W \times (\tilde{q}_x - \hat{q}_x) + h \times \tilde{q}_x^T \times K_z^T \times K_z \times \tilde{q}_x$$

Ce qui entraîne en développant :

$$M = \tilde{q}_x^T \cdot W \cdot \tilde{q}_x - 2\tilde{q}_x^T \cdot W \cdot \hat{q}_x + \hat{q}_x^T \cdot W \cdot C + h \cdot \tilde{q}_x^T \cdot K_z^T \cdot K_z \cdot \tilde{q}_x$$

$$\frac{\partial M}{\partial \tilde{q}_x} = 2 \cdot W \cdot \tilde{q}_x - 2 \cdot W \cdot \hat{q}_x + 2h \cdot K_z^T \cdot K_z \cdot \tilde{q}_x$$

La résolution de $\frac{\partial M}{\partial \tilde{q}_x} = 0$ conduit à l'expression des taux lissés suivante :

$$\tilde{q}_x = (W + h \cdot K_z^T \cdot K_z)^{-1} \cdot W \cdot \hat{q}_x$$

3.2.3 Positionnement par rapport à la loi de référence

L'utilisation d'une référence externe consiste à rechercher un positionnement de la table d'expérience par rapport à une table de référence. Dans notre cas la table de référence sera la table construite à l'aide des données de marché dans la partie 2.2.

L'approche la plus courante consiste à appliquer un taux d'abattement (ou de majoration) aux taux de la table de référence, ce qui consiste à rechercher un coefficient α tel que $q_x^{exp} = \alpha \times q_x^{ref}$.

On pense à considérer cette relation car en appliquant un coefficient α à la table de référence, on se place dans un modèle à hasard proportionnel dans lequel on suppose la relation :

$$\mu_x^{exp} = \alpha \times \mu_x^{ref} \quad (1)$$

avec μ_x^{ref} connu.

Pour obtenir une relation similaire avec les q_x rappelons que

$$q_x = 1 - p_x = 1 - \frac{S(x+1)}{S(x)} = 1 - \frac{e^{-\int_0^{x+1} \mu(s)ds}}{e^{-\int_0^x \mu(s)ds}} = 1 - e^{-\int_x^{x+1} \mu(s)ds}$$

En faisant l'hypothèse que la fonction de hasard est constante entre deux âges entiers on a donc :

$$q_x = 1 - e^{-\mu_x} \text{ et } \mu_x = -\ln(1 - q_x) \quad (2)$$

En injectant la relation (2) dans l'équation (1) on obtient :

$$-\ln(1 - q_x^{exp}) = -\ln(1 - q_x^{ref})^\alpha \text{ et } 1 - q_x^{exp} = (1 - q_x^{ref})^\alpha \quad (3)$$

La relation (3) est équivalente au premier ordre lorsque les taux sont petits à $q_x^{exp} = \alpha \times q_x^{ref}$.

Pour estimer α on peut considérer comme critère de choix l'écart entre le nombre de sorties observées et le nombre théorique de sorties associées à la table de référence abattue. On note L_x^{obs} l'effectif sous risque à l'âge x de notre population, le nombre de sorties prédit par la table abattue à l'âge x est $\alpha \times q_x^{ref} \times L_x^{obs}$. Si l'on contraint le nombre total de sorties prédit à évaluer le nombre réellement observé, on obtient la formule suivante pour estimer α .

$$\hat{\alpha} = \frac{\sum_x D_x^{obs}}{\sum_x q_x^{ref} \times L_x^{obs}}$$

3.2.4 Critère de validation des courbes lissées

Les critères de validation fournissent des aides à la décision lors de la sélection des courbes les plus pertinentes pour notre étude.

Le premier critère est motivé par l'utilisation que l'on veut faire de nos tables et sert à vérifier qu'elles sont bien adaptées aux contraintes du contrat.

Si ces tables sont utilisées pour le provisionnement, il ne faudra pas perdre de vue l'application du principe de prudence i.e. le maintien en rente éducation ne doit pas être sous-estimé.

Si ces tables sont utilisées dans une logique Solvabilité II, on devra s'employer à être le plus réaliste possible.

Notre deuxième critère de validation sera un critère plus quantitatif. Ce sera le calcul de la distance du Khi-deux. Nous mesurerons cette distance entre taux bruts et taux lissés et vérifierons que la loi lissée ne s'écarte pas trop de la loi brute.

d. Premier critère : quelle utilisation ?

Ce critère est motivé par le fait que suivant le contexte d'utilisation des tables, l'ajustement peut exiger des caractéristiques sensiblement différentes. Ce critère qualitatif reposera sur l'analyse des graphiques. On vérifiera que le positionnement de la loi lissée reste logique suivant qu'on soit dans un contexte de provisionnement ou de calcul de best estimate.

e. Deuxième critère : fidélité aux données

La première des exigences que doit satisfaire notre courbe est d'être fidèle aux données qui ont servies à la construire. Cette fidélité peut être examinée par le calcul de la distance du Khi-deux, qui va nous fournir une mesure de l'écart entre les taux bruts et les taux lissés et entraînera un rejet si cet écart est trop grand.

Pour cela il faut calculer la statistique suivante :

$$Y = \sum_{x=0}^{25} \frac{(N_x \cdot \hat{q}_x - N_x \cdot \tilde{q}_x)^2}{(N_x \cdot \tilde{q}_x)}$$

$N_x \cdot \hat{q}_x$: Nombre de sorties estimées à l'âge x .

$N_x \cdot \tilde{q}_x$: Nombre de sorties estimées à l'âge x après lissage.

Dans le cas d'un ajustement paramétrique par maximum de vraisemblance avec r paramètres, la distribution (asymptotique) de Y est un X^2_{26-r-1} ; cependant dans le cas d'un lissage non paramétrique le nombre de degrés de liberté peut être difficile à déterminer.

3.3 Estimation et lissage des taux bruts du portefeuille global

Après avoir présenté les méthodes théoriques liées à la construction d'une table d'expérience, effectuons pas à pas les différentes étapes nécessaires à la construction de la loi de survie dans l'état « enfant à charge ».

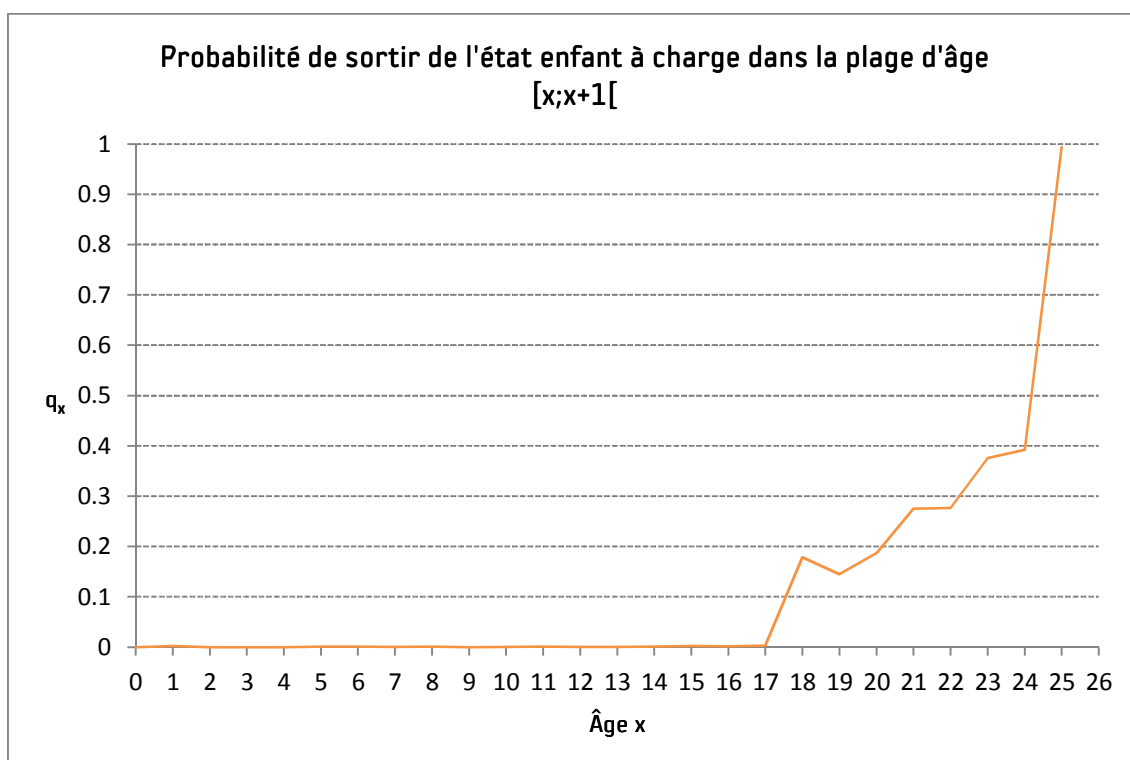
Dans la section 3.3 on considère le portefeuille dans sa globalité, c'est-à-dire que l'on considère que les individus ont des caractéristiques assez homogènes pour former une seule population.

Nous envisagerons dans le deuxième paragraphe 3.4 de réfuter l'hypothèse d'homogénéité de notre portefeuille en segmentant la population en sous-groupes plus homogènes, pour capter l'éventuelle hétérogénéité de la population.

3.3.1 Taux bruts

L'estimation des taux bruts a été réalisée grâce à la méthode de Kaplan-Meier.

Voici les probabilités estimées q_x qu'un enfant sorte de l'état « enfant à charge » dans sa $x + 1^{\text{ème}}$ année.



On observe un pic à l'âge 18, il s'explique par les caractéristiques du contrat où à partir de cet âge-là, le paiement de la rente n'est plus automatique. Des conditions supplémentaires sur le statut de l'enfant sont exigées (être scolarisé, être demandeur d'emploi), le portefeuille perd donc tous les rentiers qui ne répondent plus à ces conditions.

Cette courbe présente des irrégularités, surtout entre les âges 18 et 25 ans qu'il convient de lisser.

3.3.2 Lissage

a. Lissage non paramétrique : Méthode de Whittaker-Henderson

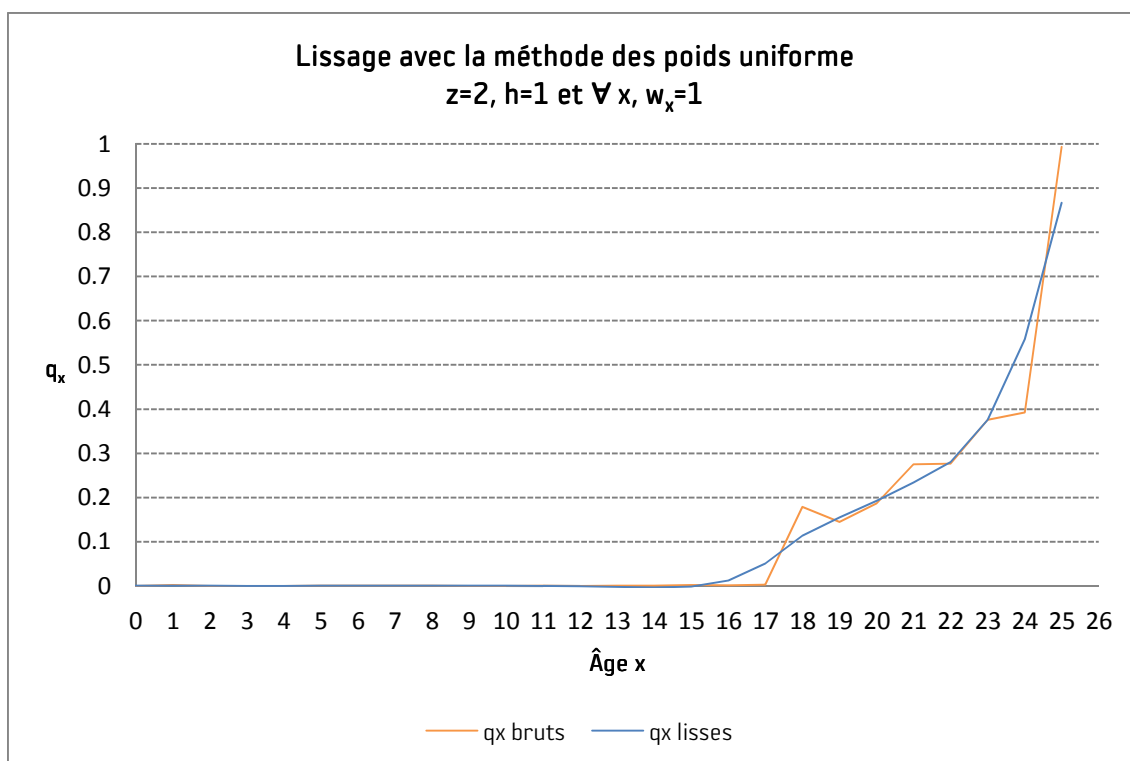
Les paramètres h et z

Les simulations réalisées permettent d'aboutir aux observations suivantes :

- Le paramètre $z = 1$ ne lisse pas suffisamment les irrégularités et nous écartons également les lissages basés sur des différences « avant » avec plus de trois valeurs consécutives ($z > 3$), les conditions de sortie peuvent être très différentes à trois années d'écart. Pour le paramètre z , nous avons le choix entre les valeurs 2 et 3.
- Le paramètre h sert à contrôler l'importance que l'on veut donner à la régularité par rapport à la fidélité. Ainsi $h = 0$ donnera une courbe qui collera exactement à la courbe brute et h très grand donnera une courbe très proche d'une droite. Comme nous l'avons déjà dit nous n'avons pas beaucoup de recul sur les durées associées au versement de rentes éducation, nous souhaitons donc ne pas trop nous éloigner des taux bruts, pour cela nous décidons d'accorder autant d'importance à la fidélité qu'à la régularité pour le lissage en fixant $h = 1$.

Les poids

Il faut affecter à chaque $\hat{q}_x$ un poids w_x . Une des méthodes consiste à affecter le même poids à tous les âges. Voici le lissage obtenu avec les paramètres suivants et la méthode des poids uniformes :



A vu d'œil ce lissage n'est pas acceptable, il sous-estime beaucoup trop les taux de sortie aux âges 18 et 25 ans et il ne présente pas une allure satisfaisante entre les âges 18 et 25 ans. La méthode consistant à accorder le même poids à toutes les observations n'est pas bonne.

Le contrat présentant des spécificités à certains âges clés, il convient d'accorder plus d'importance et donc un poids plus important à certaines observations.

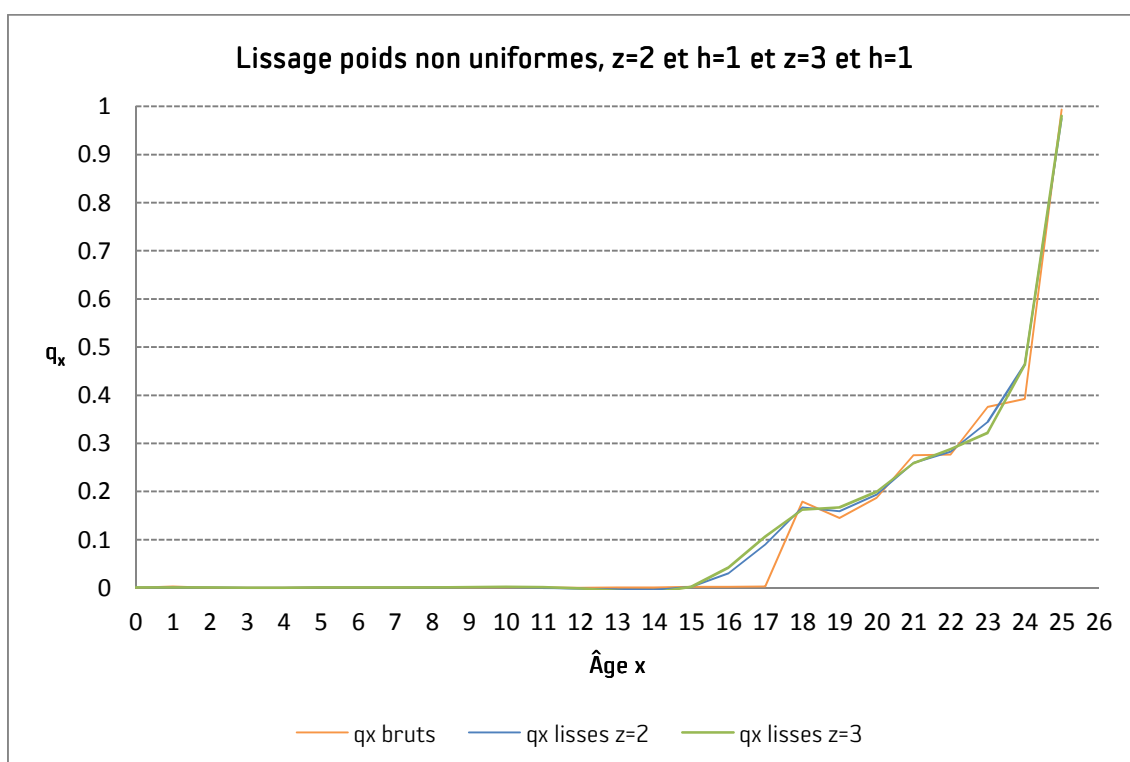
Les observations correspondantes aux âges 18 et 25 ans sont les plus importantes à nos yeux car elles doivent refléter les réalités contractuelles du contrat.

