

**Mémoire présenté le :**

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

Par : **Hamed Sékou DOUKOURE**

Titre **Mise en place d'un modèle additif généralisé (GAM) bayésien pour les  
swaps de longévité dans le contexte d'une analyse d'expérience**

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

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

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

signature

*Entreprise : PartnerRe*

Nom :

Signature :

*Membres présents du jury de l'ISFA*

*Directeur de mémoire en entreprise :*

*Nom : James ALLISON*

Signature :

Invité :

Nom :

Signature :

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

Signature du responsable entreprise



Signature du candidat



# Résumé

Le risque de longévité pose aux (ré)assureurs un problème difficile dans la mesure où les projections de flux de trésorerie futurs doivent être représentatives du risque sous-jacent à court et à long terme. En tant que tel, le besoin d'estimer correctement les taux de mortalité futurs des populations assurées est très important. Une estimation fondée sur des données permettra une meilleure estimation des réserves et une meilleure évaluation des sinistres de l'assureur ou du réassureur.

Pour ce faire, la plupart des assureurs et des réassureurs utilisent une approche non paramétrique et exogène qui consiste à ajuster et à construire leurs tables d'expérience en fonction d'une table de référence qui, au Royaume-Uni, est souvent basée sur les tables produites par le Continuous Mortality Investigation (CMI). Cette approche a été utilisée pendant longtemps en raison de sa facilité de communication et de mise en œuvre, qui consiste à calculer des ratios d'ajustement A/E qui sont directement appliqués à la table de référence choisie.

Ce mémoire vise à améliorer l'estimation de la mortalité dans les portefeuilles de swaps de longévité en proposant une méthode différente basée sur l'application de la théorie bayésienne à la construction de la table d'expérience de longévité d'un portefeuille de rentiers. Cette application se fera à travers un modèle statistique semi-paramétrique, à savoir les modèles additifs généralisés (GAM) qui utilisent la technique d'estimation bayésienne. Ces régressions font appel aux fonctions de lissage appelées splines qui permettent d'obtenir des mortalités ajustées et lisses dépendant de la covariable évaluée, qui est l'âge dans notre cas.

La première partie consiste en une contextualisation et une explication des termes essentiels utilisés dans le mémoire et sur les swaps de longévité qui sont encore peu connus en dehors de l'industrie. Dans cette partie, nous donnerons aussi un aperçu bref de la technique A/E. Ensuite, le chapitre suivant présente la théorie bayésienne dans toute sa généralité. Cette partie nous permettra de mieux comprendre le concept mathématique et statistique du GAM bayésien en l'occurrence, les P-splines bayésiennes, avant de l'implémenter sur le portefeuille. Enfin, la dernière partie est une comparaison entre les deux méthodes présentées dans les deux chapitres précédents. Cette comparaison sera faite à plusieurs niveaux ce qui nous permettra de comparer et d'opposer l'efficacité et la validité d'une méthode par rapport à l'autre.

---

**Mots clés :** Generalised additive models, Monte-Carlo par chaîne de Markov, B-splines, P-splines, Standardized Mortality Ratio, Ratio Actual/Expected (A/E).

---

# Abstract

Longevity risk poses a difficult problem for (re)insurers in that future cash flow projections must be representative of the underlying risk in both the short and long term. As such, the need to correctly estimate future mortality rates of insured populations is very important. A data-driven estimate will allow for better estimation of reserves and better valuation of the insurer's or reinsurer's claims.

To do this, most insurers and reinsurers use a non-parametric, exogenous approach of adjusting and constructing their experience tables based on a reference table which, in the UK, is often based on the tables produced by the Continuous Mortality Investigation (CMI). This approach has been used for a long time due to its ease of communication and implementation, which consists of calculating A/E adjustment ratios that are directly applied to the chosen reference table.

This thesis aims to improve the estimation of mortality in longevity swap portfolios by proposing a different method based on the application of Bayesian theory to the construction of the longevity experience table of an annuitant portfolio. This application will be done through a semi-parametric statistical model, namely the generalized additive models (GAM) that use the Bayesian estimation technique. These regressions use smoothing functions called splines that allow us to obtain adjusted and smooth mortalities depending on the covariate evaluated, which is age in our case. The first part consists of a contextualization and an explanation of the essential terms used in the thesis and on the longevity swaps which are still little known outside the industry. In this part, we will also give a brief overview of the A/E technique. Then, the next chapter presents the Bayesian theory in all its generality. This part will allow us to better understand the mathematical and statistical concept of Bayesian GAM in this case, Bayesian P-splines, before implementing it on the portfolio. Finally, the last part is a comparison between the two methods presented in the two previous chapters. This comparison will be made at several levels which will allow us to compare and contrast the efficiency and validity of one method with the other.

---

**Key words :** Generalised additive models, Markov chain Monte Carlo, B-splines, P-splines, Standardized Mortality Ratio, Ratio Actual/Expected (A/E).

---

# Remerciements

Je tiens, tout d'abord, à remercier mon maître de stage Anna SAROVA qui, par-delà son statut de tutrice dans la réalisation de cette alternance, fut pour moi un guide dans ma quête de connaissances, tant professionnelles que personnelles.

Je tiens également à remercier le manager du service Reserving Longevity, James ALLISON, pour m'avoir encadré durant cette alternance, pour ses remarques, son suivi et sa disponibilité tout au long de la réalisation de ce mémoire.

Je remercie également Patrick HOGAN pour tout le temps qu'il m'a accordé pour la mise en place et la production des résultats du modèle Bayésien.

Je tiens à adresser mes remerciements à l'ensemble des membres de mon équipe (Lukasz DMITRIEW, Stanislas BOYER, Johnathan DAVIES) pour m'avoir accueilli à bras ouverts, pour leurs bons conseils et leur bonne humeur.

Je voudrais également remercier tous les stagiaires et alternants du service Life & Health-Reserving pour les conseils et la bonne ambiance.

# Table des matières

Résumé . . . . .	I
Abstract . . . . .	II
Remerciements . . . . .	III
Introduction générale . . . . .	1
<b>1 Présentation du cadre d'étude et de l'analyse d'expérience . . . . .</b>	<b>3</b>
1.1 Swaps de longévité . . . . .	4
1.1.1 Contexte de création et définition . . . . .	4
1.1.2 Principe de couverture d'un swap de longévité . . . . .	6
1.1.3 Fonctionnement d'un swap de longévité . . . . .	6
1.2 Description du portefeuille et des données . . . . .	7
1.2.1 Présentation du portefeuille . . . . .	7
1.2.2 Retraitement des données . . . . .	7
1.3 Statistique descriptive des données . . . . .	10
1.3.1 Exposition au risque . . . . .	10
1.3.2 Nombre de décès . . . . .	11
1.3.3 Étude en fonction des années calendaires . . . . .	12
1.4 Validation des données . . . . .	13
1.5 Méthode actuelle de l'équipe . . . . .	15
1.5.1 Estimation des taux de mortalité bruts . . . . .	15
1.5.2 Ajustement et lissage des taux bruts . . . . .	17
1.5.3 Récapitulatif de la méthode . . . . .	18
<b>2 Présentation théorique de l'approche Bayésienne et motivation . . . . .</b>	<b>19</b>
2.1 Modèles additifs généralisés (GAM) . . . . .	20
2.2 Pourquoi l'estimation Bayésienne . . . . .	20
2.3 Présentation théorique de la statistique bayésienne . . . . .	21
2.3.1 Formulation mathématique et aperçu de la loi à postérieure . . . . .	21
2.3.2 Choix des lois à priori . . . . .	22
2.3.3 Implementation et calcul : Monte-carlo par chaîne de Markov . . . . .	23
2.3.4 Estimation des intervalles de confiance dans l'approche bayésienne . . . . .	25
2.4 Critères de validation des modèles Bayésiens . . . . .	26
2.4.1 Généralité . . . . .	26
2.4.2 Leave-one-out . . . . .	27
2.4.3 WAIC . . . . .	28

2.4.4	Le coefficient de détermination : $\mathbf{R}^2$	28
<b>3</b>	<b>Mise en oeuvre du modèle GAM Bayésien sur un traité de swap de longévité</b>	<b>30</b>
3.1	Définition de la mortalité	31
3.2	Spécification de notre modèle	32
3.2.1	Particularité de la modélisation	33
3.2.2	Description des paramètres du modèle	33
3.3	Définition des B-spline et P-spline	33
3.3.1	Les fonctions de lissage : splines	34
3.3.2	Les B-splines	34
3.3.3	Les P-splines	35
3.4	Réécriture du modèle	37
3.5	Introduction du modèle GAM fréquentiste	39
3.6	Retraitement des données	39
3.7	L'estimation bayésienne du modèle	40
3.7.1	Écriture de la vraisemblance des données observées	40
3.7.2	Choix et vérification des lois à priori	41
3.7.3	Lois à posteriori des paramètres	44
3.7.4	Estimation des paramètres du modèle par la méthode MCMC	45
3.8	Analyse de la convergence de la méthode d'échantillonnage	47
3.9	Premières analyses des résultats : Post predictive check	49
3.10	Fermeture de table	50
3.10.1	Traitement des âges jeunes	51
3.10.2	Traitement des âges élevés	52
<b>4</b>	<b>Méthodes d'analyse et de validation des tables du moment</b>	<b>53</b>
4.1	Analyses graphiques des forces de mortalité	54
4.2	Les graphes Gompertz log différence (Gold Chart)	56
4.3	La comparaison des valeurs modélisées et des observations : SMR ou ratio A/E	58
4.3.1	Simulation du nombre décès dans le cadre bayésien	58
4.3.2	Production du nombre décès dans le cadre fréquentiste	58
4.3.3	Analyse A/E par sexe	59
4.3.4	Analyse A/E par intervalle d'âge	60
4.3.5	Analyse A/E par année calendaire	60
4.4	Le coefficient de détermination : $\mathbf{R}^2$	61
	<b>Conclusion et perspectives</b>	<b>63</b>
	<b>Annexes</b>	<b>71</b>
	<b>A Présentation de tous les resultats graphiques</b>	<b>72</b>
	<b>B Code R utiles</b>	<b>85</b>

# Introduction générale

Avec l'augmentation de l'espérance de vie et la volatilité de la mortalité ces derniers temps, le risque de longévité, qui est le risque d'incertitude autour de l'âge du décès d'un individu qui reçoit une pension, est devenu de plus en plus difficile à prévoir. Cela a un impact direct sur les fonds de pension et les assureurs souscrivant des rentes viagères, qui sont souvent contraints ou incités à transférer ce risque aux réassureurs. C'est dans ce contexte que se développent progressivement plusieurs techniques de couverture contre ce risque, notamment les swaps de longévité qui constituent une structure de réassurance permettant à la cédante d'échanger la mortalité observée de son portefeuille contre celle prédite sur la base d'hypothèses définies au préalable.

Créés à l'origine dans les pays anglo-saxons, les swaps de longévité ont connu un essor encore plus important dans de nombreux pays, en grande partie en raison de la mise en place de certaines normes comme Solvabilité II, qui imposent d'immobiliser un montant important de capital pour faire face au risque de longévité. " Détenir un montant important de capital supplémentaire pour garantir leurs obligations en matière de rentes viagères si le risque de longévité ne peut être couvert efficacement " (Courbage, 2014). Avec l'introduction de ces normes prudentielles et comptables, les assureurs et réassureurs sont tenus d'évaluer leurs engagements à la meilleure estimation. Pour s'y conformer, ils doivent prendre en compte tous les comportements observables des assurés sur la durée de vie du traité. Pour tenir compte de l'évolution de la mortalité de leurs portefeuilles, la plupart des réassureurs mettent en œuvre des modèles pour construire leur table d'expérience. Il existe plusieurs méthodes pour établir ces tables en fonction de critères tels que : la crédibilité et la qualité des données, la complexité de la mise en œuvre, la facilité de communication, etc.

La mise en place des tables de mortalité d'expérience à partir des données observées peut se faire soit par la construction directe d'une table de mortalité prospective, soit par une construction d'une table de mortalité du moment qui sera ensuite ajustée par facteurs d'améliorations de la mortalité dans le temps. Cette seconde méthode est appropriée dans le cas où l'assureur ou le réassureur ne dispose pas d'assez de données pour caractériser la tendance de mortalité au sein de son portefeuille, ce qui est le cas lorsque le portefeuille n'a pas assez d'historique. Dans ce cas d'insuffisance de données, il construit sa table de mortalité du moment en se servant à la fois de ses données disponibles et des tables de mortalités externes de référence.

L'une des méthodes externes traditionnelles est l'analyse A/E (A : Actual qui est l'expérience réelle du portefeuille et E : Expected, qui désigne l'expérience attendue à partir de la table de mortalité de référence). Cette méthode a été utilisée pendant très longtemps en raison de sa simplicité de mise en œuvre et de sa facilité de communication.

C'est cette méthode qui est actuellement utilisée par notre équipe pour modéliser les taux de mortalité au sein de notre portefeuille.

Avec l'évolution de l'ancienneté du portefeuille qui conduit à une augmentation de la quantité et de la qualité des données, se pose la question de leur utilisation efficace. C'est dans ce contexte que des modèles statistiques permettant une utilisation efficace des données, tels que le modèle additif généralisé (GAM) avec une estimation bayésienne, ont été envisagés. Les GAM sont une extension des modèles linéaires généralisés (GLM) qui permettent uniquement de modéliser des relations linéaires entre les variables. Ils sont basés sur l'utilisation de fonctions de lissage, appelées "spline", qui permettent de modéliser des relations non linéaires de manière flexible comme la relation entre l'âge et les taux de mortalité. Il existe plusieurs types de fonctions splines dont les P-splines, qui feront l'objet d'étude dans le cadre de ce mémoire. L'utilisation d'une régression faisant appel aux P-splines permet à la fois d'estimer des taux bruts de mortalité tout en les lissant, ce qui réduit les efforts à mettre en place pour construire la table de mortalité du moment. D'autre part, l'estimation bayésienne est un outil essentiel, car elle permet d'intégrer toutes les intuitions et les informations a priori, puis de produire une distribution a posteriori qui est construite en utilisant les données et toutes les informations passées sur la mortalité du portefeuille. La caractéristique fondamentale de l'approche bayésienne est qu'elle permet une quantification directe de l'incertitude dans l'estimation des paramètres de la régression puisque les paramètres sont des lois de probabilité plutôt que des points estimés comme c'est le cas dans les estimations fréquentistes.

Dans ce mémoire, nous analyserons plus en détail la théorie bayésienne et sa combinaison avec les P-splines. La première partie sera consacrée à la mise en place du contexte d'étude et la présentation de la méthode A/E qui est la méthode d'analyse d'expérience actuellement utilisée par nos équipes. La deuxième partie, sera consacrée à l'explication théorique des éléments de la modélisations à savoir la théorie bayésienne et la présentation des P-splines. Ensuite, dans la partie suivante, il s'agira de l'application du GAM bayésien sur notre portefeuille de swap de longévité. Enfin, la dernière partie concerne l'analyse des résultats du GAM bayésien à travers des critères de validation, la comparaison de notre modèle par rapport à la méthode actuelle utilisée par l'équipe notamment sur les qualités d'ajustement, l'efficacité et la spécificité de chaque technique.

# Chapitre 1

## Présentation du cadre d'étude et de l'analyse d'expérience

### 1.1 Swaps de longévité

#### 1.1.1 Contexte de création et définition

Aujourd'hui, le sujet du vieillissement de la population constitue une problématique importante voire urgente pour la gestion du risque de longévité. En effet, la population des plus de 65 ans devrait doubler dans les prochaines décennies (INSEE-2019). Ainsi, de jours en jours, une plus grande proportion de la population atteint l'âge de la retraite et cette dite population survit de plus en plus longtemps du fait de l'augmentation de l'espérance de vie.

Le risque de longévité est, par définition, le risque qu'une population rentière survive plus longtemps qu'escompté. Autrement dit, pour les assureurs, il s'agit du risque de sous-estimer les taux de survie et qu'ainsi les provisions destinées au règlement de rentes viagères soient insuffisantes pour couvrir les promesses de futurs paiements. L'allongement de la durée de vie place le risque de longévité au cœur du débat pour les assureurs commercialisant des rentes viagères. En effet, par la commercialisation de ce type de produit viager, les assureurs s'exposent à de lourds engagements sur le très long terme. (Voir LENIAU 2021)

Le risque de longévité est particulièrement difficile à appréhender. En effet, il se compose de 5 volets principaux qui sont les suivants :

- **Le risque de base** provenant de la possibilité que les populations assurées vivent de manière générale plus longtemps que la population nationale.
- **La volatilité intrinsèque** étant les variations de la mortalité autour de la tendance centrale dues à des fluctuations purement statistiques.
- **Le risque de cohorte** étant le risque qu'une génération puisse subir une plus grande augmentation de l'espérance de vie que la génération précédente (grâce à des avancées médicales par exemple).
- **Le risque catastrophe** qui correspond à des variations de la mortalité autour d'une tendance centrale dues à un événement ponctuel, type pandémie (ex : Covid-19), canicule ou guerre...
- **Le risque de tendance** qui influence très fortement les projections notamment à moyen et long terme. Ce risque correspond au fait que les améliorations de mortalité de population dans les années futures soient, en réalité, supérieures à ce qui avait été anticipé. Ce risque est bien le plus complexe à gérer car il ne peut être atténué par la diversification car, de manière générale, l'amélioration de la mortalité concerne une population dans son ensemble.

Voilà pourquoi le transfert du risque de longévité des assureurs vers les réassureurs est en plein essor depuis le milieu des années 2000 et notamment dans les régions anglophones telles que le Royaume-Uni, le Canada et les Etats-Unis. En effet, le Royaume-Uni a été le précurseur dans le domaine du transfert du risque de longévité, principalement en raison des réglementations du pays, qui impose l'achats de rentes viagères dans le cadre de certains

plans de retraite. Toutefois, cette tendance s'accélère et gagne de nouveaux pays, d'une part, à cause de l'évolution démographique susmentionnée mais également avec la mise en place de Solvabilité II. Effectivement, cette norme prudentielle exige des assureurs "la détention d'une quantité de capital supplémentaire significative permettant de garantir leurs obligations liées au rentes viagères si le risque de longévité ne peut être couvert efficacement ou valorisé au prix du marché" (Courbage, 2014).

Il existe de nombreuses méthodes de transfert du risque de longévité.

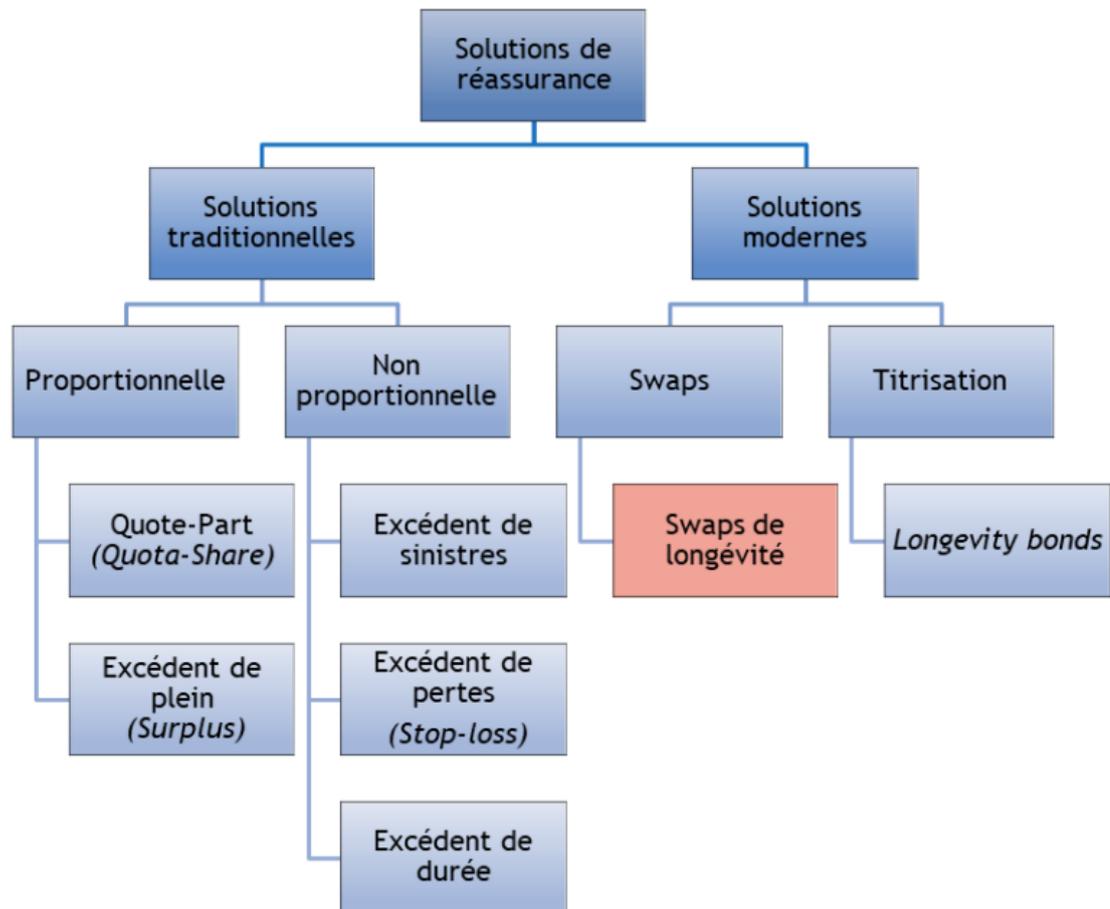


FIG. 1.1 : Organigramme des solutions de transfert du risque de longévité

### 1.1.2 Principe de couverture d'un swap de longévité

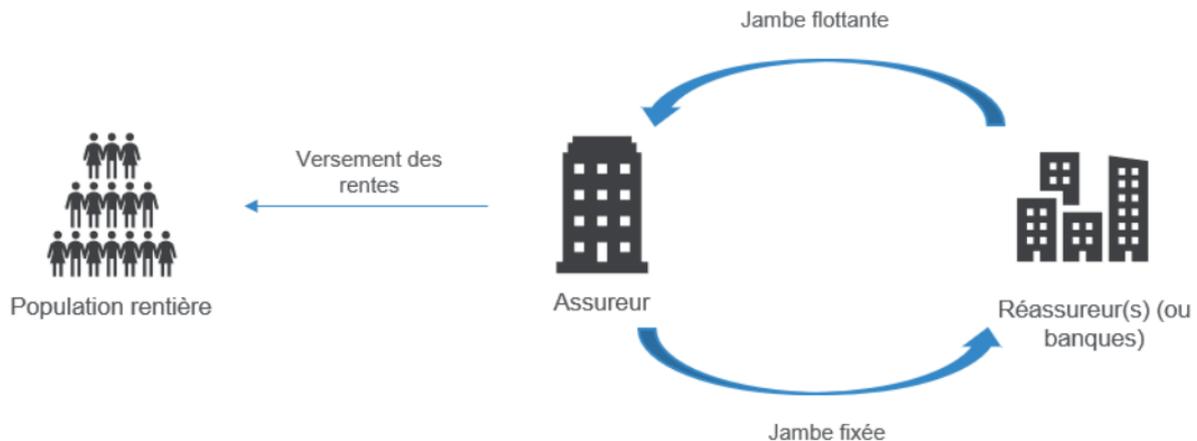


FIG. 1.2 : Fonctionnement d'un swap de longévité

Dans un traité de swap de longévité, en échange de primes appelées "jambe fixe", dont le montant est déterminé à la signature du contrat, le réassureur verse périodiquement une somme variable "jambe flottante" en fonction de la mortalité observée dans la population bénéficiaire. La différence entre la jambe flottante et la jambe fixe est donc due à l'écart de la mortalité par rapport aux prévisions. Par conséquent, les swaps de longévité sont utilisés dans une approche de transfert du risque de longévité pour les engagements de rentes d'assurance-vie ou de fonds de pension.

Le swap de longévité couvre un portefeuille fermé. En d'autres termes, cela signifie que la seule cause de sortie d'un bénéficiaire est le décès, et aussi qu'aucun rentier ne peut être ajouté à la population déjà couverte.

Les traités de longevity swaps peuvent également inclure une clause de réversion qui permet à un(e) veuf(ve) survivant(e) de percevoir un pourcentage de la rente du conjoint(e) défunt(e) selon des conditions prédéterminées. Ces conjoints bénéficiaires sont appelés les "dépendants" ou "Second lives" par opposition aux « membres » également surnommés les "First lives".

### 1.1.3 Fonctionnement d'un swap de longévité

Un recensement de la population survivante est effectué régulièrement. Depuis l'introduction de la recherche électronique, cette opération est effectuée de façon continue. Tous les membres qui n'ont pas été repérés par la recherche électronique font l'objet d'une enquête manuelle. Ainsi, les polices bénéficiaires ou leurs proches sont souvent tenus de fournir un certificat d'existence qui comprend les informations suivantes :

- Statut marital de la police

- Nom, date de naissance et sexe du conjoint actuel
- Confirmation d'existence de la police ou dans le cas échéant sa date de décès.

L'enregistrement des décès dans la population est souvent lent malgré la numérisation des certificats d'existence. Lorsqu'une police ne remplit pas le certificat d'existence pendant une période supérieure à celle définie par la convention, le paiement de sa pension est suspendu jusqu'à ce que la preuve d'existence ou la confirmation du décès soit fournie. Lors de la déclaration de décès d'une police "Life 1", si le contrat prévoit une clause de réversion, le conjoint devient bénéficiaire d'une partie de la pension précédemment versée à l'assuré décédé. Après avoir synthétisé l'ensemble des informations recueillies sur les certificats d'existence reçus, la société transférante établit un bordereau reflétant l'évolution connue de la population bénéficiaire. Ce bordereau est ensuite envoyé au(x) réassureur(s). C'est grâce à ces bordereaux que les réassureurs prennent connaissance de la mortalité observée et déclarée de la population des rentiers.

## 1.2 Description du portefeuille et des données

### 1.2.1 Présentation du portefeuille

L'ensemble de l'étude dans ce mémoire sera effectué sur un traité de swap de longévité dont les caractéristiques sont décrites ci-dessous. Par souci de confidentialité, certaines informations ne seront pas divulguées.

<b>Nom du Traité</b>	Traité A
<b>Date du début du traité</b>	01 septembre 2011
<b>Proportion réassurée</b>	100%
<b>Option de réversion</b>	Oui
<b>Cadence d'envoi des bordereaux</b>	Mensuelle

TAB. 1.1 : Présentation du traité étudié.

Le bordereau dans ce contexte est l'ensemble des informations que le réassureur reçoit de la cédante à propos du statut de chaque contrat du portefeuille (mort d'un annuitant, suspension d'un contrat, réversion au conjoint...).

La période d'investigation retenue pour l'étude de l'analyse d'expérience est celle comprise entre le 31/07/2012 et le 01/08/2019 soit une période de 7 ans plus précisément.

La durée de 7 ans a été choisie par jugement d'expert et les dates du 31 juillet et du 1er Aout ont été choisies stratégiquement dans le but de capturer de façon entière et unique les mortalités dues à la saison de rhume et de grippe au Royaume-Uni.

### 1.2.2 Retraitement des données

Les données d'expérience dont nous disposons sont des données **agrégées** après un calcul individuel de l'exposition au risque d'un contrat.

## Présentation du calcul de l'exposition par contrat

Cette partie présente comment est calculée l'exposition au risque par contrat qui est l'étape précédant l'agrégation finale des données.

En guise d'illustration, nous prenons un contrat du portefeuille qui a les caractéristiques suivants :

- **Sexe** : Masculin
- **Date de naissance** : 01/03/1934
- **Date de sortie du portefeuille** : 04/01/2016
- **Cause de sortie** : Décès

Le calcul de l'exposition pour ce contrat est résumé dans le tableau ci-dessous :

ExpPer_Name	ExpPer_Start_DT	ExpPer_End_DT	Length_Days	Length_Years	Current_Year	Current_Age	Exp_Amt	Exp_AmtSq	InGuarantee	Suspended
sys_StartDate - PolicyYear	31/07/2012	01/09/2012	32	0.087431694	2012	78	169.148852	327242.1359	No	No
PolicyYear - CalendarYear	01/09/2012	01/01/2013	122	0.333333333	2012	78	644.88	1247610.643	No	No
CalendarYear - Birthday	01/01/2013	01/03/2013	59	0.161643836	2013	78	312.722632	605005.7092	No	No
Birthday - PolicyYear	01/03/2013	01/09/2013	184	0.504109589	2013	79	975.270576	1886797.466	No	No
PolicyYear - CalendarYear	01/09/2013	01/01/2014	122	0.334246575	2013	79	646.646796	1251028.755	No	No
CalendarYear - Birthday	01/01/2014	01/03/2014	59	0.161643836	2014	79	312.722632	605005.7092	No	No
Birthday - PolicyYear	01/03/2014	01/09/2014	184	0.504109589	2014	80	975.270576	1886797.466	No	No
PolicyYear - CalendarYear	01/09/2014	01/01/2015	122	0.334246575	2014	80	646.646796	1251028.755	No	No
CalendarYear - Birthday	01/01/2015	01/03/2015	59	0.161643836	2015	80	312.722632	605005.7092	No	No
Birthday - PolicyYear	01/03/2015	01/09/2015	184	0.504109589	2015	81	975.270576	1886797.466	No	No
PolicyYear - CalendarYear	01/09/2015	01/01/2016	122	0.334246575	2015	81	646.646796	1251028.755	No	No
CalendarYear - sys_EndDate	01/01/2016	05/01/2016	4	0.010928962	2016	81	21.143608	40905.26699	No	No

FIG. 1.3 : Calcul de l'exposition par contrat

Chaque ligne du tableau représente une sous-période durant laquelle l'exposition au risque est calculée. Cette exposition est simplement le temps de présence du contrat dans le portefeuille durant la sous-période concernée. Cette sous-période varie à chaque évènement suivant : **Début de la période d'investigation, anniversaire de la police, anniversaire du contrat, date de sortie du contrat du portefeuille, date de fin de l'investigation.**

Pour chaque sous-période, nous avons les informations suivantes :

- **ExpPer\_Name** : Nom de la sous-période
- **ExpPer\_Start\_DT** et **ExpPer\_End\_DT** : Dates de debut et de fin de la sous-période
- **Length\_Days** et **Length\_Years** : Exposition calculée respectivement en jours et en année de la sous-période.
- **Current\_Year** : Année calendaire de la sous-période
- **Current\_Age** : Age de l'annuitant au cours de la sous-période
- **Exp\_Amt** : Exposition calculée en montant payé à l'annuitant
- **InGuarantee** : Prend la valeur Oui si le contrat a une période de garantie, elle prend non sinon.
- **Suspended** : Indique si la police est suspendue ou non.

## Agrégation des données et présentation des données finales

L'étape précédente est effectuée pour tous les contrats présents dans le portefeuille. L'étape suivante consiste à agréger ces données individuelles pour obtenir une table finale d'exposition.

L'agrégation est faite en fonction des facteurs de risque les plus importants en longévité qui sont : **Âge**, **Année calendaire**, **sexe** ainsi que d'autres variables secondaires telles que la durée et bien d'autres.