La plage d'âge $]18,25[$ est également importante, car c'est sur cette plage que jouent plusieurs conditions de sortie. En revanche il y a peu d'incertitude pour les âges inférieurs à 18 ans où seul le décès est une cause de sortie.

Nous décidons donc d'affecter les poids par ordre d'importance, cela nous conduit à définir le vecteur des poids comme ceci :

$$\begin{aligned} \forall i \in [0,17], w_i &= 1 \\ \forall i \in]18,25[, w_i &= 10 \\ w_{18} &= 20 \text{ et } w_{25} = 20 \end{aligned}$$

Voici les courbes obtenues avec la méthode des poids par ordre d'importance.

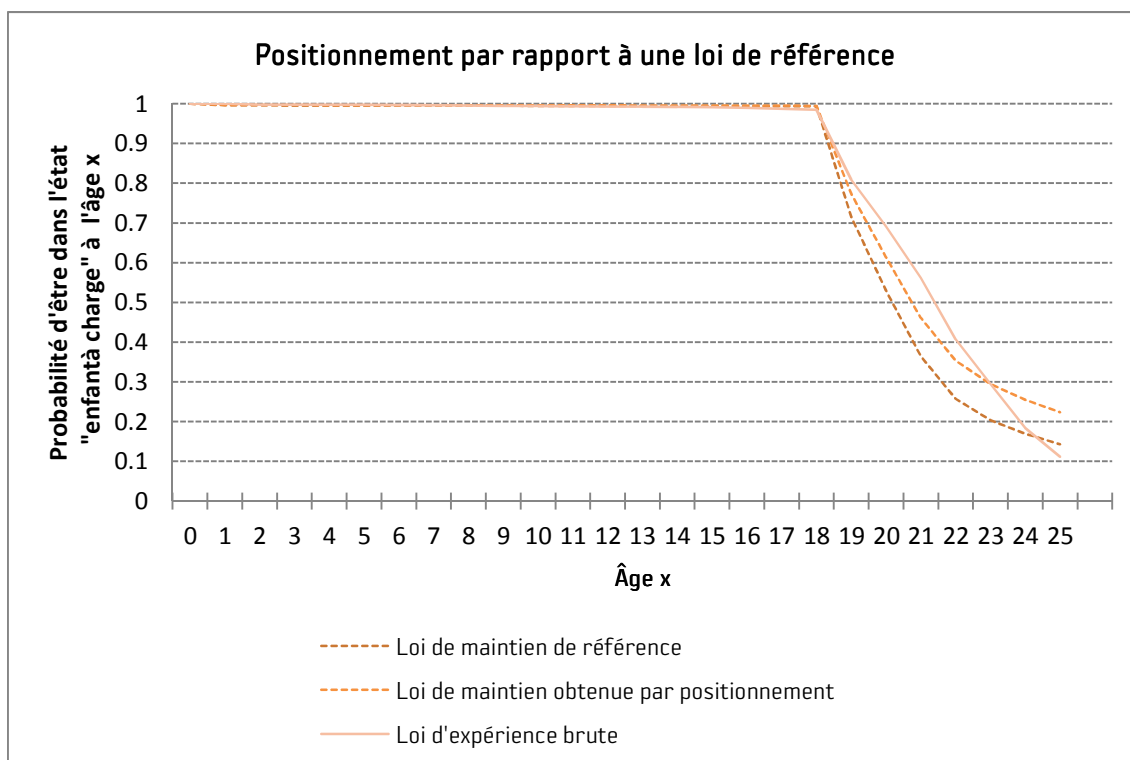


Ce lissage nous paraît plus acceptable à première vue, nous verrons si nous pouvons le valider à l'aide de critères qualitatifs et quantitatifs dans la partie validation du lissage.

b. Positionnement par rapport à une référence externe

En appliquant la méthode exposée plus haut, on trouve une estimation de $\hat{\alpha} = 0.7959$.

Sur le graphique suivant nous avons superposé notre loi de référence, notre loi d'expérience brute et notre loi de maintien obtenue par la méthode du positionnement.



La première chose qui apparaît sur ce graphique est que les courbes liées au portefeuille rente éducation (courbe brute ou courbe obtenue par positionnement) sont au-dessus de la courbe de référence obtenue grâce aux données de marché.

Il semblerait que le fait de toucher une rente encourage de manière significative les orphelins à poursuivre leurs études ou à être reconnus demandeur d'emploi.

La deuxième chose visible sur le graphique est que la méthode de lissage par positionnement sur nos données est approximative, en effet il y a des écarts non négligeables entre la loi obtenue par positionnement et la loi brute. Avant 23 ans la courbe obtenue par la méthode du positionnement sous-estime le maintien réel du portefeuille dans l'état enfant à charge, alors qu'après 23 ans, cette même courbe le surestime.

La loi positionnée ne peut être utilisée qu'à titre indicatif.

En effet nous avons considéré que la relation $q_x^{exp} = \alpha \times q_x^{ref}$ était équivalente à la relation $1 - q_x^{exp} = (1 - q_x^{ref})^\alpha$, cela est vrai pour des taux très petits, or dans notre cas plus l'âge se rapproche de 25 ans plus les taux de sortie sont grands, l'approximation ainsi faite engendre des erreurs non négligeables.

Ce positionnement nous a quand même renseignés sur les différences de comportement entre les orphelins rentiers et les jeunes français vis-à-vis des études. Dans l'ensemble on peut présupposer que les enfants ne touchant pas de rente ont des taux de sortie supérieurs aux rentiers de notre portefeuille (le paramètre $\hat{\alpha}$ nous donne un ordre d'idée pour quantifier cette différence).

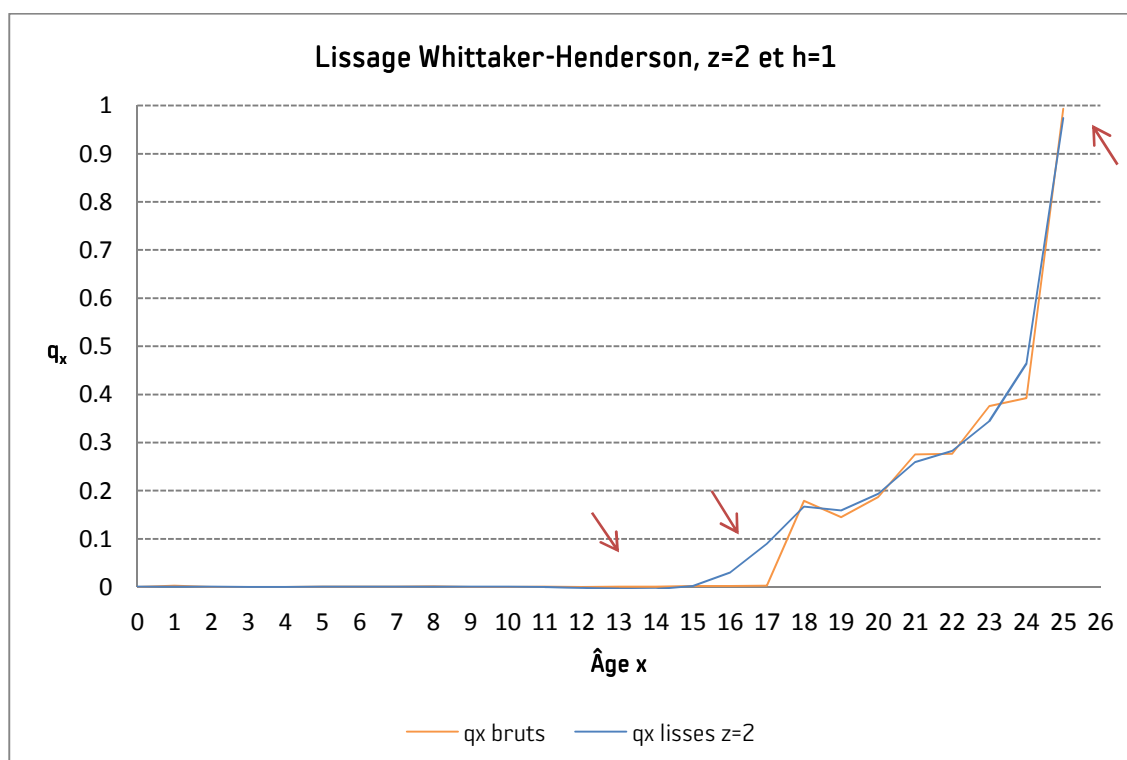
3.3.3 Adéquation de l'ajustement

Nous allons maintenant choisir à l'aide de nos critères de validation le lissage le plus adapté à notre étude. Nous écartons la méthode du « positionnement par rapport à une référence externe ». Il nous reste donc à choisir entre les deux lissages par la méthode de Whittaker Henderson.

a. Premier critère : quelle utilité

Nous souhaitons construire des tables afin d'améliorer le provisionnement des rentes éducation. Ces tables doivent être prudentes, elles ne doivent pas surestimer la sortie de l'état. Notre premier choix va donc vers la courbe obtenue avec les paramètres $z = 2$ et $h = 1$ légèrement en-dessous de la courbe obtenue avec les paramètres $z = 3$ et $h = 1$ pour les âges inférieurs à 21 ans.

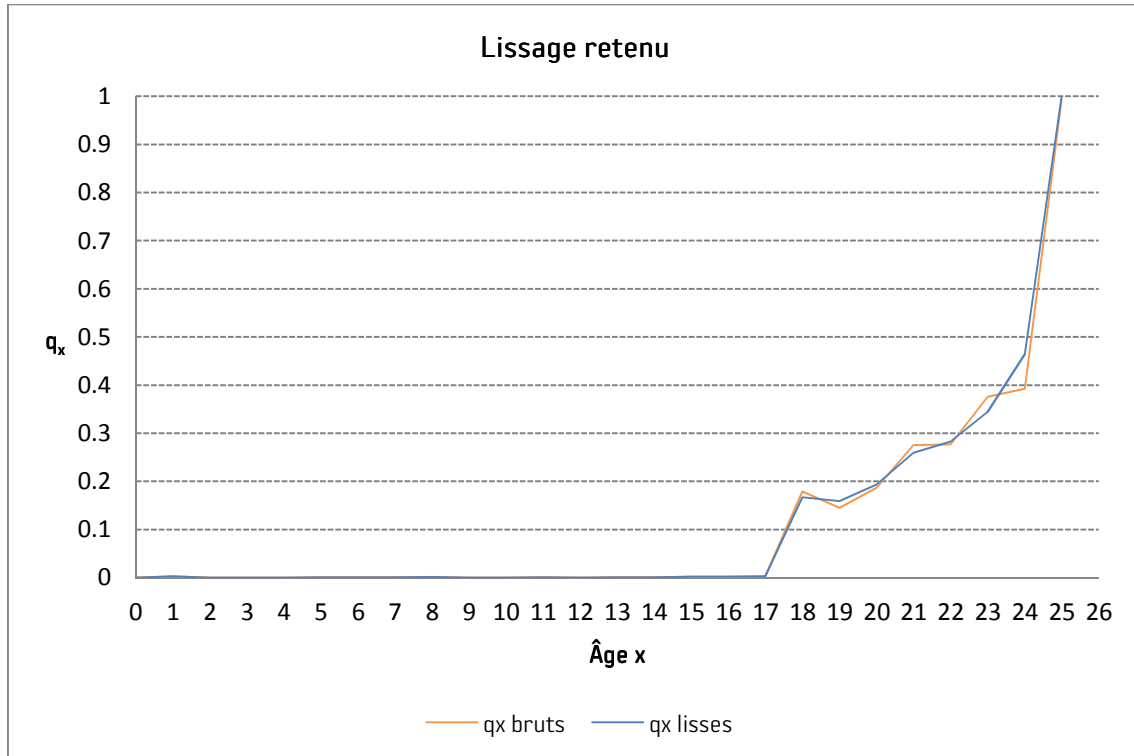
Nos tables doivent également coller aux contraintes du contrat, or en observant la courbe nous remarquons trois incohérences (marquées par des flèches sur le schéma).



Sur la plage d'âge [12, 15] ans le lissage propose des q_x négatifs, ce qui est impossible théoriquement. Il suggère également des taux de sortie entre 15 et 18 ans bien trop élevés, en effet pour les âges inférieurs à 18 ans, q_x doit représenter uniquement le taux de mortalité. Afin de corriger ces incohérences, nous décidons de forcer la valeur de chacun de ces taux pour qu'ils soient égaux aux taux

bruts avant l'âge de 18 ans. Nous forçons également q_{25} à 1, les rentiers ayant une probabilité certaine de sortir dans leur 26^{ème} année.

Le lissage finalement retenu se présente comme ceci :



b. Deuxième critère : fidélité aux données

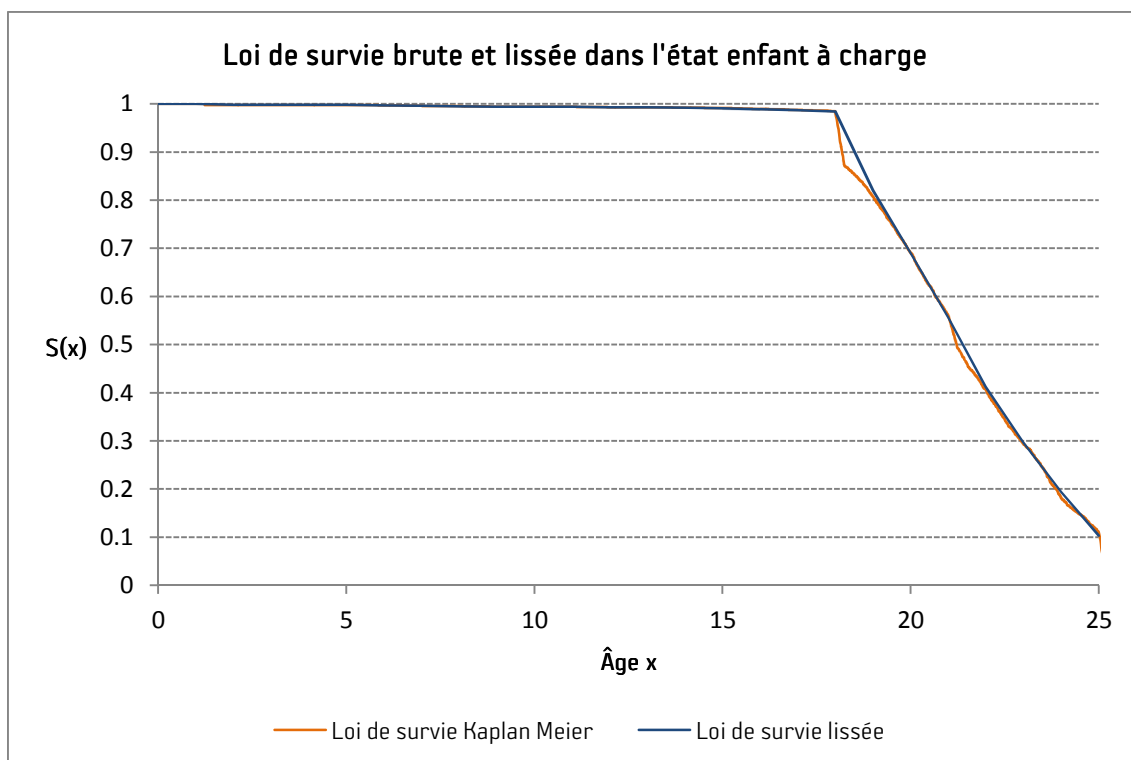
On calcule à présent la statistique suivante afin de valider de manière quantitative ce lissage.

$$Y = \sum_{x=0}^{25} \frac{(N_x \cdot \hat{q}_x - N_x \cdot \tilde{q}_x)^2}{N_x \cdot \tilde{q}_x}$$

Cette distance du Khi-deux ne nous permet pas de faire un test statistique au sens rigoureux car la loi asymptotique est difficile à déterminer dans ce cas-là. Nous comparons néanmoins cette distance à la statistique d'un Khi-deux à 25 degrés de liberté pour nous assurer que nous n'avons pas trop lissé les taux.

La valeur de 15.62 est tout à fait acceptable car elle est en dessous de la valeur limite d'un Khi-deux à 25 degré de liberté au seuil 5%.

Nous considérons comme valide cet ajustement, il nous fournit la loi de maintien suivante :



A l'issue de ce chapitre nous avons obtenu une première loi de survie dans l'état enfant à charge. Elle modélise le comportement du portefeuille dans sa globalité, en effet nous avons considéré que tous les rentiers avaient des caractéristiques homogènes vis à vis des durées passées dans l'état « enfant à charge ». Essayons à présent d'affiner cette modélisation en segmentant la population en sous-groupes de risque plus homogènes vis-à-vis de la durée étudiée.

3.4 Segmentation du portefeuille en sous-populations : estimation et lissage

Nous sommes en présence d'hétérogénéité si l'on mélange des populations ayant des caractéristiques différentes vis-à-vis d'un phénomène de durée étudiée.

L'objectif de la segmentation est de découper notre population hétérogène initiale en sous-populations plus homogènes vis-à-vis de la durée étudiée pour nous permettre de faire une estimation la plus fine possible.

Afin de détecter l'hétérogénéité d'une population, on peut faire appel à des tests d'hypothèses qui vont nous permettre de conclure sur l'éventuel besoin de segmenter la population initiale. Un des tests les plus connus est le test du Log-Rank. C'est une généralisation d'un test non paramétrique usuel au cas où la variable est censurée. Il permet de comparer des survies sans avoir à faire d'hypothèses sur leurs distributions.

Nous pensons que la variable « âge de l'enfant à la date du décès du participant » peut être une variable explicative de la variable « durée passée dans l'état enfant à charge ». En effet nous pensons que l'âge auquel l'enfant subit le décès d'un parent peut influencer sur sa durée de scolarité.

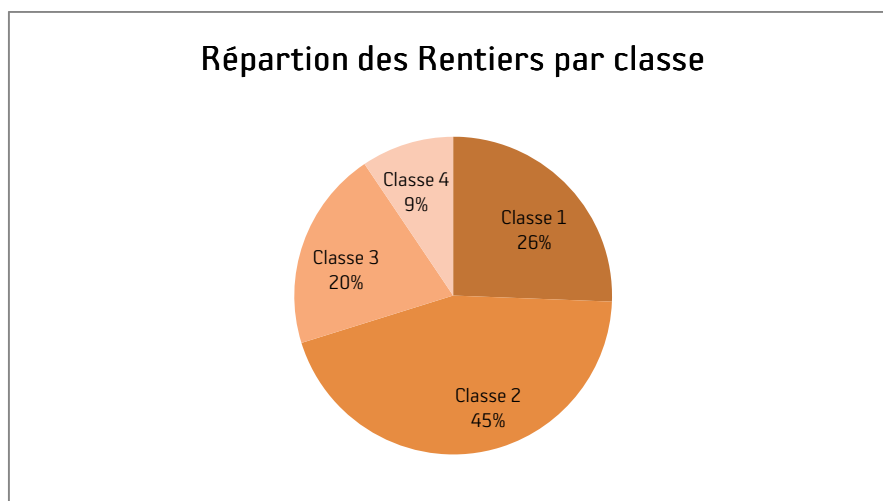
Nous aimerions donc tester dans un premier temps s'il existe une association entre ces deux variables afin de justifier une éventuelle segmentation de la population.

Nous réalisons dans un premier temps un « test de tendance » qui est une généralisation du test du Log-Rank dans le cas où l'on souhaite tester l'association entre une durée de survie et une variable quantitative (ici la variable âge).

Ce test effectué avec l'instruction « test » de la proc LIFETEST du logiciel SAS nous permet de rejeter l'hypothèse d'absence de lien entre les deux variables.

Nous pouvons donc conclure qu'il serait pertinent de segmenter la population à l'aide de cette variable. Pour cela nous créons des classes d'âges et tâchons d'établir une loi de survie dans l'état enfant à charge par classe d'âge. Ces classes d'âges ont été choisies pour être en adéquation avec les différentes étapes du système scolaire. Les classes d'âges retenues sont :

- Classe 1 « crèche/maternelle » : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]0 ; 6]. Nous supposons que ces enfants ne sont pas encore scolarisés ou sont en maternelle.
- Classe 2 « primaire/collège » : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]6; 14]. Nous supposons que ces enfants sont scolarisés à l'école primaire ou au collège.
- Classe 3 « lycée » : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]14; 18]. Nous supposons que la majeure partie du groupe est scolarisé au lycée.
- Classe 4 « études supérieures » : Population comprenant des enfants ayant vécu le décès d'un parent lorsqu'ils avaient entre]18; 25]. Nous supposons que la majeure partie du groupe est scolarisé dans un cycle d'études supérieures.



Le choix de ces classes nous permet d'isoler les âges clés en lien avec la structure du système scolaire.

La classe 1 nous permet d'observer le comportement des enfants qui sont devenus orphelin très jeunes.

La classe 2 nous permet d'observer le comportement des enfants qui sont entrés en rente en primaire ou au collège. On remarque qu'à cette période l'enfant n'est pas confronté à des choix d'orientations scolaires majeurs.

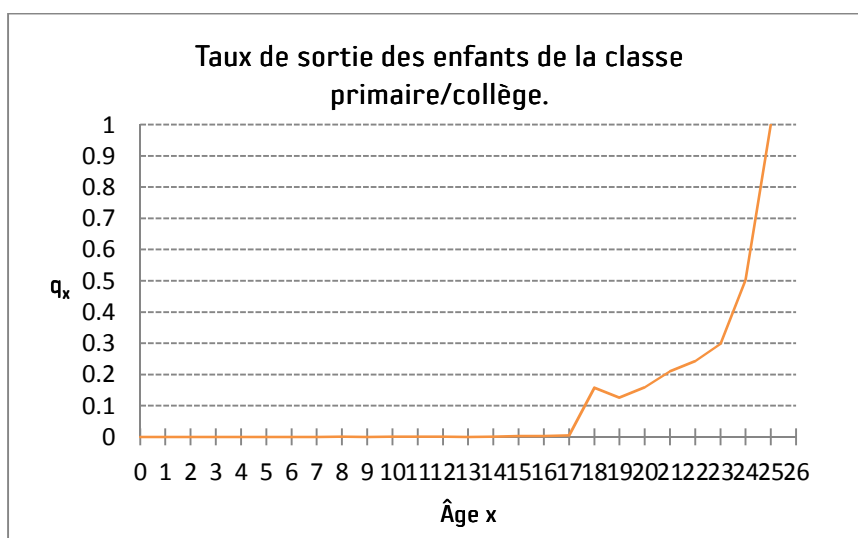
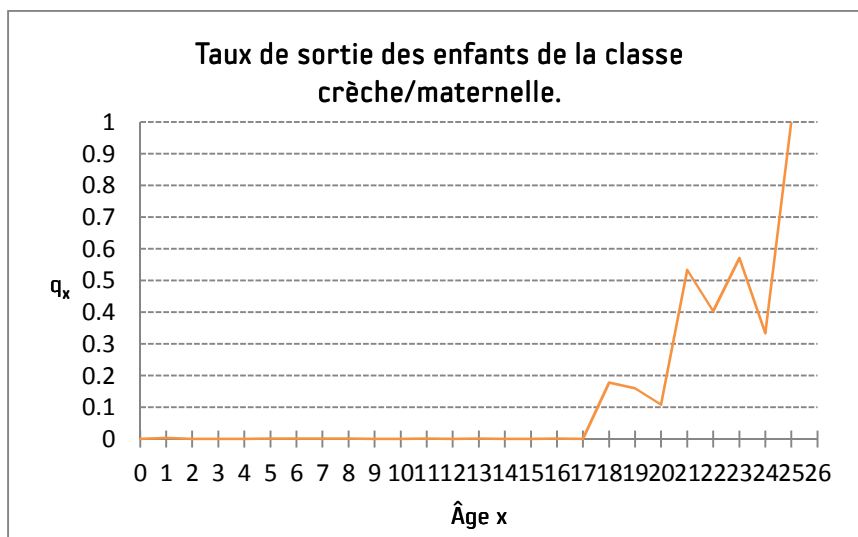
La classe 3 nous permet d'observer le comportement des enfants qui sont entrés en rente lorsqu'ils étaient en fin de collège ou au lycée. Pendant cette période l'adolescent est confronté à ses premiers choix d'orientations qui auront sans doute un impact sur sa durée de scolarisation.

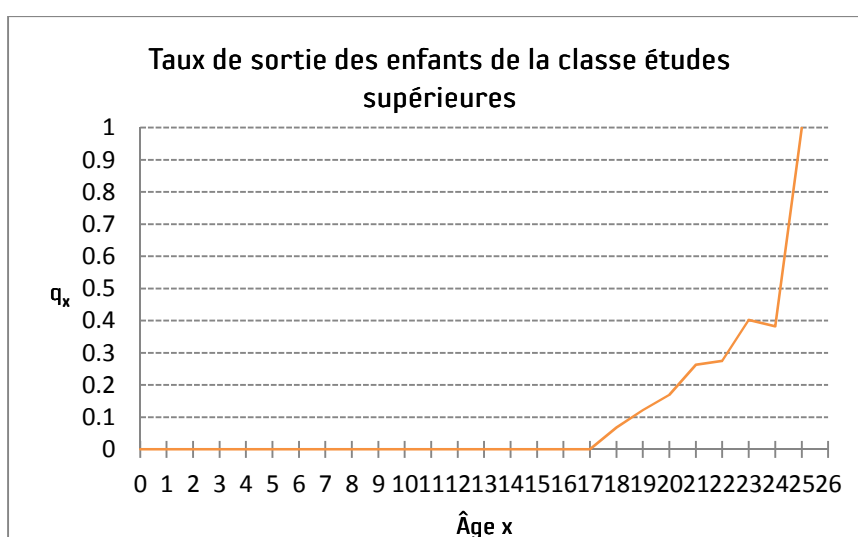
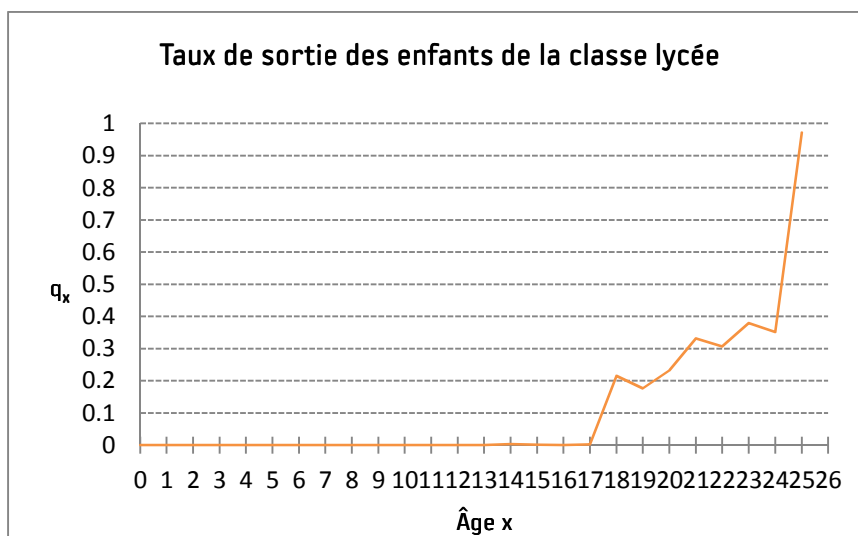
La classe 4 nous permet d'observer le comportement des enfants qui sont entrés en rente à partir de 18 ans. A cette période, la majeure partie des rentiers aura déjà fait ses choix d'orientations et entamée un cycle d'études supérieures.

Pour construire nos lois de survie, nous allons suivre la même démarche que précédemment : estimation puis lissage des courbes mais cette fois ci sur chaque sous-population. Nous obtiendrons au final cinq courbes. Nous pourrons alors les comparer et valider à posteriori le choix des classes.

3.4.1 Taux bruts

Nous estimons les taux bruts de sortie sur chaque sous-population :





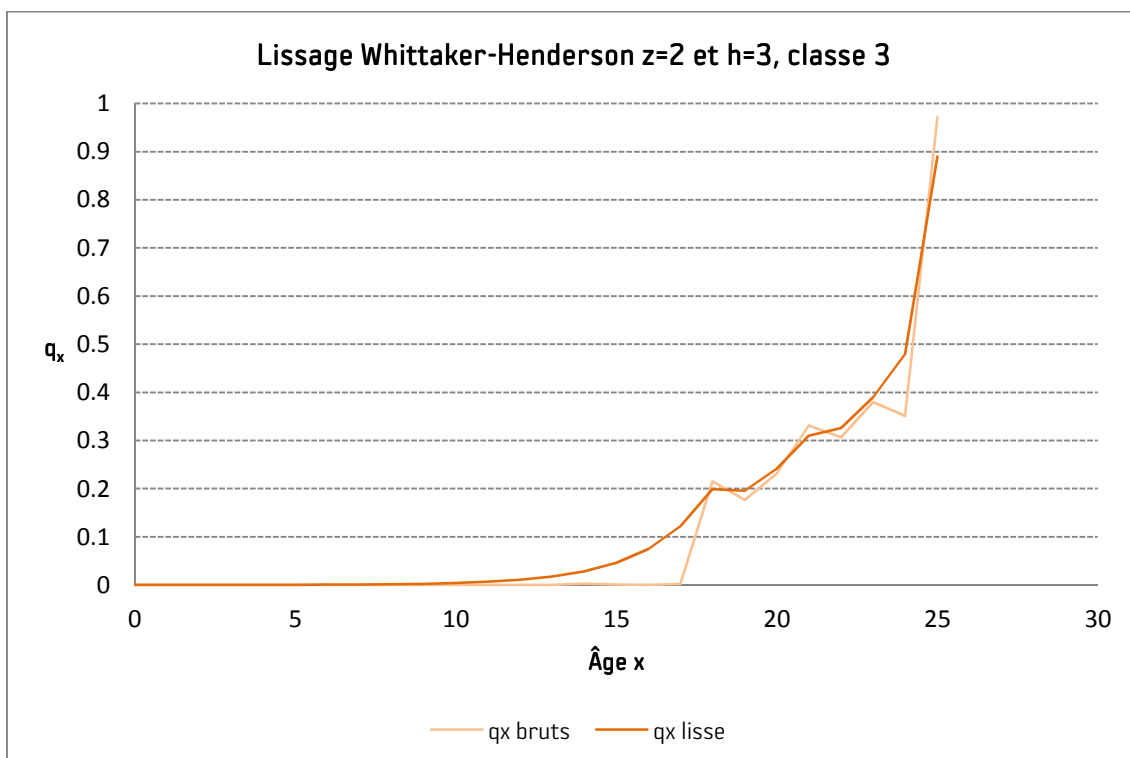
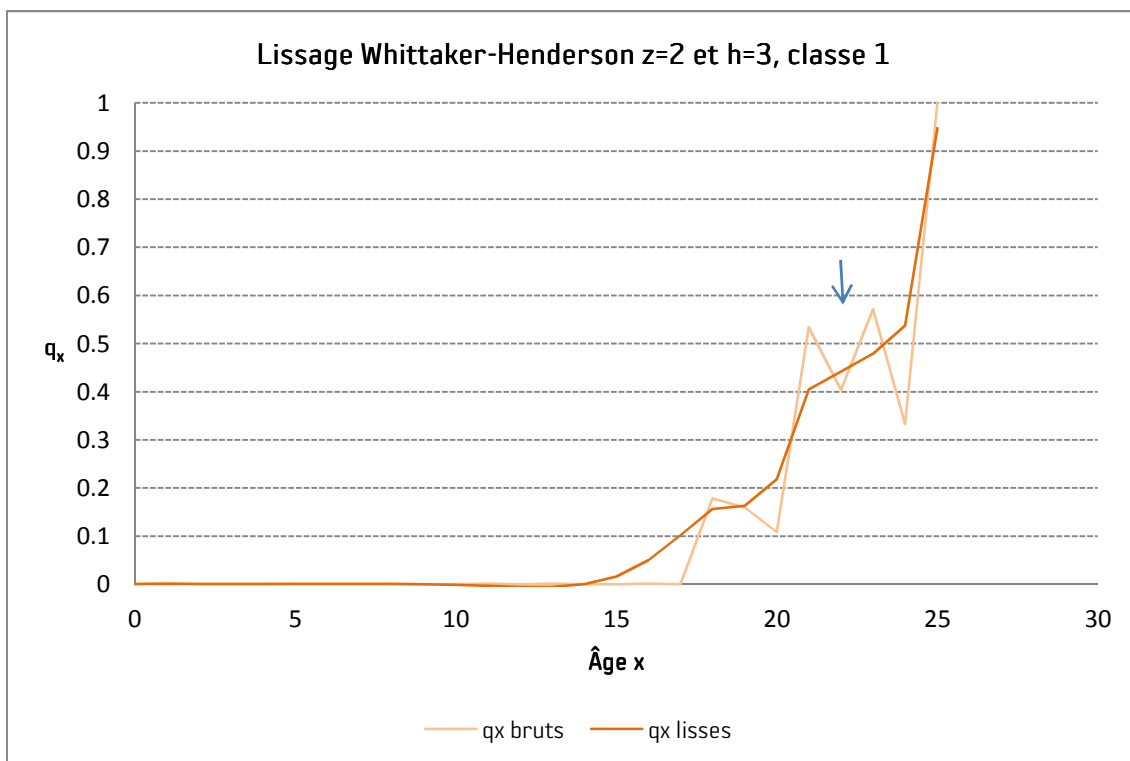
Les trois premières courbes présente un pic autour de l'âge 18, cette caractéristique est toujours liée aux conditions de versement de la rente. La classe 4 ne présente pas de pic à l'âge 18, ce qui est normal car dans ce groupe, les rentiers entrent à partir de l'âge 18 ans.