Ainsi les différentes expositions sont **additionnées** en fonction de ces variables, ce qui nous permet d'obtenir notre base de données dont les premières lignes sont :

CalendarYear	Duration	Gender	Age	InGuarantee	Suspended	ExpDays	ExpYears	AmtExp	AdjClaimCount	AdjClaimAmount	Adj_mx	Expected_L	Expected_A
2012	1	F	76	No	No	2694	7.3606557	4265.264	3	594.12	0.017645	0.129878532	75.26044549
2012	1	M	79	No	No	12089	33.030055	39577.47	4	3681.84	0.034759	1.148103375	1375.687311
2012	1	F	80	No	No	1770	4.8360656	6239.86	3	1144.32	0.029886	0.144531226	186.4851915
2012	1	M	80	No	No	20927	57.177596	124927.55	3	355.97	0.039418	2.253827964	4924.397531
2012	1	M	81	No	No	8997	24.581967	24367.869	3	2756.76	0.04493	1.104479805	1094.860249
2012	1	M	84	No	No	13132	35.879781	147143.02	4	11400.12	0.069302	2.486525206	10197.24229
2012	1	M	84	No	No	7594	20.748634	55019.451	3	9458.4	0.069302	1.437912916	3812.934387
2012	1	M	92	No	No	62	0.1693989	218.57202	3	1397.88	0.188512	0.03193372	41.2034396
2012	1	F	83	No	No	460	1.2568306	1671.2737	3	6401.4	0.046025	0.057845613	76.92035048
2012	1	M	91	No	No	547	1.4945355	3499.5198	3	2323.84	0.168034	0.251132561	588.0377911
2012	1	M	92	No	No	1004	2.7431694	3782.5875	3	4110.48	0.188512	0.517120234	713.0629774
2012	1	M	99	No	No	18	0.0491803	6.200004	3	378.2	0.372041	0.018297097	2.306655566
2012	1	M	87	No	No	3938	10.759563	26775.8	3	2863.32	0.103592	1.114602983	2773.754544
2012	2	M	61	No	No	1781	4.8661202	2344.3806	4	701.52	0.005621	0.027351313	13.17720997
2012	2	F	68	No	No	18176	49.661202	12797.71	3	394.32	0.006967	0.34601335	89.16776706
2012	2	M	70	No	No	8296	22.666667	72668.972	3	650.64	0.01241	0.281282928	901.7885867
2012	2	M	66	No	No	12098	33.054645	60670.388	3	9182.64	0.008144	0.26919125	494.089037
2012	2	F	67	No	No	18490	50.519126	29638.869	4	339.96	0.006199	0.313156257	183.7244256
2012	2	F	69	No	No	15868	43.355191	39422.628	3	1698.24	0.007837	0.339784675	308.9642688
2012	2	M	76	No	No	22512	61.508197	60763.94	3	1281.24	0.02434	1.497095887	1478.980854
2012	2	M	71	No	No	96457	263.54372	1450770.3	10	41427.96	0.013869	3.65501777	20120.34793
2012	2	F	74	No	No	2756	7.5300546	6006.8116	3	175.56	0.013988	0.105331398	84.02407314
2012	2	M	72	No	No	73048	199.5847	1307404.4	4	8140.76	0.015501	3.093762155	20266.07333
2012	2	F	73	No	No	13328	36.415301	24103.335	4	577.08	0.012474	0.454242066	300.6634192
2012	2	F	74	No	No	83993	229.48907	172760.18	4	1331.52	0.013988	3.210123415	2416.592233
2012	2	M	74	No	No	41157	112.45082	58343.834	3	2232.2	0.019377	2.178939564	1130.518115
2012	2	M	73	No	No	47785	130.56011	143425.98	4	976.8	0.017334	2.263081743	2486.094083
2012	2	F	74	No	No	54642	149.29508	117401.91	4	14797.44	0.013988	2.088359311	1642.233428
2012	2	F	74	No	No	21130	57.73224	69929.272	4	598.2	0.013988	0.807566199	978.1798929
2012	2	M	74	No	No	9380	25.628415	123022.66	3	9960.6	0.019377	0.496597252	2383.788166
2012	2	M	76	No	No	17430	47.622951	150626.01	3	11482.8	0.02434	1.159132076	3666.203612

FIG. 1.4 : Base de données finale

En plus d'avoir les variables déjà expliquées dans la section précédente, nous avons de nouvelles variables qui sont indispensables pour la suite de l'analyse. Ce sont :

- **AdjClaimcount** : Cette variable comptabilise le nombre de sorties (généralement par décès) .
- **AdjClaimAmount** : Indique le montant de pensions payées.
- **Adj\_mx** : Cette variable représente le ratio de mortalité  $m_x$  qui provient des tables de référence exogènes.
- **Expected\_L et Expected\_A** : Ces variables sont une multiplication entre le ratio de mortalité  $m_x$  et respectivement l'exposition en année **ExpYears** et l'exposition en terme de montant **AmtExp**.

Notre base de données finale compte 102 224 enregistrements.

## 1.3 Statistique descriptive des données

Dans cette partie, il s'agit d'étudier les facteurs de risques et de nos quantités d'intérêts de façon univariée et bivariée. Elle permettra d'avoir un aperçu de manière synthétique sur les caractéristiques du portefeuille de swap de longévité.

### 1.3.1 Exposition au risque

Le graphique suivant montre la répartition des expositions en fonction des âges et des sexes des rentiers du portefeuille sur la période d'investigation ( 31/07/2012 - 01/08/2019).

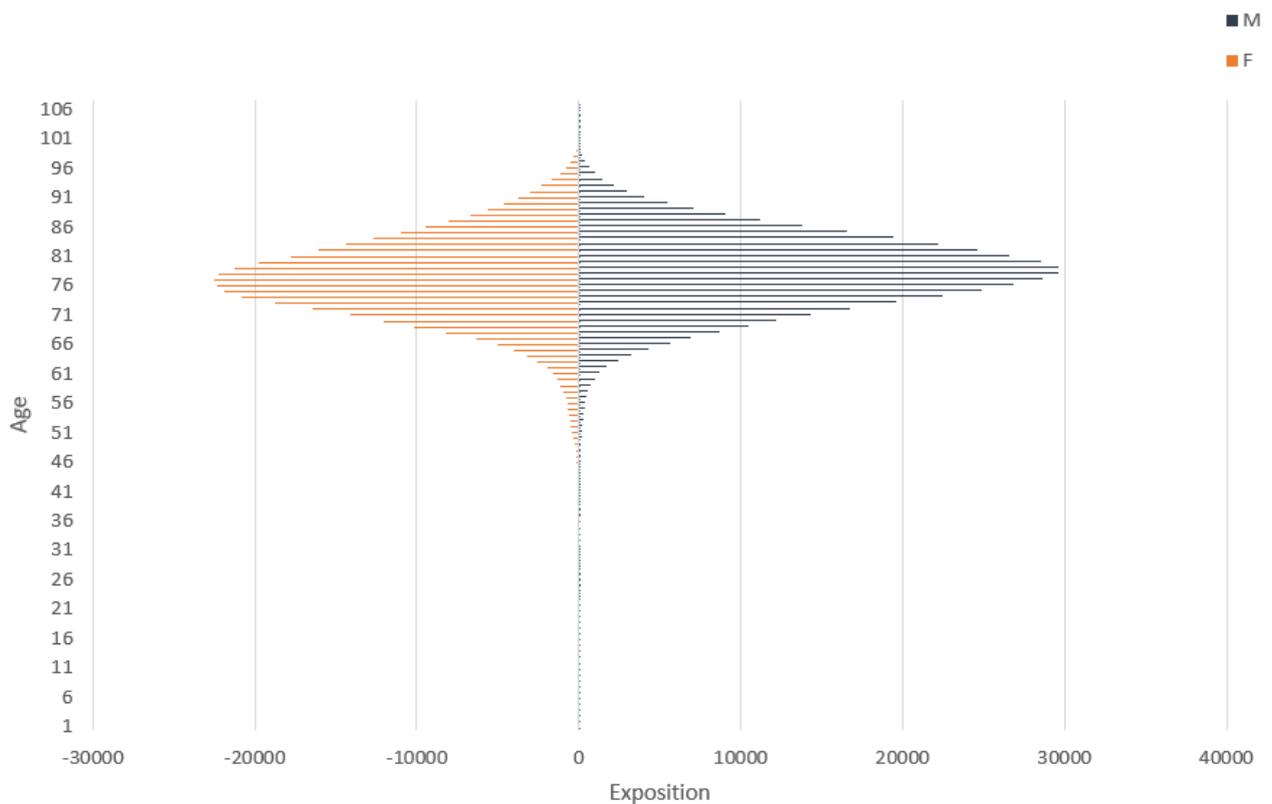


FIG. 1.5 : Exposition au risque par âge et par sexe

Comme on peut le constater, la population du portefeuille est très vieille, l'exposition au risque est la plus importante pour la tranche d'âge 70-90 ans pour les deux sexes. Ce qui est logique au vue de l'âge de départ à la retraite (65 ans). Par conséquent dans la suite de l'analyse, **l'âge de référence (Core Age en anglais)** sera l'intervalle d'âge 65-96 ans. Logiquement, les résultats de notre étude sera beaucoup plus crédible sur cet intervalle.

Par ailleurs, nous avons une plus grande exposition pour les hommes comparée à celle des femmes. Ce tableau ci-dessous résume l'analyse sur l'exposition au risque :

	Population complète		
	Femmes	Hommes	Total
<b>Exposition totale</b>	384544.6916	472260.3698	856805.0614
<b>% Exposition totale</b>	44.88%	55.12%	100.00%
<b>Age Min</b>	19	23	19
<b>Age Max</b>	107	106	107

FIG. 1.6 : Tableau récapitulatif des expositions au risque par sexe

### 1.3.2 Nombre de décès

Comme précédemment, le graphique suivant montre la répartition du nombre de décès en fonction des âges et des sexes des rentiers du portefeuille sur la période d'investigation (31/07/2012 - 01/08/2019).

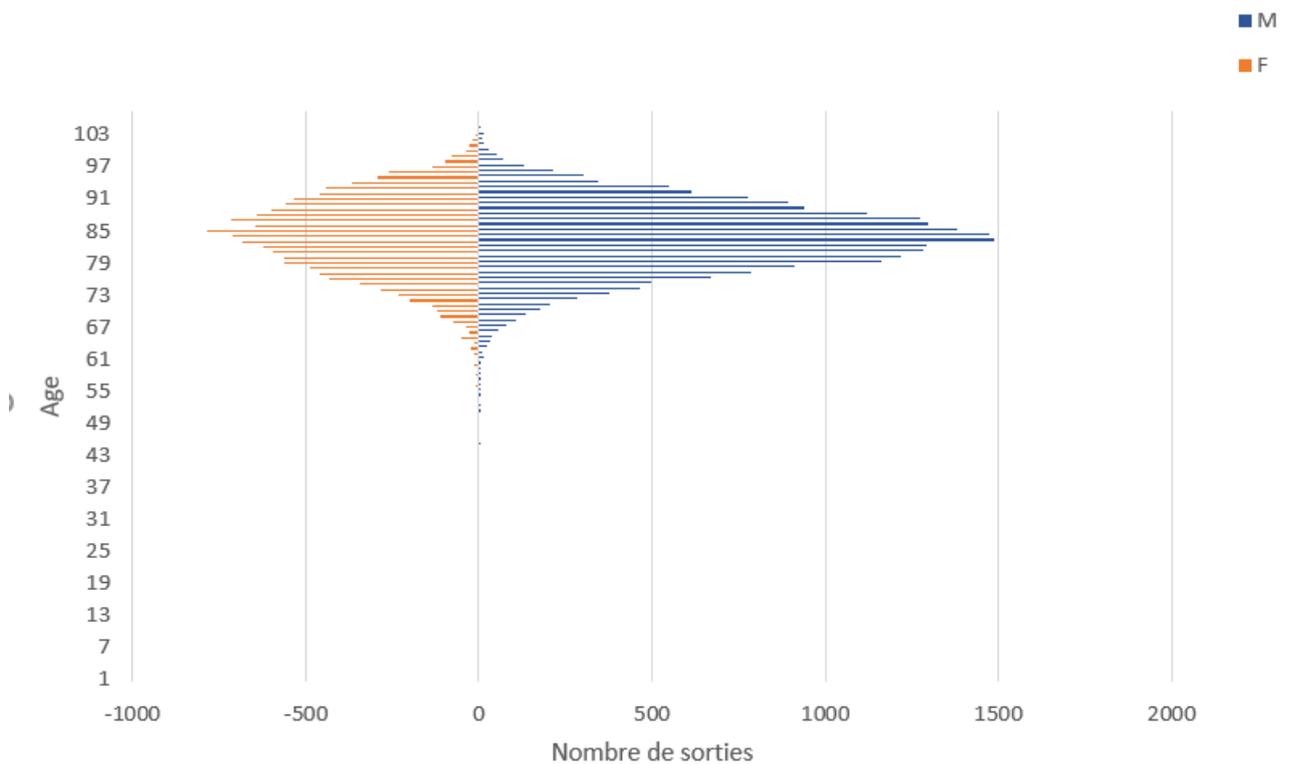


FIG. 1.7 : Nombre de sorties par age et par sexe

Ce graphe vient en complément renforcer notre interprétation sur la crédibilité autour des différents intervalles d'âge de notre portefeuille.

La mortalité dans la population masculine est plus grande que celle de la population féminine même si cette population masculine a une exposition au risque plus grande. Le tableau suivant récapitule l'analyse graphique du nombre de décès.

	Population complète		
	Femmes	Hommes	Total
<b>Nombre de sorties</b>	13563	22832	36395
<b>% Nombre de sorties</b>	37.27%	62.73%	100.00%
<b>Age Min</b>	19	23	19
<b>Age Max</b>	107	106	107

FIG. 1.8 : Tableau récapitulatif du nombre de sorties par sexe

### 1.3.3 Étude en fonction des années calendaires

Le but de cette partie est d'étudier l'impact et la corrélation entre les années calendaires et nos différentes réponses notamment le nombre de décès et des expositions au risque.

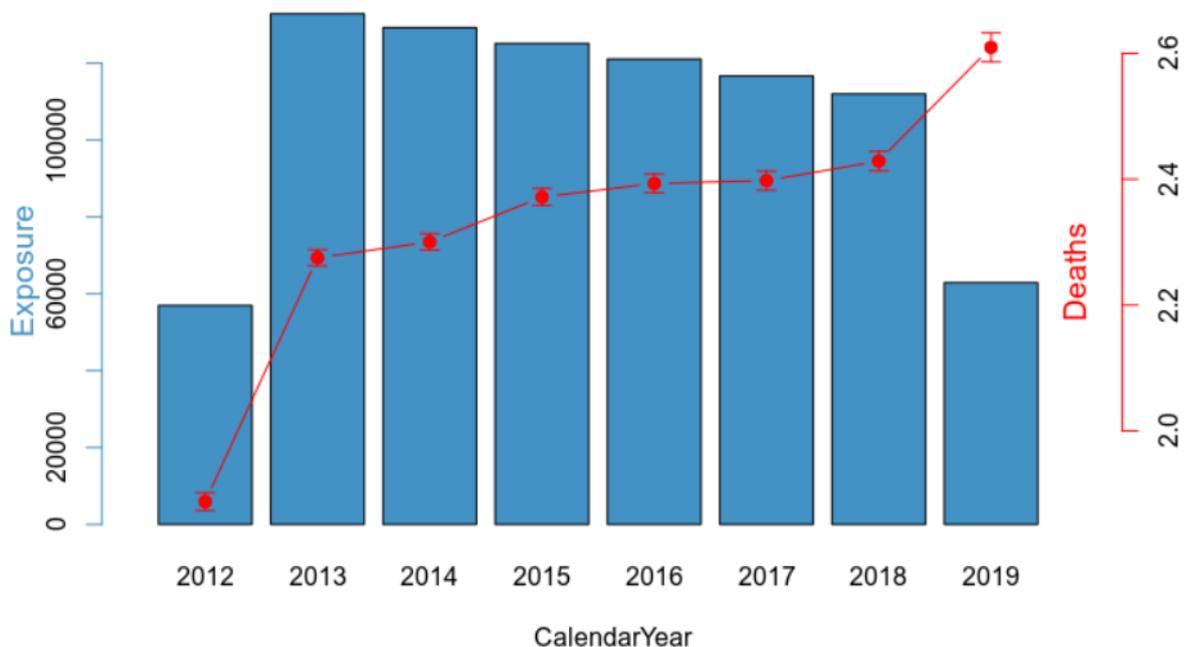


FIG. 1.9 : Exposition et nombre de décès par année calendaire

Sur le graphe ci-dessus, les barres représentent les expositions au risque par année tandis que les points d'estimation en rouge représentent les moyennes des nombres de décès par année et leur intervalle de confiance normale à 95%.

Comme prévu, nous avons beaucoup moins d'expositions au risque pour les années 2012 et 2019 vu que ces années ne sont pas complètes. Entre les années complètes 2013-2018, nous constatons une tendance inverse entre le nombre de décès moyen qui croît et l'exposition au risque qui décroît. Le tableau ci-dessous permet de mieux appréhender ces tendances :

Année calendaire	Exposition au risque	Nombre de décès
<b>2012</b>	56 981.17	1812
<b>2013</b>	132 835.67	4625
<b>2014</b>	129 226.15	4785
<b>2015</b>	125 132.01	5142
<b>2016</b>	121 047.42	5343
<b>2017</b>	116 667.43	5511
<b>2018</b>	112 016.11	5678
<b>2019</b>	62 899.10	3499

FIG. 1.10 : Tableau récapitulatif des expositions et du nombre de décès par année calendaire

## 1.4 Validation des données

Avant de créer une table de mortalité d'expérience de qualité, il est important de vérifier la suffisance des données de mortalité utilisées d'autant plus que dans le cadre bayésien les données en plus des lois à priori sont indispensables pour la détermination des lois à posteriori de la quantité d'intérêt.

Comme précisé dans la section 1.2.1, pour la création de notre table d'expérience du moment, nous disposons d'un historique de données de sur 7 années (31/07/2012 au 01/08/2019), sachant que les observations sur les années 2012 et 2019 ne sont pas entières.

Pour vérifier la suffisance de nos données, nous utilisons le critère de Cochran. Ce dernier indique qu'au minimum 80% des classes d'âge doivent avoir au moins cinq sinistres ou décès et toutes les classes d'âge au moins un sinistre :

$$\begin{cases} N_x \times \hat{q}_x \geq 5; \\ N_x \times (1 - \hat{q}_x) \geq 5; \\ d_x > 0 \end{cases} \quad (1.1)$$

où

$d_x$  représente le nombre de décès à l'âge  $x$

$N_x$  est la population d'âge  $x$  exposée au risque ;

$q_x = \frac{d_x}{N_x}$  représente l'estimateur brut de la mortalité à l'âge  $x$ .

Bien que l'estimation des mortalités dans le cadre de la construction des tables d'expérience du moment ou transversales ne prenne pas en compte les années calendaires, nous décidons, tout de même, d'appliquer le test par année calendaire. Cette façon de faire nous permettra d'une part d'avoir un aperçu des années qui influencent le plus les estimations et celles qui y contribuent le moins.

## Chapitre 1. Présentation du cadre d'étude et de l'analyse d'expérience

Les expositions à notre disposition sont des expositions centrales au risque alors que le critère de Cochran fait appel aux expositions initiales au risque. En effet, les données étant agrégées, nous n'avons pas la possibilité d'obtenir le nombre de personnes dans le portefeuille exposées au risque.

Les termes expositions centrales et initiales sont expliquées dans la section suivante (3.1). Ainsi nous adaptons le critère et nous nous basons plus sur la dernière condition du critère à savoir la stricte positivité du nombre de décès par âge.

Le tableau suivant nous présente le nombre de décès et les expositions centrales au risque du portefeuille par année calendaire et par âge (jusqu'à 67 ans).

Age	2012		2013		2014		2015		2016		2017		2018		2019	
	Dx	Ex														
19	0,4208	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0,4208	0	1,2466	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0,7534	0	1,2466	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0,7534	0	1,2466	0	0	0	0	0	0	0	0	0
23	0,4208	0	0	0	0	0	0,7534	0	1,2486	0	0	0	0	0	0	0
24	0,8415	0	1,8274	0	0	0	0	0	0,7514	0	1,2466	0	0	0	0	0
25	0	0	1,1726	0	1,8274	0	0	0	0	0	0,7534	0	1,2466	0	0	0
26	0	0	0	0	1,1726	0	1,8274	0	0	0	0	0	0,7534	0	0,8274	0
27	0	0	0	0	0	0	1,1726	0	1,8306	0	0	0	0	0	0,3342	0
28	0	0	0	0	0	0	0	0	1,1694	0	1,8274	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	1,1726	0	1,8274	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	1,1726	0	1,4082	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,3342	0
35	0,8415	0	0,3233	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0,1694	0	1,6767	0	0,3233	0	0	0	0	0	0	0	0	0	0	0
37	0,6803	0	0,7479	0	1,6767	0	0,3233	0	0	0	0	0	0	0	0	0
38	1,2541	0	2,9945	0	0,7479	0	1,6767	0	0,3279	0	0	0	0	0	0	0
39	0,6776	0	1,5041	0	2,9945	0	0,7479	0	1,6721	0	0,3233	0	0	0	0	0
40	2,6995	0	3,4055	0	1,5041	0	2,9945	0	0,7486	0	1,6767	0	0,3233	0	0	0
41	4,2077	0	7,0795	0	3,4055	0	1,5041	0	2,9973	0	0,7479	0	1,6767	0	0,3233	0
42	4,623	0	10,638	0	7,5918	0	4,1534	0	1,5027	0	2,9945	0	0,7479	0	0,8384	0
43	6,8333	0	12,666	0	11,504	0	7,7452	0	4,1585	0	1,5041	0	2,9945	0	0,5808	0
44	17,178	0	21,644	0	13,307	0	11,918	0	7,7514	0	4,1534	0	1,5041	0	2,3233	0
45	18,342	1	43,501	1	21,644	0	14,153	0	12,246	0	7,7452	0	4,1534	0	0,3726	0
46	17,992	0	42,499	0	43,049	0	21,907	0	15,568	0	12,247	0	7,7452	0	3,5671	0
47	28,413	0	43,068	0	42,499	0	43,049	0	22,077	0	15,559	0	12,247	0	4,8082	0
48	41,262	0	83,416	0	43,068	0	42,499	0	43,06	0	22,063	0	15,559	0	7,4685	0
49	45,298	0	102,36	0	84,934	0	43,068	0	42,497	0	41,586	3	22,063	0	9,7096	0
50	51,563	0	111,83	0	103,12	0	84,981	0	43,123	0	41,241	0	41,307	0	14,386	2
51	60,243	0	121,39	1	112,05	0	104,02	0	86,612	0	43,23	0	41,241	0	24,375	0
52	67,667	0	152,16	1	120,22	0	113,93	0	105,06	0	87,329	1	42,901	0	24,732	0
53	57,967	0	157,83	2	151,55	1	119,59	2	113,92	0	105,03	0	87,773	0	24,962	0
54	79,052	0	147,24	1	155,81	1	149,83	0	118,5	5	113,83	0	105,66	0	59,197	0
55	84,016	0	202,21	0	147,51	0	155,32	1	148,85	4	114,28	1	113,91	0	62,937	0
56	78,372	0	179,13	0	200,75	5	148	3	155,79	0	147,82	3	113,52	1	67,134	0
57	121,83	0	219,96	0	180,43	0	202,73	0	147,47	7	155,47	1	145,99	0	64,132	1
58	175	1	322	2	220,44	2	182,69	0	200,8	1	144,27	1	154,21	4	89,501	1
59	212,27	0	439,45	4	323,09	1	219,58	2	183,19	1	200,86	1	144,54	1	90,159	0
60	281,07	3	568,38	4	440,21	3	323,02	5	218,56	1	184,46	0	198,29	3	87,03	1
61	345,58	6	680,45	5	568,34	2	439,17	4	322,54	2	220,17	1	183,4	0	122,62	1
62	466,62	1	894,24	13	679,96	0	569,57	4	438,35	1	320,55	4	221,46	0	98,564	4
63	642,61	2	1251,6	12	885,65	8	675,6	11	570,61	0	437,98	5	316,83	8	140,36	0
64	803,39	7	1602,3	10	1246,7	7	879,67	7	675,67	6	571,08	9	433,85	1	194,86	0
65	1081,6	3	2066,2	23	1596,5	15	1240,1	21	875,53	6	675,82	12	571,02	3	265,23	1
66	1161,9	9	2702,1	25	2055	16	1585,9	19	1237,3	4	875,12	5	676,53	2	361,43	4
67	1516,5	16	2854,7	30	2699,5	24	2047,8	15	1585,7	6	1238,9	14	877,2	9	394,61	4

FIG. 1.11 : Tableau récapitulatif des expositions et du nombre de décès par âge

Comme nous pouvons le voir, le nombre de décès est positif sur l'ensemble des années calendaires à partir de 65 ans. Nous avons donc une insuffisance de données sur la tranche d'âge [19, 64], ce qui est normal sachant que notre portefeuille est composé, en majorité, de pensionnaires.

Nous avons le même problème d'insuffisance de données sur les âges supérieurs à 95 ans. Cette dernière remarque relève plus d'un jugement d'expert de notre part que sur le critère de Cochran.

Ainsi les tranches d'âge [19, 64] et [96, 120] seront hors de notre scope de modélisation avec le GAM bayésien et seront modélisées différemment en se basant sur des données

externes que nous développerons dans la section "Fermeture de table".

A ce stade, nous vérifions que la suffisance des données à travers ce critère, plus tard, après avoir calibré le modèle, d'autres critères seront mis en place pour vérifier les qualités d'ajustement dans le cadre bayésien et la crédibilité des estimations.

### 1.5 Méthode actuelle de l'équipe

L'analyse d'expérience dans l'équipe pour la construction des tables de mortalité prospective se fait en deux étapes.

La première concerne la mise en place d'une table de mortalité transversale ou du moment appelé *base mortality table* en anglais.

La construction d'une table du moment est l'étude statique de la mortalité qui permet d'obtenir des taux de mortalité par âge sur une période de courte durée et figée (3 à 7 ans en moyenne) sans prendre en compte les améliorations générationnelles. Cette analyse permet uniquement de capter la mortalité niveau du portefeuille. L'analyse d'expérience de cette table est faite chaque année à partir des données du portefeuille et d'une table de référence externe.

La seconde étape permet de donner l'aspect prospectif à la table statique à travers la mise en place des taux de mortalité d'amélioration (*Improvement tables*).

La mise en place d'une table de mortalité prospectif requiert un grand historique de données ce qui n'est pas le cas pour notre portefeuille dont l'année de création est 2011. De ce fait, la table d'amélioration des mortalités utilisée par l'équipe est entièrement externe et sa révision se fait chaque 3 ans.

Cette façon de splitter les différentes tables a plusieurs avantages. En effet, la mise en place de la table du moment permet d'avoir un premier aperçu sur les taux de mortalité du portefeuille sachant que l'estimation taux de mortalités fait appel, en partie, aux données du portefeuille.

Dans ce mémoire, l'analyse d'expérience traitée est la première notamment la mise en place d'un table de mortalité transversale.

Cette section présente donc la méthode utilisée par l'équipe dans le cadre de l'analyse d'expérience du moment.

#### 1.5.1 Estimation des taux de mortalité bruts

La méthode utilisée pour le calcul des taux de mortalité de la table transversale est l'approche par rapprochement. Cette méthode consiste à appliquer de façon multiplicative un pourcentage aux taux de mortalité d'une table de référence. Dans notre cas la table de référence est une table du Continuous Mortality Investigation (CMI).

Ce pourcentage est connu sous le nom de *Standardized Mortality Ratio* ou SMR ou encore ratio "Actual-to-Expected" (A/E). Actual est l'expérience observée du portefeuille de l'entreprise conductrice de l'analyse et E (expected) se décline comme le nombre de décès attendus calculé à partir des taux de mortalité provenant de la table de mortalité de référence ou Benchmark. Le ratio A/E appliqué à la table de référence permet d'obtenir

des taux de mortalité qui reflètent "plus ou moins" l'expérience du portefeuille. L'étude A/E est une méthode efficace et simple à implementer vu que les résultats sont des ratios et des métriques de crédibilité calculées autour de ces pourcentages. De plus, elle ne nécessite pas une grande quantité de données.

Dans notre cas, ce ratio est calculé par tranches d'âge (intervalle de 5 ans) et par sexe. Ce choix permet d'arbitrer entre la crédibilité et la diversification des ratios sur l'ensemble des âges du portefeuille. En effet, la crédibilité est réduite lorsque les ratios sont calculés par âge unitaire du fait du faible nombre de décès ou d'exposition pour certains âges.

Dans le procédé, l'on distingue deux façons de calculer les ratios, qui conduisent à des résultats plus ou moins différents.

L'on peut conduire une analyse pondérée par le montant de la rente ou par la survie de l'annuitant. (Voir MACDONALD et al. 2018).

- **Ratio basé sur la survie de l'annuitant**

Dans le cas où la pondération est portée sur la vie ou la survie de l'annuitant, le ratio agrégé A/E **pour une tranche d'âge  $I$** , pour une année calendaire donnée, est exprimée sous la forme :

$$A/E_I^{survies} = \frac{\sum_{x \in I} D_x}{\sum_{x \in I} (E_x^c * m_x^{ref})} \quad (1.2)$$

$D_x$  représente le nombre total de décès observés dans le portefeuille parmi les personnes d'âge  $x$ ,

$E_x^c$  est l'exposition centrale de l'individu sur une année exprimée en années.

Le terme  $m_x^{ref}$  représente le taux brut de mortalité pour les individus d'âge  $x$ , pour l'année calendaire choisie, provenant de la table de référence.

- **Ratio pondéré par le montant de l'annuité**

L'autre approche est de tenir compte du montant de l'annuité du rentier. En introduisant le montant de la rente, nous introduisons donc un proxy socio-économique selon l'hypothèse : montant élevé de la rente = classe socio-économique plus élevée = espérance de vie plus élevée.

Dans ce cas, le ratio A/E peut s'écrire comme suit :

$$A/E_I^{montant} = \frac{\sum_{x \in I} M_x}{\sum_{x \in I} (E_x^c * \mathbf{A}_x * m_x^{ref})} \quad (1.3)$$

Où  $M_x$  représente le coût des sinistres réellement payés au sein de la population d'âge  $x$ ,  $\mathbf{A}_x$  représente la valeur contractuelle de l'annuité de l'individu.

On peut remarquer que  $E_x^c * \mathbf{A}_x = E_x^{montant}$  représente l'exposition en montant de la population d'âge  $x$  qui est une multiplication entre l'exposition centrale et le montant de l'annuité contractuelle.

Dans l'équipe, c'est la seconde approche qui est utilisée, qui permet de prendre en compte le montant de l'annuité lors de l'estimation des taux bruts de mortalité.

En calculant ces ratios pour la maille d'agrégation choisie, leurs intervalles de confiance à 95% sont mis en place dans le but de jauger la crédibilité autour de ses ratios.

Ces intervalles de confiance sont calculés à l'aide d'une méthodologie définie à la section 3.2.1 de l'analyse de l'expérience de mortalité du régime de retraite autogéré (SAPS) du CMI pour la période 2005-2012. La formule s'énonce comme suit :

$$I_{A/E} = \left[ A/E^{montant} - z_{\frac{\alpha}{2}} * \sigma^{montant}; A/E^{montant} + z_{\frac{\alpha}{2}} * \sigma^{montant} \right] \quad (1.4)$$

où

$z_{\frac{\alpha}{2}}$  est le quantile d'ordre  $1 - \frac{\alpha}{2}$  de la loi centrée réduite qui est égale à 1.96 en prenant  $\alpha = 5\%$ . et  $\sigma^{montant}$  est tel que :

$$\sigma^{montant} = \sqrt{A/E^{montant} \times \frac{Expected_{SQ}}{Expected^2}}$$

avec :

$$Expected = \sum_{x \in I} (E_x^c * A_x * m_x^{ref}) \text{ qui est le dénominateur de l'équation 1.3,}$$

et

$$Expected_{SQ} = \sum_{x \in I} (E_x^c * A_x^2 * m_x^{ref})$$

Après avoir calculé les différents ratios par tranche d'âge et par sexe, le taux brut de mortalité (s'il doit être mis à jour) est calculé comme suit :  $m_x = A/E_I^{montant} \times m_x^{ref}$  si  $x \in I$ .

### 1.5.2 Ajustement et lissage des taux bruts

Après avoir calculé les différents ratios, un ensemble de jugements d'expert est considéré qui décident de si oui ou non les taux de mortalité doivent être mis à jour à partir des ratios. Si oui, comment ils doivent être appliqués de sorte à obtenir des taux de mortalité lissés et ajustés.

Concernant l'ajustement, les taux de mortalités sont automatiquement ajustés si la valeur du ratio A/E par sexe est en dehors de l'intervalle  $[100 - 3\%; 100 + 3\%]$ .

L'ajustement est réalisé sur les deux taux de mortalité des sexes, si au moins l'un des ratios ne respecte pas le critère de  $-/+ 3\%$ .

Pour ce qui est du lissage des taux, il est fait à partir des ratios calculés par tranches d'âge et est entièrement basé sur un jugement d'expert dont l'objectif est d'atteindre une table lissée et acceptable selon la réglementation sur la construction des tables Best Estimate.

### 1.5.3 Récapitulatif de la méthode

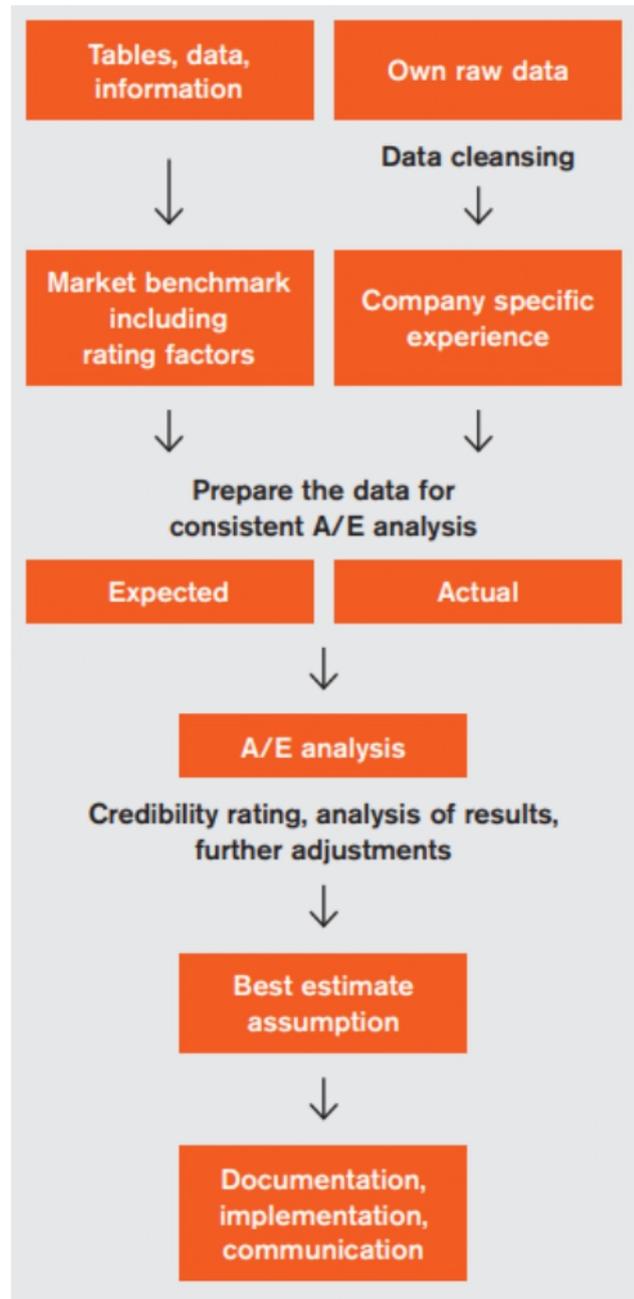


FIG. 1.12 : Présentation du procédé Actual-to-expected

## Chapitre 2

# Présentation théorique de l'approche Bayésienne et motivation

## 2.1 Modèles additifs généralisés (GAM)

Un modèle additif généralisé, initialement développé par Trevor Hastie et Robert Tibshirani, est un modèle linéaire généralisé avec un prédicteur linéaire utilisant une somme de fonctions de lissage de covariables WAKEFIELD 2012.