Nous remarquons dès à présent que ce sont les classe 1 et 3 qui possèdent les taux de sortie les plus élevés et volatiles. Une attention particulière sera portée pour le lissage de ces courbes.

3.4.2 Lissage et adéquation

Sur les courbes brutes on observe quelques incohérences qui doivent être corrigées par le lissage. La population ayant été segmentée, les estimations sont faites sur des volumes de données moins importants et on s'expose à des erreurs plus importantes. Par exemple sur la courbe 1 on voit deux pics

majeurs à 21 ans et 23 ans avec une chute à 22 ans et d'après nos connaissances sur la structure du système scolaire il n'y a aucune raison d'observer un tel phénomène, nous souhaitons avoir des taux qui progressent avec l'âge pour cette période.

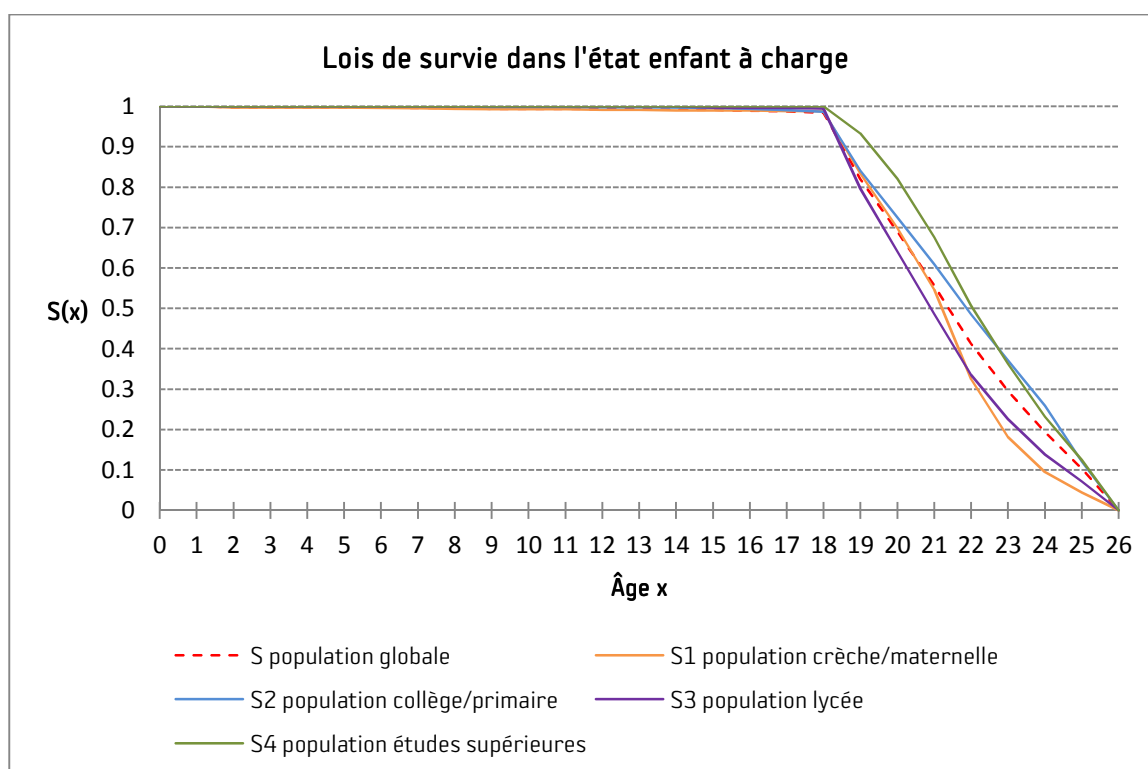


3.4.3 Les 5 courbes

Sur ce graphique, les cinq lois de survie dans l'état enfant à charge ont été reconstituées et nous voyons bien que les courbes sont différentes.

Les lois représentant la survie des enfants dans l'état « enfant à charge » dont le décès survient lorsqu'ils ont entre]0 ; 6] ans (classe 1) ou entre]14 ; 18] ans (classe 3) semblent se positionner en dessous de la loi globale.

A l'inverse les lois représentant le maintien des enfants dont le décès survient lorsqu'ils ont entre]6 ; 14] ans (classe 2) ou entre]18 ; 25] ans (classe 4) semblent se positionner au-dessus de la loi globale.



Nous aimerions pouvoir comparer et quantifier les différences entre ces courbes, c'est l'objet de la section suivante.

3.5 Comparaison des courbes de survie et analyse de l'impact sur le provisionnement

Pour confirmer et quantifier les différences entre les courbes de survie, nous allons faire des tests d'homogénéité et calculer des indicateurs permettant de résumer chaque table.

3.5.1 Comparaison des courbes de survie

a. Le test du Log-Rank

Nous souhaitons comparer la survie de deux populations ayant vécu le décès d'un parent à des âges différents. La survie est certes une variable quantitative, nous ne pouvons cependant pas utiliser un simple test qui permettrait de comparer la survie moyenne dans les deux populations. En effet la distribution de survie étant asymétrique, l'hypothèse de normalité n'est pas vérifiée.

Notre deuxième idée se porte sur un test de rang, test non paramétrique permettant de comparer deux moyennes sans avoir à faire l'hypothèse de normalité. Ce n'est toujours pas possible car nous sommes en présence de données censurées. Nous utilisons alors le test du Log-Rank qui est une généralisation d'un test non paramétrique usuel cette fois-ci en présence de données censurées.

Ce test s'intéresse à la différence entre les deux courbes à chaque événement e_i .

L'hypothèse nulle H_0 que l'on cherche à tester, est celle de l'égalité des probabilités de sorties de l'état « enfant à charge » pour chaque classe, à chaque instant i.e. l'égalité des fonctions de survies pour chaque classe.

A chaque instant, on prédit le nombre de sorties de chaque classe, comme si chacune était issue de la même population.

Le nombre de sorties prédites à l'instant i dans la classe j se calcule comme ceci :

$$\hat{d}_{ij} = \frac{d_i \cdot e_j(i)}{e_i}$$

d_i : nombre de sorties dans la population totale,

e_i : nombre de rentiers sous risque à l'instant t_i ,

$e_j(i)$: nombre de rentiers sous risque dans la classe j à l'instant t_i .

Nous tenons compte de la censure dans le calcul de $e_j(i)$ et de e_i et le nombre total de sorties prédites pour chaque classe s'écrit alors :

$$\hat{d}_j = \sum \hat{d}_{ij}$$

On réalise alors un test de χ^2 avec le nombre de sorties réellement observées :

$$\chi^2 = \frac{\sum_j (d_j - \hat{d}_j)^2}{\sum_j (\hat{d}_j)^2}$$

Ce test permet de valider si l'hypothèse H_0 est vérifiée ou non.

Nous décidons de tester les courbes deux à deux au niveau 95%. Ci-dessous le récapitulatif des p-value obtenues grâce au logiciel SAS et la PROC LIFETEST.

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1				
Classe 2	0,0135			
Classe 3	0.0131	0.0018		
Classe 4	<.0001	<.0001	<.0001	

Avec un risque de première espèce de 5%, l'hypothèse nulle d'absence de différence de survie entre chaque couple est rejetée, l'hypothèse d'égalité des fonctions de survie est donc rejetée.

Les résultats de ce tests sont cependant à prendre à titre indicatif, car habituellement le Khi-deux utilisé pour calculer le seuil de significativité du test nécessite de nombreux temps de sorties ou de nombreuses sorties à chaque temps de sortie. Ce qui n'est pas le cas dans notre étude car nous présentons un fort taux de censure.

b. Espérance de vie résiduelle

Nous aimerions comparer les caractéristiques de nos tables grâce à des indicateurs. L'espérance de vie résiduelle dans l'état enfant à charge est un indicateur assez concret pour résumer nos tables et pour se rendre compte du phénomène de sortie de l'état.

L'espérance de vie résiduelle à l'âge x noté e_x^0 s'interprète comme la durée de vie moyenne au-delà de l'âge x .

Si l'on considère la variable aléatoire $T_x = [T - x | T > x]$ comme la durée résiduelle dans l'état enfant à charge d'un individu conditionnellement au fait qu'il soit dans l'état à l'âge x . L'espérance de vie restante e_x^0 est définie par :

$$e_x^0 = E[T_x] = \frac{1}{l_x} \cdot \int_{h \geq 0} l_{x+h} \cdot dh = \int_{h \geq 0} {}_h p_x \cdot dh$$

L'intégrale nous donne le nombre moyen d'années vécues dans l'état au-delà de l'âge x par les individus ayant atteint cet âge, et la division par l_x donne donc le nombre moyen d'années restant à vivre pour chacun des l_x individus ayant atteint l'âge x .

Le calcul de e_x^0 nécessite la formulation d'une hypothèse à propos de la répartition des sorties au sein de chaque année d'observation.

On suppose donc que les sorties sont uniformément réparties entre deux âges entiers x et $x + 1$. Cela veut dire qu'en moyenne les sorties se produisent en milieu d'années. Cette hypothèse n'est pas incohérente avec notre étude où les fins d'années scolaires arrivent en milieu d'année.

L'hypothèse de répartition uniforme des sorties peut s'écrire :

$${}_tq_x = t \times q_x \text{ ou } {}_tp_x = 1 - t + t \times p_x \text{ car } {}_tq_x = 1 - {}_tp_x$$

On a ainsi

$$\begin{aligned} e_x^0 &= \int_{t \geq 0} {}_tp_x dt \\ &= \sum_{k \geq 0} \int_{t=k}^{t=k+1} {}_kp_x \cdot {}_{t-k}p_{x+k} dt \\ &= \sum_{k \geq 0} {}_kp_x \cdot \int_{t=0}^{t=1} {}_tp_{x+k} \cdot dt \end{aligned}$$

On utilise l'hypothèse ${}_tp_{x+k} = 1 - {}_tq_{x+k} = 1 - t \cdot q_{x+k}$

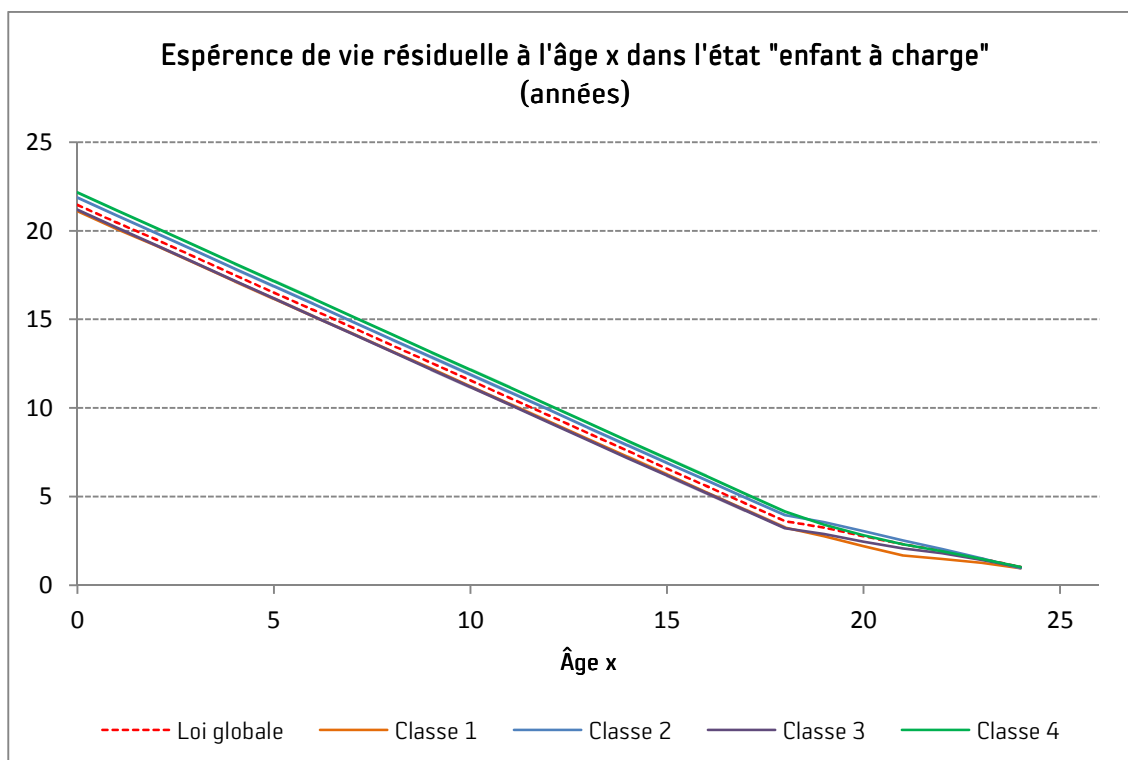
$$\begin{aligned} &= \sum_{k \geq 0} {}_kp_x \cdot \int_{t=0}^{t=1} (1 - t \cdot q_{x+k}) \cdot dt \\ &= \sum_{k \geq 0} {}_kp_x \cdot \left(1 - \frac{q_{x+k}}{2}\right) \\ &= \sum_{k \geq 0} {}_kp_x - \frac{1}{2} \cdot \sum_{k \geq 0} {}_kp_x \cdot q_{x+k} \\ &= \sum_{k \geq 1} {}_kp_x + 1 - \frac{1}{2} \cdot \sum_{k \geq 0} {}_kp_x \cdot q_{x+k} \end{aligned}$$

On a $\sum_{k \geq 0} {}_kp_x \cdot q_{x+k} = 1$ et $\sum_{k \geq 0} {}_kp_x = 1 + \sum_{k \geq 1} {}_kp_x$

Donc finalement

$$e_x^0 = -\frac{1}{2} + \sum_{k \geq 0} {}_kp_x$$

Ci-dessous un graphique résumant les espérances de vie à chaque âge dans l'état enfant à charge pour chacune des cinq tables :



Comme on pouvait s'y attendre la classe 1 et la classe 3 sont les classes qui proposent une durée de vie moyenne dans l'état enfant à charge la plus faible.

3.5.2 Impact sur le provisionnement

Il est intéressant de mesurer l'impact que peut avoir l'utilisation de tables segmentées sur le provisionnement. Pour chaque classe, nous allons calculer la provision associée à un rentier avec la loi d'expérience globale puis avec la loi spécifique de la classe d'appartenance. Nous pourrions avoir une première indication sur la nécessité ou non d'utiliser les tables spécifiques à chaque classe notamment si les écarts de provisions sont importants.

Pour le calcul de la provision représentant les engagements de l'assureur vis-à-vis de l'enfant du participant, nous allons prendre dans un premier temps des hypothèses simplifiées qui ne correspondent pas aux réalités du contrat. Nous affinerons ces hypothèses dans la dernière partie de cette étude accordée au provisionnement réel de notre client.

Supposons que la garantie prévoit le versement d'une rente annuelle jusqu'à 25 ans sous condition d'être dans l'état enfant à charge. La valeur actuelle des engagements de l'assureur vis-à-vis d'un rentier d'âge x peut s'écrire :

$$VAP = R \cdot \sum_{k=1}^{25-x} v^k \cdot \frac{l_{x+k}}{l_x}$$

Le calcul de cette valeur est effectué pour les âges correspondant à l'âge moyen de chaque classe, une fois avec les l_x provenant de notre loi globale, une fois avec les l_x provenant de la loi spécifique.

Les résultats du tableau ont été obtenus en prenant $R = 2000 \text{ €}$ et $i = 2.75\%$.

On mesure l'écart relatif entre ces deux valeurs afin de quantifier l'impact de la segmentation.

	Loi de survie globale (1)	Loi de survie par classe d'âge (2)	Ecart relatif $\frac{(1)-(2)}{(1)}$
Classe 1 : simulation avec un rentier de 3 ans	28004.3 €	27607.6 €	1.42%
Classe 2 : simulation avec un rentier de 10 ans	18742.1 €	19205.3 €	-2.47%
Classe 3 : simulation avec un rentier de 16 ans	9269.7 €	8600.6 €	7.22%
Classe 4 : simulation avec un rentier de 20 ans	4257.2 €	4360.2 €	-2.42%

Pour les simulations effectuées avec les rentiers des classes 2 et 4, la loi globale propose un montant de provision inférieur à celui calculé avec les lois issues de la segmentation et inversement pour les classes 1 et 3.

Au vu de ces chiffres et si les effectifs restent constants, un provisionnement avec la loi globale paraît suffisant car les classes se compensent entre elles. Il faudra tout de même veiller à ce que le portefeuille ne se déforme pas dans le temps car si les enfants appartenant à la classe 2 et 4 viennent à occuper la majorité du portefeuille, la compagnie peut se trouver en sous-provisionnement si elle utilise la loi globale. Il sera conseillé dans ce cas-là d'utiliser les tables spécifiques à chaque classe.

Conclusion de la partie 3

Après avoir rappelé la théorie liée aux méthodes d'estimation et de lissage des taux bruts de sortie, nous les avons appliquées sur le portefeuille dans sa globalité puis sur le portefeuille segmenté en quatre sous-populations.

Les taux bruts ont été estimés dans les deux cas par la méthode de Kaplan-Meier et nous avons étudié deux pistes de lissage : le positionnement par rapport à une référence externe (la loi de marché construite en partie 2) et le lissage non paramétrique de Whittaker-Henderson, le dernier ayant été retenu.

La segmentation du portefeuille suivant la variable « âge de l'enfant au décès du parent » nous a ensuite permis de construire quatre lois d'expérience correspondant chacune à un profil de risque différent.

Nous avons ensuite comparé entre elles les lois issues de la segmentation pour valider l'intérêt de celle-ci et mesuré l'impact de l'utilisation de ces lois sur le provisionnement. Nous obtenons ainsi les premières pistes de réflexion qui seront reprises en dernière partie pour déterminer la méthode la mieux adaptée à notre étude à savoir utiliser la loi globale ou les lois segmentées pour provisionner cette garantie.

4 Une autre approche, la modélisation semi-paramétrique : le modèle de Cox

Jusqu'ici nous avons souhaité expliquer la durée passée dans l'état « enfant à charge » à l'aide de la variable explicative « âge de l'enfant à la survenance du décès », nous avons donc estimé les distributions de survie sur les différentes tranches de cette variable et construit des tables de survie appropriées à chaque classe.

Nous aimerions trouver un modèle permettant de hiérarchiser les sous-populations : de la moins risquée à la plus risquée vis-à-vis de la survie dans l'état enfant à charge.

Nous allons essayer de modéliser nos durées à l'aide d'un modèle à hasard proportionnel. Avec ces modèles, les variables explicatives ont un effet multiplicatif par rapport à une fonction de hasard de base et c'est cette propriété qui retient notre attention. Cela nous permettra de quantifier les différences entre les sous-populations même si la taille des données reste limitée. Avec cette modélisation nous espérons pouvoir dire suivant l'âge auquel est survenu le décès, si le risque de sortir prématurément de l'état se trouve augmenté ou diminué et de combien par rapport aux autres populations.

Parmi les modèles à hasard proportionnel, deux familles s'offrent à nous :

- Celle où l'on suppose la fonction de hasard de base « connue » : par exemple lorsque l'on veut positionner la sortie d'un groupe spécifique par rapport à un groupe référence. Nous aurions pu ici reprendre notre raisonnement consistant à positionner la loi de nos rentiers par rapport à la loi de référence obtenue grâce aux chiffres de l'Education Nationale et de l'INSEE, mais il nous est impossible d'exprimer précisément le taux de hasard instantané de celle-ci.
- Celle où l'on ne suppose pas de forme particulière pour la fonction de hasard de base : en effet si nous ne sommes intéressés que par la mesure de l'effet des variables explicatives alors la spécification de la fonction de hasard de base n'apporte rien au modèle.

Nous sommes ici dans un contexte où l'objectif est le positionnement de différentes populations les unes par rapport aux autres, sans considération du niveau absolu du risque. Nous allons donc choisir d'étudier un modèle à hasard proportionnel où la fonction de base est inconnue, ce modèle se nomme le modèle de Cox.

4.1 Présentation du modèle

4.1.1 Généralités

Le modèle de Cox est un outil semi-paramétrique d'estimation de la fonction de survie, permettant la prise en compte de l'hétérogénéité des données en faisant intervenir des variables explicatives. Il peut également intégrer les phénomènes de censure et de troncature.

Le modèle de Cox définit la fonction de risque instantané d'un évènement, conditionnelle à un ensemble de covariables fixes, comme le produit d'une fonction positive déterministe inconnue $h_0(t)$ et d'une

fonction de régression sur l'ensemble des covariables. Pour que $h_0(t)$ soit positive, la fonction de régression choisie est la fonction exponentielle.

Le modèle appliqué à la durée passée dans l'état « enfant à charge » peut être formulé de la façon suivante :

$$h(t, Z) = h_0(t) \times e^{\beta^T \cdot Z}$$

Où $\beta^T = (\beta_1, \dots, \beta_p)$ est un vecteur de p coefficients inconnus et Z est le vecteur des variables explicatives caractérisant l'individu.

La fonction de survie s'exprime alors par la formule suivante :

$$S(t) = \exp\left(-\int_0^t h(s, Z) ds\right) = \exp\left(-\int_0^t h_0(s) ds \cdot e^{\beta^T \cdot Z}\right)$$

La fonction $h_0(t)$ est appelée fonction de risque instantanée de base et est considérée comme un paramètre de nuisance, l'intérêt résidant en général dans l'estimation du paramètre de régression β car on s'intéresse principalement à l'association entre Z et la survenance de la sortie.

Cox a proposé de baser l'inférence statistique pour les modèles contenant un paramètre de nuisance, sur une vraisemblance partielle ne contenant pas ce paramètre.

Prenons un vecteur aléatoire X de densité $f_X(x; \theta)$, où θ est un vecteur de paramètres décomposables en $\theta = (\Phi, \beta)$. Supposons que l'on s'intéresse à β et que Φ est le paramètre de nuisance.

Supposons également que X peut se décomposer en (V, W) et que la densité de X puisse être réécrite comme un produit :

$$f_X(x, \theta) = f_{W/V}(w/v; \theta) \cdot f_V(v; \theta)$$

Si l'un des facteurs du produit ne dépend pas de Φ , on peut penser baser l'inférence statistique pour β sur ce seul facteur. Ceci entraîne une perte d'information si le terme ignoré dépend de β . Néanmoins, s'il ne contient que peu d'informations sur β , cette perte peut apparaître comme un inconvénient mineur au regard de la simplicité de l'inférence développée sur le terme retenu. C'est sur ce principe qu'est basée l'estimation par maximum de vraisemblance partielle proposée par Cox.

Pour être en accord avec ce principe, nous allons donc écrire la vraisemblance de notre modèle et la décomposer afin de pouvoir isoler dans un terme que l'on négligera, l'incidence de la fonction de hasard de base. L'estimation de β se fera alors par la méthode du maximum de vraisemblance sur la vraisemblance partielle.

4.1.2 De la vraisemblance du modèle de Cox à la vraisemblance partielle de Cox

Nous considérons le triplet (X_i, D_i, Z_i) , où $X_i = T_i \wedge C_i$, $D_i = 1_{\{T_i \leq C_i\}}$ et Z_i est le p -vecteur de covariables qui explique l'individu i .

X_i est la durée dans l'état enfant à charge de l'individu i que l'on a pu observer.

T_i est la durée réelle passée dans l'état.

C_i est la durée que l'on a pu observer avant que l'individu ne sorte par censure.

D_i nous indique si X_i représente toute la durée ou si celle-ci a été censurée.

Nous supposons que :

- T et C sont indépendants conditionnellement à Z .
- La censure est non informative : i.e. les instants de censure et les covariables qui leur sont associées ne contiennent pas d'information sur β .
- La distribution de Z ne dépend pas des paramètres β et $h_0(t)$.

En ignorant les termes impliquant les distributions de C et de Z , la vraisemblance obtenue au vue d'un n -échantillon $(X_i, D_i, Z_i)_{1 \leq i \leq n}$ s'écrit donc :

$$L_n(\beta, \Lambda_0) = \prod_{i=1}^n [h_0(X_i) \cdot e^{\beta^T \cdot Z_i} \cdot e^{-\Lambda_0(X_i) \cdot \exp(\beta^T \cdot Z_i)}]^{D_i} \cdot [e^{-\Lambda_0(X_i) \cdot \exp(\beta^T \cdot Z_i)}]^{1-D_i}$$

$$\text{où } \Lambda_0(X_i) = \int_0^{X_i} h_0(s) ds$$

Estimer le paramètre $\Lambda_0(X_i) = \int_0^{X_i} h_0(s) ds$ relève d'un problème d'estimation fonctionnelle. Il n'existe pas de maximum de la vraisemblance lorsque $\Lambda_0(X_i) = \int_0^{X_i} h_0(s) ds$ appartient à l'ensemble des fonctions positives. La méthode d'estimation du maximum de vraisemblance ne peut pas être appliquée.

Cox a donc proposé d'estimer β à partir d'une vraisemblance partielle qui s'exprime comme ceci :

$$L^*(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta' \cdot Z_i)}{\sum_{j=1}^n \exp(\beta' \cdot Z_j) \cdot 1_{T_i \leq T_j}} \right]^{D_i}$$

Essayons de comprendre le sens de cette expression : les observations peuvent s'écrire sous la forme d'une suite ordonnée d'événements C_i et D_i avec :

C_i : ensemble des censures et troncatures intervenants entre $]t_{i-1}, t_i]$

D_i : ensemble des sorties intervenues en t_i .

La vraisemblance des observations s'écrit alors :

$$L = L(C_1, D_1, \dots, C_k, D_k)$$

En décomposant la vraisemblance pas à pas en conditionnant par les événements déjà passés, la vraisemblance peut se réécrire :

$$L = L(C_1, D_1, \dots, C_k, D_k) = L(C_1) \cdot L(D_1|C_1) \cdot L(C_2|C_1, D_1) \cdot \dots \cdot L(D_k|C_1, \dots, C_{k-1}, D_1, \dots, D_{k-1})$$

En regroupant les événements relatifs aux sorties d'une part et ceux relatifs aux censures d'autre part, notre vraisemblance peut s'écrire :

$$L = \left\{ \prod_{i=1}^k P(D_i|C_1, \dots, C_i, D_1, \dots, D_{i-1}) \right\} \times \left\{ \prod_{i=1}^k P(C_i|C_1, \dots, C_i, D_1, \dots, D_{i-1}) \right\}$$

L'idée de base de la vraisemblance partielle de Cox consiste à ignorer dans la vraisemblance le terme associé aux censures d'après l'hypothèse de censure non informative.

$$L \text{ devient } L^* = \left\{ \prod_{i=1}^k P(D_i|C_1, \dots, C_i, D_1, \dots, D_{i-1}) \right\}$$

L'évènement $R_i = \{C_1, \dots, C_i, D_1, \dots, D_{i-1}\}$ décrit la population sous risque juste avant t_i et il reste donc à évaluer $P(D_i|R_i)$.

C'est-à-dire la probabilité que ce soit pour le sujet expliqué par D_i que se produise la sortie à l'instant t_i sachant qu'un évènement se produit en t_i et connaissant l'ensemble des sujets à risque à cet instant. Cette probabilité s'exprime comme suit si on suppose l'absence d'ex aequo :

$$P(D_i|R_i) = \frac{h(t_i, Z_i)}{\sum_{j \in R_i} h(t_i, Z_j)}$$

$$= \frac{h_0(t_i) \cdot \exp(\beta^T \cdot Z_i)}{\sum_{j \in R_i} h_0(t_i) \cdot \exp(\beta^T \cdot Z_j)}$$

Le paramètre nuisible $h_0(t)$ s'élimine, ce qui conduit à retomber sur l'expression de la vraisemblance partielle de Cox énoncée plus haut.

4.1.3 Estimation des paramètres du modèle

Toujours dans le cas d'absence aequo, l'estimateur β est obtenu en maximisant la vraisemblance partielle ou de façon équivalente la log vraisemblance partielle.

Cette dernière s'écrit :

$$\log(L^*(\beta)) = \sum_{i=1}^n \sum_{j=1}^p \beta_j \cdot Z_{ji} - \sum_{i=1}^n \log \left(\sum_{l \in R_i} \left(\exp \left(\sum_{j=1}^p \beta_j \cdot Z_{jl} \right) \right) \right)$$

En dérivant p fois cette expression par rapport à β_p nous obtenons un vecteur U tel que

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_p \end{pmatrix} \text{ avec } u_j = \left(\frac{\partial \log(L^*(\beta))}{\partial \beta_j} \right) = \sum_{i=1}^n Z_{ji} - \sum_{i=1}^n \frac{\sum_{l \in R_i} Z_{jl} \cdot \exp(\beta^T \cdot Z_l)}{\sum_{l \in R_i} \exp(\beta^T \cdot Z_l)}$$

Le vecteur β est alors estimé en résolvant le système d'équation engendré par $U = 0$.

Pour cela il nécessaire d'avoir recours à des méthodes numériques d'optimisation comme l'algorithme de Newton-Raphson.

4.1.4 Tests statistiques

Les tests statistiques ont pour objet de vérifier la pertinence de l'utilisation du modèle de Cox. Les vérifications à effectuer peuvent se décomposer en deux étapes :

- Test de significativité des variables explicatives.
- Vérification de l'hypothèse de hasard proportionnel.

a. Tests de significativité des variables explicatives

Ces tests aussi appelés test de nullité des coefficients consistent à tester l'absence d'influence des variables explicatives. Cela revient à tester l'hypothèse que les coefficients de régression sont nuls.

Pour tester l'hypothèse $H_0: \beta = 0$, trois tests sont à notre disposition :

Test du score qui calcule la statistique $U^T(0) \cdot I^{-1}(0) \cdot U(0)$ avec I matrice d'information de Fischer.

Test de Wald qui calcule la statistique $\hat{\beta}^T \cdot I(\hat{\beta}) \cdot \hat{\beta}$.

Test du rapport de vraisemblance qui calcule la statistique $-2[L^*(0) - L^*(\hat{\beta})]$.

Ces trois statistiques suivent asymptotiquement sous l'hypothèse H_0 , des distributions de χ^2 à p degrés de liberté (p : nombre de variables explicatives).

Pour tester le coefficient de régression de chacune des variables explicatives nous utiliserons le test de Wald avec un seuil en dessous duquel l'hypothèse est rejetée de 5%.

b. L'hypothèse de hasard proportionnel

Le modèle de Cox repose sur l'hypothèse forte de proportionnalité. Cette hypothèse peut être vérifiée en pratique graphiquement. En effet du fait de la constance des effets des covariables au cours du temps pour deux groupes homogènes expliqués par Z_1 et Z_2 , le rapport des taux de hasard cumulés est indépendant du temps et s'écrit :

$$\frac{\int_0^t h(s, Z_1) \cdot ds}{\int_0^t h(s, Z_2) \cdot ds} = \frac{\int_0^t h_0(s) \cdot ds \cdot \exp(\beta^T \cdot Z_1)}{\int_0^t h_0(s) \cdot ds \cdot \exp(\beta^T \cdot Z_2)} = \exp(\beta^T \cdot (Z_1 - Z_2))$$

Ainsi si le modèle de hasard proportionnel est adéquat, pour tous les groupes j et k , les points $\left(t, \ln\left(\int_0^t \widehat{h(s, Z_1)} \cdot ds - \int_0^t \widehat{h(s, Z_2)} \cdot ds\right)\right)$ sont alignés sur une droite constante.

On peut également vérifier de manière équivalente que les $\ln(-\ln(S(t)))$ vs $\ln(t)$ de chaque groupe sont parallèles.

4.2 Mise en œuvre

4.2.1 Le modèle avec nos données

Nous souhaitons toujours étudier l'impact de la variable âge de l'enfant à la survenance du décès sur la durée passée dans l'état « enfant à charge ».

Nous discrétisons pour cela cette variable de la même manière que précédemment, c'est-à-dire en fonction de la structure de système scolaire et de façon à mettre en évidence les périodes clés de la vie d'un enfant comme sa petite enfance ou son adolescence.

Nous souhaitons choisir notre population de référence parmi les classes présentant les risques de maintien les plus élevés. Nous décidons donc de choisir la classe 2 « primaire/collège » représentant la tranche d'âge]6,14] car c'est elle qui présente le volume de données le plus important.

Notre modèle s'écrit :

$$h(t, Z) = h_0(t) \cdot \exp(\beta_1 \cdot \text{classe}_1 + \beta_2 \cdot \text{classe}_2 + \beta_3 \cdot \text{classe}_3)$$

$h_0(t)$: fonction de hasard de base estimée grâce à la tranche [6,14[.

β_i : coefficient de la régression.

classe_i : variable binaire prenant les valeurs 0 ou 1 suivant que l'enfant appartienne ou non à la tranche d'âge correspondante.

$\text{classe}_1 \rightarrow]0,6] \text{ ans}$

$\text{classe}_2 \rightarrow]14,18] \text{ ans}$

$classe_3 \rightarrow]18,25] \text{ ans}$

Pour connaître le taux de hasard de la population de référence, i.e. de la tranche d'âge] 6,14], on reprend l'équation de notre modèle avec $\beta_i = 0 \forall i \in \{1; 2; 3\}$.

4.2.2 L'estimation des paramètres

Nous effectuons cette modélisation grâce à la procédure PHREG du logiciel SAS qui prend en charge toutes les étapes du processus de modélisation. Voici les sorties de l'estimation des paramètres du modèle et des tests de significativité de chacune des variables explicatives :

Estimations par l'analyse du maximum de vraisemblance						
Paramètre	DDL	$\hat{\beta}$	Erreur type	Khi-2	Pr > Khi-2	Rapport de risque
$classe_1$	1	0.24119	0.10995	4.8118	0.0283	1.273
$classe_2$	1	0.32836	0.04432	54.8800	<.0001	1.389
$classe_3$	1	0.10344	0.05413	3.6516	0.0560	1.109

On peut rejeter l'hypothèse H_0 de nullité des coefficients pour les variables $classe_1$ et $classe_2$ par contre on ne peut pas rejeter l'hypothèse de nullité pour la variable $classe_3$.

Ainsi cette variable n'est pas significative, on décide donc de supprimer cette classe en insérant les rentiers de cette classe à la population de référence qui a manifestement un comportement proche.