Mathématiquement, le modèle peut être écrit comme suit :

$$g(\mu_i) = \mathbf{X}_i^* \boldsymbol{\theta} + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots \quad (2.1)$$

où

$$\mu_i = E(Y_i) \text{ and } Y_i \sim \text{une loi de la famille exponentielle}$$

$Y_i$  est la variable réponse,  $\mathbf{X}_i^*$  est un vecteur contenant les covariables,  $\boldsymbol{\theta}$  est le vecteur des paramètres correspondants aux covariables et  $f_i$  sont les fonctions de lissage des variables  $x_j$ .

Un GAM est plus flexible qu'un GLM, en effet la relation entre les prédicteurs et la réponse n'est pas restrictive à une fonction monomiale. Nous avons donc plus de liberté pour modéliser l'impact d'une variable sur la réponse à travers les fonctions de lissage alors que le GLM suppose un impact linéaire des variables sur la réponse, même si ce n'est pas ce qui est observé DENUIT et al. 2019.

Comme nous l'avons mentionné, nous aimerions trouver un modèle qui utilise efficacement et minutieusement les données car nous en avons en quantité suffisante. Dans ce cas, les GAM sont considérés comme un modèle plus orienté données car nous pouvons les utiliser pour définir complètement la forme des lisseurs, alors que les GLM sont plus orientés modèle (définis par les fonctions monomiales additives).

## 2.2 Pourquoi l'estimation Bayésienne

La théorie bayésienne est un concept ancien qui n'était pas très utilisé dans le passé en raison de la limite des calculs informatiques. Aujourd'hui, avec l'évolution de la technologie, cette approche commence à être répandue et son efficacité continue à être prouvée. Ayant déjà choisi le type de notre modèle (GAM au lieu de GLM), nous pouvons maintenant évaluer la meilleure façon de l'ajuster. Dans l'usage courant, nous avons l'approche fréquentiste et l'approche bayésienne.

La comparaison de ces deux approches fait depuis longtemps l'objet des discussions. Même si les deux approches font des inférences à partir des données, dans l'approche fréquentiste, les paramètres et les hypothèses sont considérés comme des quantités inconnues mais fixes (non aléatoires), et il n'est donc pas possible de faire des déclarations de probabilité sur ces inconnues. Francisco J. SAMANIEGO 2010b. Dans cette approche, nous choisissons d'abord un modèle parmi un candidat de modèles pertinents et les paramètres de ce modèle sont fixés et estimés par la méthode du maximum de vraisemblance ou par la méthode des moindres carrés.

Ainsi, dans ce cas, nous impliquons que le modèle et l'estimation des paramètres sont fixés, corrects et que le comportement des procédures inférentielles est évalué dans le cadre d'un

échantillonnage hypothétique répété des données WAKEFIELD 2012. En revanche, les méthodes bayésiennes fournissent une distribution complète pour chaque paramètre et des probabilités pour différents choix de modèles, permettant ainsi une modélisation explicite de l'incertitude du modèle et des paramètres Voir ULLER 2019.

Et comme en actuariat en général et dans l'équipe en particulier, nous faisons très souvent faire appel au jugement d'experts pour ajuster ou valider les résultats des modèles, le modèle bayésien permet d'incorporer pleinement ce processus de supposition et de croyance en fixant une distribution aux lois à priori. Le rôle de ces lois à priori dans l'analyse bayésienne est de capturer des informations "pré-données" sur les paramètres, puis d'utiliser l'expérience antérieure qui a été recueillie pour mettre à jour les informations "pré-données" sur les paramètres en informations "postérieures" sur les paramètres.

### 2.3 Présentation théorique de la statistique bayésienne

Afin de bien comprendre l'approche bayésienne, cette partie va décrire les différents concepts de la statistique bayésienne. Elle a pour but d'expliquer tous les termes théoriques qui seront utilisés dans la partie modélisation.

Comme indiqué ci-dessus, tous les paramètres sont considérés comme des variables aléatoires. Ces distributions sont déterminées en fixant d'abord une distribution à priori qui sera ajustée par les données pour obtenir finalement la distribution postérieure.

#### 2.3.1 Formulation mathématique et aperçu de la loi à posteriori

Soit  $\theta = [\theta_1, \theta_2, \dots, \theta_p] \in \Theta \subseteq \mathbb{R}^p$  l'ensemble des informations inconnues du modèle, essentiellement les paramètres et  $\mathbf{x} = [x_1, x_2, \dots, x_n]$  l'ensemble de nos données observées. En faisant intervenir le théorème de Bayes, la distribution de la loi à posteriori est donnée comme suit :

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta) * \pi(\theta)}{\int_{\Theta} f(\theta|\mathbf{x})\pi(\theta)d\theta} \quad (2.2)$$

Où  $\pi(\theta)$  représente l'ensemble des informations connues à priori sur les lois de probabilités des paramètres  $\theta_p$ .

$f(\mathbf{x}|\theta)$  représente la densité de la loi conjointe des données observées sachant  $\theta$  fixé autrement dit la vraisemblance des données observées.

Sachant que le dénominateur  $\int_{\Theta} f(\theta|\mathbf{x})\pi(\theta)d\theta$  sert de constante de normalisation, nous pouvons écrire de façon simplifiée :

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta)$$

Ce qui implique la distribution à posteriori, conditionnellement aux données d'observation, est proportionnelle au produit de la vraisemblance et de la loi à priori.

En général, le nombre de paramètres à estimer  $p$  est strictement supérieur à 1. Admettons que ce soit le cas ici, la densité  $\pi(\theta|\mathbf{x})$  est celle d'une loi de probabilité jointe de dimension  $p$ .

Par conséquent, les densités marginales se calculent en intégrant la densité jointe, ainsi pour un paramètre  $k \leq p$  on aura :

$$\pi(\theta_k|\mathbf{x}) = \int_{\Theta \setminus \theta_k(\Omega)} \pi(\boldsymbol{\theta}|\mathbf{x}) d\theta_{-k} \quad (2.3)$$

Où  $\theta_{-k} = [\theta_1, \theta_2, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_p]$  et  $\theta_k(\Omega)$  représente le support de la variable aléatoire  $\theta_k$ .

### 2.3.2 Choix des lois à priori

La spécification des distributions à priori des paramètres inconnus est une étape très importante voir indispensable dans la construction. Ces informations à priori permettent de tenir compte de l'expérience passée et d'incorporer les opinions d'expert directement dans la construction du modèle ce qui constitue une force dans l'approche bayésienne. Une première remarque est de constater que quelques soit les données en notre disposition  $y$ , si  $\pi(\theta) = 0$  alors  $\pi(\theta|y) = 0$ . Ce point souligne l'importance qu'occupe le choix de la loi à priori des paramètres lors de la construction du modèle. Cependant, il peut être difficile de réduire les informations à priori en une loi de probabilité. Une approche générale est d'assigner des probabilités sur l'ensemble des valeurs possibles du paramètre inconnu. Si on suppose que notre paramètre suit une loi normale, les moyenne et écart-type seront déterminés en fonction des nos opinions après s'être posés une série de questions.

#### Loi non-informatives

Il existe des cas où l'on ne dispose pas d'informations sur le paramètre ou parfois l'on veut mener une étude plutôt objective en voulant que l'analyse soit, en très grande partie, menée par les données. Dans ce cas, l'on peut spécifier une loi à priori qui apporte peu d'information au modèle pour ne pas perturber l'objectivité du modèle. Une intuition est d'attribuer une distribution tel que  $\pi(\boldsymbol{\theta}) \propto 1$ . Cependant, la construction de ce genre de loi à priori est parfois difficile surtout si l'ensemble des valeurs possibles du paramètre n'est pas fini.

En effet, le cas décrit juste haut appliqué sur un intervalle non borné conduit à des loi à priori impropres (i.e  $\int_{\Theta} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \infty$ ). Si de plus l'intégrale de la fonction vraisemblance est impropre également, alors la loi à posteriori sera impropre. Les lois à posteriori impropres conduisent très souvent à des résultats aberrants.

Plusieurs méthodes ont été développées dans le but de mettre en place des lois à priori sans conduire à des lois à posteriori impropres notamment la loi à priori de Jeffreys (1961) défini comme :

$$\pi(\boldsymbol{\theta}) \propto [I(\boldsymbol{\theta})]^{-\frac{1}{2}},$$

où  $I(\boldsymbol{\theta})$  est l'information de Fisher de la vraisemblance.

#### Lois à priori conjuguées

Dans l'approche bayésienne le terme de lois à priori conjuguées ou conjugate prior définit l'ensemble des loi de probabilité telles que les lois à priori et à posteriori ( $p(\boldsymbol{\theta}|\mathbf{y})$ )

et  $p(\boldsymbol{\theta})$ ) appartiennent à la même famille de probabilité.

Pour illustration, nous prenons le cas de la famille exponentielle.

**Définition :** Un modèle statistique  $(E, \mathcal{E}, \mathcal{P})$  sur un espace des observations  $E$  est dit de famille exponentielle générale, s'il existe des fonctions  $\eta, T, C$  et  $h$  telles que les vraisemblances, pour tout  $\boldsymbol{\theta} \in \Theta$ , s'écrivent :

$$f(x_i|\boldsymbol{\theta}) = C(\boldsymbol{\theta})h(x_i)\exp(\eta(\boldsymbol{\theta})^T T(x_i))$$

avec les conditions que :

- $T$  soit une fonction mesurable à valeurs dans  $\mathbb{R}^p$  ;
- $\eta$  soit une fonction à valeurs dans  $\mathbb{R}^p$  ;
- $C$  soit une fonction réelle positive qui ne dépend pas de  $x$  ;
- $h$  soit une fonction borélienne positive qui ne dépend pas de  $\boldsymbol{\theta}$ .

Cette famille est utile en statistique et probabilité de façon générale parce qu'elle englobe presque toutes les lois usuelles de probabilité : loi de poisson, loi normale, loi gamma... Pour  $n$  observations indépendantes et identiquement distribuées ayant la forme de densité plus haut.

$$f(\mathbf{x}|\boldsymbol{\theta}) = \left[ \prod_{i=1}^n h(x_i) \right] C(\boldsymbol{\theta})^n \exp [\eta(\boldsymbol{\theta})^T t(\mathbf{x})]$$

avec :

$$t(\mathbf{x}) = \sum_{i=1}^n T(x_i)$$

Ces genres de loi à priori sont très utiles dans l'analyse, en effet, elles permettent de facilement calculer certaines intégrales qui sont en général difficile à calculer analytiquement vu que la forme de loi à posteriori est déjà connue.

### 2.3.3 Implementation et calcul : Monte-carlo par chaîne de Markov

#### Généralité et problème d'implementation

Le modèle bayésien a depuis longtemps été ignoré à cause de sa difficulté d'implementation qui nécessite de puissante machine et de grandes ressources.

En rappelant de façon détaillée, la distribution de la loi a posteriori de l'équation 2.2 on a :

$$\pi((\theta_1, \theta_2, \dots, \theta_p)|\mathbf{x}) = \frac{f(\mathbf{x}|(\theta_1, \theta_2, \dots, \theta_p)) * \pi((\theta_1, \theta_2, \dots, \theta_p))}{\int \int \dots \int_{\Theta} f(\mathbf{x}|(\theta_1, \theta_2, \dots, \theta_p)) * \pi((\theta_1, \theta_2, \dots, \theta_p)) d\theta_1, d\theta_2, \dots, d\theta_p}$$

Sachant que la loi à posteriori est proportionnelle à la vraisemblance des données et de la loi à priori, le calcul exact de cette loi nécessite le calcul explicite de la constante de normalisation qui se pose ici comme l'intégrale de dimension  $p$ .

Une mauvaise évaluation de cette grandeur fausse le calcul de la loi à posteriori et donc de toutes les inférences qui en découlent. En plus de la loi conjointe à posteriori, nous nous intéresserons aussi aux lois marginales de chaque paramètre  $\theta_i$ . Dans la plupart des cas, les calculs analytiques et numériques de cette constante ne sont pas évidents et nécessitent de grands moyens.

Plusieurs méthodes ont déjà été mises en place pour ce faire, cependant dans cette partie nous décrivons seulement la méthode que nous utiliserons pour le calibrage de nos différents modèles.

### MCMC : Monte-carlo par chaîne de Markov

Le but fondamental de cette technique est de produire des réalisations ou échantillons de la loi à posteriori  $\pi(\boldsymbol{\theta}|\mathbf{x})$  avec  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p] \in \Theta \subseteq \mathbb{R}^p$ , peu importe la difficulté de calcul de cette dernière.

Par souci de simplicité, dans la suite de cette partie, la densité  $\pi(\boldsymbol{\theta}|\mathbf{x})$  de la loi à posteriori, se notera  $f(x)$  avec  $x = [x_1, x_2, \dots, x_p]$

On pose la quantité :

$$u_h = E_f[h(x)] = \int_{\Theta} h(x)f(x)dx$$

où  $h$  est une fonction déterministe connue.

Soit  $X^{(0)}, X^{(1)}, \dots$  une séquence de vecteurs aléatoires générés à partir d'une chaîne de Markov à pas de temps discret homogène avec un espace d'états  $\Theta$  avec comme noyau de transition  $P(y|x)$ , ce qui implique :

$$\begin{aligned} P(y|x) &= P(X^{(t)} \leq y | X^{(t-1)} = x, \dots, X^{(0)} = x^{(0)}) \\ &= P(X^{(t)} \leq y | X^{(t-1)} = x) \end{aligned}$$

et

$$P(X^{(t)} \leq y | X^{(t-1)} = x) = P(X^{(1)} \leq y | X^{(0)} = x)$$

avec :  $x, y \in \Theta$  et  $t \in \mathbb{N}$ .

Le choix de la chaîne est fait de sorte qu'elle soit irréductible. Ce qui implique qu'elle admet une unique mesure de probabilité invariante dont on note  $f$  la densité. Après un temps et un nombre adéquat d'itérations de la chaîne, soit  $m$  fois par exemple, la séquence de vecteurs aléatoires  $X^{(m+1)}, X^{(m+2)}, \dots$  admet une loi qui a approximativement pour densité  $f$ . Sachant cela nous pouvons alors calculer la quantité  $u_h$  par :

$$\hat{u}_h = E_f[h(x)] = \frac{1}{n} \sum_{t=m+1}^{m+n} h(X^{(t)})$$

Bien évidemment les  $X^{(m+1)}, X^{(m+2)}, \dots$  ne sont pas indépendants, ici  $n$  est assez grand pour minimiser la variance de  $\hat{u}_h$ .

Tout étant théoriquement défini, le principe de cette méthode est de trouver cette chaîne de Markov respectant toutes ses hypothèses (irréductibilité, homogénéité...) et qui admet comme unique mesure de probabilité invariante la loi de probabilité de notre loi à posteriori.

### L'algorithme de Metropolis–Hastings

Cet algorithme répond à cette préoccupation à savoir de créer une chaîne de Markov telle que définie dans la partie précédente. Il marche comme suit : on suppose qu'à l'itération  $t$  on a le vecteur  $X^{(t)} = x$  en  $t + 1$  on aura le vecteur  $X^{(t+1)}$  de cette façon :

1. On génère un vecteur  $y$  à partir de la densité d'une loi de proposition  $q(\cdot|x)$ .
2. Ensuite on calcule la probabilité d'acceptation de  $y$  comme valeur de l'itération en  $t + 1$  :

$$\alpha(x, y) = \min\left(1, \frac{f(y)}{f(x)} * \frac{q(x|y)}{q(y|x)}\right) \quad (2.4)$$

3. La dernière étape est d'attribuer  $X^{(t+1)}$  tel que :

$$X^{(t+1)} = \begin{cases} y & \text{avec la probabilité } \alpha(x, y) \\ x & \text{sinon.} \end{cases}$$

Il s'ensuit que la séquence produite  $X^{(0)}, X^{(1)}, \dots$  est une chaîne de Markov homogène ayant comme noyau de transition  $P(y|x)$  dont la densité s'écrit :

$$p(y|x) = \alpha(x, y)q(y|x) \quad \forall y, x \in \Theta$$

On peut montrer que  $f$  est une mesure invariante de cette chaîne. Et si la chaîne est ergodique alors  $f$  est unique ce qui nous ramène à dire que la séquence de vecteurs aléatoires  $X^{(m+1)}, X^{(m+2)}, \dots$  admet une loi qui a approximativement pour densité  $f$  avec  $m$  adéquatement choisi.

Une remarque importante est que dans l'équation 2.4, le ratio  $\frac{f(y)}{f(x)}$  est le ratio de la loi à postériori, ce ratio permet de se débarrasser de la constante de normalisation qui comme on l'avait vu dans les sections précédentes, n'est pas évidente à calculer.

### L'algorithme d'échantillonnage de Gibbs

Cet algorithme est une forme particulière de l'algorithme de Metropolis–Hastings. La différence ici est qu'on accepte toujours la valeur de proposition. Cet algorithme est utilisé dans le cas où l'on a plusieurs paramètres à estimer. Depuis bien longtemps, la méthode MCMC se basait essentiellement sur les deux algorithmes (Metropolis–Hastings et de l'échantillonnage de Gibbs). Plusieurs packages et outils de calcul informatique sont basées sur ces algorithmes, nous pouvons citer WinBUGS, OpenBUGS, MCMCglmm.

#### 2.3.4 Estimation des intervalles de confiance dans l'approche bayésienne

Lors de l'estimation des paramètres d'une régression, il est important d'avoir un aperçu sur la crédibilité autour de l'estimation du paramètre donc d'avoir des intervalles de

confiance. Dans l'estimation fréquentiste commune, la notion d'intervalle de confiance à  $\alpha\%$  se formule comme suit : "nous sommes sûrs à  $\alpha\%$  que l'intervalle  $[L, U]$  contient la vraie valeur du paramètre". Cependant, dès son introduction initiale, Neyman (1938) a précisé que l'interprétation d'un intervalle de confiance n'impliquait aucune conclusion probabiliste sur le paramètre inconnu, mais devait plutôt être interprétée en termes de fréquence relative à laquelle le processus par lequel l'intervalle a été généré capturerait le paramètre inconnu lors d'essais répétés de l'expérience. La spécification de ces intervalles se base sur des données non observées plutôt que sur des données expérimentales disponibles ce qui peut conduire à des conclusions indéfendables lorsqu'elles sont appliquées à l'expérience en question.

Dans l'estimation Bayésienne, la notion d'intervalle de confiance est plus simple et parvient à éviter les conflits potentiels avec les données observées. Comme nous l'avons dit, toutes les inférences sur un paramètre  $\theta$  sont basées sur sa loi à postériori. L'équivalence dans le cadre bayésien d'un intervalle de confiance pour le paramètre  $\theta$  est appelée *intervalle de crédibilité*, en anglais *credibility interval* et est obtenue à partir de la distribution postérieure en sélectionnant un intervalle correspondant au niveau de probabilité souhaité.

Par exemple, tout intervalle  $[\theta_L; \theta_U]$  tel que :

$$\int_{\theta_L}^{\theta_U} g(\theta|\mathbf{y})d\theta = 1 - \alpha \quad (2.5)$$

est un intervalle de crédibilité à  $100(1 - \alpha)\%$  pour le paramètre  $\theta$  où  $g(\theta|\mathbf{y})$  est la densité à postériori de la loi du paramètre  $\theta$  (Francisco J SAMANIEGO 2010a).

## 2.4 Critères de validation des modèles Bayésiens

### 2.4.1 Généralité

Lors du jugement et de la sélection d'un modèle, la simplicité et le "good-fit" sont les critères les plus importants. En effet, nous voulons choisir le modèle le plus simple et à la fois qui décrit bien les variations de nos données. Une modèle trop simple aura tendance à faire du "under-fitting" tandis qu'un compliqué fera du "overfitting".

L'évaluation de la pertinence d'un modèle repose sur essentiellement les critères d'information et la cross-validation. La validation croisée mesure l'ajustement d'un modèle sur l'ensemble des données de test, qui n'est pas utilisé pour ajuster le modèle, tandis que les critères d'information qui sont une forme modifiée de la somme des carrés des résidus en ajoutant une pénalité pour la complexité du modèle ou pour l'overfitting.

De façon classique, les critères d'information utilisés sont Akaike information criterion (AIC) ou Bayesian information criterion (BIC), dans le cadre Bayésien, le critère d'information utilisé est le Watanabe–Akaike information criterion (WAIC) qui est un critère plus généralisé et plus adapté dans le cadre bayésien que les autres critères. De l'autre côté, pour la cross-validation, la méthode basée sur le PSIS-LOO (Pareto smoothed importance sampling-Leave-one-out cross-validation) est une méthode robuste qui utilise la log-vraisemblance de la loi à postériori des paramètres.

Comme dans la section 2.3.1, posons comme données  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_p] \in \Theta \subseteq \mathbb{R}^p$  l'ensemble des informations inconnues du modèle, essentiellement les paramètres et  $\mathbf{x} = [x_1, x_2, \dots, x_n]$  l'ensemble de nos données observées. Ainsi  $f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i|\boldsymbol{\theta})$ . Si de plus on suppose que nous disposons d'une distribution à priori  $\pi(\boldsymbol{\theta})$ , la distribution à posteriori sera donc  $\pi(\boldsymbol{\theta}|\mathbf{x})$  et la densité de la loi à posteriori predictive est, pour une donnée  $\mathbf{y} = [y_1, y_2, \dots, y_n] : g(\mathbf{y}|\mathbf{x}) = \int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$ . Dans le cadre bayésien, la mesure de la précision d'un modèle sur une donnée  $\mathbf{y}$  se fait à travers une quantité **lppd** (log point wise predictive density) calculée comme suit :

$$\text{lppd} = \sum_{i=1}^n \log g(y_i|\mathbf{x}) \quad (2.6)$$

$$= \sum_{i=1}^n \log \left[ \int_{\Theta} f(y_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} \right] \quad (2.7)$$

Si nos données sont divisées en train et en test alors nous aurons un  $\text{lppd}_{train}$  et un  $\text{lppd}_{test}$ . De façon pratique, pour calculer cette quantité, l'on peut utiliser la moyenne empirique des échantillons de la loi à posteriori simulés.

Si on a un nombre  $S$  d'échantillons simulés qu'on note  $\boldsymbol{\theta}^s, s = 1, \dots, S$ , la quantité lppd sera :

$$\widehat{\text{lppd}} = \sum_{i=1}^n \log \left[ \frac{1}{S} \sum_{s=1}^S f(y_i|\boldsymbol{\theta}^s) \right]$$

### 2.4.2 Leave-one-out

Avant la construction d'un modèle, les données sont séparées en données de d'apprentissage qui permettent de mettre en place le modèle (train) et en données de validation qui permettent d'évaluer la précision du modèle(test) PERNA et al. 2012.

Le Leave-one-out est un type de cross validation où la repartition de nos données  $\mathbf{x} = [x_1, x_2, \dots, x_n]$  est telle que un nombre  $n - 1$  sont considérées comme données d'apprentissage et l'individu restant est utilisé pour valider le modèle construit à travers les données train. Le LOO-CV est estimée comme suit :

$$\text{lppd}_{loo} = \sum_{i=1}^n \log g(y_i|\mathbf{x}_{-i})$$

Avec

$$g(y_i|\mathbf{x}_{-i}) = \int_{\Theta} f(y_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x}_{-i})d\boldsymbol{\theta},$$

qui est la densité de la loi à posteriori prédictive sans la  $i$ eme observation (VEHTARI et al. 2017).

De façon similaire à la section plus haut, avec un nombre  $S$  de simulation, en pratique, cette quantité peut être calculée comme :

$$\widehat{\text{lppd}}_{loo} = \sum_{i=1}^n \log \left[ \frac{1}{S} \sum_{s=1}^S f(y_i|\boldsymbol{\theta}^{is}) \right]$$

où  $\boldsymbol{\theta}^{is} s = 1, \dots, S$  sont les simulations de la loi à posteriori  $\pi(\boldsymbol{\theta}|\mathbf{x}_{-i})$ .

### 2.4.3 WAIC

Avant la description de ce critère, nous prenons la peine de rappeler les critères classiques comme le AIC, BIC et le DIC GAO et al. 2018.

Le critère AIC (Akaike information criterion) par Hirotugu Akaike en 1973 s'écrit :

$$AIC = -2 \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}_{EMV}) + 2p$$

Le critère BIC quant à lui, formulé par Gideon E. Schwarz (1978) :

$$BIC = -2 \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}_{EMV}) + p \log n,$$

où

- $p$  est le nombre de paramètres du modèle ;
- $\boldsymbol{\theta}_{EMV}$  est l'estimateur du maximum de vraisemblance du paramètre ;  $\boldsymbol{\theta}$
- $f(y_i | \boldsymbol{\theta}_{EMV})$  est la vraisemblance des données.

Le premier terme en commun  $-2 \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}_{EMV})$  mesure la divergence entre le modèle ajusté et les données. Le second terme mesure la complexité du modèle.

Dans le cadre bayésien, nous définissons le critère WAIC comme :

$$WAIC = 2 \sum_{i=1}^n \left( \log \mathbb{E}(f(y_i | \boldsymbol{\theta})) - \mathbb{E}(\log f(y_i | \boldsymbol{\theta})) \right)$$

Dans le cas où l'on ne dispose que des échantillons de la loi à postériori, le critère WAIC basé sur les échantillons s'écrit :

$$\widehat{WAIC} = 2 \sum_{i=1}^n \left( \log \left[ \frac{1}{S} \sum_{s=1}^S f(y_i | \boldsymbol{\theta}^{is}) \right] - \frac{1}{S} \sum_{s=1}^S \log f(y_i | \boldsymbol{\theta}^{is}) \right)$$

où  $\boldsymbol{\theta}^{is}$   $s = 1, \dots, S$  sont les simulations de la loi à posteriori  $\pi(\boldsymbol{\theta} | \mathbf{x}_{-i})$ .

### 2.4.4 Le coefficient de détermination : $\mathbf{R}^2$

Le coefficient de détermination noté  $\mathbf{R}^2$  correspond à la part de variance expliquée d'un modèle par rapport à la variance totale des données. Étant un ratio, sa valeur est comprise entre 0 et 1. Une valeur de la statistique proche de 1 est signe d'une bon ajustement du modèle.

Considérons une régression où la réponse  $y$ , les variables explicatives  $X$ , le vecteur contenant les paramètres  $\theta$  et les valeurs estimées  $\mathbb{E}(y|X, \theta)$  visant à ajuster les données observées  $(X, y)_n, n = 1, \dots, N$ . On note  $\hat{\theta}$  la valeur estimée des paramètres du modèle par maximum de vraisemblance ou par les moindres carrés. Les valeurs prédites sont notées

alors  $\hat{y} = \mathbb{E}(y|X, \hat{\theta})$ .

Par ailleurs, nous définissons la fonction variance telle que :

$$V_{n=1}^N z_n = \frac{1}{N-1} \sum_{i=1}^N (\hat{z}_n - \bar{z})^2 \text{ pour tout vecteur } z.$$

La formule du  $\mathbf{R}^2$  s'écrit :

$$\mathbf{R}^2 = \frac{V_{n=1}^N \hat{y}}{V_{n=1}^N y} \quad (2.8)$$

Mais cette définition habituelle du  $R^2$  pose un problème pour les modèles bayésiens. En effet, dans certains cas où les données sont insuffisantes et que les lois à priori sont très informatives, il arrive le numérateur  $V_{n=1}^N \hat{y}$  (variance expliquée par le modèle) soit plus grand que le dénominateur  $V_{n=1}^N y$  (variance totale). Par conséquent le ratio devient supérieur à 1. (TJUR 2009).

Pour résoudre ce problème, les auteurs GELMAN et al. 2019 ont proposé une retraitement du critère dans le but de l'adapter dans les cas bayésiens.

Sachant que les paramètres  $\theta$  dans le cadre bayésien, sont des lois de probabilités. Si nous disposons des échantillons de taille  $S$  de notre loi à posteriori du vecteur  $\theta$ , le retraitement consiste à adapter la variance estimée par le modèle et à décomposer la variance totale :

$${}_s \mathbf{R}_{bayes}^2 = \frac{\text{var}_{fit}}{\text{var}_{fit} + \text{var}_{res}} \quad (2.9)$$

où

$$\begin{aligned} \text{var}_{fit} &= V_{n=1}^N \tilde{y}^s, \\ \text{var}_{res} &= \mathbb{E} \left( V_{n=1}^N (y - \tilde{y}^s) | \theta^s \right) \\ \text{avec } \tilde{y} &= \mathbb{E}(y|X, \theta^s) \quad s = 1, \dots, S \end{aligned}$$

Ainsi, nous aurons un  ${}_s \mathbf{R}^2$  pour chaque échantillon, le  $\mathbf{R}^2$  final peut être choisi de sorte à être la moyenne empirique ou la médiane de ces  ${}_s \mathbf{R}^2$ .

Cette solution proposée a consisté à décomposer le dénominateur comme la somme de la variance prédite et la variance d'erreur. Cette mesure est plus adaptée et est basée sur le modèle : tous les termes de la variance proviennent du modèle et non directement des données.

Nous rappelons que, dans le cadre de notre étude, le  $\mathbf{R}^2$  utilisé pour le modèle bayésien sera une moyenne empirique des différents  $\mathbf{R}^2$  de chaque échantillon de la loi à posteriori du vecteur des paramètres.

## Chapitre 3

Mise en oeuvre du modèle GAM  
Bayésien sur un traité de swap de  
longévité

### 3.1 Définition de la mortalité

Il est important de préciser les notations sur certaines quantités de mortalité ainsi que la quantité d'intérêt qui sera modélisée à travers notre modèle.

Soit  ${}_t p_x$  la probabilité de survie entre les période  $x$  et  $x+1$  et  ${}_t q_x = 1 - {}_t p_x$  la probabilité de décès sur la même période, où  $x > 0$  représente l'âge et  $0 \leq t \leq 1$  fait référence à un point dans le temps. La probabilité de décès  ${}_t q_x$  est calculée en rapportant un nombre de décès sur la période à l'effectif en début de période.

Nous pouvons calculer également une autre quantité importante qui le ratio central de mortalité  ${}_t m_x$  (central mortality ratio en anglais) qui est obtenu en rapportant le nombre de décès à l'effectif moyen sur la période soit :

$${}_t m_x = \frac{{}_t d_x}{{}_t L_x}, \quad (3.1)$$

où  ${}_t d_x$  est le nombre attendu de décès au cours de l'année calendaire  $t$  chez les personnes âgées de  $\lfloor x \rfloor$  au dernier anniversaire.

Et  ${}_t L_x$  la population moyenne durant l'année calendaire  $t$  chez les personnes âgées de  $\lfloor x \rfloor$  au dernier anniversaire.

Ce dénominateur est généralement estimé par le temps total vécu par les personnes âgées de  $\lfloor x \rfloor$  durant l'année calendaire  $t$ . On parle alors d'**exposition central au risque** notée  $E_{xt}^c$  contrairement au dénominateur dans l'expression de  ${}_t q_x$  qui lui est le nombre total de personnes en vie de personnes âgées de  $\lfloor x \rfloor$  durant l'année calendaire  $t$ . On parle dans ce dernier cas d'**exposition initial au risque** notée  $E_{xt}^i$ . le nombre attendu de décès au cours de l'année calendaire  $t$  chez les personnes âgées de  $\lfloor x \rfloor$  au dernier anniversaire est estimé par le nombre de décès observée qu'on note  $D_{xt}$ .

L'estimation résultante  $D_{xt}/E_{xt}^c$  du ratio central de mortalité  ${}_t m_x$  est appelé **taux brut de mortalité pour l'âge  $x$  en  $t$**  (crude death rate en anglais.)