Notre nouveau modèle s'écrit :

$$h(t, Z) = h_0(t) \cdot \exp(\beta_1 \cdot classe_1 + \beta_2 \cdot classe_2)$$

Nous lançons de nouveau la procédure pour obtenir les résultats suivant :

Estimations par l'analyse du maximum de vraisemblance						
Paramètre	DDL	$\hat{\beta}$	Erreur type	Khi-2	Pr > Khi-2	Rapport de risque
$classe_1$	1	0.21165	0.10868	3.7927	0.0415	1.236
$classe_2$	1	0.28812	0.03869	55.4493	<.0001	1.334

L'hypothèse H_0 est rejetée pour les coefficients de régression associés aux variables $classe_1$ et $classe_2$. Ils sont donc non nuls et la segmentation s'avère pertinente dans l'implémentation du modèle de Cox.

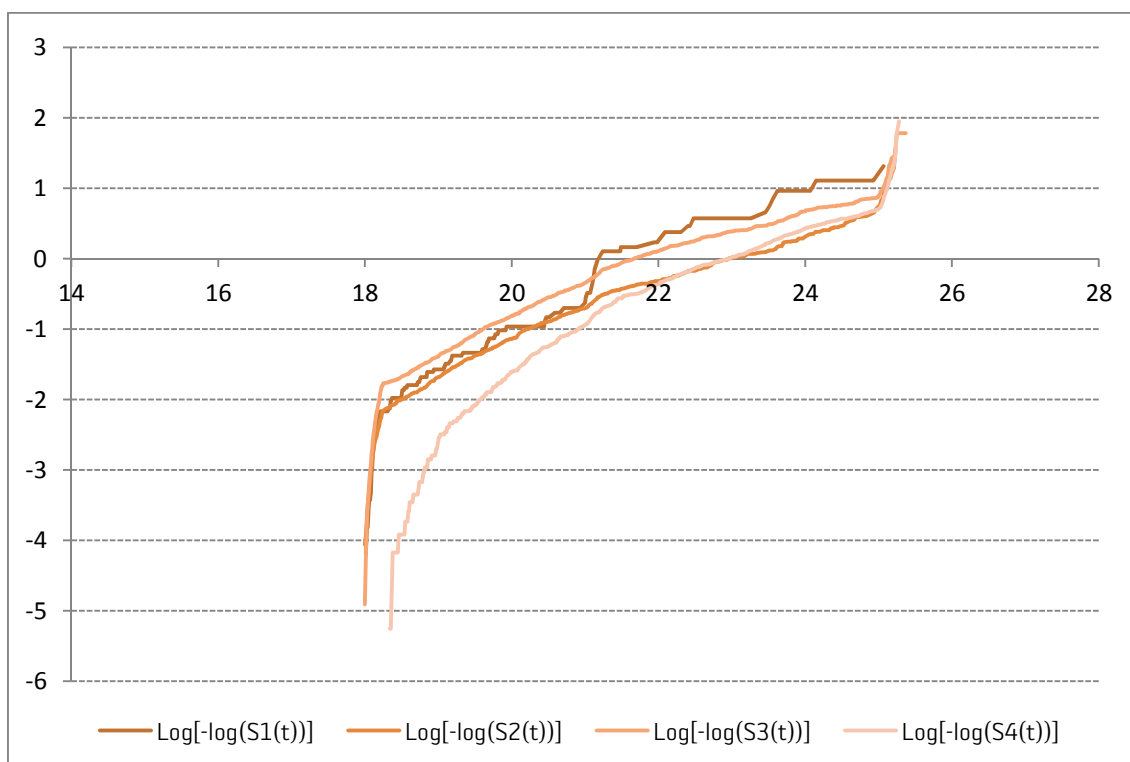
La dernière colonne du tableau nous indique les rapports de risque avec la population de référence. Par exemple pour les enfants appartenant à la classe 2, il y a une augmentation du taux de hasard d'environ 33% par rapport à la tranche de référence.

4.3 Validité du modèle

Le modèle repose sur une hypothèse forte de proportionnalité qui n'est pas toujours vérifiée en pratique. Nous allons vérifier graphiquement cette hypothèse par deux procédés :

1. Analyse du graphique présentant les courbes $\log(-\log(S(t)))$ vs $\log(t)$ de chaque tranche d'âges, avec $S(t)$ estimée par la méthode de Kaplan-Meier. Si les courbes sont parallèles, l'utilisation du modèle de Cox est envisageable.
2. Comparaison des courbes $S(t)$ estimée par Kaplan-Meier et des courbes obtenues grâce au modèle de Cox.

Courbes des $\log(-\log(S(t)))$ vs $\log(t)$:



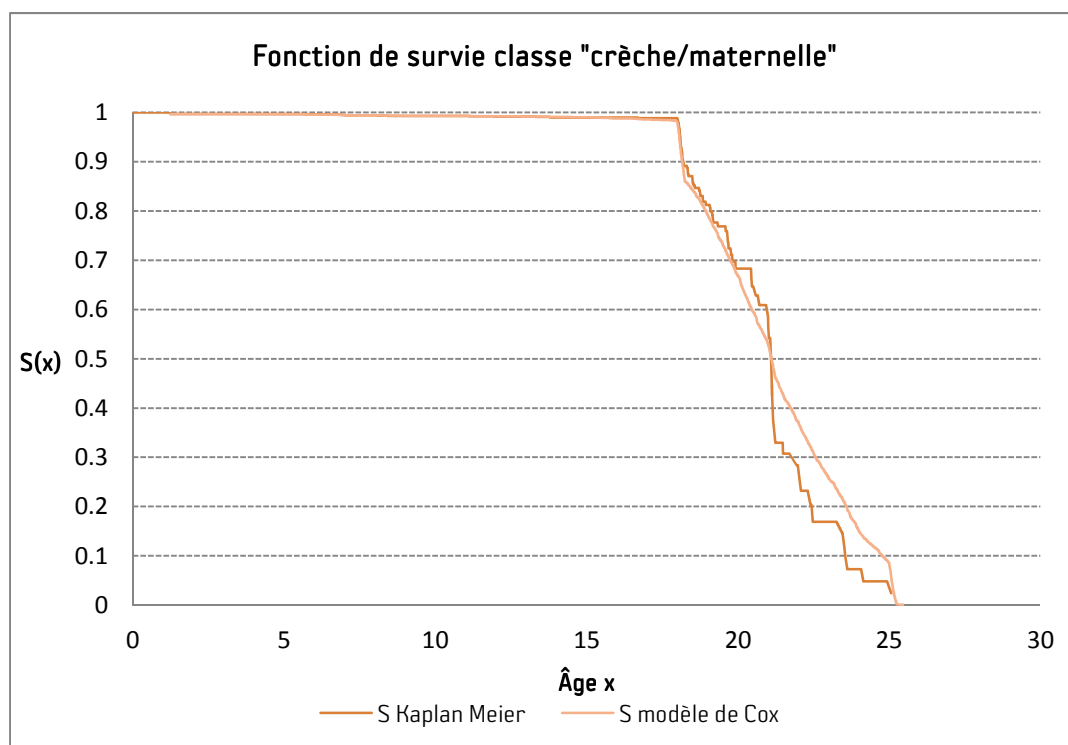
D'après ces courbes, on voit que l'hypothèse de proportionnalité n'est pas acceptable de manière générale car les courbes ne sont pas toutes parallèles entre elles.

La courbe marron et la courbe rose ne s'adaptent pas bien à l'hypothèse car elles ne sont parallèles à aucune autre courbe. Ce sont des classes présentant un petit volume de données, cette raison peut en partie expliquer les difficultés de modélisation.

En revanche on peut valider l'hypothèse graphique de proportionnalité entre les classes 2 et 3 des tranches d'âges] 6,14] et]14,18] car les courbes correspondantes sont parallèles entre elles.

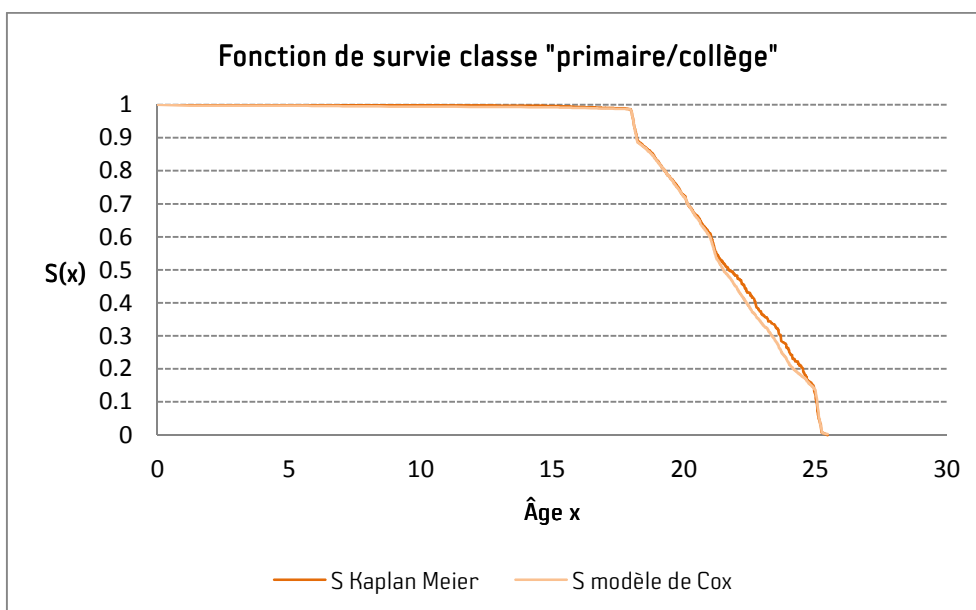
Effectuons à présent la deuxième vérification graphique qui consiste à tracer pour chaque segment de la population, les courbes de survie estimées par Kaplan-Meier et les courbes ajustées avec le modèle de Cox :

Classe « crèche/maternelle » :



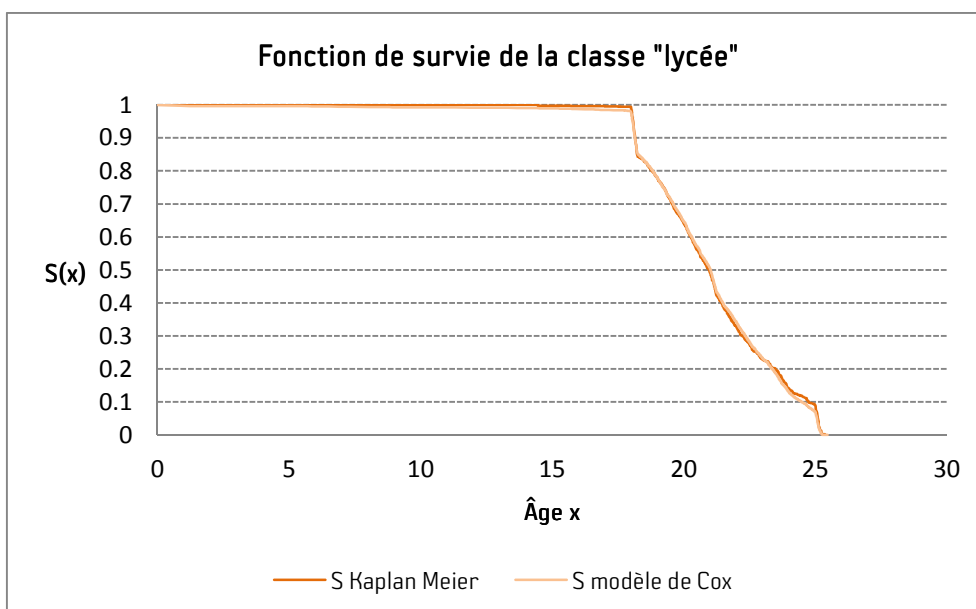
En accord avec le graphique des $\log(-\log S)$, la courbe du modèle de Cox est différente de la courbe brute, on ne peut donc pas valider l'utilisation de ce modèle pour cette classe. Notons tout de même que la courbe du modèle de Cox surestime la survie de cette classe à partir de 20 ans et que dans un contexte de provisionnement cela ajouterait juste un peu plus de prudence vis-à-vis du risque.

Classe « primaire/collège » :



La courbe s'ajuste plutôt bien aux estimations brutes, cela est normal puisque la population de référence de notre modèle est constituée en majeure partie des rentiers de cette tranche d'âge.

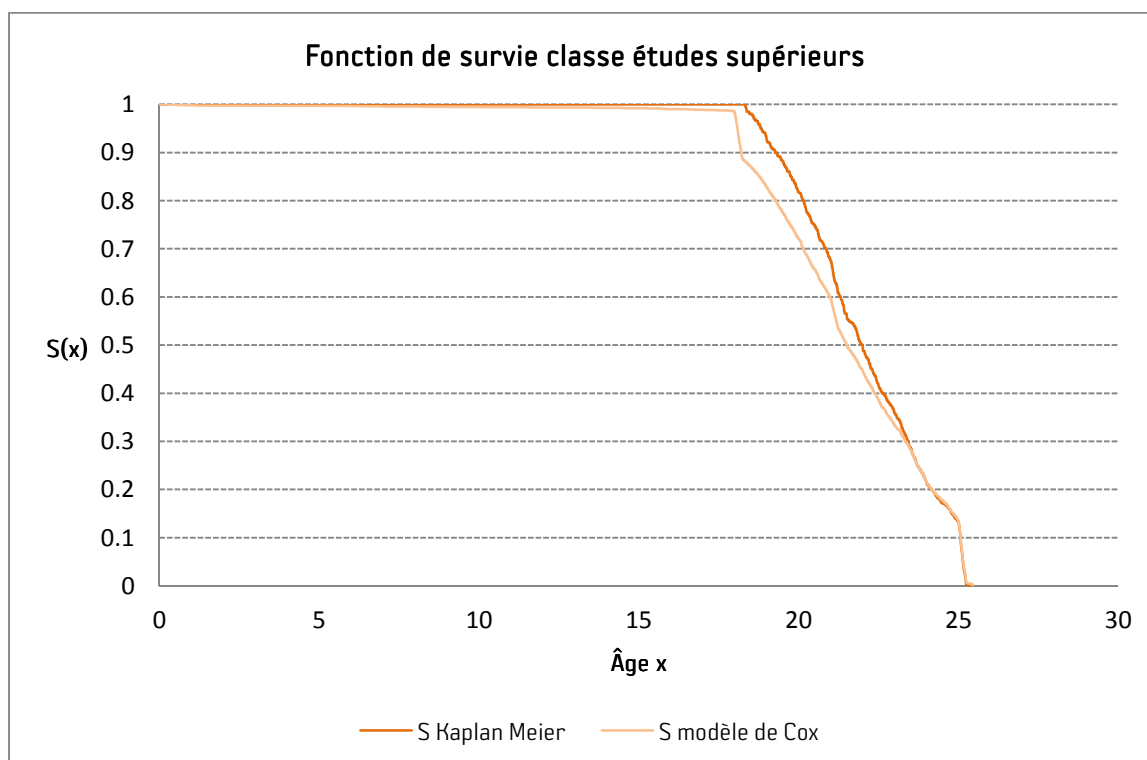
Classe « Lycée » :



Conformément au graphique des LLS, la courbe colle bien aux estimations brutes. Notre modèle est relativement efficace pour prédire le comportement de cette tranche d'âge. Il ne peut cependant pas

prendre en compte les spécificités du contrat, il faudra donc retravailler cet ajustement pour les âges inférieur à 18 ans.

Classe « études supérieurs » :



La courbe ne colle pas bien aux données et sous-estime le maintien réel de cette tranche. Il est donc à bannir dans une logique de provisionnement.

Conclusion de la partie 4 :

Le modèle de Cox n'est pas utilisable en pratique car il ne respecte pas l'hypothèse de proportionnalité sur toutes les classes, il nous a cependant aidé à quantifier l'impact de l'âge de l'enfant au décès pour certaines classes. Le modèle s'adaptant bien aux classes « primaire/collège » et « lycée ». Nous pouvons retenir globalement qu'un enfant appartenant à la classe « lycée » a un risque instantané de sortir de l'état d'enfant à charge d'environ 30 % supérieur à celui d'un enfant appartenant à la classe « primaire/collège ».

5 Validation et utilisation des tables

Nous avons maintenant le choix entre deux méthodes pour le provisionnement du portefeuille :

- Soit utiliser la loi unique représentant le portefeuille dans sa globalité pour calculer les provisions de chaque individu.
- Soit utiliser les lois issues de la segmentation du portefeuille suivant le profil de chaque individu.

Afin de choisir la démarche la mieux adaptée pour répondre à notre problématique, il convient de s'intéresser aux risques inhérents à l'utilisation de chaque loi d'expérience construite.

5.1 Les risques liés aux tables d'expérience

Lors de la construction et de l'utilisation de tables d'expérience, on s'expose naturellement à des risques de natures systématiques. Il faut savoir les identifier et les quantifier afin de pouvoir piloter en toute transparence l'utilisation de ces lois.

Les risques systématiques peuvent provenir principalement de deux sources :

- Si le portefeuille se déforme dans le temps, nous courrons le risque que la table d'expérience ne modélise plus correctement le comportement de la population.
- Plus le niveau de segmentation retenu est fin, plus le risque d'estimation augmente, car les lois sont construites avec des volumes de données de plus en plus fins.

5.1.1 **Risque de déformation**

Dans notre étude nous avons déterminé quatre profils de risque suivant l'âge auquel l'enfant subit le décès d'un parent. La loi unique a été construite avec une certaine répartition des profils de risque propre au portefeuille au moment de l'élaboration.

Si cette répartition venait à évoluer de manière significative, la loi ne serait sûrement plus en adéquation avec la structure du portefeuille.

Pour utiliser la loi unique pour le provisionnement, il faudra surveiller dans le temps l'évolution des effectifs de chaque classe et vérifier que les classes présentant un maintien plus élevé que la loi globale (classe 2 et classe 4) ne prennent pas trop d'importance dans l'effectif total.

Pour réduire le risque de déformation dans le temps, il est conseillé de segmenter le portefeuille en groupe de risques homogènes et de construire une loi d'expérience pour chaque groupe.

Cependant suivant le niveau de segmentation retenu, un nouveau risque apparaît et doit être pris en compte : le risque d'estimation.

5.1.2 Risque d'estimation

Le risque d'estimation résulte de la diminution du volume des données lors de la construction de lois d'expérience. Ainsi plus le niveau de segmentation est fin, plus le risque d'estimation est important et plus les résultats obtenus sont volatiles. Il convient donc de quantifier ce risque pour pouvoir conclure sur la fiabilité et sur l'utilisation des résultats obtenus.

Nous présentons ici une mesure du risque d'estimation consistant à quantifier le risque via les intervalles de confiance des tables brutes.

Intervalle de confiance :

Les taux de sortie bruts ont été obtenus grâce à l'estimateur de Kaplan-Meier de la fonction de survie. La variance de Greenwood donne une estimation de la variance de la fonction de survie, point par point.

$$Var(\hat{S}(x)) = \hat{S}(x)^2 \cdot \gamma(x)^2$$

Où :

- $\hat{S}(x)$ est l'estimateur de Kaplan-Meier de la fonction de survie à l'âge x
- $\gamma(x) = \sqrt{\sum_{i=1}^x \frac{d_i}{n_i \cdot (n_i - d_i)}}$, avec d_i le nombre de sorties et n_i le nombre d'individus exposés au risque de sortir à l'ancienneté x

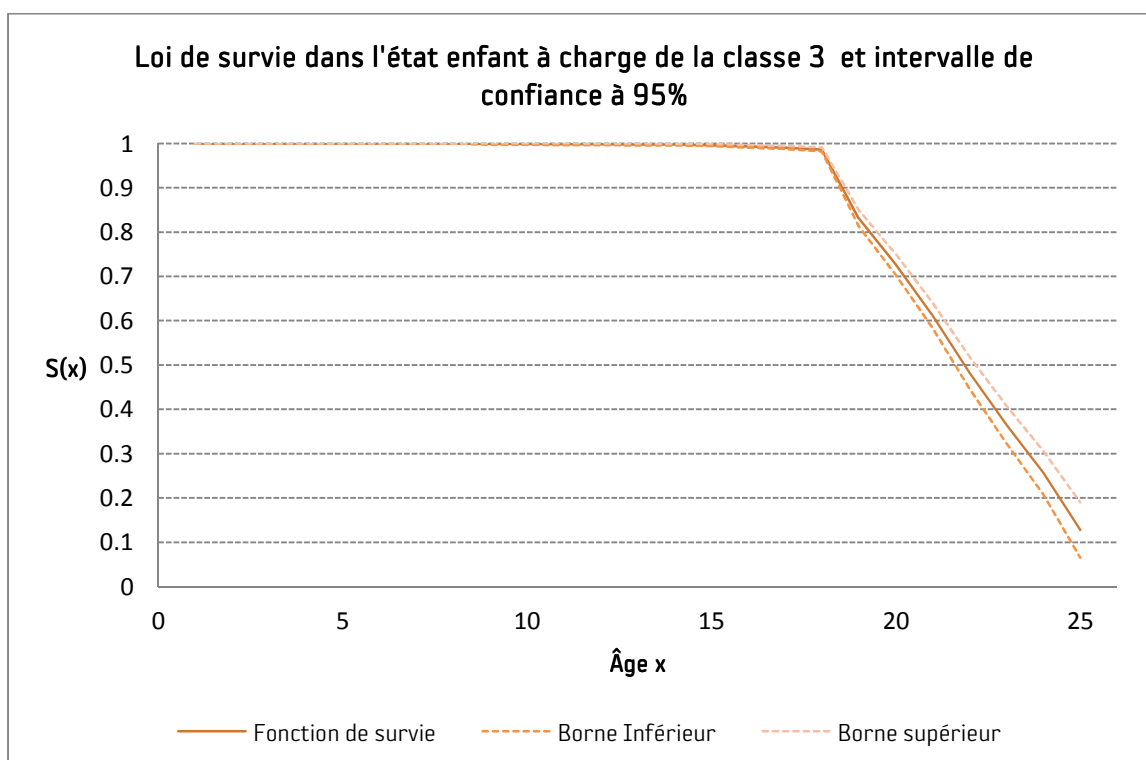
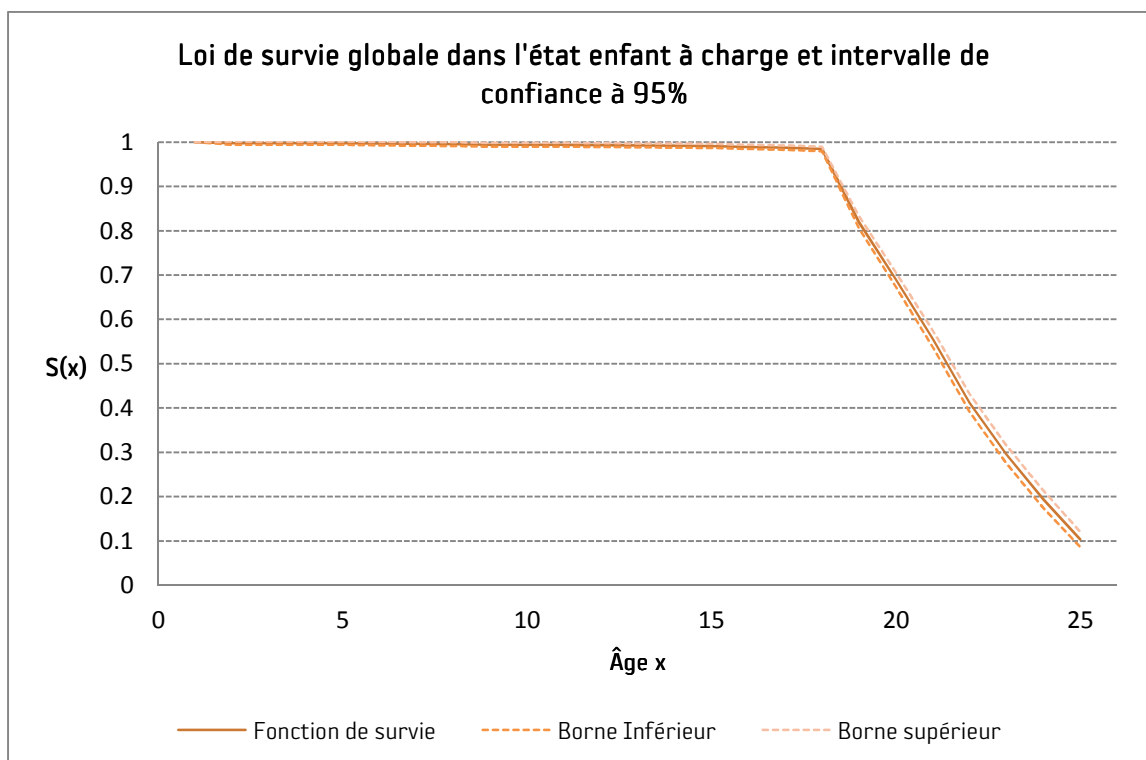
Il est alors possible de déterminer l'intervalle de confiance ponctuel linéaire de la fonction de survie :

$$\left[S(x) - u_{\frac{\alpha}{2}} \cdot S(x) \cdot \gamma(x); S(x) + u_{\frac{\alpha}{2}} \cdot S(x) \cdot \gamma(x) \right]$$

Où $u_{\frac{\alpha}{2}}$ correspond au quantile d'ordre $\frac{\alpha}{2}$ de la loi centrée réduite et $1-\alpha$ désigne le niveau de confiance de l'intervalle.

Le calcul de l'intervalle de confiance à 95% a été réalisé sur la loi représentant le portefeuille global et sur les quatre lois issues de la segmentation.

Les graphiques ci-dessous présentent les fonctions de survie ainsi que leurs intervalles de confiance pour la loi unique et pour la loi représentant les rentiers de la classe 3.



Une simple analyse visuelle permet de constater que l'intervalle de confiance est plus large pour la loi issue du processus de segmentation que pour la loi unique. Les largeurs des intervalles de confiance augmentent également sur les deux graphiques avec l'âge.

Ainsi en toute logique plus l'échantillon qui sert à construire une loi est faible plus l'intervalle de confiance est grand et moins bonne est l'estimation.

D'après nos résultats, seule la loi globale nous paraît être une candidate potentielle pour estimer les probabilités de paiement des flux futurs du portefeuille. Nous allons la soumettre à deux derniers tests de validation basés sur des principes « backtesting ».

5.2 Une méthode de validation : Le backtesting

Les contrôles de type backtesting (ou ex-post) consistent à comparer rétrospectivement les prévisions (ce qui se serait passé si les tables sont pertinentes) avec les réalisations (ce qui s'est réellement passé). Ces contrôles permettent d'évaluer la qualité et la précision d'une table d'expérience et plus généralement d'un modèle.

Une table d'expérience permet non seulement de prédire le nombre de sorties à chaque âge, mais c'est également un des paramètres nécessaire à l'évaluation des engagements de l'assureur puisqu'elle représente la probabilité de payer le flux à chaque instant.

Nous allons donc effectuer deux contrôles ex-post pour valider la table d'expérience construite.

5.2.1 Vérification de l'équilibre « sorties observées / sorties attendues »

Une vérification assez simple de la fidélité de la table construite consiste à appliquer notre table au portefeuille étudié : ceci permet de calculer le nombre de sorties prédites par la table sur la période d'étude. Ce nombre de sorties attendues, noté D_{exp} est ensuite comparé au nombre de sorties effectivement observées, noté D_{obs} . Un ratio $\frac{D_{obs}}{D_{exp}}$ proche de 1 montre une bonne capacité de la table à prédire les sorties réelles.

5.2.2 Vérification de l'équilibre « charge constatée / charge prévisionnelle »

La période d'observation ayant été fixée du 1^{er} janvier 2000 au 31 décembre 2008, nous n'avons pas utilisé les sorties se produisant au cours de l'année 2009 pour notre estimation. Nous allons donc tester la table d'expérience sur des données du portefeuille non utilisées pour la construction et ainsi pouvoir dire si la table reflète bien le portefeuille dans sa globalité.

Ce contrôle consiste à utiliser la table d'expérience pour calculer la provision mathématique au 31 décembre 2008, notée provision à l'ouverture et la provision mathématique au 31 décembre 2009, notée provision à la clôture.

La provision mathématique est l'espérance actualisée des charges futures, ainsi la différence de ces deux provisions peut être interprétée comme la charge prévisionnelle que devra payer l'assureur au cours de l'année 2009. Nous pouvons donc comparer cette différence avec les prestations réellement payées au cours de l'année 2009.

Théoriquement on devrait vérifier l'équation suivante :

$$PM_{Ouv} + IT = PM_{CLO} + Prestation_{2009}$$

Dans la réalité :

- Si $PM_{Ouv} + IT > PM_{CLO} + Prestation_{2009}$, nous sommes en présence d'un sur-provisionnement et il y a dégagement d'un Boni.
- Si $PM_{Ouv} + IT < PM_{CLO} + Prestation_{2009}$, nous sommes en présence d'un sous-provisionnement et il y a dégagement d'un Mali.

5.3 Comparaison des résultats du backtesting sur la loi d'expérience construite et sur la loi utilisée actuellement

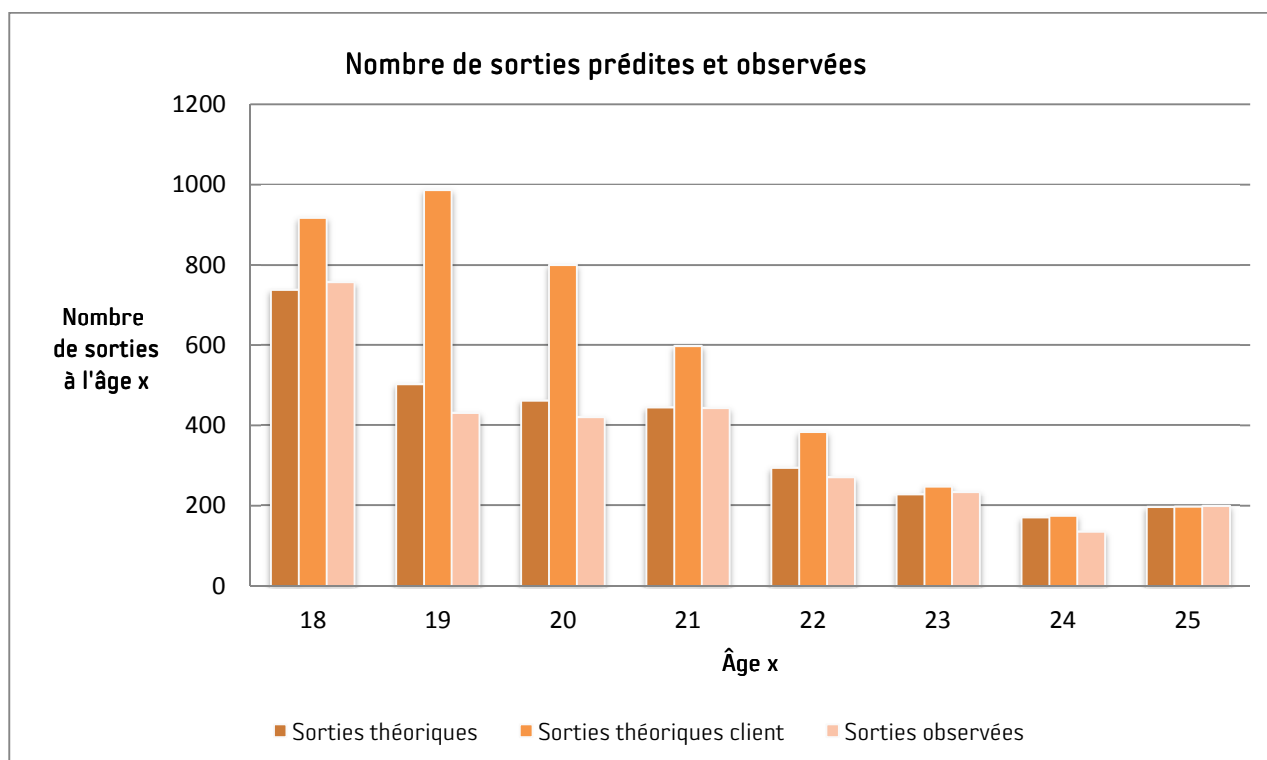
La table d'expérience construite a pour objectif d'être utilisée pour le provisionnement de ce produit. Actuellement le client utilise pour provisionner la table TPG93 abattue d'un certain pourcentage. Nous testons donc les deux méthodes de backtesting exposés plus haut sur la table construite et sur la table actuellement utilisée par notre client.

5.3.1 Vérification de l'équilibre « sorties observées / sorties attendues »

	Loi expérience globale construite	Loi utilisée actuellement
Ratio = $\frac{D_{obs}}{D_{exp}}$	0.95	0.68

Avec un ratio proche de 1, la table d'expérience nous paraît acceptable. La loi utilisée actuellement semble surestimer de façon non négligeable les sorties, cela conduit très certainement à sous-estimer le maintien en rente éducation. Ce résultat est néanmoins logique car la loi d'expérience a été construite directement à l'aide des données alors que la loi abattue non.