Les probabilités de décès et de survie  ${}_t q_x$  et  ${}_t p_x$  sont des nombres sans dimension alors que les taux de décès sont exprimés en inverse de l'unité de temps et décomptent des décès par unité de temps. Cette différence conduit aux relations suivantes avec la fonction de hasard, appelée dans ce contexte "taux instantané de mortalité" :

$$\mu_{x+t} = \frac{1}{{}_t p_x} \frac{\partial}{\partial t} q_x, \quad (3.2)$$

En général,  $\mu_{x+t}$  n'est pas directement observable, car les données de mortalité ne sont pas collectées de façon continue.

Dans le but d'obtenir ces quantités à partir des données discrètes, il est crucial de poser une hypothèse de constance de  $\mu_{x+t}$  entre chaque année entière et d'année calendaire entière, soit :

$$\mu_{x+\xi} = \mu_x, \text{ pour } 0 \leq \xi < 1 \text{ et } x \text{ un nombre entier.} \quad (3.3)$$

Cette hypothèse permet de simplifier considérablement la mise en place de nos modèles concernant nos quantités d'intérêt.

La dérivée directe est que la fonction de hasard est égale au taux brut de mortalité  ${}_t m_x$ .

Par ailleurs, la relation entre la probabilité de survie et la fonction de hasard normalement s'exprimant sous la forme :

$${}_t p_x = \exp \left[ - \int_0^t \mu_{x+s} ds \right], \text{ avec } 0 \leq t < 1 \quad (3.4)$$

s'écrit de façon simplifiée sous l'hypothèse :

$${}_t p_x = \exp(-t\mu_x), \text{ avec } 0 \leq t < 1 \quad (3.5)$$

et donc :

$$q_x = 1 - \exp(-\mu_x). \quad (3.6)$$

qui est une relation très importante pour la suite de notre analyse vu qu'elle permettra de passer d'une quantité à une autre très facilement.

## 3.2 Spécification de notre modèle

Comme dans la section précédente, nous posons  $D_{x,g}$  et  $E_{x,g}$  comme respectivement, le nombre de décès observés des personnes chez les personnes âgées de  $x$  et de sexe  $g \in [M, F]$  et l'exposition central au risque soit le temps total en années vécus par ces mêmes personnes.

Dans le cadre de la modélisation, nous modélisons la variable réponse : le nombre de décès comme une variable aléatoire suivant une loi de poisson :

$$D_{x,g} \sim \mathcal{P}(\mu_{x,g} E_{x,g}) \quad (3.7)$$

En se basant sur l'hypothèse faite dans la section précédente sur la constance de la force de mortalité entre deux âges entiers, on obtient :

$$D_{x,g} \sim \mathcal{P}(-\log(1 - q_{x,g}) E_{x,g}) \quad (3.8)$$

L'hypothèse que le nombre de décès suit une loi de poisson est très utilisée lors des analyses de la mortalité. Cette loi est celle qui décrit le mieux, de façon naturelle la mortalité d'après les arguments de Czado et al. (2005).

Ainsi, nous modélisons notre modèle additif généralisé tel que :

$$\mathbf{log}(\mathbb{E}(D_{x,g})) = \mathbf{log}(E_{x,g}) + \mathbf{log}(\mu_{x,g}), \quad (3.9)$$

où la force de mortalité  $\mu_{x,g}$  est modélisée comme suit :

$$\mathbf{log}(\mu_{x,g}) = \beta_0 + \beta_g + f_g(x) \quad (3.10)$$

Ce qui revient à modéliser un GAM avec variable offset qui est l'exposition au risque composée de la fonction **log**.

En utilisant la fonction **log** comme fonction lien, le modèle se rapproche du modèle log-poisson.

### 3.2.1 Particularité de la modélisation

Par ailleurs, nous avons décidé de modéliser de façon jointe les mortalités des femmes et celles des hommes plutôt que de le faire séparément. Cette modélisation jointe a plusieurs avantages, en effet, d'abord elle simplifie le processus de modélisation qui évitera de calibrer deux fois le modèle qui est déjà coûteux en temps et en moyen de calcul. Le but recherché de l'équipe étant de mettre en place une méthode d'analyse d'expérience simple à implémenter, cette modélisation s'impose comme appropriée.

De plus, cette modélisation conjointe évite d'avoir des croisements entre les mortalités féminines et masculines. En effet, d'après les résultats de la statistique descriptive du portefeuille dans le chapitre précédent, les mortalités féminines sont moins élevées que celles des hommes. Cette remarque s'inscrit dans le cadre général qui reste un indicateur clé dans l'analyse de la cohérence des mortalités modélisées.

Les taux de mortalité masculins et féminins présentent souvent des tendances convergentes aux âges élevés. En modélisation jointe, les informations pourraient être partagées entre les deux sexes. Cela produira des estimations cohérentes, surtout aux âges où l'exposition est faible, puisque nous mettons en commun les deux ensembles de données.

### 3.2.2 Description des paramètres du modèle

Les fonctions  $f_g, g \in [M, F]$  représentent les fonctions de lissage utilisées dans la modélisation qui définissent la différence principale entre un GLM et un GAM. L'indice  $g$  indique que dans la modélisation nous obtiendrons deux fonctions splines calibrées selon chaque sexe. Ces fonctions dépendantes de l'âge  $x$ , permettent de capturer la forme sous-jacente de la courbe de log-mortalité tout en obtenant une forme lissée. Cela constitue la principale motivation de l'utilisation d'un GAM dans le cadre de notre analyse d'expérience.

Par ailleurs, il existe plusieurs types de splines pouvant être utilisées pour la calibration d'un GAM. Dans notre cas, nous utilisons une famille de fonctions de lissage appelée P-splines que nous verrons dans la section suivante.

Le paramètre  $\beta_g$  est un paramètre qui permet de tenir compte de la différence de mortalité entre les populations masculines et féminines.

Le fait que nous ayons décidé de calibrer deux fonctions de lissage dans notre modèle (une pour chaque sexe) tiendra compte de cette spécificité. Mais l'introduction de ce paramètre viendra l'amplifier dans le cas où les deux fonctions de lissage calibrées sont proches.

Le paramètre  $\beta_0$  désigne l'intercept du modèle qui s'applique sur toutes les observations des données.

## 3.3 Définition des B-spline et P-spline

Étant donnée que la modélisation des mortalités dans notre étude s'articulent autour de fonctions de lissage appelées splines, il se pose essentiel de définir les splines en général et les P-splines en particulier.

### 3.3.1 Les fonctions de lissage : splines

Une spline est une fonction continue dans un intervalle choisi, formée par des fonctions de base simples qui sont connectées entre elles. (HASTIE et al. 2009). Chaque point où les fonctions se connectent est appelé un nœud de la spline. Les fonctions de base peuvent être sous plusieurs formes notamment être des polynômes.

Une spline est d'ordre  $K$  si elle est formée en reliant des fonctions de base polynomiales de degré  $K - 1$ , qu'elle est dérivable  $K - 2$  fois aux nœuds et qu'elle est continue.

Pour une variable donnée  $x$ , son évaluation par une spline  $s$  est faite d'abord en divisant sa plage de valeurs en intervalles contigus séparés par les noeuds et ensuite en définissant des polynômes distincts sur chaque intervalle.

La fonction spline peut s'écrire alors :

$$s(x) = \sum_{m=1}^M \beta_m b_m(x)$$

où  $b_m$  sont les fonctions de base qui sont déjà choisies et fixes,  $M$  le nombre de fonctions de base et les  $\beta_m$  sont les paramètres à estimer. Il est à noter que le nombre de nœuds limite le nombre de fonctions de base,  $M$ .

### 3.3.2 Les B-splines

Les fonctions de base peuvent prendre différentes formes. Dans notre modélisation, nous avons décidé d'utiliser la fonction B-spline. Parmi les nombreux choix possibles, la fonction B-spline est particulièrement utile dans notre cadre en raison de sa structure creuse que nous verrons plus bas dans cette partie. Cette caractéristique nous est particulièrement utile, car elle améliore considérablement la stabilité numérique du processus d'estimation lorsque le nombre de coefficients utilisés est important (ZHU et al. 2022).

En se basant sur le livre WOOD 2006 comme référence principale. Pour définir une fonction B-spline avec  $M$  fonctions de base, il convient de définir préalablement  $M + m + 1$  noeuds tels que :  $x_1 < x_2 < \dots < x_{M+m+1}$  où  $m + 1$  est l'ordre de la spline.

L'intervalle sur lequel la fonction B-spline est évalué est  $[x_{m+2}; x_M]$  de sorte que les premiers et derniers noeuds  $m + 1$  soient essentiellement arbitraires.

Ainsi l'expression d'une fonction B-spline d'ordre  $m + 1$  est donnée comme :

$$f(x) = \sum_{i=1}^M \beta_i B_i^m(x), \quad (3.11)$$

où les fonctions de base de la B-spline sont définies de façon récurrente telles que :

$$B_i^m(x) = \frac{x - x_i}{x_{i+m+1} - x_i} B_i^{m-1}(x) + \frac{x_{i+m+2} - x}{x_{i+m+2} - x_{i+1}} B_{i+1}^{m-1}(x), \text{ pour } i = 1, \dots, M$$

Et

$$B_i^{-1}(x) = \begin{cases} 1 & \text{si } x_i < x < x_{i+1} \\ 0 & \text{sinon.} \end{cases}$$

Comme on peut le voir chaque fonction de base de la B-spline n'a qu'un support local, ce qui signifie que l'impact de la  $i$ ème fonction de base devient nul lorsque  $x \geq x_{i+m+2}$  ou  $x \leq x_i$ . Cette caractéristique rend la structure de la fonction B-spline creuse. Comme on l'avait dit, elle procure une stabilité numérique lors de l'estimation des coefficients de la fonction. D'autre part, elle est particulièrement adaptée à la modélisation des effets temporaires à court terme. Ce qui est le cas dans notre analyse où nous voulons estimer la mortalité de base du moment sans tenir compte des effets long-terme entre les différents âges et générations (effets cohorts). Par ailleurs, il s'impose de choisir l'ordre de la fonction B-spline.

#### 3.3.3 Les P-splines

Dans une régression, le choix de du nombre de noeuds et de leur emplacement peut avoir un grand impact sur la forme de la fonction B-spline après calibration. L'estimation des coefficients peut conduire à un overfitting si la forme de la fonction est trop déformée. Inversement, la qualité d'ajustement est réduite si la forme de la fonction tend vers celle d'une fonction linéaire.

Ainsi le challenge est de trouver un équilibre entre la qualité d'ajustement de du modèle et la capacité de lissage de la fonction obtenue.

L'utilisation des P-splines permet de résoudre ce problème. Les P-splines sont des lisseurs utilisant des fonctions de base B-spline, généralement définies sur des noeuds uniformément espacés, et une pénalisation appliquée directement aux paramètres,  $\beta_i$ , dans le but contrôler l'irrégularité de la fonction tout en optimisant la qualité d'ajustement du modèle.

Pour notre jeu de données  $(\mathbf{d}, \mathbf{e}, \mathbf{x})$  défini respectivement comme le vecteur des décès observés, des expositions et des variables explicatives, composé de  $N$  observations.

L'enjeu est de trouver une fonction  $f$  qui minimise :

$$\sum_{i=1}^N \left( d_i - f(x_i) \right)^2 + \lambda \int f''(x)^2 dx \quad (3.12)$$

En effet, la première partie de l'expression représente l'erreur quadratique entre le nombre de décès observés et celui estimé par la fonction. Le second membre de l'expression, qui désigne la pénalisation, est l'intégrale de la dérivée seconde au carré de la fonction recherchée. Ce dernier mesure le degré de lissage de la fonction  $f$ , plus il est grand, plus la fonction est courbée. Dans le cas où  $f$  est une fonction linéaire, il est évidemment égal à zéro.

Le terme  $\lambda$  est un paramètre réglable, utilisé pour contrôler le poids relatif à accorder aux degré de lissage lors de la recherche de la fonction  $f$ . Ainsi en minimisant l'expression 3.12, nous réduisons l'erreur d'estimation tout en cherchant à minimiser le degré de courbure de la fonction.

Des travaux de GREEN et al. 1993 ont montré que les fonctions qui minimisent l'expression sont des splines cubiques. Ce qui justifie notre choix d'utilisation d'une P-spline dont la dimension de la B-spline est choisie de sorte à obtenir des fonctions cubique, donc

$m + 1 = 3$  donc  $m = 2$ .

De façon générale, la pénalisation peut s'écrire sous forme matricielle telle que :

$$\lambda \int f''(x)^2 dx = \lambda \boldsymbol{\beta}^T S \boldsymbol{\beta} \quad (3.13)$$

où  $\boldsymbol{\beta}$  est le vecteur contenant l'ensemble des paramètres la fonction spline et  $S$  une matrice constante servant à pénaliser les paramètres et  $\lambda$  est le paramètre de lissage à estimer qui contrôle l'influence de la pénalisation.

Dans le cas des p-splines, la pénalisation s'exprime sous une forme différentielle entre les différents paramètres de la B-spline. La pénalisation a la forme :

$$\lambda \int f''(x)^2 dx = \lambda \boldsymbol{\beta}^T S \boldsymbol{\beta} = \lambda \sum_{j=k+1}^M \left( \Delta^k \beta_j \right)^2 \quad (3.14)$$

où  $\Delta \beta_j = \beta_j - \beta_{j-1}$ .

Nous prenons un exemple pour mieux comprendre le principe : dans le cas où  $k = 1$ , nous aurons comme pénalisation :

$$\mathcal{P} = \sum_{i=1}^{M-1} (\beta_{i+1} - \beta_i)^2 = \beta_1^2 - 2\beta_1\beta_2 + 2\beta_2^2 - 2\beta_2\beta_3 + \dots + \beta_M^2$$

Nous pouvons récrire la pénalisation sous forme matricielle telle que :

$$\mathcal{P} = \boldsymbol{\beta}^T \begin{bmatrix} 1 & -1 & 1 & \dots & \dots \\ -1 & 2 & -1 & \dots & \dots \\ 0 & -1 & 2 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \boldsymbol{\beta}$$

### Pénalisation de la vraisemblance

L'introduction de cette pénalisation sur les coefficients  $\beta_i$  conduit à obtenir une log vraisemblance et une vraisemblance pénalisées.

En effet, la log vraisemblance pénalisée est alors :

$$l_p(\boldsymbol{\beta}, \lambda) = l(\boldsymbol{\beta}) - \frac{1}{2} \lambda \boldsymbol{\beta}^T S \boldsymbol{\beta} \quad (3.15)$$

Il s'agit bien d'une jauge entre la qualité d'ajustement (une log-vraisemblance élevée donnant des lissages proches des données) et la pénalisation (une grande pénalisation indiquant que les lissages sont trop ondulés).

Les P-splines sont faciles à mettre en place même dans le cadre bayésien. Il existe plusieurs librairies dans les différents logiciels de calibrage qui tiennent compte des P-splines.

### 3.4 Réécriture du modèle

Notre modèle peut dès lors s'écrire sous forme matricielle.

On définit la matrice du modèle  $\mathbf{X}$  de taille  $N * p$  où  $N$  est le nombre d'observations dans notre base de données et  $p$  est le nombre total de paramètres à estimer soit  $2 * K + 2$  paramètres où  $K$  est le nombre de fonctions de base de la B-spline.

L'objectif de la matrice du modèle est de représenter l'ensemble des variables du modèle de sorte à être cohérent avec le vecteur des paramètres  $\boldsymbol{\alpha} = (\beta_0, \beta_g, \beta_1^F, \dots, \beta_K^F, \beta_1^M, \dots, \beta_K^M)$  dans l'écriture de la régression.

Avant d'écrire la matrice de notre modèle, nous définissons la matrice de la B-spline notée  $\mathbf{X}_{bs}$ , qui consiste à écrire la fonction B-spline définie à l'équation 3.11 sous forme matricielle telle que :

$$\mathbf{X}_{bs}(i, j) = B_j^m(x_i) \text{ où } m = 2.$$

Dans la spécification de notre modèle, nous avons décidé d'appliquer les mêmes fonctions de base pour les populations masculines et féminines. Évidemment, les paramètres estimés associés à ces fonctions de base seront différents à la fin du calibrage selon les sexes.

Donc, notre matrice du modèle s'écrira :

$$\mathbf{X}(i, j) = \begin{cases} 1 & \text{pour } j = 1 \\ 1_{\{g_i=F\}} & \text{pour } j = 2 \text{ où } g_i \text{ est le sexe de l'observation } i. \\ 1_{\{g_i=F\}} \mathbf{X}_{bs} & \text{pour } 2 < j \leq K + 2 \\ 1_{\{g_i=M\}} \mathbf{X}_{bs} & \text{pour } K + 2 < j \leq 2 * K + 2 \end{cases}$$

où  $x_i$  est désigne l'âge de l'observation  $i$ .

Comme nous l'avons vu plus haut, l'on peut définir la valeur des B-splines seulement après avoir choisi un ensemble de noeuds.

Le choix du nombre de noeuds et de leur emplacement impactent la valeur finale de la fonction de lissage.

Le vecteur comprenant l'ensemble des noeuds est composé de deux sous-vecteurs. En effet, il est composé d'un ensemble de noeuds internes et d'un ensemble de noeuds "augmentés". Les noeuds internes sont les plus importants car ils définissent les supports des fonctions de base contrairement aux noeuds augmentés qui sont choisis de façon arbitraire.

Dans notre cas, nous choisissons nos noeuds internes de façon uniforme sur l'ensemble des valeurs des âges avec comme nombre de noeuds défini par défaut  $M - m$  où  $m + 1$  est l'ordre de la B-spline.

Explicitement, le nombre de fonction de bases que nous utiliserons est  $M = 10$  qui est la valeur par défaut dans les régressions faisant appel aux splines et la B-spline est cubique donc  $m = 2$ . Ce qui nous un total de 8 noeuds internes.

Explicitement, pour respectivement les populations masculines et féminines, nous auront comme valeur du  $k^{ieme}$  noeud :

$$k_i = x_{min} + (i - 1) \frac{x_{max} - x_{min}}{7}, \forall i \in [1, 8] \quad (3.16)$$

où  $x_{min}$  et  $x_{max}$  sont respectivement les valeurs minimale et maximale des âges dans le portefeuille.

La matrice du modèle étant définie, la version matricielle de notre modèle 3.10 peut s'écrire sous la forme :

$$\log(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha}. \quad (3.17)$$

Explicitement :

$$\log(\boldsymbol{\mu}) = \begin{bmatrix} \log(\mu_1) \\ \log(\mu_2) \\ \vdots \\ \log(\mu_N) \end{bmatrix},$$

et

$$\mathbf{X} = \begin{bmatrix} 1 & 1_F^{(1)} & 1_F^{(1)} B_1(x_1) & \dots & 1_F^{(1)} B_{10}(x_1) & 1_M^{(1)} B_1(x_1) & \dots & 1_M^{(1)} B_{10}(x_1) \\ 1 & 1_F^{(2)} & 1_F^{(2)} B_1(x_2) & \dots & 1_F^{(2)} B_{10}(x_2) & 1_M^{(2)} B_1(x_2) & \dots & 1_M^{(2)} B_{10}(x_2) \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & 1_F^{(N)} & 1_F^{(N)} B_1(x_N) & \dots & 1_F^{(N)} B_{10}(x_N) & 1_M^{(N)} B_1(x_N) & \dots & 1_M^{(N)} B_{10}(x_N) \end{bmatrix}$$

où  $1_G^{(i)} = 1_{\{g_i=G\}}$  qui est littéralement l'indicatrice qui prend la valeur 1 si l'observation  $i$  est de sexe  $G$  et 0 sinon.

Sous cette forme, le GAM peut être considéré comme un GLM standard avec un grand nombre de paramètres à estimer.

La solution du modèle n'est pas unique, en effet, les paramètres de la fonction de lissage ne peuvent être estimés qu'à une constante additive près. Donc il est indispensable d'imposer des contraintes sur la fonction de lissage dans le but d'éviter des problèmes d'identifiabilité.

Nous avons le choix entre plusieurs types de contraintes à imposer. Il peut s'agir soit de fixer la valeur d'un coefficient par exemple  $\beta_1^g = 0$  ou d'appliquer une contrainte centrale telle que :

$$\sum_{i=1}^N f_g(x_i) = 0 \quad (3.18)$$

C'est cette dernière que nous utiliserons lors de l'ajustement de notre modèle.

### Récapitulatif du modèle

$$\begin{cases} \mathbf{log}(\mu_{x,g}) = \beta_0 + \beta_g + f_g(\mathbf{x}), g \in \{F, M\} \\ \mathcal{P} = \lambda \sum_{i=2}^{10} (\beta_i - \beta_{i-1})^2 \\ \text{s.c. } \sum_{i=1}^N f_g(x_i) = 0 \end{cases}$$

où

- $f_g(x_i) = \sum_{i=1}^{10} \beta_i^g B^i(x_i)$
- $\alpha = (\beta_0, \beta_g, \beta_1^F, \dots, \beta_{10}^F, \beta_1^M, \dots, \beta_{10}^M)$  qui est le vecteur contenant l'ensemble des paramètres.
- $\mathcal{P}$  est la pénalisation utilisée qui correspond à  $\lambda \sum_{j=k+1}^M \left( \Delta^k \beta_j \right)^2$  de l'équation 3.14 pour  $k = 1$ .

### 3.5 Introduction du modèle GAM fréquentiste

Les modèles GAM avec une estimation bayésienne même s'ils s'avèrent efficaces dans certains domaines, ils ne sont pas assez répandues dans le domaine de la réassurance encore moins en longévité.

Dans un souci d'interprétation et de facilité de compréhension, nous introduisons pendant l'étude du modèle bayésien, le même modèle GAM mais avec une estimation fréquentiste des paramètres. Les variables réponses, paramètres du modèle, fonction lien et structure du modèle restent inchangés. Dans ce dernier cas, nos paramètres ne sont pas supposés être des variables aléatoires, ce qui ne nécessite pas de grands algorithmes pour leurs estimations.

### 3.6 Retraitement des données

L'estimation bayésienne a depuis longtemps existé mais peu utilisé du fait de sa longue durée de calcul même avec l'évolution de la technologie. Cependant, il existe plusieurs algorithmes notamment les méthodes MCMC qui raccourcissent la durée.

Cette durée de calcul varie beaucoup en fonction de la quantité de données et du nombre de covariables dans le modèle.

Dans notre cas, avec 102 224 observations et 2 covariables utilisées dans le modèle (le sexe et l'âge) et une covariable supplémentaire qui est l'année calendaire. Nous rappelons que cette dernière covariable n'intervient pas dans le modèle.

La durée d'exécution de l'algorithme est estimée à environ 20h ce qui est très considérable si nous gardons notre base de données telle qu'elle est.

Par ailleurs, l'un des avantages fondamentaux de modéliser la variable réponse comme une variable suivant une loi de poisson est que cette loi dispose d'une propriété de stabilité par la somme. Ce qui revient à dire :

Soit des variables  $(X_i)_{i=1, \dots, n}$ ,  $n \in \mathbb{N}$  étant i.i.d.

Si :

$$X_i \sim \mathcal{P}(\lambda_i)$$

Alors,

$$Y = \sum_{i=1}^n X_i \sim \mathcal{P}\left(\sum_{i=1}^n \lambda_i\right) \quad (3.19)$$

De cette propriété, il est possible d'agréger nos données similaires en fonction de l'exposition au risque et du nombre de décès de sorte à obtenir une combinaison unique de nos 2 covariables du modèle.

En effet pour un groupe d'individus fixé  $x, g$  comprenant  $n_{x,g}$  individus :

$$\sum_{i=1}^{n_{x,g}} d_{x,g}^{(i)} \sim \mathcal{P}\left(\mu_{x,g} \sum_{i=1}^{n_{x,g}} e_{x,g}^{(i)}\right) \quad (3.20)$$

Avec

$$D_{x,g} = \sum_{i=1}^{n_{x,g}} d_{x,g}^{(i)} \text{ et } E_{x,g} = \sum_{i=1}^{n_{x,g}} e_{x,g}^{(i)}$$

Après avoir agrégé nos données, nous obtenons une nouvelle table bidimensionnelle en fonction de nos deux covariables et qui sera directement utilisée pour calibrer les différents modèles.

Cette nouvelle table contient désormais environ 500 lignes, ce qui réduit considérablement la durée d'exécution des algorithmes de résolution.

## 3.7 L'estimation bayésienne du modèle

Après la spécification de l'équation du modèle et des différents paramètres à estimer, cette section présente plus en détail la procédure d'estimation bayésienne de notre GAM. D'abord il s'agit de déterminer explicitement la vraisemblance des données de mortalité observées et de spécifier les lois à priori de chaque paramètre. Ensuite, à travers les méthodes MCMC décrites dans la partie 2.3.3, les lois à posteriori sont déterminées.

### 3.7.1 Écriture de la vraisemblance des données observées

Étant donné notre modèle décrit dans la section 3.4 avec les contraintes établies pour le rendre identifiable. Ayant supposé que le nombre décès du portefeuille suit une loi de poisson :  $D_{x,g} \sim \mathcal{P}(\mu_{x,g} E_{x,g})$ , nous avons donc :

$$P(D_{x,g} = d) = \frac{(\mu_{x,g} E_{x,g})^d}{d!} \mathbf{exp}(-\mu_{x,g} E_{x,g})$$

où

$$\mu_{x,g} = \mathbf{exp}(\beta_0 + \beta_g + s_g(x))$$

En supposant que le nombre de décès est indépendant entre les âges et selon le sexe, la

vraisemblance du modèle associée à l'ensemble des données  $\mathbf{D} = (D_{x,g}, E_{x,g}, X, G)$  s'écrit :

$$\mathcal{L}(\mathbf{D}|\boldsymbol{\alpha}) = \prod_x \prod_g \frac{\exp\left(-E_{x,g} \exp(\beta_0 + \beta_g + s_g(x))\right) \left(E_{x,g} \exp(\beta_0 + \beta_g + s_g(x))\right)^{D_{x,g}}}{D_{x,g}!} \quad (3.21)$$

avec le vecteur des paramètres  $\boldsymbol{\alpha} = (\beta_0, \beta_g, \beta_1^F, \dots, \beta_{10}^F, \beta_1^M, \dots, \beta_{10}^M)$ .

Comme discuté dans la section 3.3.3, la vraisemblance retenue est celle pénalisée dans le but d'éviter l'overfitting issue de l'estimation directe des coefficients de la B-spline. Nous notons  $\mathcal{L}_p$  cette vraisemblance pénalisée dont l'expression est :

$$\mathcal{L}_p(\mathbf{D}|\boldsymbol{\alpha}, \lambda) = \mathcal{L}(\mathbf{D}|\boldsymbol{\alpha}) \exp\left(-\frac{1}{2} \lambda \boldsymbol{\beta}^T S \boldsymbol{\beta}\right) \quad (3.22)$$

où  $\boldsymbol{\beta} = (\beta_1^F, \dots, \beta_{10}^F, \beta_1^M, \dots, \beta_{10}^M)$  qui est l'ensemble des paramètres des splines.

$S$  est la matrice de pénalisation qui est construite de telle sorte qu'elle pénalise les coefficients de chaque spline liée aux deux populations, masculines et féminines, de façon similaire.

### 3.7.2 Choix et vérification des lois à priori

Le choix et la spécification des lois à priori des paramètres est une étape indispensable dans la modélisation bayésienne, c'est à travers elle que l'on incorpore toutes croyances à priori sur l'espace et la valeur intuitive des paramètres avant même de faire appel aux données.

#### Choix des lois à priori

Cette étape permet de prendre en compte les jugements d'expert et les intuitions que l'on a sur l'influence des covariables sur la variable réponse.

Dans notre modèle, les paramètres à estimer sont l'intercept  $\beta_0$ , le paramètre d'impact lié au sexe (une variable qualitative)  $\beta_g$ , les différents paramètres de la fonction de lissage de chaque sexe  $\beta_1^F, \dots, \beta_{10}^F, \beta_1^M, \dots, \beta_{10}^M$  et le paramètre de pénalisation  $\lambda$  qui est un hyperparamètre.

De façon générale, dans la spécification du GAM Bayésien, il y a deux types d'effet (variable) dans la modélisation : les variables qui sont définies pour avoir un effet au niveau de chaque individu et celles qui ont leur effet au niveau d'un groupe d'individus.

Dans la régression, chaque type d'effet a son paramètre dont il est possible de spécifier la loi à priori. En particulier dans notre modèle, nous n'avons que des variables ayant un effet au niveau de chaque individu : l'âge et le sexe.

Vu que nous avons défini la variable d'impact du sexe  $\beta_g$  comme une variable qualitative binaire dont la modalité de référence est le sexe masculin, il est possible de définir la loi à priori de la modalité qui n'est pas en référence en l'occurrence la modalité du sexe féminin.

### Chapitre 3. Mise en oeuvre du modèle GAM Bayésien sur un traité de swap de longévité

---

Dans notre modèle, nous avons décidé d'utiliser la loi normale comme loi sous jacente de tous nos paramètres. En effet, cette loi permet de tenir compte des valeurs négatives que peuvent prendre les paramètres.

La spécification de chaque paramètre se fera par le choix des paramètres de la loi normale à savoir la moyenne et la variance.

Le paramètre  $\beta_g$  est un paramètre important du modèle qui permet de tenir compte des différences observées entre les mortalités masculines et féminines. Comme on a pu le constater dans la chapitre sur l'étude des données, les femmes ont un mortalité plus faible que les hommes. Ainsi, la loi du paramètre doit être choisie de sorte à tenir compte de cette particularité.

La variable est une variable qualitative binaire dont la modalité de référence est la modalité masculine. Ainsi la loi à priori de la modalité féminine doit être choisie de sorte à être négative. En effet l'interprétation de l'estimation du paramètre se faisant toujours par rapport à la variable de référence, la négativité indiquera au modèle que l'effet des populations féminines sur le taux instantané de mortalité doit être plus faible que celui des populations masculines.

Ainsi nous choisissons comme paramètres :

$$\beta_F \sim \mathcal{N}(m, \sigma^2) \text{ avec } m = -3 \text{ et } \sigma^2 = 1 \quad (3.23)$$

La moyenne est choisie négative et la variance proche est unitaire dans le but de forcer le paramètre à tendre vers le négatif.

Le choix des lois à priori des paramètres des fonctions de lissage et du coefficient de pénalisation dépend de la covariable et permet d'incorporer les certitudes à priori que nous avons sur la forme du lissage de l'âge par rapport aux forces de mortalité.

Ces croyances sont que la structure du modèle doit être à la fois conduite par les données et être lisse au point où nous choisissons de pénaliser les modèles qui sont, d'une certaine manière, trop instables ou ondulés.

Par ailleurs, le coefficient de pénalisation  $\lambda$  est estimé simultanément avec les paramètres de la fonction de lissage. Dans l'estimation bayésienne, il est exprimé de sorte à être la variance de la loi des paramètres de lissage. Ainsi, nous posons  $\lambda = 1/\tau^2$ .

De façon générale, la loi à priori des  $\beta_i^g$  conditionnellement à  $\tau^2$  est donnée par une loi gaussienne multivariée telle que :

$$p(\boldsymbol{\beta}|\tau^2) \propto \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\beta}^T S \boldsymbol{\beta}\right)$$

La loi à priori de la variance est très faiblement informative et positive. Cette loi choisie est une loi Gamma Inverse de sorte à avoir  $\tau^2 \sim IG(a, b)$  avec  $a = b = 0.01$ . Ces hyperparamètres  $a$  et  $b$  peuvent être variables dépendant des paramètres des fonctions de lissage. Mais dans notre cas, ils seront fixes pour simplifier l'estimation.

Cette loi à priori du lissage est appropriée car elle rend explicite le fait que nous pensons que les modèles lisses sont plus probables que les modèles irréguliers.

Comme mentionné dans les sections précédentes, le modèle calibre deux fonctions de lissage dépendant du sexe masculin ou féminin. Les informations à priori sont les mêmes

pour chaque fonction de lissage. En effet, l'hypothèse de faible mortalité chez les populations féminines par rapport aux populations masculines est déjà prise en compte dans les informations du paramètre  $\beta_g$ . Par conséquent, nous donnons la liberté aux données de définir les formes finales des splines selon les sexes.

La matrice de pénalisation  $S$ , comme définie dans la section sur les splines, n'est pas inversible, par conséquent la loi à priori est impropre, ce qui la rend "quasi-gaussienne multivariée". Pour résoudre ce problème, une transformation basée sur les valeurs propres est effectuée.

#### Prior predictive check

L'objectif de cette partie est de s'assurer de la pertinence et la légitimité du choix de nos différentes loi à priori. Le choix de nos lois à priori est-il approprié? Le modèle est-il raisonnable bien avant d'avoir observé les données?

Sachant que la loi à posteriori est conduite à la fois par la vraisemblance des données et les lois à priori tel que :

$$p(\boldsymbol{\alpha}|\mathbf{D}) \propto \mathcal{L}_p(\mathbf{D}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}), \quad (3.24)$$

Nous décidons de faire fi de la vraisemblance des données i.e.  $\mathcal{L}_p(\mathbf{D}|\boldsymbol{\alpha}) = 1$ , ce qui équivaut à construire un modèle qui ignore complètement les données observées et se base entièrement sur les informations des loi à priori.

Après avoir implémenté le modèle à travers le solveur, nous obtenons nos échantillons de la loi à posteriori basés que sur les spécifications de la loi à priori.