Cet histogramme détaille le nombre prédit et observé de sorties pour chaque âge :

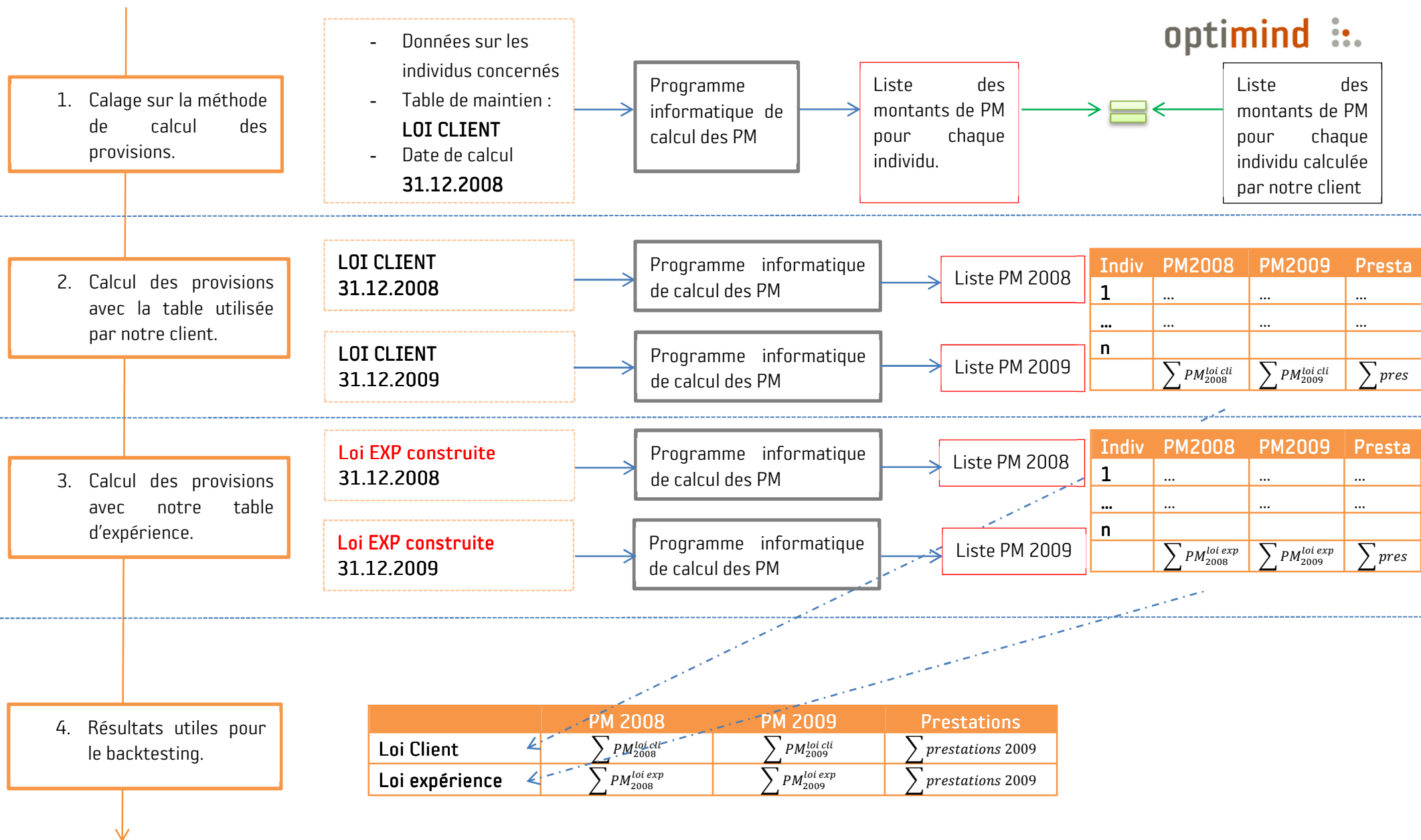


5.3.2 Vérification de l'équilibre « charge constatée / charge prévisionnelle »

La première étape est de déterminer le périmètre d'individus sur lequel nous allons baser notre calcul. Nous souhaitons baser notre calcul sur des individus présents au 31 décembre 2008 et dont nous connaissons le montant de prestations versées au cours de l'année 2009. Nous ne prenons pas en compte les nouveaux arrivants au cours de l'année 2009 car ils n'ont pas de provision au 31 décembre 2008. Nous pourrions ainsi suivre la variation de la provision mathématique pour ces individus au cours de l'année 2009. Les individus sortant au cours de l'année 2009 auront une provision nulle au 31 décembre 2009 et les autres auront un montant de provision clôture inférieur au montant de provision ouverture, car une année se sera écoulée.

Afin de mesurer l'impact des tables sur les montants de provisions, notre client nous a fourni les programmes informatiques nécessaires au calcul des PM. Après s'être calé « informatiquement » afin d'être sur les mêmes bases de calcul, nous effectuons 2 backtesting, le seul paramètre changeant est la table utilisée.

Voici un schéma récapitulatif des opérations effectuées :



Résultats du backtesting :

Chaque montant de ce tableau est exprimé en pourcentage de la PM à l'ouverture.

Table	PM OUV au 31.12.2008	PM CLO au 31.12.2009	IT	Prestations Année 2009	Charge prévisionnelle	Charge constatée	Boni ou Mali
Loi expérience	100.00%	87.76%	1.88%	14.72%	100.00%	100.60%	-0.60%
Loi client	100.00%	87.36%	1.87%	16.35%	100.00%	101.84%	-1.84%

Avec

- IT: Intérêts techniques calculés avec la formule suivante:

$$IT = \left(\frac{PM_{ouv} + PM_{clo}}{2} \right) \times 2\%$$

- Charge prévisionnelle : PM_{ouv}
- Charge constatée : $PM_{CLO} + prestation - IT$
- Boni ou Mali = $PM_{ouv} - PM_{CLO} + prestation - IT$.

Il y a dégagement d'un mali dans les deux situations. Ce test nous permet de dire que les deux lois ont tendance à sous-provisionner le maintien dans l'état enfant à charge. Il prouve également que la loi d'expérience construite avec les données du portefeuille est « meilleure » car c'est elle qui se rapproche le plus de l'équilibre de l'équation.

Conclusion de la partie 5

L'objet de cette dernière partie a été de choisir une loi candidate, puis de lui faire subir des tests sur des principes backtesting afin de la valider définitivement.

Le choix de la loi à tester a été fait en prenant compte différents risques de natures systématiques apparaissant lors de la création ou de l'utilisation des tables. Dans notre cas nous avons choisi d'utiliser la loi unique afin de minimiser le risque d'estimation.

Nous avons ensuite élaboré deux tests consistant à comparer les prévisions avec les réalisations. L'adéquation et la prudence d'une loi d'expérience doit être mesurée sur le phénomène de sortie mais également sur l'impact de celui-ci sur la situation de l'assureur. Ainsi notre premier test compare les sorties prédites avec les sorties observées et le deuxième fait le rapprochement avec des éléments comptables en suivant les boni-mali de liquidation des provisions.

Nous avons testé la loi unique d'expérience et la loi utilisée actuellement par l'organisme et dans les deux cas notre loi s'est avérée être la plus prudente et la plus proche de la réalité.

Conclusion

Les intérêts de ce mémoire ont été de plusieurs natures :

Tout d'abord le partenariat R&D nous a incité à aborder la problématique aussi bien de manière théorique qu'opérationnelle. Notre client souhaitant remettre en question sa propre méthode d'estimation des probabilités de versements futurs pour le provisionnement de son portefeuille de rentes éducation, l'objectif a été de fournir une expertise fiable et concrète.

D'autre part, l'objet du mémoire consistant à construire des tables d'expérience pour modéliser la durée passée dans l'état « enfant à charge », nous avons pu étudier différentes techniques offrant la possibilité d'aborder les choses de façon différente. Nous avons fait le choix d'utiliser des méthodes non paramétriques ou semi-paramétriques n'ayant pas assez de recul pour utiliser les modèles paramétriques.

Enfin la possibilité de construire les tables sur un risque original a ajouté un attrait supplémentaire à cette étude, les recherches annexes sur la scolarisation en France ou sur le statut des orphelins en France se sont révélées utiles et intéressantes.

Dans un premier temps, nous nous sommes assurés de maîtriser les mécanismes du produit, les durées en jeu et les données. Ce travail a impliqué une réflexion critique sur chacun de ces points. La loi dite de référence et le processus mis en place permettant de juger et de fiabiliser les données en est le résultat.

Après une étude sur les risques liés à l'utilisation et à la création de chaque table, nous avons préféré utiliser la loi globale construite avec la méthode de Kaplan-Meier et le lissage de Whittaker-Henderson. Notre portefeuille présentant un fort taux de censure, l'utilisation de cet estimateur est justifiée. Cette loi a également été validée par des tests statistiques classiques, comme celui du Khi-deux mais également par des contrôles « ex-post » qui nous fournissent des éléments de validation supplémentaires.

Globalement, nos résultats montrent que la loi utilisée par notre client se positionne en dessous de la table d'expérience construite dans ce mémoire, notre loi applique ainsi mieux le principe de prudence nécessaire à toute logique de provisionnement.

La segmentation du portefeuille a également été utile puisqu'elle nous a permis de mieux comprendre le risque et de démontrer des différences de survie dans l'état enfant à charge suivant l'âge auquel l'enfant subit le décès d'un parent.

Si le portefeuille ne se déforme pas dans le temps, ces différences ne remettent cependant pas en cause l'utilisation de la loi unique dans notre étude.

Sous conditions de surveiller la composition du portefeuille, provisionner avec notre loi permettra de mieux saisir le comportement réel des rentiers du portefeuille.

Dans notre étude le risque d'estimation étant non négligeable, nous n'avons pas pu utiliser les lois issues de la segmentation mais de façon plus générale, l'utilisation de tables segmentées par profil de risque homogènes permet un provisionnement plus stable et plus fidèle dans le temps.

Cette notion est d'ailleurs largement reprise par la directive Solvabilité II qui définit entre autre les règles de calcul des provisions : ceux-ci doivent être effectués par groupes de risques homogènes. Toute la difficulté est de définir la juste segmentation qui doit être un compromis entre niveau de finesse suffisant et risques d'estimation ou de modèle minimum.

Annexe 1 : Données de marché utilisées pour construire la loi de référence

Population scolaire et universitaire en 2007/2008 par âge :

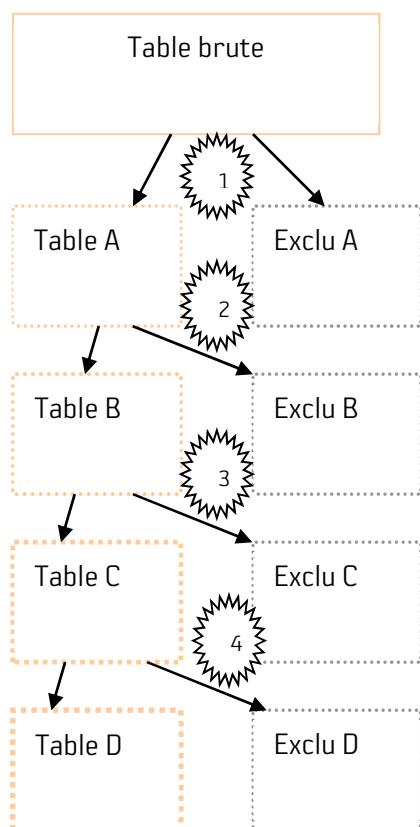
Âge	Population totale	Population scolarisée
3	795873	795873
4	793841	793841
5	796199	796199
6	808730	808730
7	816426	816426
8	791499	791499
9	770803	770803
10	760074	760074
11	771652	771652
12	763787	763787
13	751857	751857
14	749810	749810
15	786330	779897
16	803515	769614
17	812323	741991
18	819883	638389
19	821173	530068
20	817617	415793
21	824056	325810
22	809499	255019
23	802499	181281
24	787262	125697
25	833515	82192

Taux de chômage depuis la sortie de formation initiale, année 2009.	
Sortis depuis 1 à 4 ans de formation initiale	20.2
Sortis depuis 5 à 10 ans de formation initiale	10.5
Sortis depuis 11 ans et plus de formation initiale	7.0

Source : INSEE, enquêtes Emploi

Annexe 2 : Traitement des données

Nous vous avons présenté dans la partie 2.3.2 le processus de traitement des données (voir schéma ci-dessous), les chiffres ont été donnés uniquement pour l'arrêtés n°1, voici un tableau récapitulant les chiffres relatifs au traitement de chaque arrêté.



Etape n°1 :

On sélectionne les enregistrements concernant la branche professionnelle et les types de prestations

Etape n°2:

Fusion des lignes concernant le même rentier pour un même type de prestation en seule ligne.

Etape n°3 :

Fusion des lignes concernant le même individu pour des prestations différentes en une seule ligne

Etape n°4 :

Contrôle de cohérence et suppression des lignes aberrantes.

	Arrêté 1		Arrêté 2		Arrêté 3		Arrêté 4		Arrêté 5	
	Garde	Exclu	Garde	Exclu	Garde	Exclu	Garde	Exclu	Garde	Exclu
Etape 1	71.34%	28.66%	70.01%	29.99%	70.16%	29.84%	70.62%	29.38%	70.46%	29.54%
Etape 2	96.81%	3.19%	97.10%	2.90%	96.91%	3.09%	96.69%	3.31%	96.53%	3.47%
Etape 3	99.46%	0.54%	99.37%	0.63%	99.36%	0.64%	99.33%	0.67%	99.30%	0.70%
Etape 4	94.27%	5.73%	94.77%	5.23%	94.72%	5.28%	94.56%	5.44%	94.49%	5.51%

Annexe 3 : Estimateur de Greenwood

Dans la partie 5.1.2, pour calculer les intervalles de confiance de la fonction de survie estimée par Kaplan-Meier, nous avons besoin d'une estimation de la variance de cette fonction de survie.

Pour cela nous avons utilisé l'estimateur de Greenwood :

L'expression de $\hat{S}(t) = \prod_{i/a_i} \left(1 - \frac{d_i}{n_i}\right)$ permet d'écrire :

$$\ln(\hat{S}(t)) = \sum_{i/a_i} \ln\left(1 - \frac{d_i}{n_i}\right) = \sum_{i/a_i} \ln(1 - \hat{q}_i)$$

On suppose que les variables $\ln(1 - \hat{q}_i)$ sont indépendantes.

La loi de $n_i * \hat{p}_i$ est binomiale de paramètres $(n_i, \hat{p}_i)$.

En employant la delta-méthode qui consiste à considérer que si X est approximativement égal à $\mu + \sigma Z$, où Z est centrée et réduite et σ petit, alors on a :

$$Var(f(X)) = Var(f(\mu + \sigma Z)) = Var(f(\mu) + \sigma Z f'(\mu)) = \sigma^2 \cdot f'(\mu)^2$$

On obtient en prenant $f = \ln$ et $X = \hat{p}_i$ que

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

$$\text{Et que } Var(\ln(\hat{S}(t))) = Var(\sum_{i/a_i} \ln(1 - \hat{q}_i)) = \sum_{i/a_i} Var(\ln(1 - \hat{q}_i)) = \sum_{i/a_i} \frac{\hat{q}_i}{n_i \cdot (1 - \hat{q}_i)}$$

On utilise une nouvelle fois la delta méthode pour dire que :

$$Var(\ln(\hat{S}(t))) = Var(\hat{S}(t)) \cdot \frac{1}{\hat{S}(t)^2}$$

Et en remplaçant $Var(\ln(\hat{S}(t)))$ par son expression calculée précédemment on obtient la formule utilisée dans la partie 5.1.2

$$\widehat{Var}(\hat{S}(t)) = \hat{S}(t)^2 \cdot \sum_{i/a_i} \frac{\hat{q}_i}{n_i \cdot (1 - \hat{q}_i)} = \hat{S}(t)^2 \cdot \sum_{i/a_i} \frac{d_i}{n_i \cdot (n_i - d_i)}$$

Bibliographie

Cours et livres

F.PLANCHET et P.THEROND. *Modèle de durée. Application actuarielle.*

A.DELWARDE et M.DENUIT. *Construction de tables de mortalité périodiques et prospectives.*

O.LOPEZ. *Cours sur les modèles de durée.*

Mémoires d'actuariat

A.GAUMET (ISFA 2001). *Construction de tables d'expérience pour l'entrée et le maintien en incapacité.*

C.MOSSE (ISFA). *Construction d'un indicateur de maintien en arrêt de travail – Apport des simulations à la tarification et à la décision.*

V.LE NEUDER. *Construction de tables d'entrée et de maintien en chômage au Chili.*

M.BEN DBABIS. *Mortalité d'expérience et impacts sur le solde technique de l'assureur.*

T.HAGER. *Etude du maintien en incapacité sur un portefeuille d'assurance emprunteur.*

E.JOUANNIGOT. *Etude comparative de modèles de durée en présence de données hétérogènes et censurées.*

Articles et études

M.CHOUKROUN. *Le modèle additif d'Aalen, une alternative au modèle de Cox dans le cadre de la construction d'une loi de maintien en incapacité de travail.* Bulletin Français d'actuariat, Vol. 8, n°16, juillet-décembre 2008, pp. 107-138.

D. RULLIERE, D. SERANT (1997). *Généralisation de l'estimateur 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, pp. 97-114.

Sites Internet

Ministère de l'éducation nationale : <http://www.education.gouv.fr/pid316/reperes-et-references-statistiques.html>.

Institut nationale de la statistique et des études économiques : http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=natnon03314.

Centre technique des institutions de prévoyance : <http://www.ctip.asso.fr/>

<http://www.ressources-actuarielles.net/>

Lois et réglementation

Loi Evin du 31 décembre 1989 (J.O. 2/01/1990) renforçant les garanties offertes aux personnes assurées contre certains risques.

Commission d'Agrément de l'Institut des Actuaire, *Lignes directrices mortalité*.