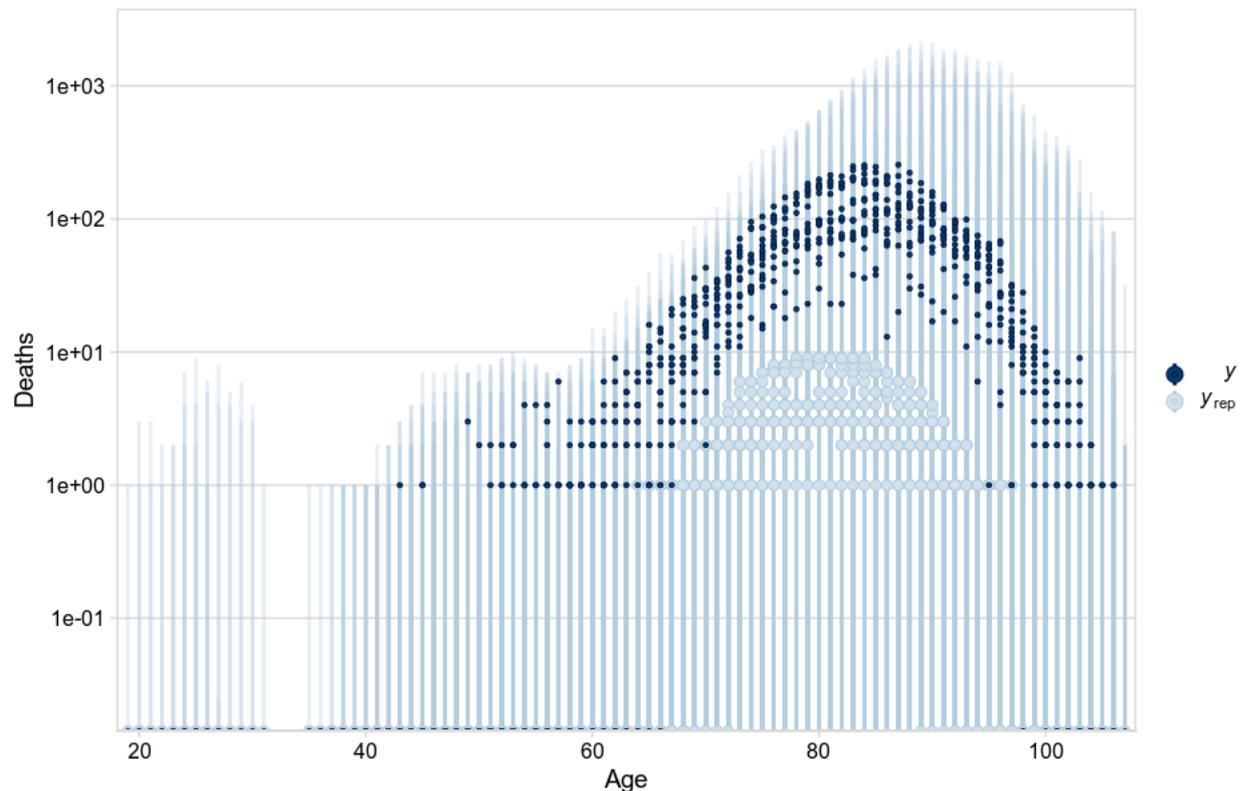


FIG. 3.1 : Prior predictive check

Ce graphe représente le nombre de décès en fonction de l'âge et du sexe. Les points en bleu ciel sont liés au modèle prior et les points en bleu pur représentent les données réellement observées.

Sachant que le modèle a été construit sans avoir vu les données, la qualité d'ajustement n'est évidemment pas bonne, mais le point le plus important à propos de ce graphe est l'ensemble des incertitudes autour de chaque point du "prior" qui est représenté par les lignes bleu ciel. En effet, cette incertitude est assez grande pour contenir la plus grande partie de nos données comme on peut le voir. Ce qui implique que l'espace d'incertitude des lois à priori est assez grand pour ne pas contraindre l'espace de la loi à posteriori.

### 3.7.3 Lois à posteriori des paramètres

Après avoir choisi nos lois à priori, la prochaine étape est déterminer les lois à posteriori.

Dans notre cas, en reprenant toutes les informations précédentes sur l'équation du modèle, les paramètres à déterminer et les lois à priori, la loi jointe de l'ensemble des paramètres  $\alpha = (\beta_0, \beta_g, \beta)$  où  $\beta = (\beta_1^F, \dots, \beta_{10}^F, \beta_1^M, \dots, \beta_{10}^M)$  :

$$p(\alpha, \tau^2 | \mathbf{D}) \propto \mathcal{L}(\mathbf{D} | \alpha) p(\beta_0) p(\beta_g) p(\beta | \tau^2) p(\tau^2) \quad (3.25)$$

$$\propto \mathcal{L}(\mathbf{D} | \alpha) p(\beta_0) p(\beta_g) \mathbf{exp}\left(-\frac{1}{2\tau^2} \beta^T S \beta\right) (\tau^2)^{-a-1} \mathbf{exp}\left(-\frac{b}{\tau^2}\right) \quad (3.26)$$

Comme nous l'avons déjà dit, la détermination explicite de la loi à posteriori des paramètres est très difficile voire impossible. Ainsi la solution préconisée est l'estimation

numérique grâce à l'une des méthodes MCMC que nous avons vues qui permet de produire des échantillons de la loi à posteriori sur lesquels nous basons toutes les inférences.

### 3.7.4 Estimation des paramètres du modèle par la méthode MCMC

La méthode MCMC utilisée pour calibrer les paramètres du modèle est la méthode algorithmique de Metropolis-Hastings.

Comme nous l'avons brièvement vu dans la section 2.3.3, les algorithmes de Metropolis-Hastings produisent des chaînes de Markov dont la distribution stationnaire est précisément celle à partir de laquelle les échantillons doivent être tirés. Ces algorithmes sont basés sur une chaîne de Markov dont la dépendance vis-à-vis de l'état précédent est divisée en deux parties : une proposition et une acceptation de la proposition. Les propositions suggèrent une prochaine étape arbitraire dans la trajectoire de la chaîne et l'acceptation assure que la direction limite appropriée est maintenue en rejetant les mouvements non désirés de la chaîne.

#### Algorithmes de Metropolis-Hastings des paramètres

Sachant que notre modèle peut se mettre sous la forme d'un GLM, nous utilisons une des méthodes de résolution d'un GLM que nous combinons à la méthode MCMC.

L'idée de base est de combiner le score de Fisher ou la méthode *Iteratively weighted least squares* (IWLS) qui est une méthode utilisée pour l'estimation des paramètres de régression dans les modèles linéaires généralisés (FAHRMEIR et al. 1994) et l'algorithme de Metropolis-Hastings.

Plus précisément, l'objectif est d'approximer les lois conditionnelles des paramètres  $\beta_j^g$  et  $\tau^2$  par une distribution gaussienne, obtenue en effectuant un scoring de Fisher à chaque itération de l'échantillonnage.

On suppose que nous sommes à une itération  $t$  de la chaîne telle que la valeur actuelle des paramètres du modèle est  $\alpha_c^{(t)}$  et que nous voulons estimer la valeur des paramètres de l'itération suivante, soit,  $\alpha_c^{(t+1)}$ . On note  $\eta_c^{(t)} = \exp\left(\mathbf{X}\alpha_c^{(t)}\right)$  qui est le prédicteur de notre modèle sur les valeurs à l'itération  $t$ . On pose aussi  $\mu^{(t)} = \mathbf{X}\alpha_c^{(t)}$  où  $\mathbf{X}$  est la matrice du modèle.

Ainsi, en se basant sur l'IWLS, une nouvelle proposition  $\alpha_p^{(t)}$  est extraite de la loi de proposition  $q(\alpha_p^{(t)}, \alpha_c^{(t)})$  qui est un loi normale multidimensionnelle de paramètres,  $m$  le vecteur moyenne et  $P$  la matrice de covariance :

$$P(\alpha_c^{(t)}) = \mathbf{X}^T W(\alpha_c^{(t)}) \mathbf{X} + \frac{1}{\tau^2} S \text{ et } m(\alpha_c^{(t)}) = P^{-1} \mathbf{X}^T W(\alpha_c^{(t)}) \tilde{y}(\alpha_c^{(t)}) \quad (3.27)$$

avec :

$$W(\alpha_c^{(t)}) = \mathbf{diag}(w_1, \dots, w_N),$$

$$\tilde{y}(\alpha_c^{(t)}) = (y - \mu^{(t)}) g'(\mu^{(t)}) + \eta_c^{(t)}$$

où  $w_n^{-1} = (g'(\mu_n^{(t)}))^2$ ,  $g$  est la fonction lien **log** et  $y$  est le taux de mortalité instantané réellement observé.

La méthode d'échantillonnage de Metropolis-Hastings requiert une probabilité d'acceptation du nouvel état proposé. Dans notre cas cette probabilité est telle que :

$$\alpha(\boldsymbol{\alpha}_p^{(t)}, \boldsymbol{\alpha}_c^{(t)}) = \min \left\{ 1, \frac{\mathcal{L}(\mathbf{D}|\boldsymbol{\alpha}_p)p(\boldsymbol{\beta}_p|\tau^2)q(\boldsymbol{\alpha}_p^{(t)}, \boldsymbol{\alpha}_c^{(t)})}{\mathcal{L}(\mathbf{D}|\boldsymbol{\alpha}_c)p(\boldsymbol{\beta}_c|\tau^2)q(\boldsymbol{\alpha}_c^{(t)}, \boldsymbol{\alpha}_p^{(t)})} \right\}$$

A chaque itération, il s'impose de recalculer les valeurs des paramètres de la loi normale multivariée  $P(\boldsymbol{\alpha}_c^{(t)})$  et  $m(\boldsymbol{\alpha}_c^{(t)})$ .

Pour ce qui est des autres paramètres  $\beta_0$  et  $\beta_g$ , ils sont estimés utilisant la même technique que celle décrite ci-dessus.

#### Algorithme de résolution

1. *Initialisation*  $t = 0$  :

Initialiser tous les paramètres  $\boldsymbol{\alpha}^{(0)}$  à partir de leur loi à priori et l'hyperparamètre  $\tau^2$  grâce à sa loi Inverse-gamma avec les paramètres présentés dans la section 3.7.2.

2. *Pour*  $t = 1, \dots, N_{samples}$  *mettre à jour l'état*  $\boldsymbol{\beta}^{(t+1)}$  :

- Calculer la vraisemblance  $\mathcal{L}(\mathbf{D}|\boldsymbol{\alpha}_c^{(t)})$  où  $\boldsymbol{\alpha}_c^{(t)} = (\beta_0^{(t)}, \beta_g^{(t)}, \boldsymbol{\beta}_c^{(t)})$
- Calculer le prédicteur du modèle  $\eta_c^{(t)} = \mathbf{exp}\left(\mathbf{X}\boldsymbol{\alpha}_c^{(t)}\right)$
- Tirer aléatoirement un vecteur de proposition  $\boldsymbol{\beta}_p^{(t)}$  de la loi normale mutidimensionnelle  $q(\boldsymbol{\alpha}_p^{(t)}, \boldsymbol{\alpha}_c^{(t)})$  dont les paramètres sont  $m(\boldsymbol{\alpha}_c^{(t)})$  qui est le vecteur de moyenne et  $P(\boldsymbol{\alpha}_c^{(t)})$  qui la matrice de variance-covariance.
- Calculer la vraisemblance du vecteur de proposition  $\mathcal{L}(\mathbf{D}|\boldsymbol{\alpha}_p^{(t)})$  où  $\boldsymbol{\alpha}_p^{(t)} = (\beta_0^{(t)}, \beta_g^{(t)}, \boldsymbol{\beta}_p^{(t)})$
- Calculer la probabilité d'acceptation  $\alpha(\boldsymbol{\alpha}_p^{(t)}, \boldsymbol{\alpha}_c^{(t)})$
- Accepter  $\boldsymbol{\beta}_p^{(t)}$  avec la probabilité  $\alpha(\boldsymbol{\alpha}_p^{(t)}, \boldsymbol{\alpha}_c^{(t)})$  et si c'est le cas, on pose  $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}_p^{(t)}$  dans le cas contraire  $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}_c^{(t)}$ .

3. *Mettre à jour*  $\beta_0^{(t+1)}, \beta_g^{(t+1)}$  :

Utiliser le même processus 2. pour calculer les valeurs de ces paramètres à l'itération  $t + 1$ .

4. *Mettre à jour le paramètre de la variance*  $\tau^2$  :

Tirer une valeur du paramètre à partir de la loi inverse-gamma dont les paramètres sont connus et donnés dans la section 3.7.2.

### 3.8 Analyse de la convergence de la méthode d'échantillonnage

Après avoir défini les éléments nécessaires pour le calibrage, nous mettons en place la méthode présentée ci-dessus. Le calcul se fait directement dans le logiciel Rstudio qui fait appel au logiciel solveur STAN grâce à des packages.

Nous utilisons 4 chaînes indépendantes et s'exécutant parallèlement qui contiennent chacune 2000 échantillons dont la moitié sert de "warm-up" durant laquelle le solveur fait tourner l'algorithme dans le but de bien caractériser la loi à posteriori et avoir une version stable de l'échantillonnage, une fois fait, ces échantillons sont écartés. Les échantillons retenus sont les 1000 autres par chaîne. Ce qui nous fait un total de 4000 échantillons, qui est un nombre suffisant pour baser les inférences.

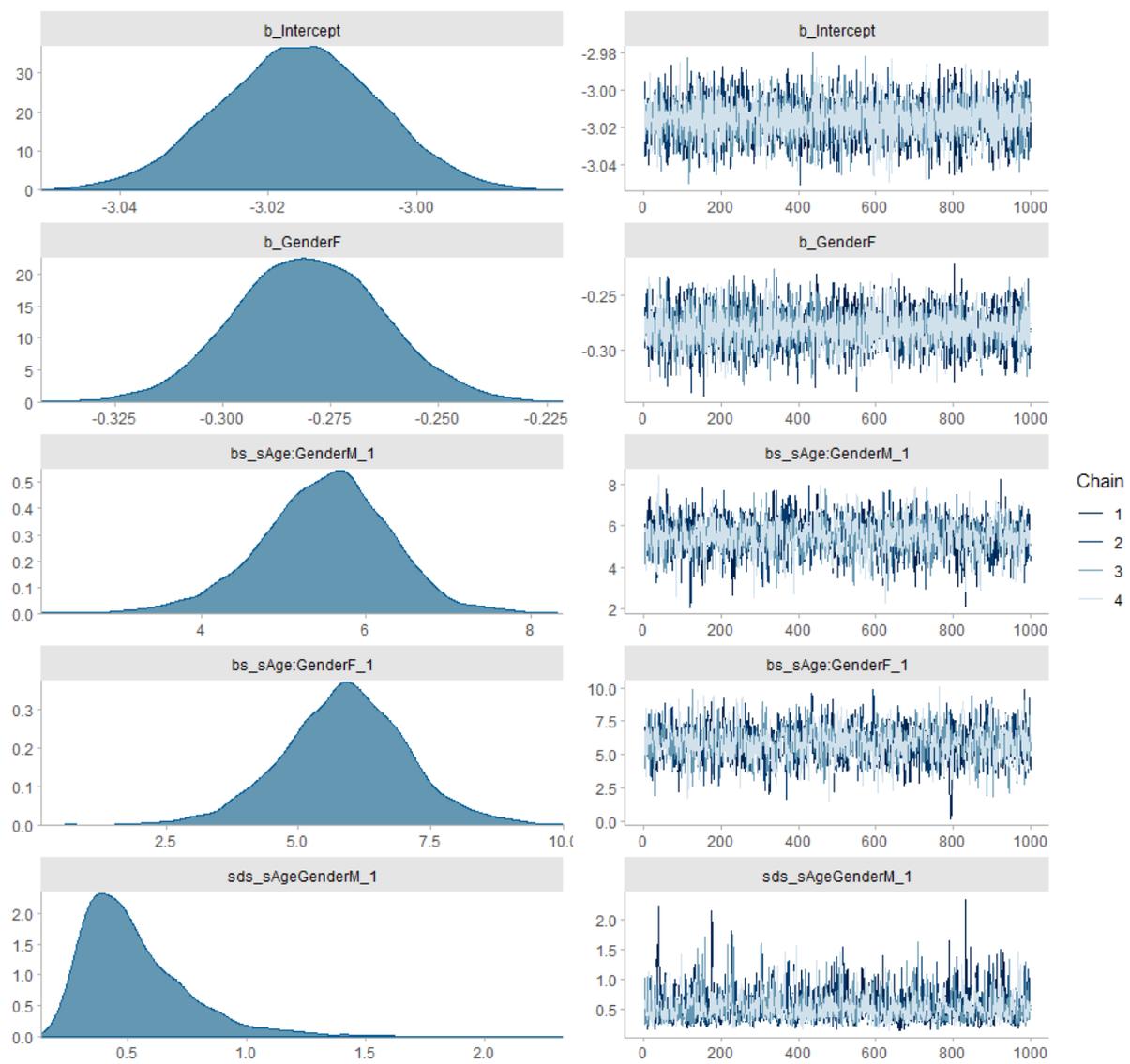


FIG. 3.2 : Densité et itérations des paramètres du modèle

Les graphes ci dessus représentent les itérations de l'algorithme MCMC pour chaque

### Chapitre 3. Mise en oeuvre du modèle GAM Bayésien sur un traité de swap de longévité

---

paramètre du modèle à droite et la densité de la chaîne de Markov utilisée pour estimer la loi du paramètre à gauche. Ces graphes indiquent que les chaînes ont convergé vers la véritable la distribution à postériori et se sont bien mélangées. Ceci est visible dans les partie droite de la figure. Si les chaînes ont convergé et se sont bien mélangées, alors ce graphique ressemble à une "grosse chenille poilue". Les figures de gauches représentent les densités des lois de chaque paramètre, ce qui nous donne une idée graphique sur les paramètres et le type des lois à postériori.

De plus un autre indicateur de la convergence des chaînes est le  $\hat{R}$ . Ce indicateur permet de vérifier que les paramètres ont convergé à un degré acceptable.

Après avoir calibré le modèle, il s'agit de vérifier que  $\hat{R} < 1.01$  comme le préconise les auteurs VEHTARI et al. 2021.

Dans notre cas, nous avons les resultats suivants :

```
Population-Level Effects:
      Estimate Est.Error 1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept      -3.02      0.01   -3.04   -3.00 1.00    6288    2750
GenderF        -0.28      0.02   -0.31   -0.25 1.00    4939    2779
sAge:GenderM_1  5.50      0.79    3.80    6.97 1.00    1835    2168
sAge:GenderF_1  5.87      1.19    3.46    8.25 1.00    1959    2052

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

FIG. 3.3 :  $\hat{R}$  du modèle

Comme on peut le voir, les  $\hat{R}$  du modèle sont tous égaux à 1, ce qui veut dire que notre calibrage a été correctement exécuté.

### 3.9 Premières analyses des résultats : Post predictive check

Bien avant d'analyser les différents ajustements produits par le modèle bayésien, il est essentiel de vérifier la façon dont le modèle s'adapte aux données observées ce qui fait l'objet de cette partie. Celle-ci se pose comme un complément de la partie prior predictive model. En effet, cette fois-ci nous construisons le modèle avec la vraisemblance des données. Le graphe ci-dessous est le même dans la section prior prédictive check à la différence que la vraisemblance des données est utilisée.

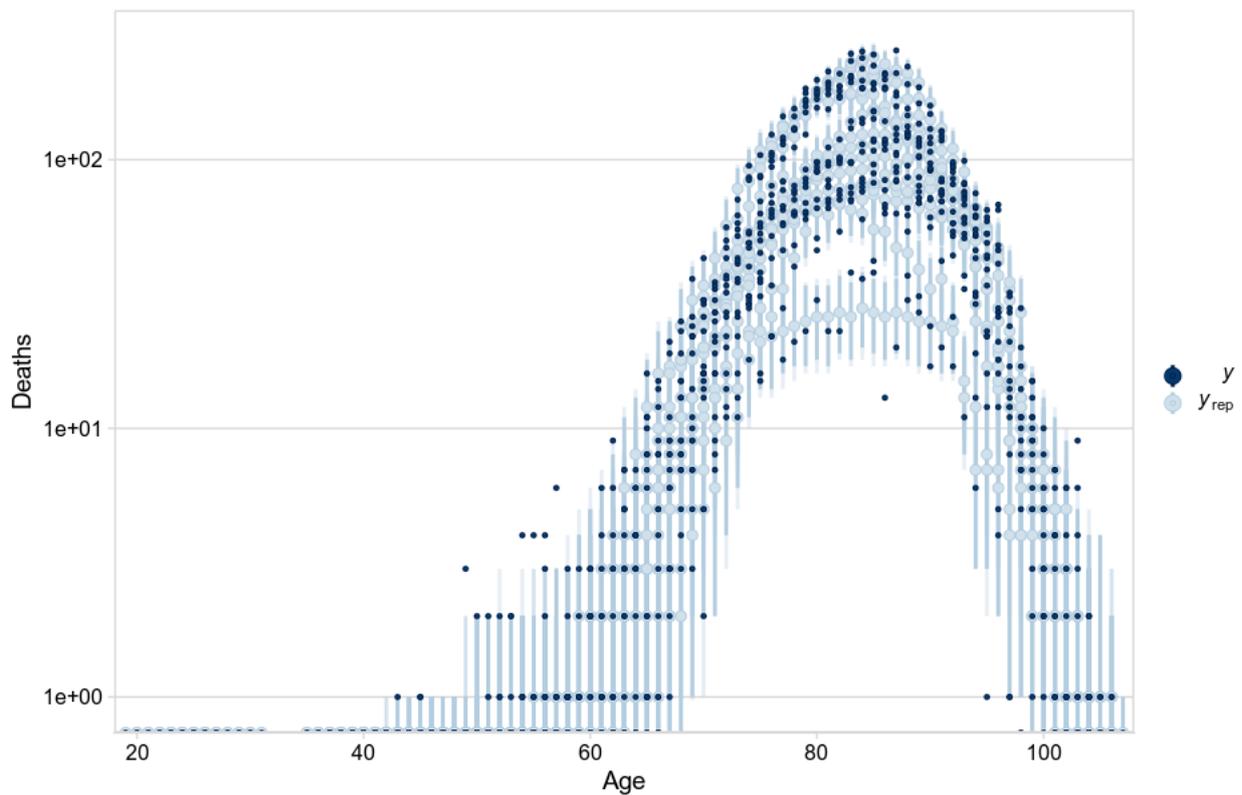


FIG. 3.4 : Prosterior predictive check

Sur ce nouveau graphe, la plupart des points bleu ciel et pur sont confondus, ce qui implique que le modèle à su s'adapter aux données. Nous pouvons aussi constater la réduction de l'incertitude autour des points estimés dépendant de la quantité d'exposition de chaque âge.

Le graphe ci-dessous représente la densité du nombre total de décès issue de la loi à postériori en bleu et le nombre total de décès observés par la ligne orange.

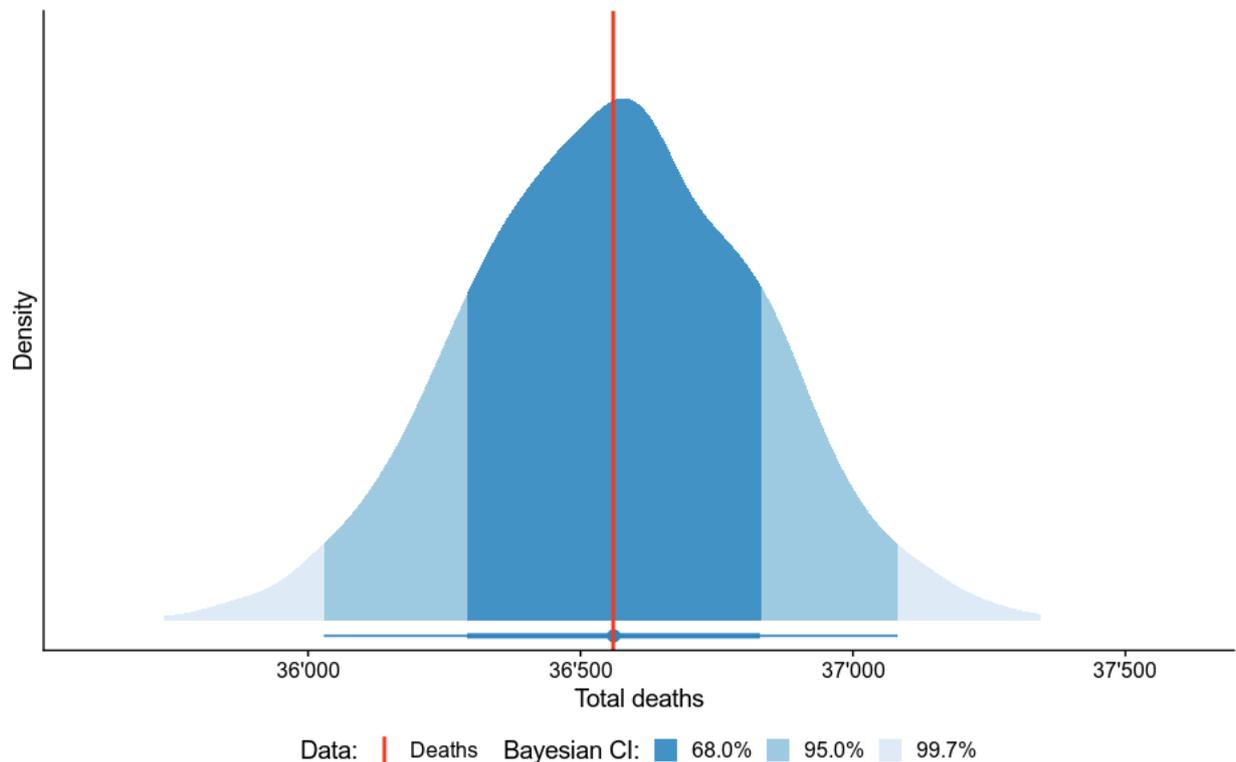


FIG. 3.5 : Posterior predictive check 2

Le point bleu au bas du graphe est la médiane de cette distribution avec des intervalles de confiance qui s'étendent sur l'axe des abscisses. Il est remarquable de voir que la médiane correspond presque exactement au nombre total de décès observés (la ligne orange passe par le point). Ce qui nous montre encore une fois la qualité d'ajustement du modèle face aux données.

### 3.10 Fermeture de table

Comme nous avons pu le voir dans la section Validation des données et dans les sections précédentes, les tranches d'âge jeunes et au delà de 95 ans n'ont pas suffisamment de données pour créer une crédibilité conséquente autour de l'estimation de nos mortalités même si l'estimation bayésienne est plus adaptée dans ces cas de figure.

Nous décidons de modéliser les mortalités sur ces tranches d'âge de façon différente.

La question de la fermeture de la table est importante dans le cas de la construction d'une table pour des provisionnements de rentes viagères comme dans notre cas.

Cette importance est moindre vu que nous avons très peu d'expositions sur ces tranches d'âge, la fermeture de table reste tout de même indispensable dans le cas où nous voulons avoir des tables du moment complètes.

### 3.10.1 Traitement des âges jeunes

Pour estimer les mortalités sur les tranches jeunes i.e [19, 64], nous appliquons une méthode similaire à celle qu'utilise l'équipe à savoir celle d'appliquer un abattement sur des tables réglementaires de référence sur les tranches d'âge jeunes.

Les tables de référence utilisées sont des tables de mortalité transversales produites par le CMI ce qui est cohérent avec la localisation de la population de notre portefeuille (Grande-Bretagne). Ces tables sont équivalentes aux tables de référence TH 00-02 et TF 00-02 de l'INSEE.

En rappelant le fonctionnement de la méthode de rapprochement décrite dans la section 1.5.1, le ratio SMR pour un intervalle d'âge  $I$  est tel que :

$$A/E_I^{montant} = \frac{\sum_{x \in I} M}{\sum_{x \in I} E_x^c * A_x * m_x^{ref}} \quad (3.28)$$

Où  $M$  représente le coût des sinistres réellement payés au sein de la population d'âge  $x$ ,  $A_x$  représente la valeur contractuelle de l'annuité de l'individu.

L'intervalle de confiance 95% autour ce ratio est :

$$I_{A/E} = [A/E^{montant} - 1.96 * \sigma^{montant}; A/E^{montant} + 1.96 * \sigma^{montant}] \quad (3.29)$$

où

$$\sigma^{montant} = \sqrt{A/E^{montant} \times \frac{Expected_{SQ}}{Expected^2}}$$

avec :

$$Expected = \sum_{x \in I} E_x^c * A_x * m_x^{ref}$$

et

$$Expected_{SQ} = \sum_{x \in I} E_x^c * A_x^2 * m_x^{ref}$$

Nous décidons d'agréger tous les âges jeunes en une seule tranche [19, 64] de sorte à avoir un seul ratio qui lui est assigné ainsi que son intervalle de confiance à 95%.

Par souci de prudence, il est naturel de choisir la borne supérieure de l'intervalle de confiance comme abattement sur les taux de référence. L'objectif étant d'avoir des taux supérieurs à ceux de la table de référence.

Ainsi nous aurons :

$$m_x = \left[ A/E^{montant} + 1.96 \sqrt{A/E^{montant} \times \frac{Expected_{SQ}}{Expected^2}} \right] m_x^{ref} \quad (3.30)$$

Nous appliquons cette formule respectivement sur les taux bruts de mortalités masculins et féminins.

### 3.10.2 Traitement des âges élevés

Pour les âges élevés, contrairement aux âges jeunes, nous ne disposons d'aucune statistique à partir de 107 ans, il n'est donc pas possible de faire un ajustement par une table de référence.

L'approche utilisée est une extrapolation, nous décidons de prolonger le taux de décès par les méthodes d'extrapolation aux grands âges.

La modélisation des taux aux âges élevés est nécessaire. En effet, les grands âges représentent l'enjeu du risque de longévité donc ils jouent un rôle important dans le calcul de provision des rentes.

Plusieurs méthodes existent mais dans notre cas, nous utilisons la méthode d'extrapolation de **Coale et Kisker** (QUASHIE et al. 2005).

La méthode consiste à extrapoler les taux de mortalité aux grands âges (jusqu'à  $x = 110$  ans par exemple) en se basant sur la formule :

$$\hat{\mu}_x = \hat{\mu}_{65} \times e^{g_x(x-65)} \quad (3.31)$$

où  $g_x$  désigne le taux moyen de croissance de  $\hat{\mu}_x$  entre 65 et  $x$  ans.

On calcule ainsi les coefficients  $g_x$  jusqu'à un certain âge, puis nous les extrapolons afin de pouvoir recomposer les taux  $\hat{\mu}_x$ . Coale et Kisker ont en effet remarqué empiriquement que les courbes des  $g_x$  possèdent en général un pic aux alentours de 80 ans avant de décroître linéairement. Ils ont par conséquent proposé l'équation :

$$g_x = g_{80} + s * (x - 80) , x \geq 80. \quad (3.32)$$

Finalement, on peut utiliser la formule suivante pour extrapoler au-delà de 80 ans les taux instantanés de mortalité :

$$\hat{\mu}_x = \hat{\mu}_{x-1} \times e^{g_{80} + s*(x-80)} , x \geq 80.$$

On utilise les valeurs de paramètres suivantes :

$$s = -\frac{\ln(\hat{\mu}_{79} + 31 \times g_{80})}{465} \text{ et } g_{80} = \frac{\ln\left(\frac{\hat{\mu}_{80}}{\hat{\mu}_{65}}\right)}{15}.$$

Ainsi nous appliquons cette méthode respectivement sur les taux de mortalités masculins et féminins.

## Chapitre 4

### Méthodes d'analyse et de validation des tables du moment

Les critères de validation de modèle fournissent des aides à la décision dans le cadre de la sélection du modèle le plus pertinent. Après la construction des tables d'expérience, nous nous devons de vérifier leur cohérence et sélectionner la méthode la plus robuste s'adaptant à notre portefeuille d'étude. Pour ce faire, nous commençons par des analyses graphiques des forces de mortalité et ensuite mettre en place des tests quantitatifs.

### 4.1 Analyses graphiques des forces de mortalité

Après avoir vérifié la fiabilité et la crédibilité des lois à postériori, cette partie permettra d'analyser graphiquement les résultats des différents modèles.

Lors de la validation de nos données, nous avons décidé de modéliser autrement les mortalités des âges jeunes et très élevés, cependant, il est intéressant de voir comment se comporte la calibration bayésien lorsque les données sont insuffisantes.

Dans notre analyse, nous avons à notre disposition 3 modélisations de la force de mortalité :

- Les forces de mortalité issues de la méthode actuelle utilisée par l'équipe : *Last year mu*;
- Notre modèle 3.10 calibrée avec une estimation fréquentiste : *Poisson frequentist GAM*;
- Enfin, notre modèle 3.10 faisant appel à l'estimation bayésienne : *BGM*.

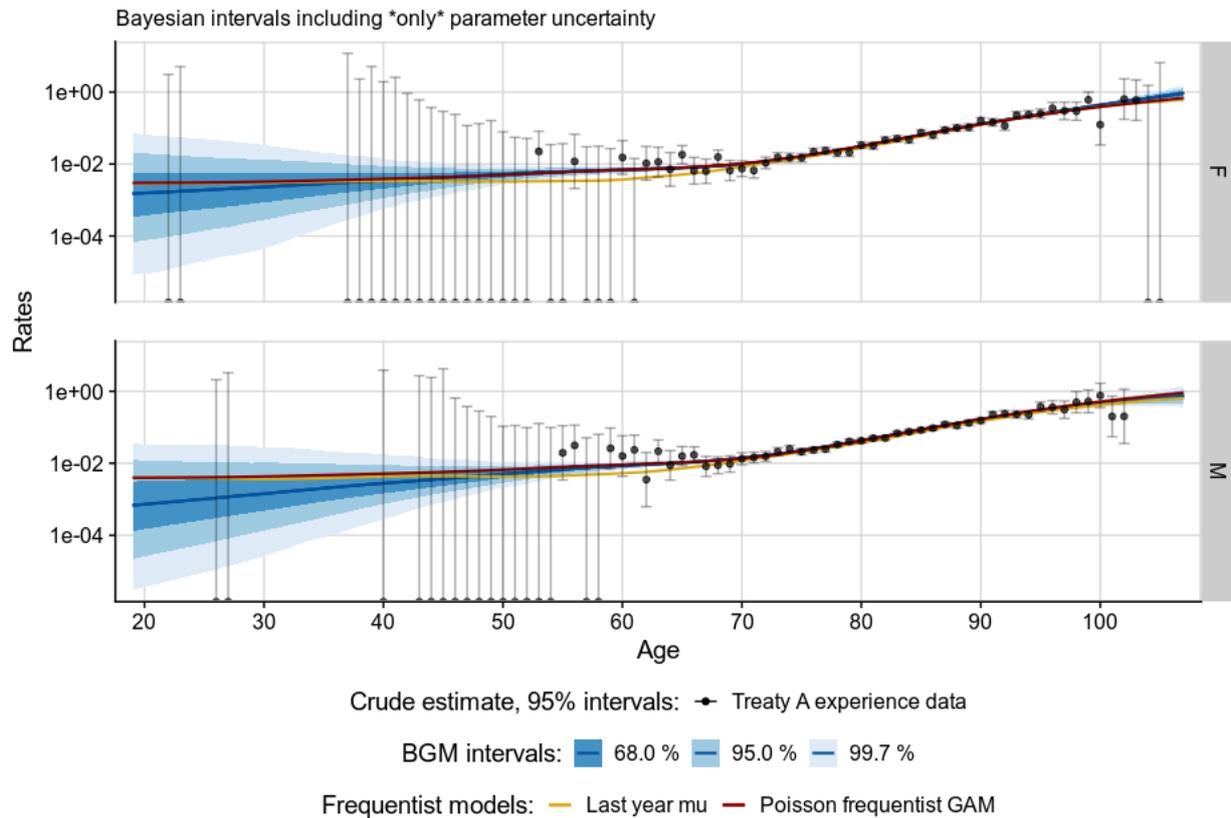


FIG. 4.1 : Représentation des forces de mortalité pour tout âge

Les points en noir désignent les taux bruts des données réellement observées avec un intervalle de confiance issu d'une loi de poisson à 95%. Le graphe ci-dessus, nous montre clairement que les 3 modèles sont très proches des données de façon globale, cette proximité est d'autant plus grande pour les âges de référence où la crédibilité est plus marquée.

La crédibilité autour du modèle bayésien sont des intervalles de confiance issue de la loi à postériori des paramètres à différents niveaux 68%, 95% et 99.7%. La façon dont sont construits les intervalles de confiance en se basant sur la loi à postériori est présentée dans la section 2.3.4.

Comme on pourrait le voir, les intervalles de confiance s'élargissent au fur et à mesure que nous nous éloignons de notre tranche d'âge de référence soit [65, 96].

Néanmoins, il est intéressant de voir comment notre BGM se comporte aux jeunes âges, les mortalités estimées sont nettement inférieures à celles des deux autres modèles. Cela est plus visible sur les mortalités des populations masculines.

Ce faible taux de mortalité répond quelque part au critère de prudence que nous avons mentionné dans la section de fermeture de table pour les âges jeunes. Ce qui implique que nous aurions pu utiliser les taux de mortalité que nous propose le modèle BGM même pour les âges aux nous avons très peu d'expositions.

C'est l'une des avantages clés du modèle bayésien qui a tendance à utiliser pleinement toutes les informations contenues dans les données sans les extrapoler et de proposer des crédibilités plus ou moins grandes dépendant de la quantité de ces données.

## 4.2 Les graphes Gompertz log différence (Gold Chart)

Nous pouvons voir sur le graphe de la section précédente que les ajustements et les taux bruts ont tous une forme largement log-linéaire ce qui les rend difficiles à distinguer notamment pour les âges de référence. Les graphiques conventionnels des taux de mortalité sur une échelle linéaire ou logarithmique peuvent rendre difficile la distinction entre les taux bruts des données et les taux provenant des modèles.

Pour mieux visualiser les différences entre les modèles, une notion importante entre en jeu, à savoir les graphes “Gompertz log différence”.

Au lieu de représenter directement les taux de mortalités non ajustés, nous représentons graphiquement  $\log(\mu_x) - \log(g(x))$  où  $g(x)$  est une estimation du taux de mortalité provenant du modèle de Gompertz qui est de la forme :  $\log(g(x)) = A + Bx$ .

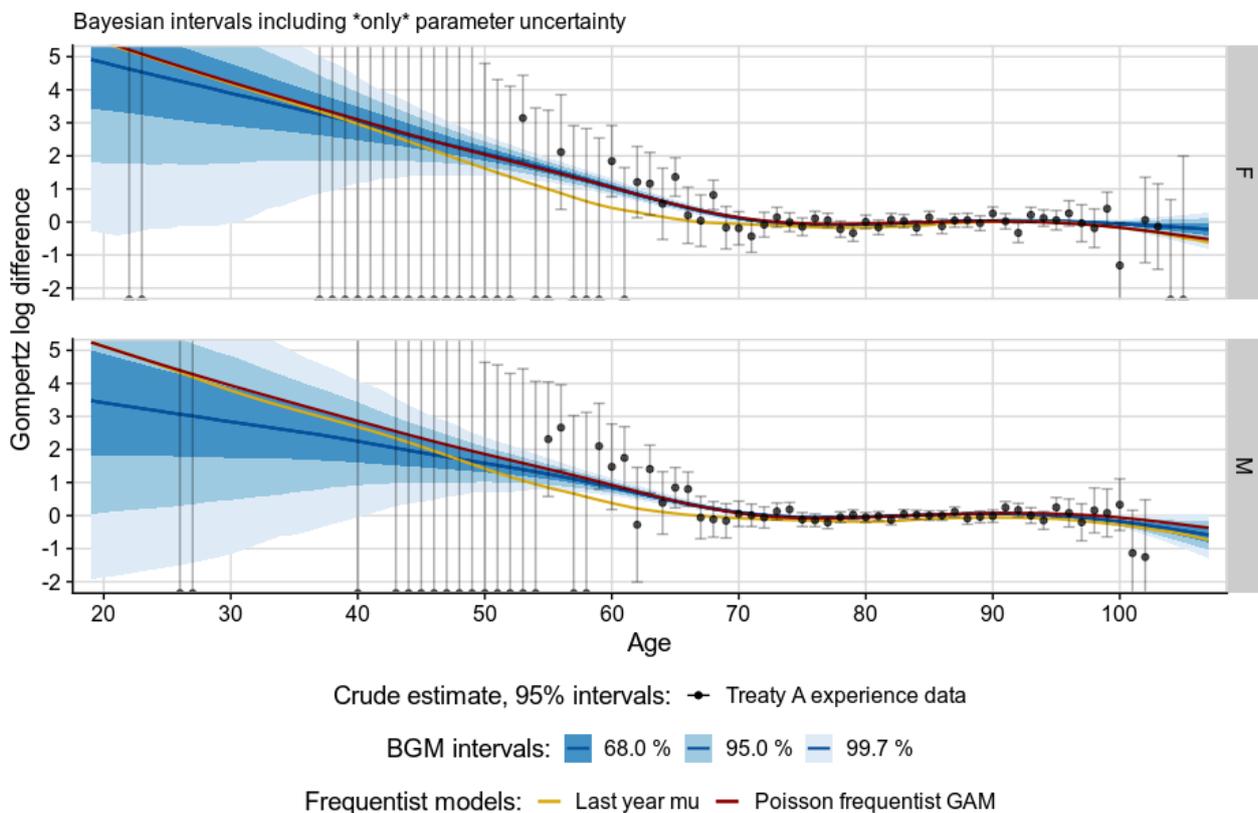


FIG. 4.2 : Graphes Gompertz log différence pour tout âge

Ce graphe ci-dessus représente exactement les mêmes données que sur le graphe 4.1 mais avec des graphes Gompertz log différence. Nous pouvons maintenant clairement apercevoir la différence nette entre les ajustements et les taux bruts mais surtout les différences entre les ajustements des 3 modèles.

La visualisation de ce graphe permet de confirmer nos analyses de la section précédente. En effet, nous constatons que les deux modèles GAMs sont très proches l'un de l'autre parmi les âge de référence, la différence est remarquable pour les populations jeunes (moins de 50 ans) et très vieilles (plus de 100 ans), où l'on a moins de données.

Quant au modèle interne de l'entreprise, il décrit également assez bien les données de ce qu'on voit sur les graphes ci-dessous pour les âges compris entre 70 et 100 ans. Cependant, avant cet intervalle, notamment entre 48 et 70 ans, le modèle est tiré vers le bas par une valeur très basse à l'âge 62, il ne reste pas moyennement fidèle aux données à ce niveau surtout quand il s'agit des mortalités des populations féminines.

Pour plus creuser dans la recherche de différence graphique entre les modèles, nous zoomons les graphes précédents de sorte à avoir un bon aperçu sur les âges de référence. Avec ces graphes zoomés, nous voyons mieux les capacités d'ajustement des différents modèles pour les âge de référence.

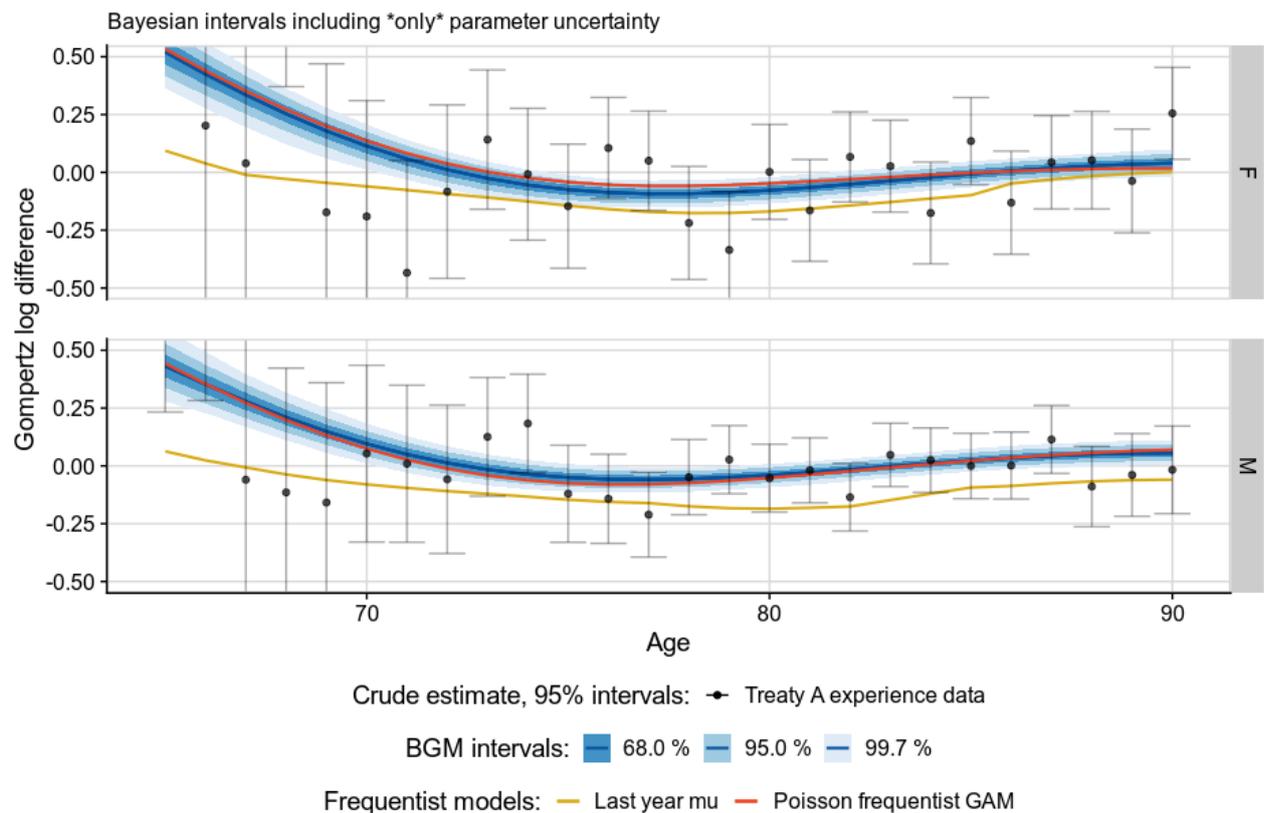


FIG. 4.3 : Graphes Gompertz log différence pour les âges de référence

Ce graphe zoomé nous permet de mieux voir les différences entre nos 3 modèles. Encore une fois, nous remarquons que les deux modèles GAM (Fréquentiste et Bayésien) sont presque confondus sur cette tranche d'âge de référence. Par contre, le modèle actuellement utilisé par l'équipe a tendance à être influencé par les valeurs presque extrêmes des mortalités. Cette remarque est plus visible sur le graphe des populations masculines.

### 4.3 La comparaison des valeurs modélisées et des observations : SMR ou ratio A/E

Les ratios A/E ou SMR peuvent servir d'ajustement des taux de mortalité comme le fait l'équipe. Dans cette partie nous présenterons différentes analyses A/E pas dans un but d'ajustement mais pour mieux analyser les mortalités prédites par nos différents modèles et voir dans quelle mesure ils se rapprochent de nos données.

Le numérateur : Actual étant une grandeur provenant des données, i.e nombre de décès réellement observées, les différences entre les modèles se feront autour de l'estimation du dénominateur : Expected i.e le nombre de décès prédit par nos différents modèles.

Cette analyse A/E est pondérée que par le nombre de décès ou de vie dans le portefeuille en contraste de l'analyse qui prend en compte le montant des annuités des différents contrats.

#### 4.3.1 Simulation du nombre décès dans le cadre bayésien

Dans notre modèle GAM Bayésien, le nombre de décès suivant une loi de Poisson avec pour espérance  $\mu_{x,g}E_{x,g}$ , sa prédiction nécessite dans un premier temps la prédiction de nos différentes forces de mortalité que nous avons déjà obtenues grâce à notre algorithme de résolution.

Pour chaque observation  $x, g$  fixée :

- Nous avons donc 4000 prédictions de  $\mu_{x,g}$ ,
- Chaque prédiction est utilisée comme paramètre dans la simulation de notre loi de Poisson comme dans l'équation 3.10
- Le nombre de décès sera au final la moyenne empirique de nos 4000 simulations.

Ce qui se résume à :

$$\widehat{D}_{x,g} = \frac{1}{N} \sum_{i=1}^N X_{x,g}^{(i)}$$

Avec

$$\text{Avec } N = 4000 \text{ et } X_{x,g}^{(i)} \sim \mathcal{P}(\widehat{\mu}_{x,g}^{(i)} E_{x,g}), i \in [1, N]$$

#### 4.3.2 Production du nombre décès dans le cadre fréquentiste

Dans le cas fréquentiste, l'obtention du nombre de décès attendu est grandement similaire à celle du cas Bayésien à la différence que nous n'avons besoin d'avoir aucune distribution d'aucune loi à posteriori. La réponse du modèle est une estimation ponctuelle et non une distribution de probabilité.

Comme présentée dans la partie (Expérience analysis), dans l'analyse A/E de l'équipe longévité de l'entreprise, l'estimation du nombre de décès attendu (Expected) se base entièrement sur le Benchmark qui est externe aux données. Le nombre de décès dans ce cas ne suit aucune loi, son estimation ressort d'une multiplication entre les différentes forces de mortalités et l'exposition au risque. On a donc :

$$\widehat{D}_{x,g} = \widehat{\mu}_{x,g} E_{x,g}$$

Par ailleurs, dans le but d'obtenir une crédibilité autour de ces estimations, nous supposons que ce nombre de décès suit une loi poisson comme pour nos deux modèles GAMs.

Disposant du nombre de décès par toutes les observations de notre base de données, nous pouvons à présent calculer les différents ratios A/E, que nous effectuerons par sexe et par âge.

Tous les graphes dans cette partie, représentent les ratios A/E, les points sont les ratios avec un intervalle de confiance à 95%.

### 4.3.3 Analyse A/E par sexe

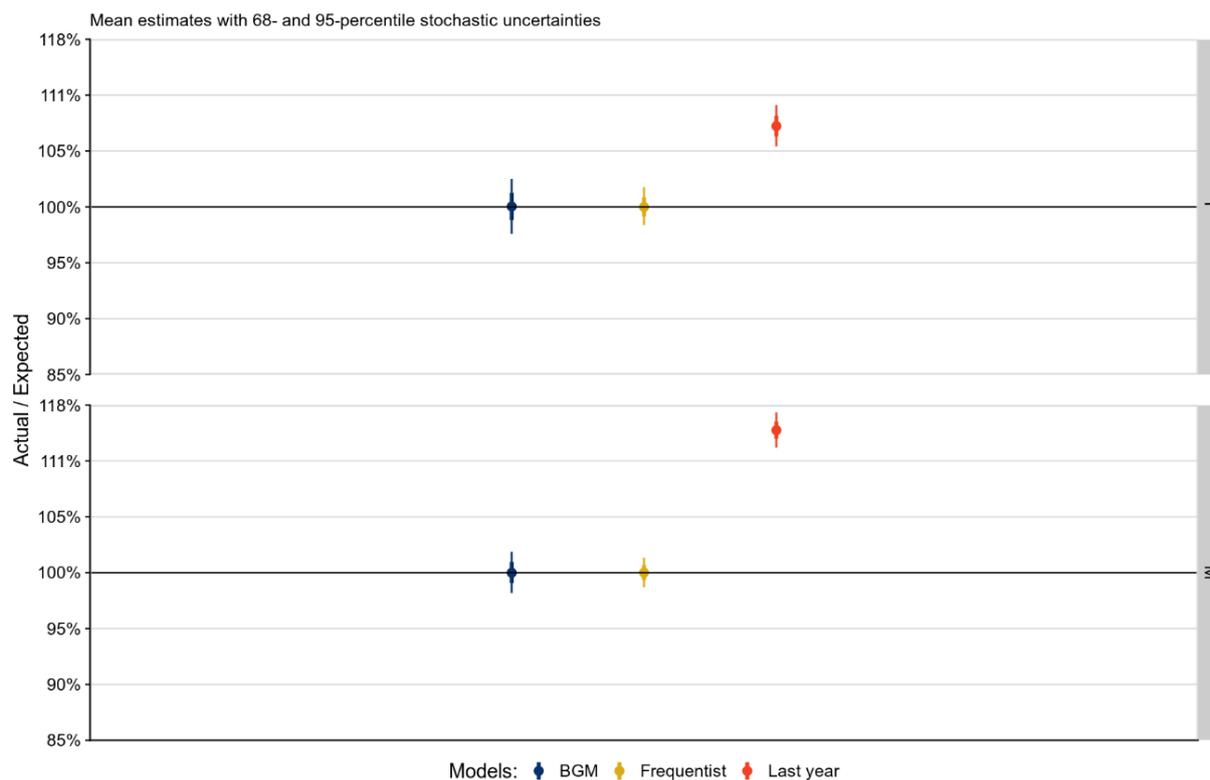


FIG. 4.4 : Ratios A/E par sexe

Ces graphes montrent une nette différence entre les modèles GAM et le utilisé utilisé par l'équipe. Les ratios des GAMS sont pratiquement à 100 % avec une crédibilité plus marquée par les hommes qui s'explique par une plus grande exposition au risque du côté masculin.

### 4.3.4 Analyse A/E par intervalle d'âge

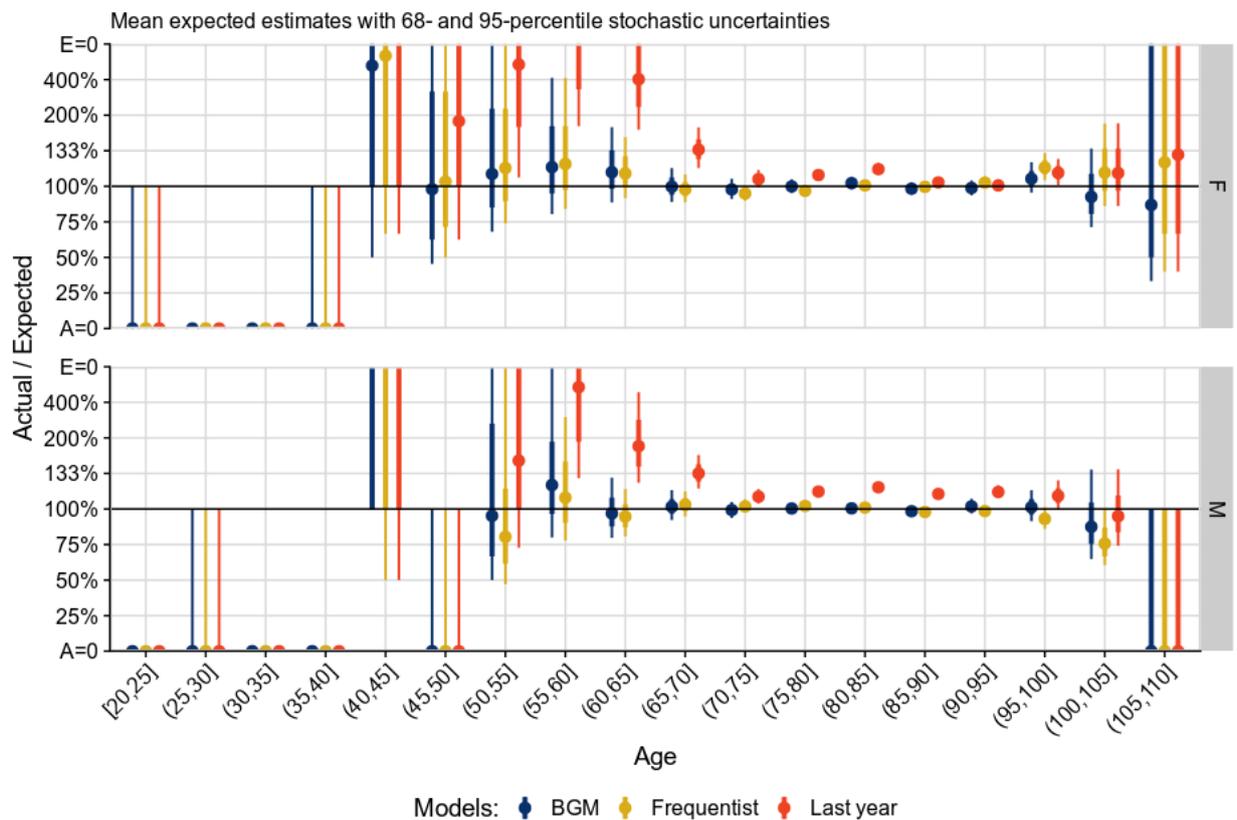


FIG. 4.5 : Ratios A/E par intervalle d'âge toutes années calendaire

Par intervalle d'âge, il est une fois de plus notable de voir la proximité des points A/E à la ligne de 100 % pour les deux GAMs par rapport au modèle utilisé par l'entreprise. Par ailleurs, ce modèle "Last Year" est mieux adapté pour les hommes que pour les femmes, en effet, à travers les âges de référence, il sous-estime clairement le nombre observé de décès, contrairement aux deux autres modèles GAMs. Cependant, nous voyons aussi un meilleur ajustement du modèle Bayésien par rapport au modèle fréquentiste. Les points A/E du GAM Bayésien restent beaucoup plus près de 100% que dans le cas fréquentiste.

### 4.3.5 Analyse A/E par année calendaire

Même si la construction des tables du moment ne tient pas compte des années calendaires, nous décidons tout de même d'analyser nos ratios par année calendaire dans le but de visualiser l'ajustement de notre modèle bayésien par année.

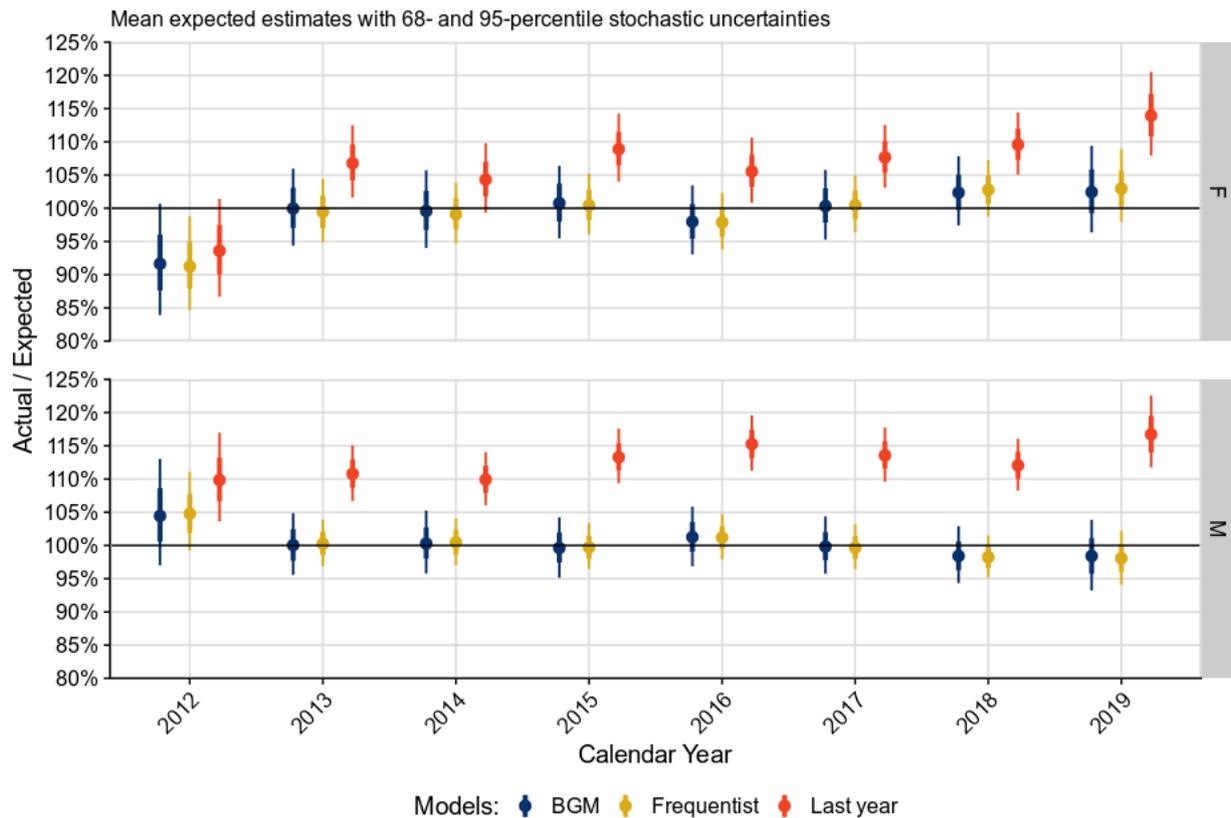


FIG. 4.6 : Ratios A/E par année calendaire

La première remarque est de constater que les graphes sont différents d'une année à une autre. Les résultats des années 2012 et 2019 sont beaucoup plus volatiles par rapport aux autres années, parce que ces années n'ont pas été complètement observées. Encore une fois, nous voyons la difficulté pour le modèle "Last Year" à décrire le nombre de décès chez les hommes.

A part l'année 2012, le modèle utilisé par l'entreprise a une forte tendance à sous estimer le nombre de décès dans le portefeuille. Il le fait encore plus chez les hommes.

#### 4.4 Le coefficient de détermination : $R^2$

Un autre indicateur important est le coefficient de détermination noté  $R^2$  qui correspond à la part de variance expliquée d'un modèle par rapport à la variance totale des données.

Nous avons pu voir dans la section 2.4.4 que ce coefficient a dû être adapté dans le cadre bayésien vu que la valeur de la formule standard peut être supérieur à 1 lorsqu'elle est appliquée à des modèles bayésiens.

Ainsi grâce à la nouvelle formule 2.9 nous calculons le coefficient pour 3 modèles afin de pouvoir les comparer.

	R2
<b>BGM</b>	0.9942487
<b>Poisson frequentist model</b>	0.9936156
<b>Last year</b>	0.82381

FIG. 4.7 :  $R^2$

Le critère des coefficients de détermination vient encore une fois confirmé les différences des qualités d'ajustement entre les 3 modèles que nous avons relevées dans les sections précédentes. Nous pouvons voir que le BGM a le  $R^2$  le plus grand et le modèle utilisé par l'équipe a le minimum.

Le modèle fréquentiste est très proche du BGM, les différences d'ajustement entre ces deux GAMs se dessinent en fonction de la quantité de données utilisées pour calibrer la régression.

# Conclusion et perspectives

### Conclusion générale

En conclusion, nous avons montré que notre modèle GAM, faisant appel aux P-splines avec une estimation bayésienne est une méthode efficace pour construire des tables de mortalité du moment dans le contexte de l'analyse d'expérience d'une portefeuille de swap de longévité. En utilisant cette méthode, nous avons pu modéliser la mortalité de manière flexible et précise pour notre portefeuille de swaps de longévité, ce qui est essentiel pour une gestion efficace des risques de longévité.

Cette combinaison entre P-splines et estimation bayésienne, permet de mieux exploiter les données dans leur entièreté. Les fonctions de lissage sont calibrées par les données qui pénalisent les coefficients pour éviter un overfitting. L'incertitude de chaque coefficient est mesurée de façon probabiliste à travers l'estimation bayésienne.

En comparant, notre modèle et la méthode utilisée par l'équipe, notre approche permet d'incorporer directement les jugements d'experts à travers les lois à priori, ce qui était jusque là fait en ad-hoc. En plus, la modélisation ne fait appel à aucune table de mortalité externe, ce qui permet de mieux refléter les expériences de notre portefeuille et de base notre table de mortalité entièrement sur les populations du portefeuille.

Par ailleurs, le modèle permet d'allier estimation de taux bruts, lissage et ajustement en une seule étape à travers les fonctions de lissage. Cette caractéristique permet d'alléger les tâches à remplir pour la mise en place d'une table de mortalité du moment.

Toutefois, notre modèle présente quelques inconvénients qui sont pertinents à relever. D'abord, la mise en place du nécessite une certaine expertise. En effet, l'estimation bayésienne exige une certaine expertise pour spécifier les distributions a priori des paramètres et hyperparamètres, ce qui peut être difficile pour les personnes qui ne sont pas familières à cette approche. En plus, les résultats de l'estimation bayésienne peuvent être plus difficiles à communiquer que les résultats des méthodes d'estimation classiques (fréquentistes), notamment autour de la crédibilité car les intervalles de confiance peuvent être difficiles à comprendre. Ensuite, un autre inconvénient est le besoin de mettre en place de grand moyens. En effet, comme nous l'avons vu tout au long de l'étude, les méthodes de calibrage sont lourdes et complexes à mettre en place d'où la nécessité des algorithmes MCMC, qui ont une longue durée de calcul dépendant du volume des données. Les P-splines, quant à elles, nécessitent un certain nombre de données pour bien caractériser la forme de relation entre l'âge et la force de mortalité estimée.

En terme de perspective, nous comptons étendre l'étude de sorte à incorporer les effets cohortes au sein des populations afin de tenir compte de l'amélioration de la mortalité au sein du portefeuille. Cet effet nécessitera l'incorporation de nouvelles variables explicatives dans le modèle comme l'année de naissance et l'année calendaire par exemple. Cette analyse permettra de ne plus modéliser la mortalité niveau et tendance séparément.

Nous gardons en tête que la mise en place de ce genre de modèle requiert une quantité très importante de données d'observation de notre portefeuille, ce qui n'est pas le cas en ce moment.

# Bibliographie

- DENUIT, Michel et Julien TRUFIN (2019). “Effective statistical learning methods for actuaries”. In.
- FAHRMEIR, Ludwig et al. (1994). *Multivariate statistical modelling based on generalized linear models*. T. 425. Springer.
- GAO, Guangyuan, GAO et ZHANG (2018). *Bayesian claims reserving methods in non-life insurance with STAN*. Springer.
- GELMAN, Andrew et al. (2019). “R-squared for Bayesian regression models”. In : *The American Statistician*.
- GREEN, Peter J et Bernard W SILVERMAN (1993). *Nonparametric regression and generalized linear models : a roughness penalty approach*. Crc Press.
- HASTIE, Trevor et al. (2009). *The elements of statistical learning : data mining, inference, and prediction*. T. 2. Springer.
- LENIAU, Morgane (2021). “Prise en compte des déclarations tardives dans l’évaluation des swpas de longévité sous Solvabilité II.” Memoire. DUAS.
- MACDONALD, Angus S, Stephen J RICHARDS et Iain D CURRIE (2018). *Modelling mortality with actuarial applications*. Cambridge University Press.
- PERNA, Cira et Marilena SIBILLO (2012). *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Springer.
- QUASHIE, Aki et Michel DENUIT (2005). “Modèles d’extrapolation de la mortalité aux grands âges”. In : *Institut des Sciences Actuarielles et Institut de Statistique, Université Catholique de Louvain, WP*.
- SAMANIEGO, Francisco J (2010a). *A comparison of the Bayesian and frequentist approaches to estimation*. T. 24. Springer.
- (2010b). *A Comparison of the Bayesian and Frequentist Approaches to estimation*. en. Springer Series in Statistics.
- TJUR, Tue (2009). “Coefficients of determination in logistic regression models—A new proposal : The coefficient of discrimination”. In : *The American Statistician* 63.4, p. 366-372.
- ULLER, MARIA. M (2019). *Computational Bayesian Statistics*. en. Cambridge.
- VEHTARI, Aki, Andrew GELMAN et Jonah GABRY (2017). “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC”. In : *Statistics and computing* 27.5, p. 1413-1432.
- VEHTARI, Aki et al. (2021). “Rank-normalization, folding, and localization : An improved  $\hat{R}$  for assessing convergence of MCMC (with discussion)”. In : *Bayesian analysis* 16.2, p. 667-718.

- WAKEFIELD, Jon (2012). *Bayesian and Frequentist Regression Methods*. en. Sous la dir. de David W. PEARSON, Nigel C. STEELE et Rudolf F. ALBRECHT. Departments of Statistics et Biostatistics University of Washington Seattle, Washington USA : Springer Series in Statistics.
- WOOD, Simon N (2006). *Generalized additive models : an introduction with R*. chapman et hall/CRC.
- ZHU, Xiaobai et Kenneth Q ZHOU (2022). “Smooth projection of mortality improvement rates : a Bayesian two-dimensional spline approach”. In : *European Actuarial Journal*, p. 1-29.

# Table des figures

1.1	Organigramme des solutions de transfert du risque de longévité . . . . .	5
1.2	Fonctionnement d'un swap de longévité . . . . .	6
1.3	Calcul de l'exposition par contrat . . . . .	8
1.4	Base de données finale . . . . .	9
1.5	Exposition au risque par âge et par sexe . . . . .	10
1.6	Tableau récapitulatif des expositions au risque par sexe . . . . .	11
1.7	Nombre de sorties par age et par sexe . . . . .	11
1.8	Tableau récapitulatif du nombre de sorties par sexe . . . . .	12
1.9	Exposition et nombre de décès par année calendaire . . . . .	12
1.10	Tableau récapitulatif des expositions et du nombre de décès par année calendaire . . . . .	13
1.11	Tableau récapitulatif des expositions et du nombre de décès par âge . . . . .	14
1.12	Présentation du procédé Actual-to-expected . . . . .	18
3.1	Prior predictive check . . . . .	44
3.2	Densité et itérations des paramètres du modèle . . . . .	47
3.3	$\hat{R}$ du modèle . . . . .	48
3.4	Prosterior predictive check . . . . .	49
3.5	Posterior predictive check 2 . . . . .	50
4.1	Représentation des forces de mortalité pour tout âge . . . . .	55
4.2	Graphes Gompertz log différence pour tout âge . . . . .	56
4.3	Graphes Gompertz log différence pour les âges de référence . . . . .	57
4.4	Ratios A/E par sexe . . . . .	59
4.5	Ratios A/E par intervalle d'âge toutes années calendaire . . . . .	60
4.6	Ratios A/E par année calendaire . . . . .	61
4.7	$\mathbf{R}^2$ . . . . .	62
A.1	Representation des forces de mortalité pour tout âge et pour l'année 2012 .	72
A.2	Representation des forces de mortalité pour tout âge et pour l'année 2013 .	73
A.3	Representation des forces de mortalité pour tout âge et pour l'année 2014 .	74
A.4	Representation des forces de mortalité pour tout âge et pour l'année 2015 .	75
A.5	Representation des forces de mortalité pour tout âge et pour l'année 2016 .	75
A.6	Representation des forces de mortalité pour tout âge et pour l'année 2017 .	76
A.7	Representation des forces de mortalité pour tout âge et pour l'année 2018 .	77
A.8	Representation des forces de mortalité pour tout âge et pour l'année 2019 .	77
A.9	Graphes Gold pour tout âge et pour l'année 2012 . . . . .	78
A.10	Graphes Gold pour tout âge et pour l'année 2013 . . . . .	78
A.11	Graphes Gold pour tout âge et pour l'année 2014 . . . . .	79

## Table des figures

---

A.12 Graphes Gold pour tout âge et pour l'année 2015 . . . . .	79
A.13 Graphes Gold pour tout âge et pour l'année 2016 . . . . .	80
A.14 Graphes Gold pour tout âge et pour l'année 2017 . . . . .	80
A.15 Graphes Gold pour tout âge et pour l'année 2018 . . . . .	81
A.16 Graphes Gold pour tout âge et pour l'année 2019 . . . . .	81
A.17 Ratios A/E par année calendaire pour l'intervalle d'âge 60-70 . . . . .	82
A.18 Ratios A/E par année calendaire pour l'intervalle d'âge 70-80 . . . . .	83
A.19 Ratios A/E par année calendaire pour l'intervalle d'âge 80-90 . . . . .	83
A.20 Ratios A/E par année calendaire pour l'intervalle d'âge 90+ . . . . .	84

# Liste des tableaux

1.1	Présentation du traité étudié. . . . .	7
-----	--	---

# Liste des algorithmes

# Annexes

# Annexe A

## Presentation de tous les resultats graphiques

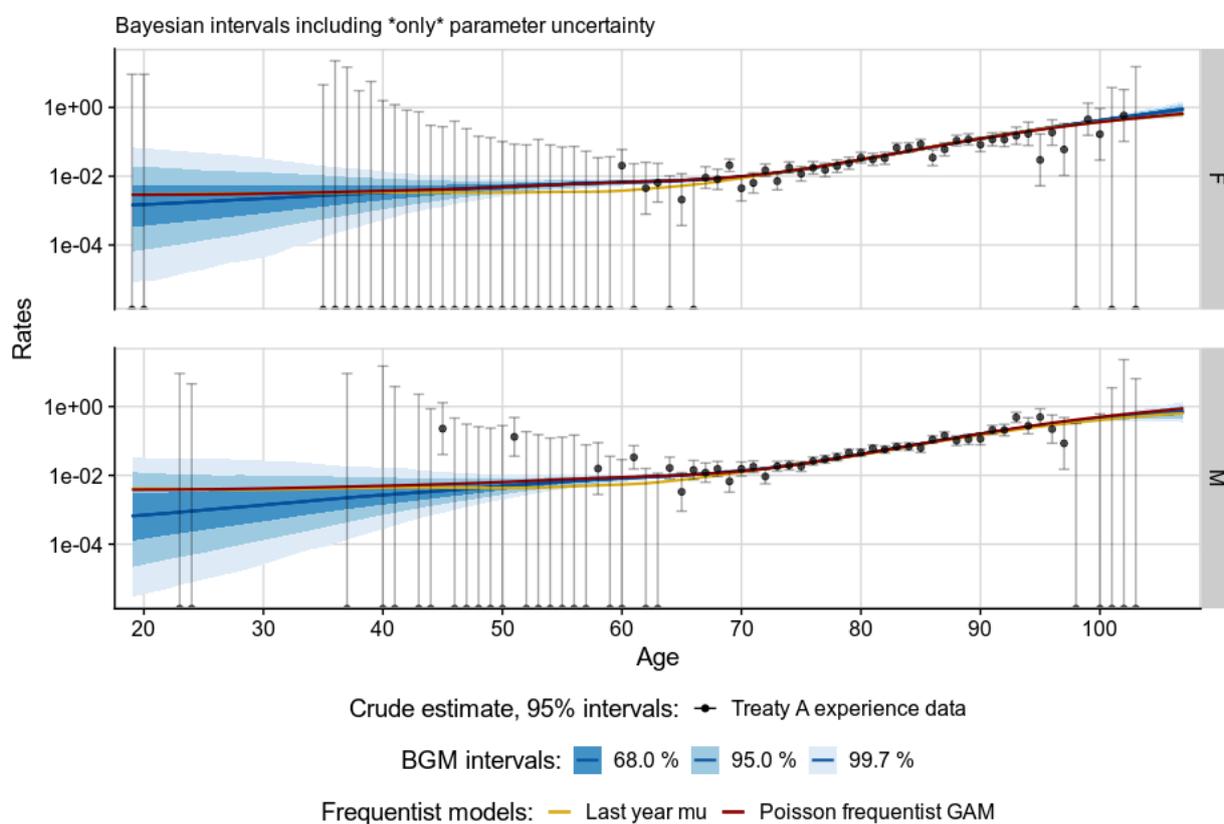


FIG. A.1 : Representation des forces de mortalité pour tout âge et pour l'année 2012

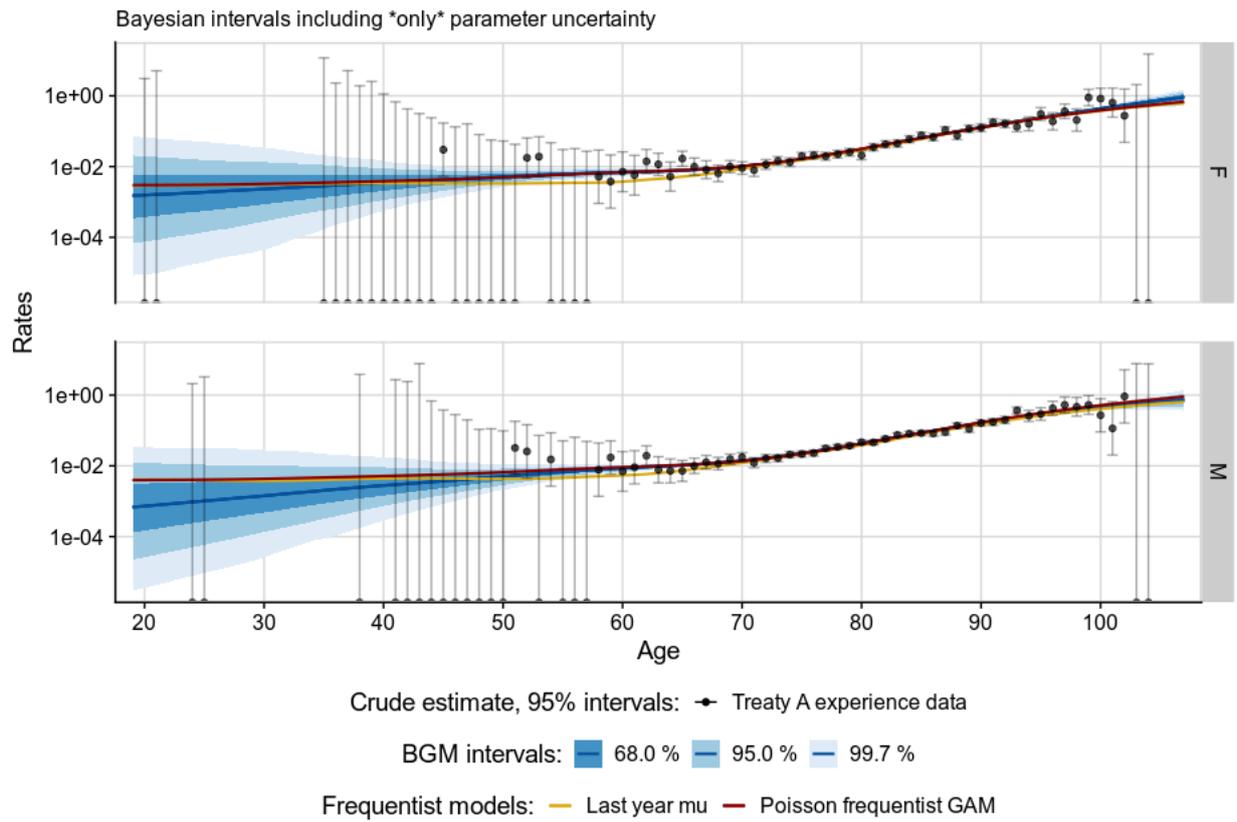


FIG. A.2 : Representation des forces de mortalité pour tout âge et pour l'année 2013

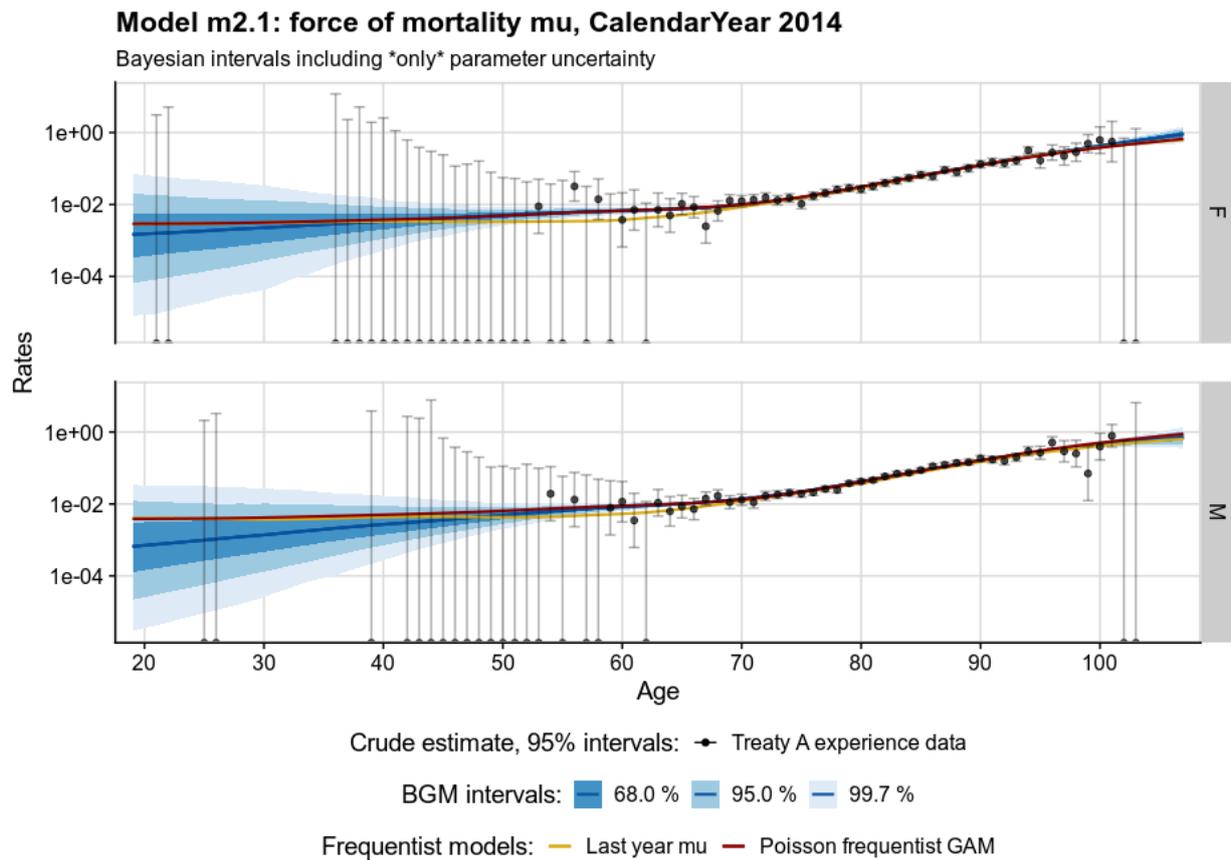


FIG. A.3 : Representation des forces de mortalité pour tout âge et pour l'année 2014

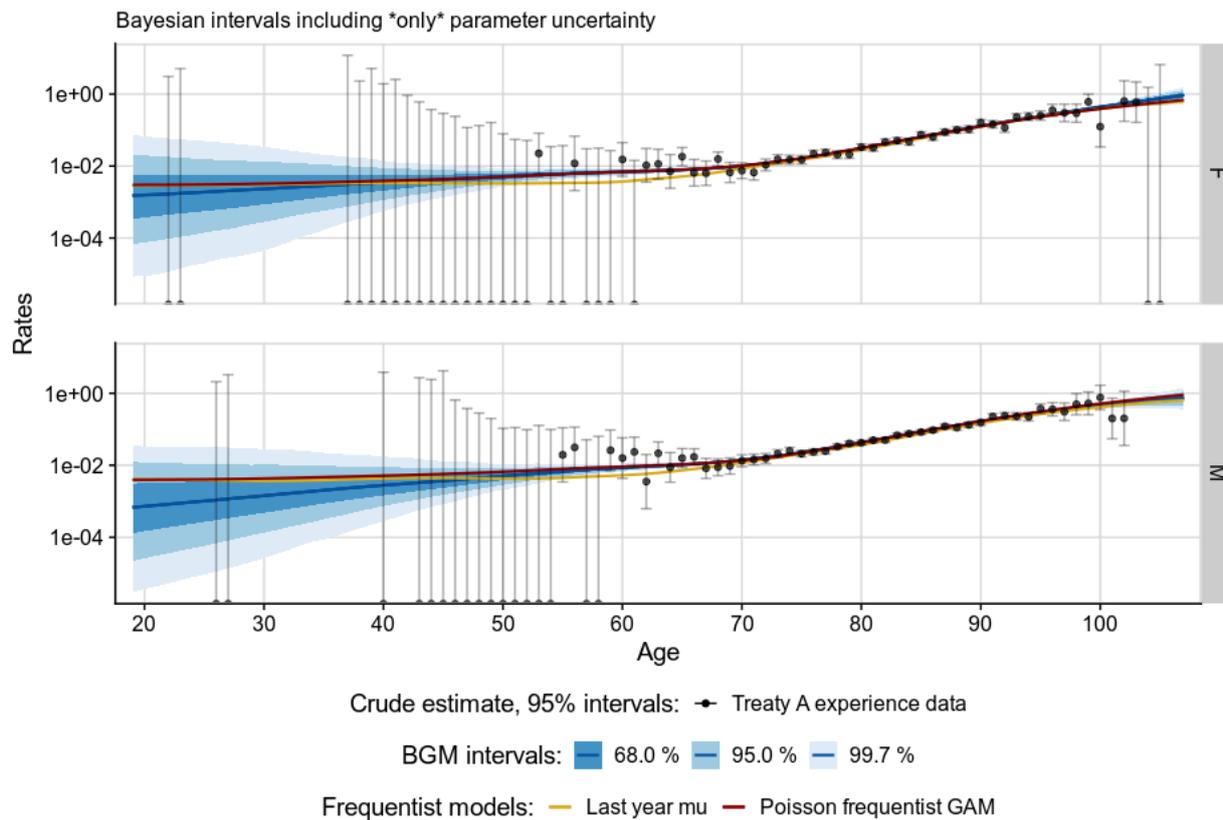


FIG. A.4 : Representation des forces de mortalité pour tout âge et pour l'année 2015

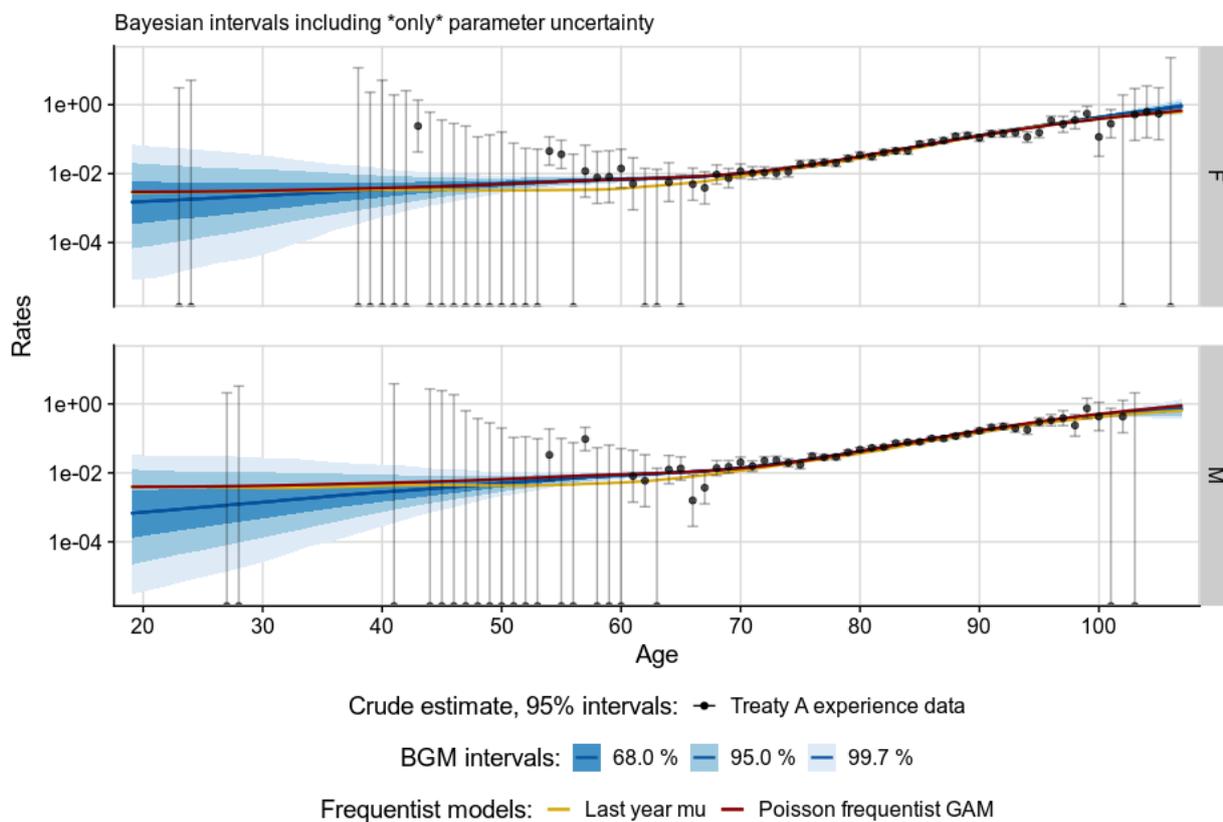


FIG. A.5 : Representation des forces de mortalité pour tout âge et pour l'année 2016

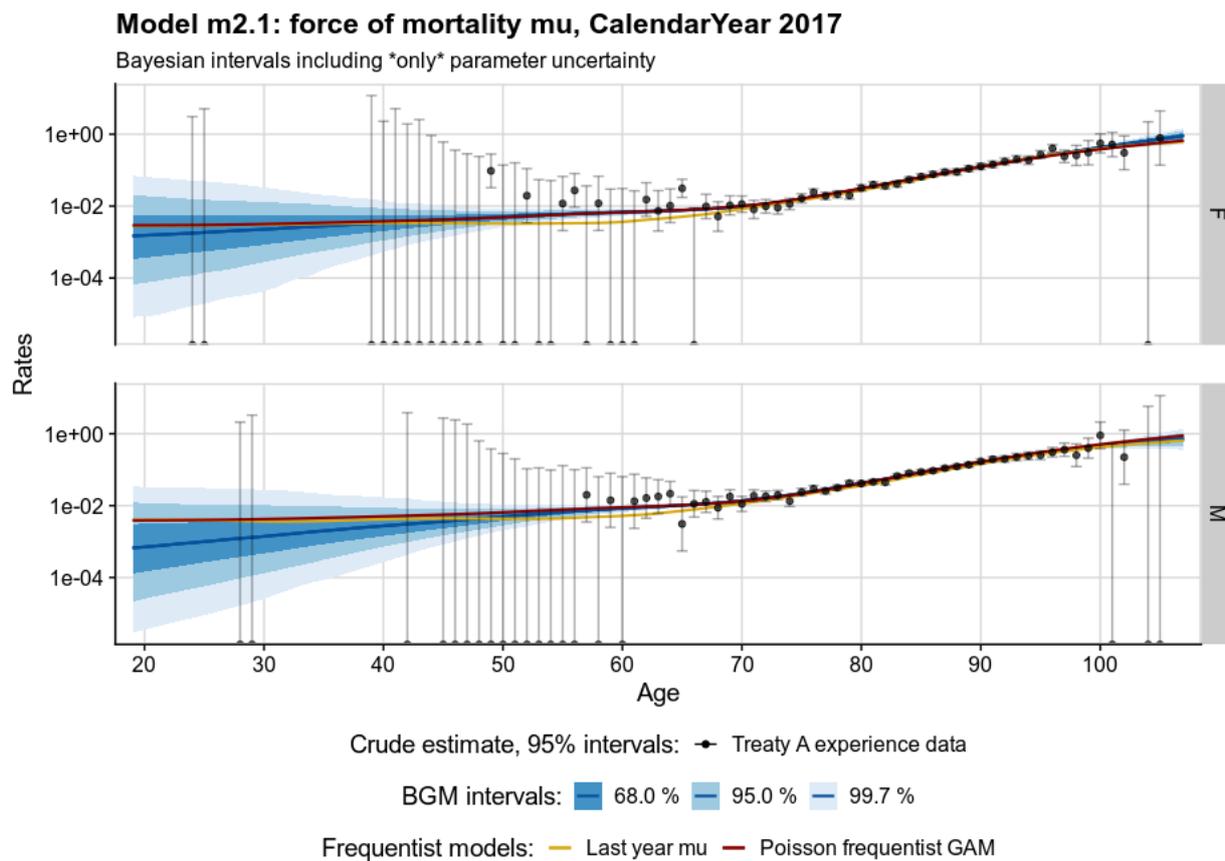


FIG. A.6 : Representation des forces de mortalité pour tout âge et pour l'année 2017

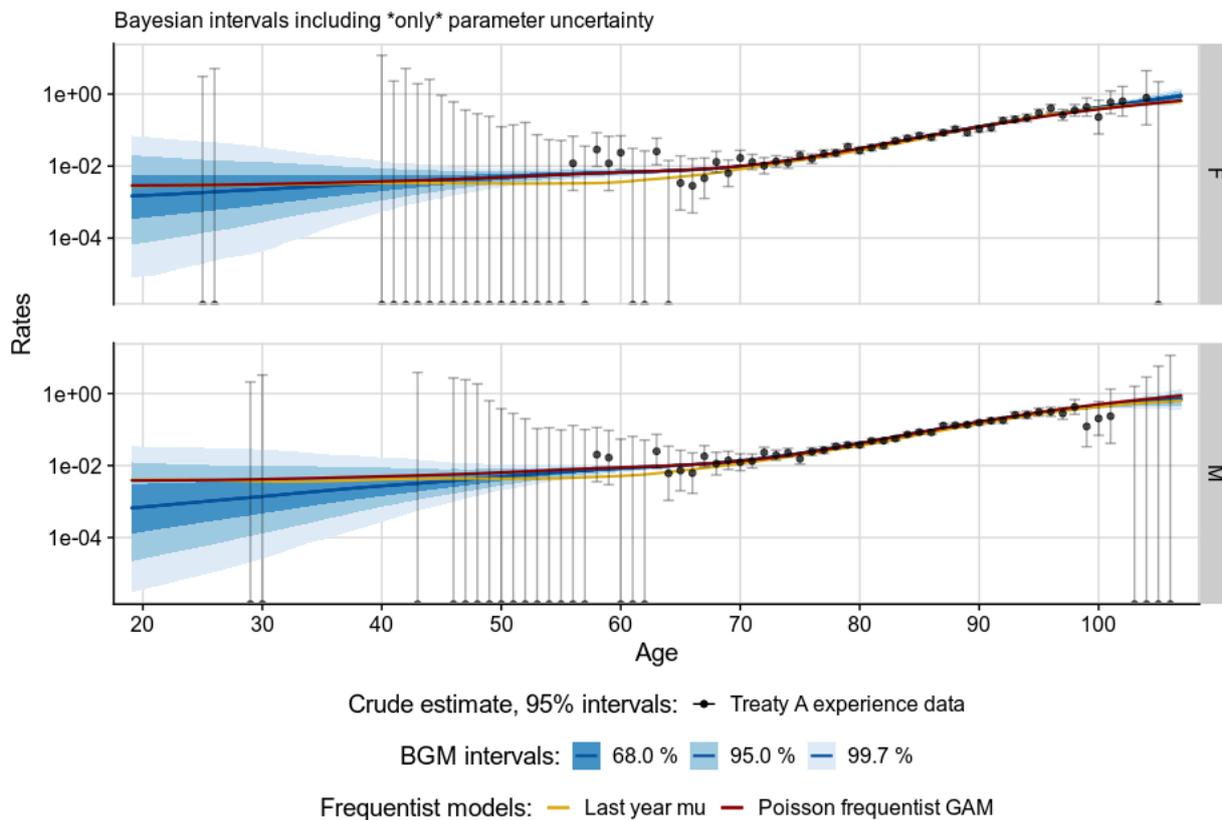


FIG. A.7 : Representation des forces de mortalité pour tout âge et pour l'année 2018

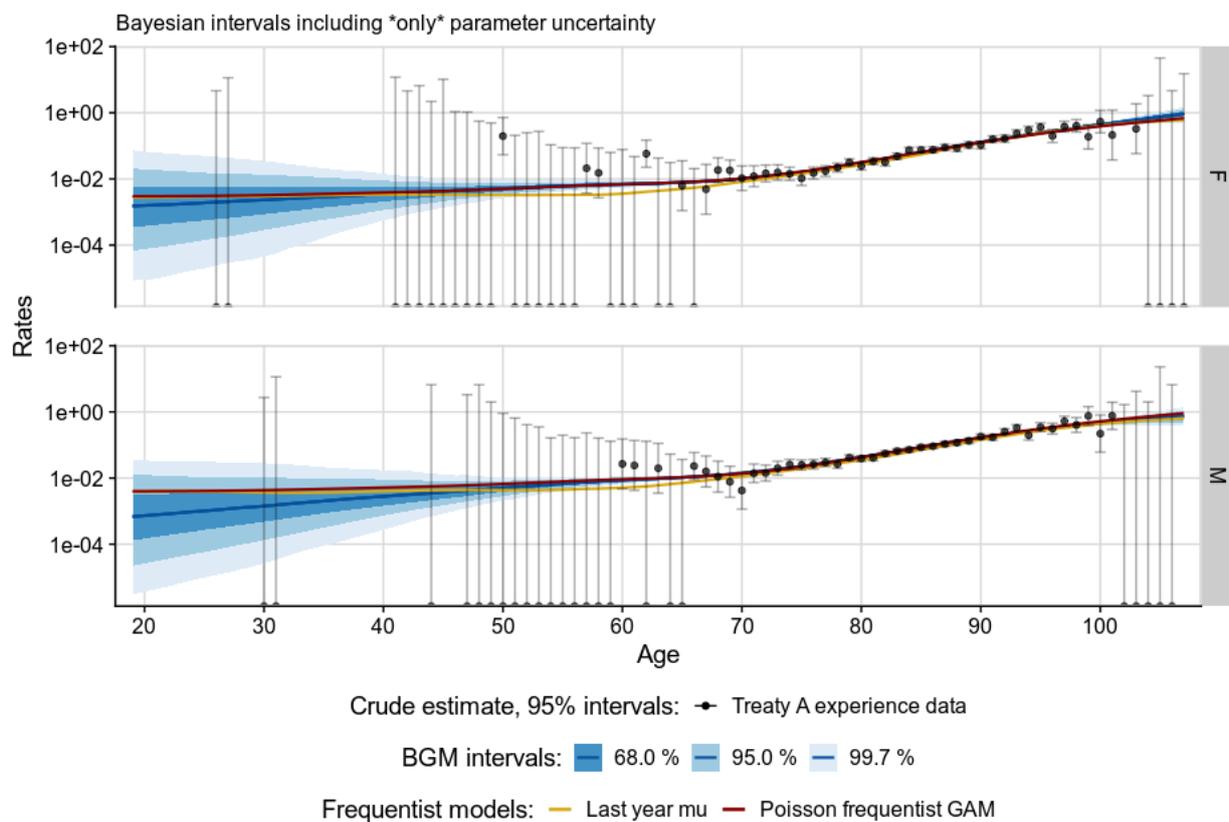


FIG. A.8 : Representation des forces de mortalité pour tout âge et pour l'année 2019

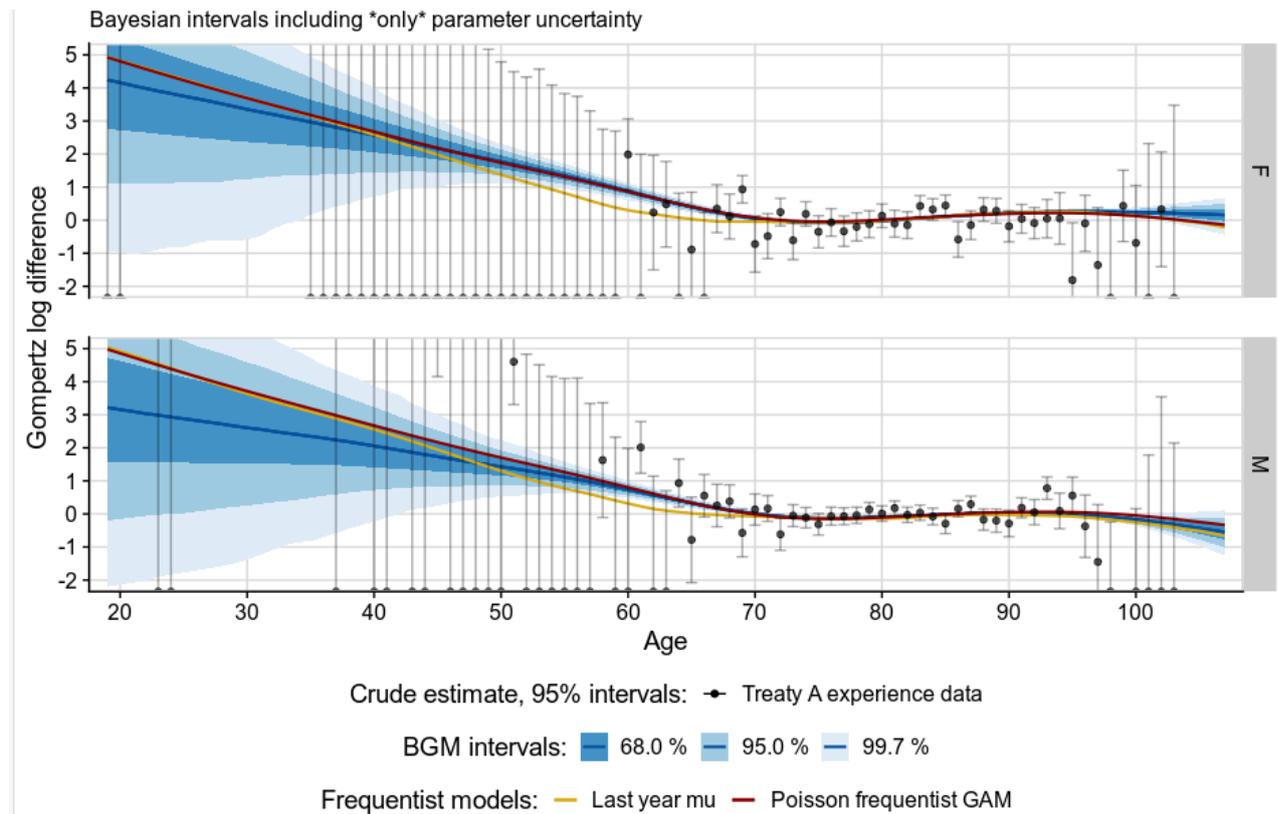


FIG. A.9 : Graphes Gold pour tout âge et pour l'année 2012

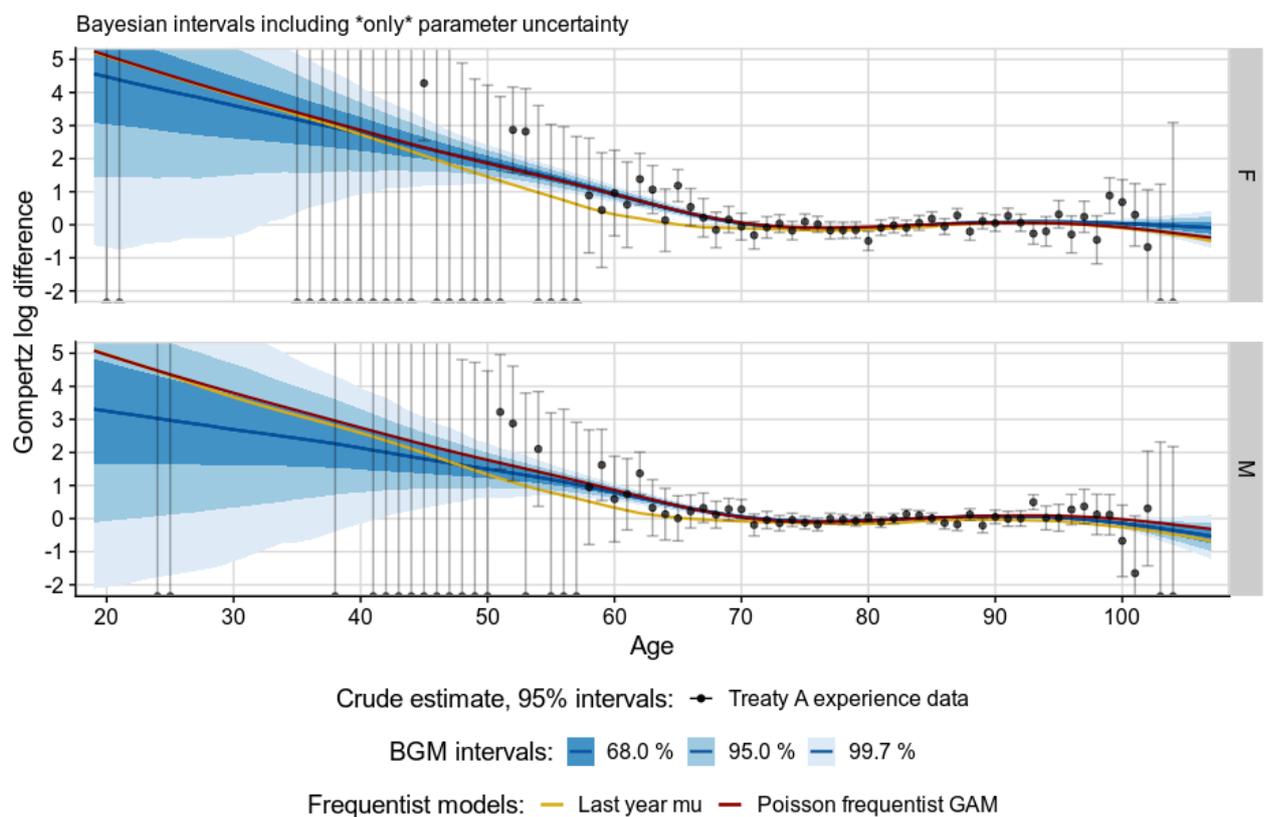


FIG. A.10 : Graphes Gold pour tout âge et pour l'année 2013

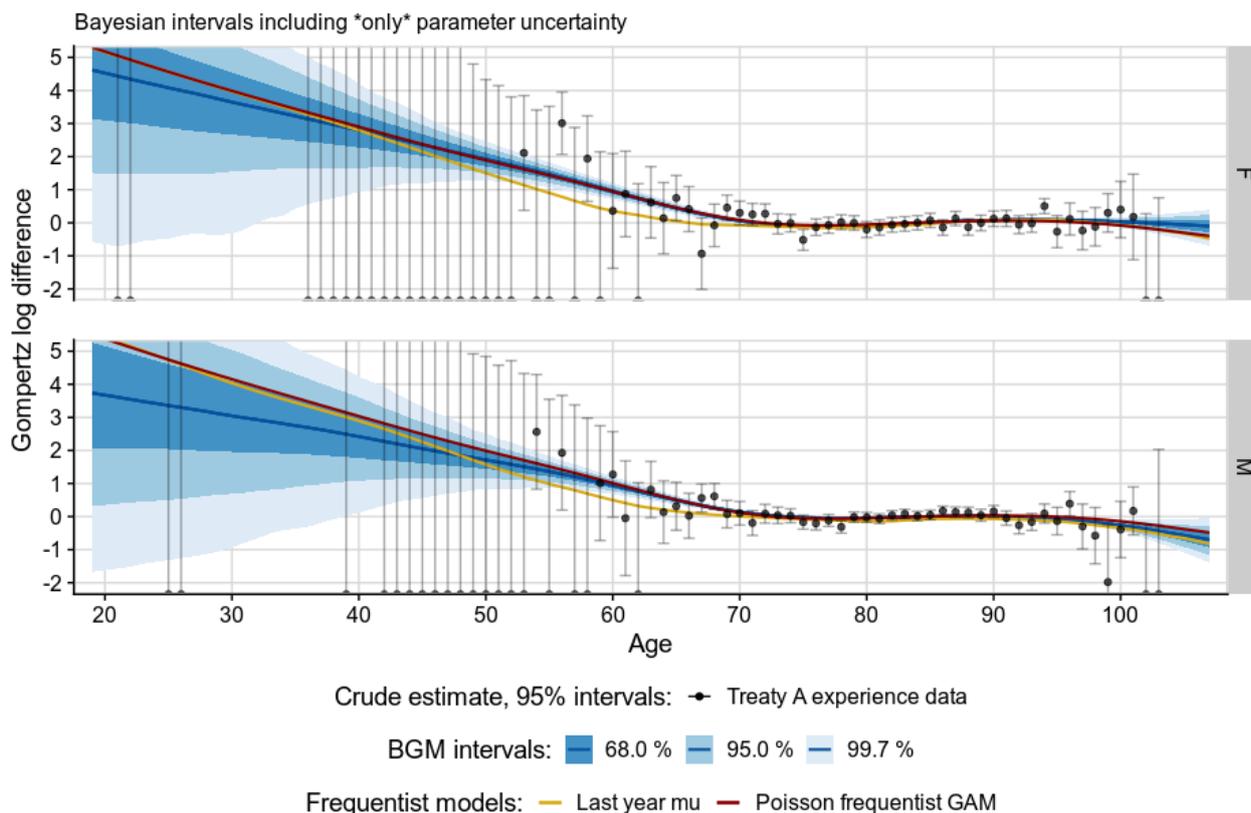


FIG. A.11 : Graphes Gold pour tout âge et pour l'année 2014

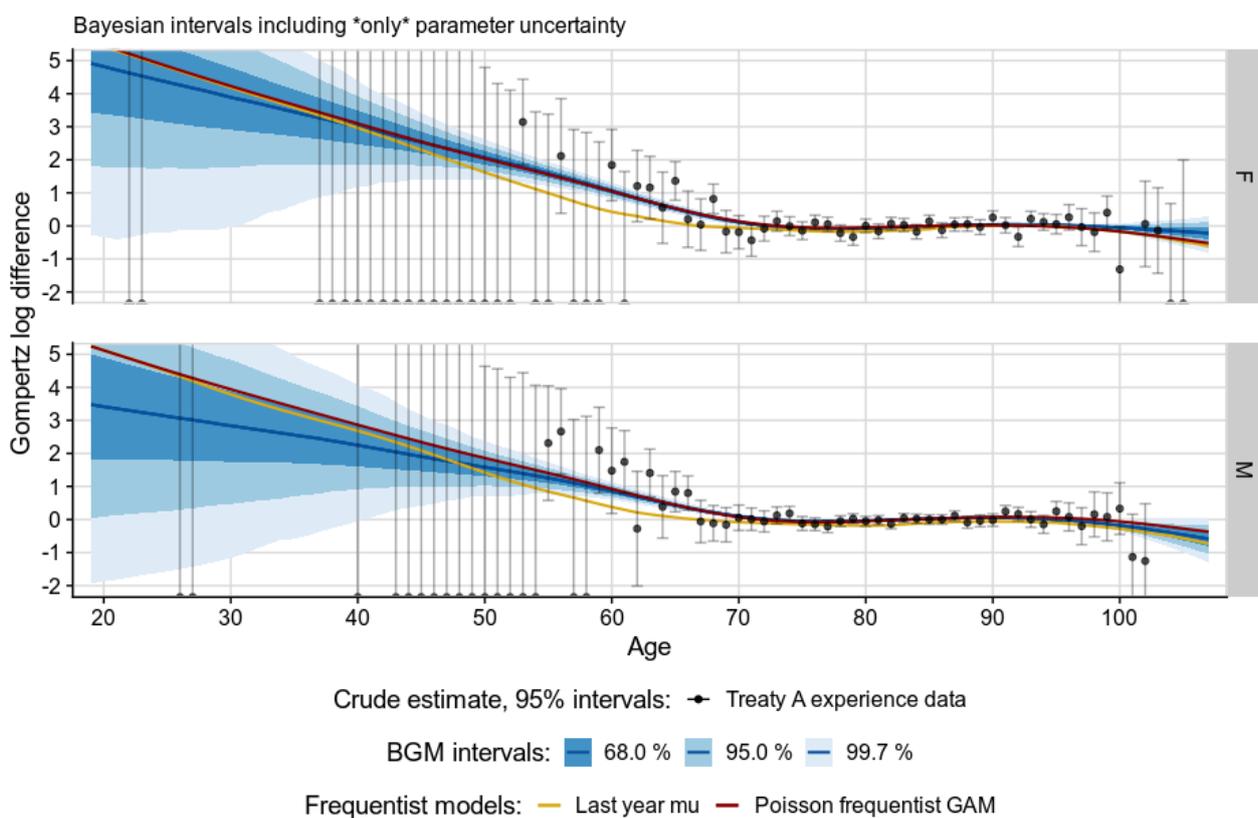


FIG. A.12 : Graphes Gold pour tout âge et pour l'année 2015

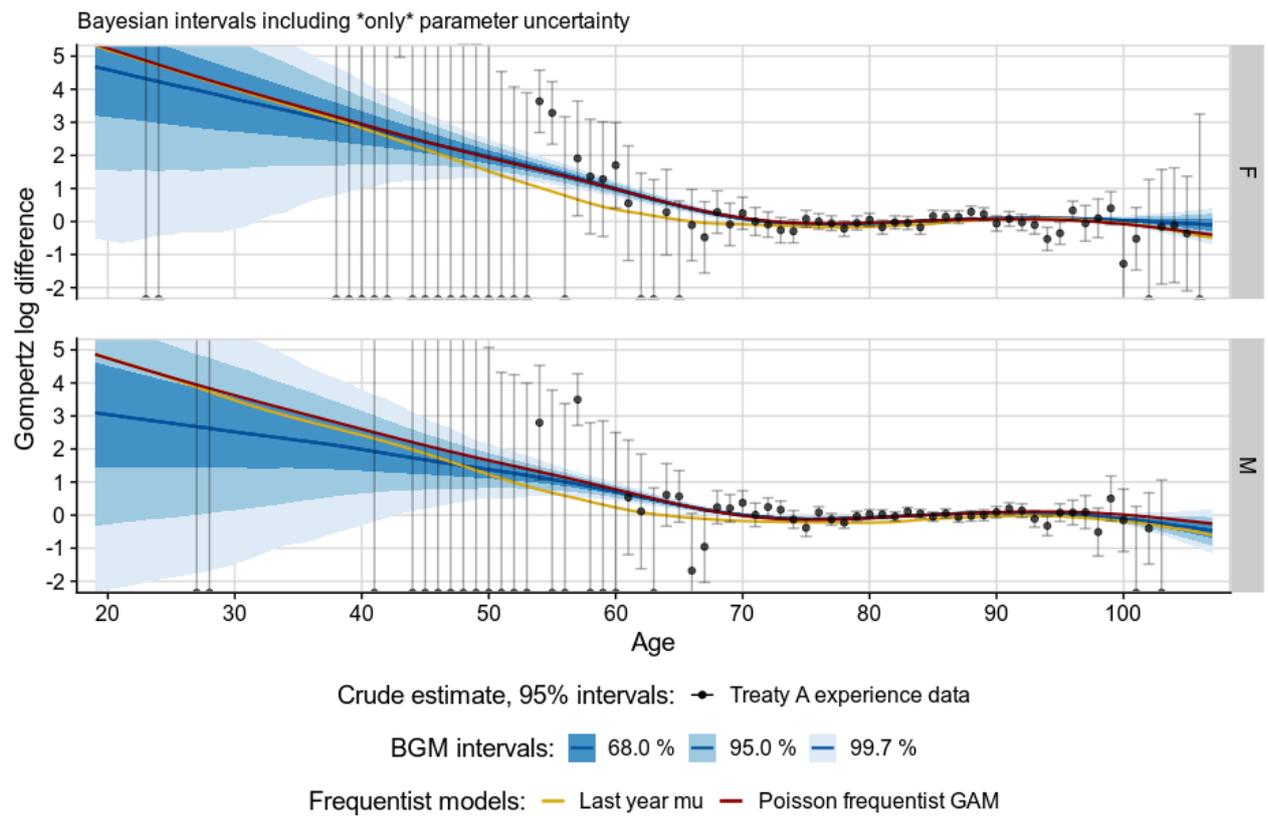


FIG. A.13 : Graphes Gold pour tout âge et pour l'année 2016

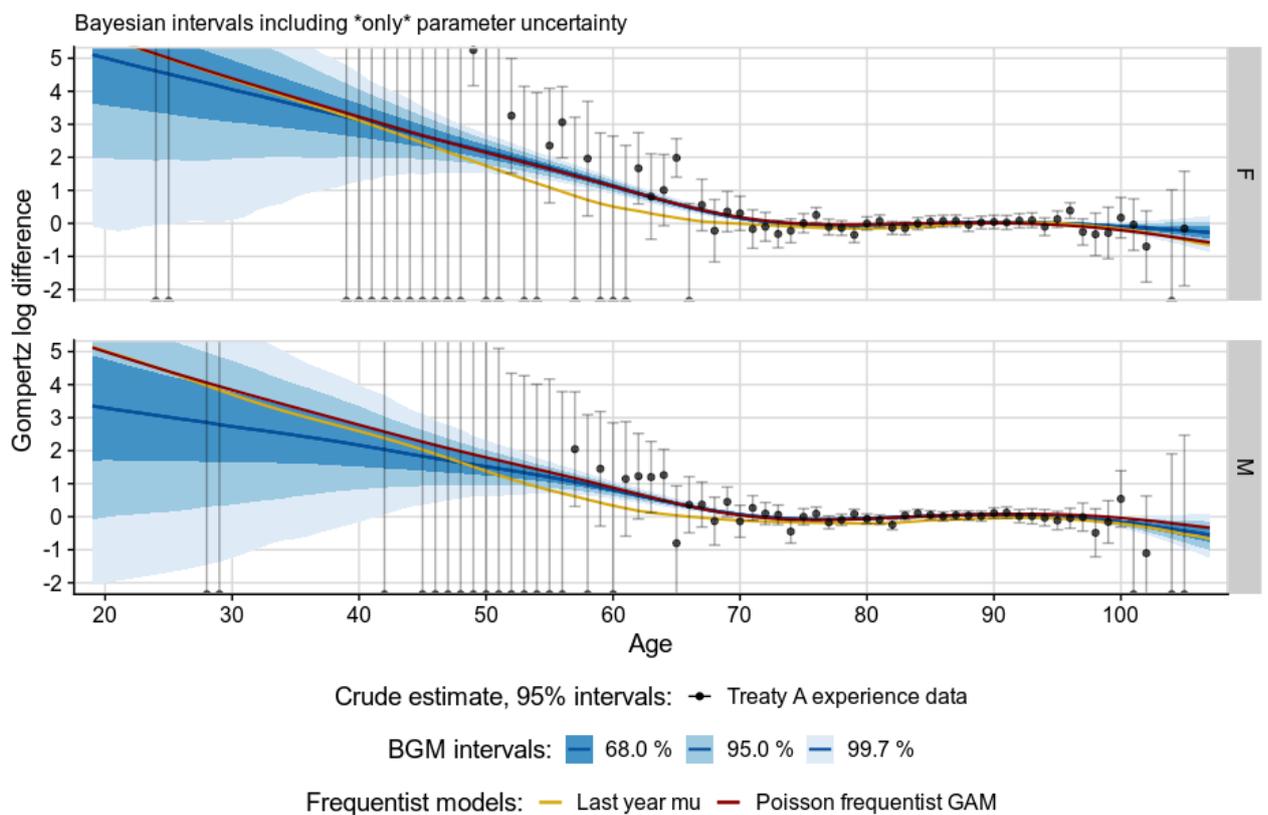


FIG. A.14 : Graphes Gold pour tout âge et pour l'année 2017

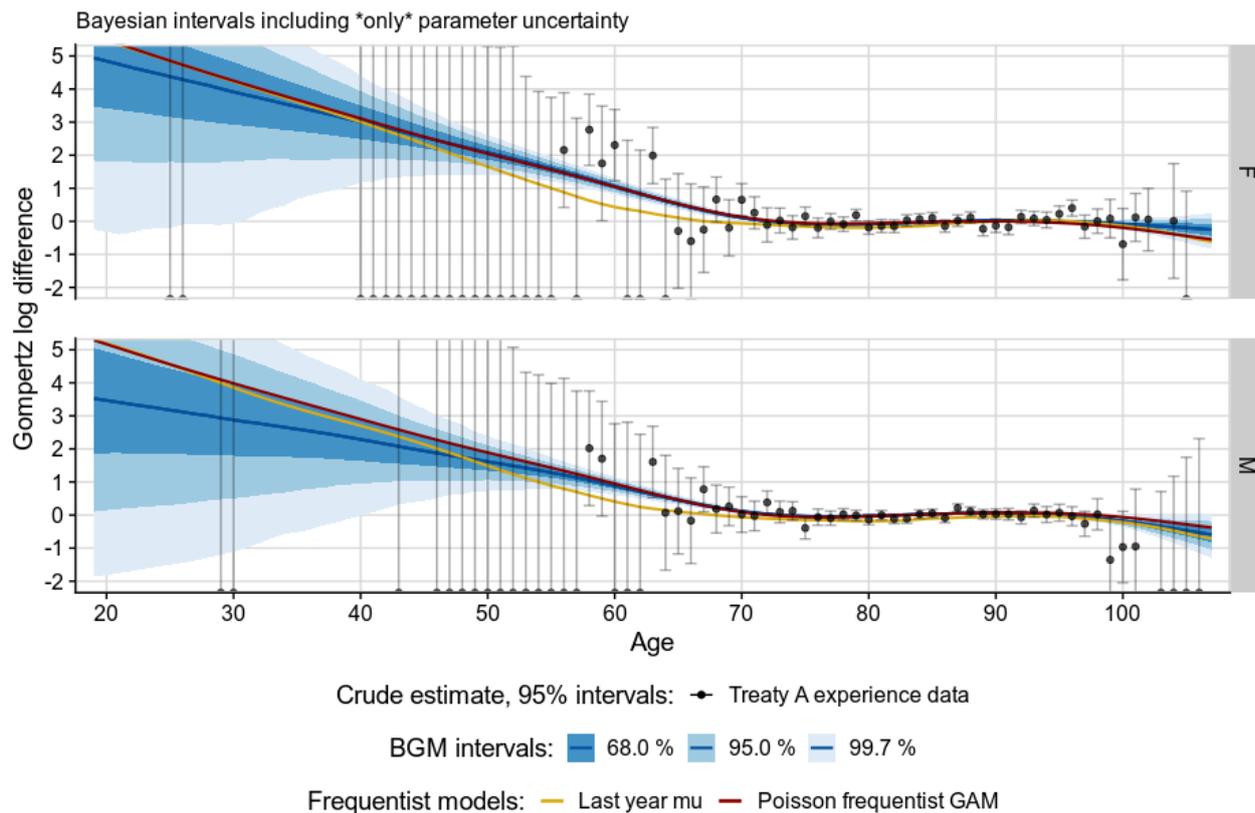


FIG. A.15 : Graphes Gold pour tout âge et pour l'année 2018

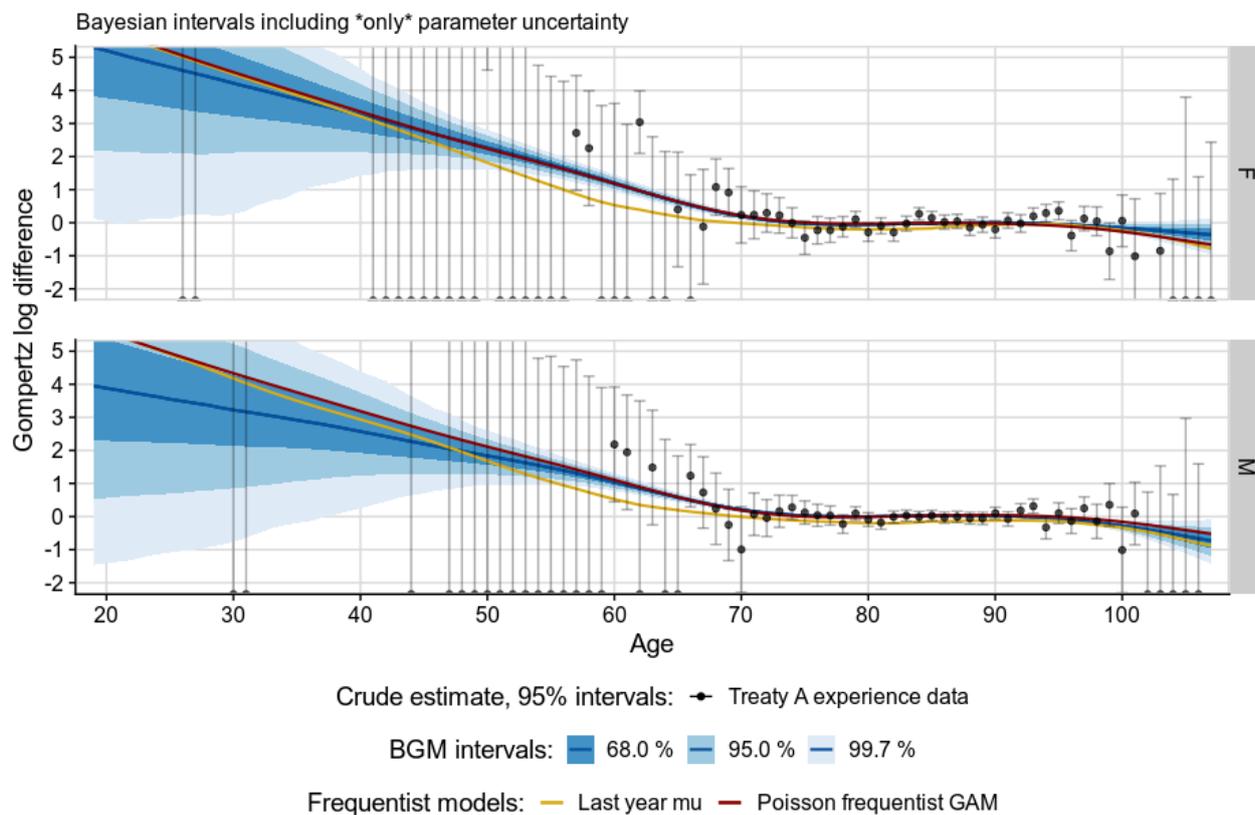


FIG. A.16 : Graphes Gold pour tout âge et pour l'année 2019

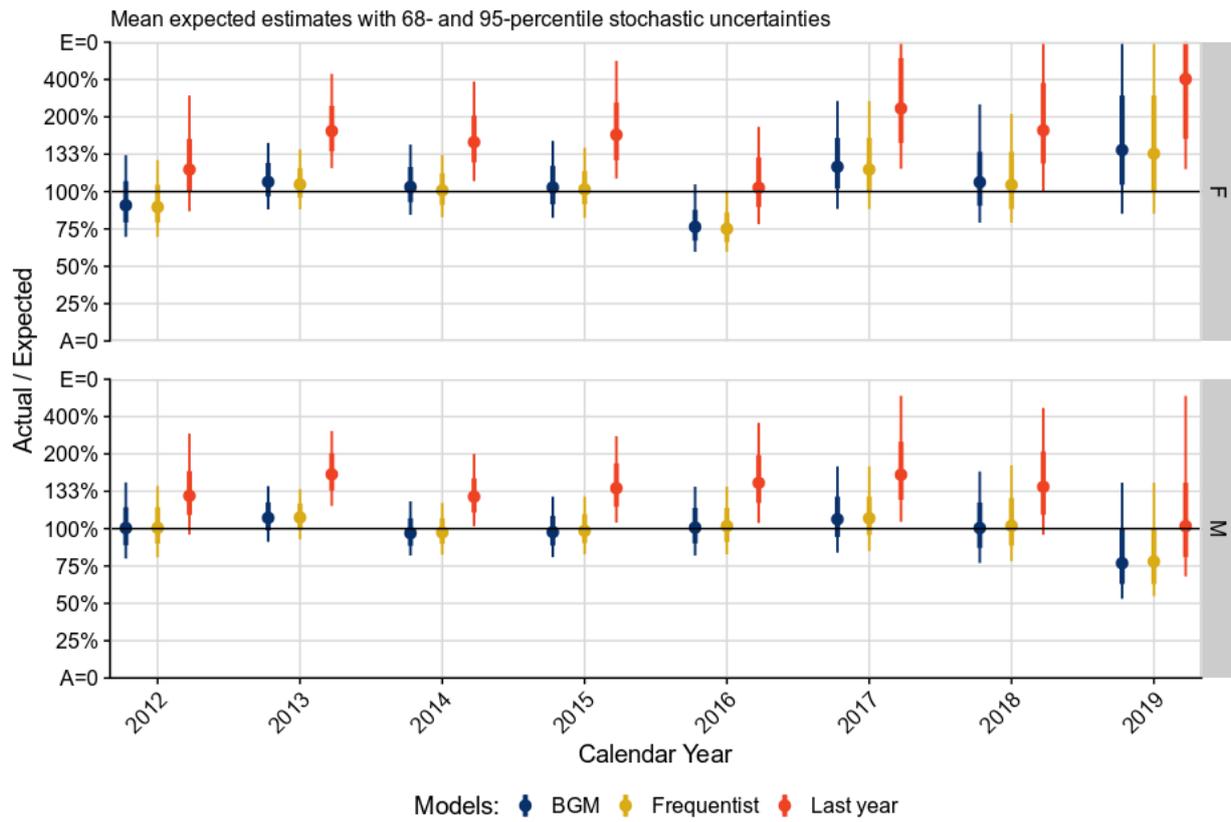


FIG. A.17 : Ratios A/E par année calendaire pour l'intervalle d'âge 60-70

Toutes les courbes ci-dessous ont une échelle linéaire.

## Annexe A. Presentation de tous les resultats graphiques

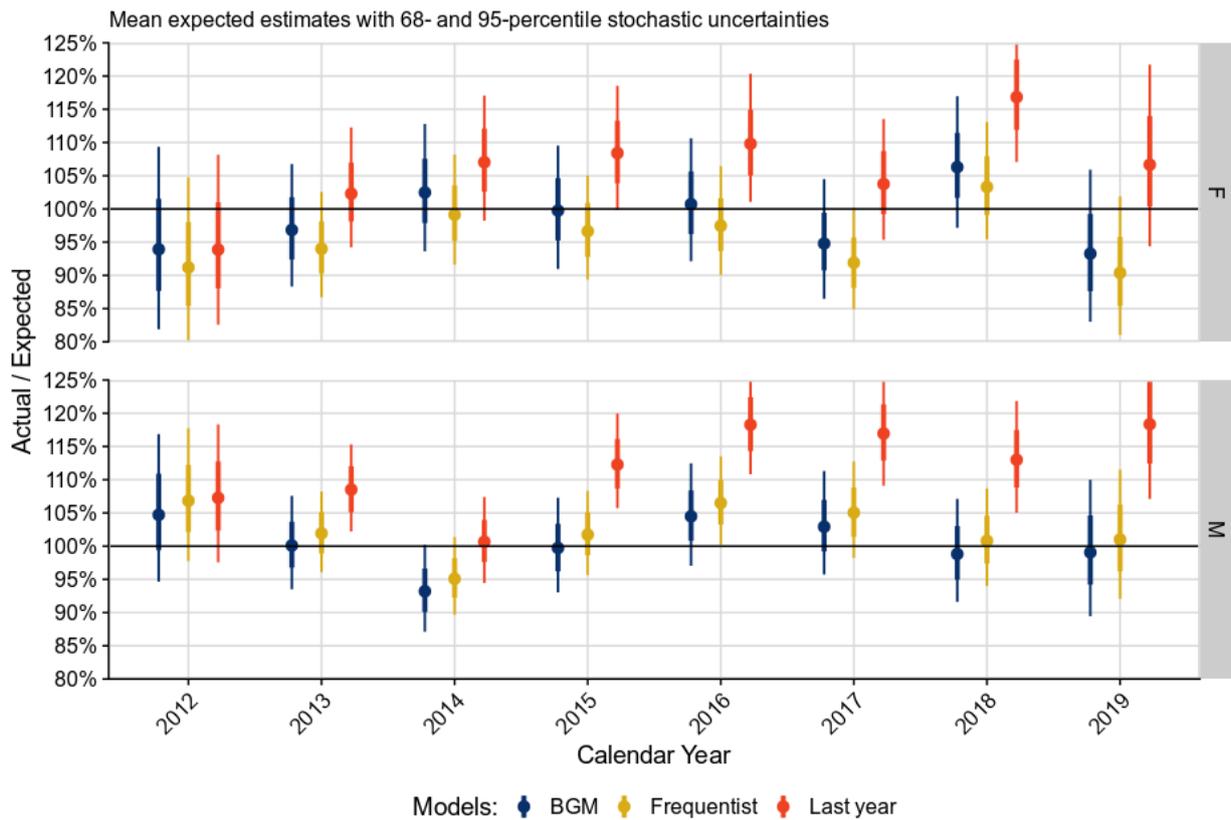


FIG. A.18 : Ratios A/E par année calendaire pour l'intervalle d'âge 70-80

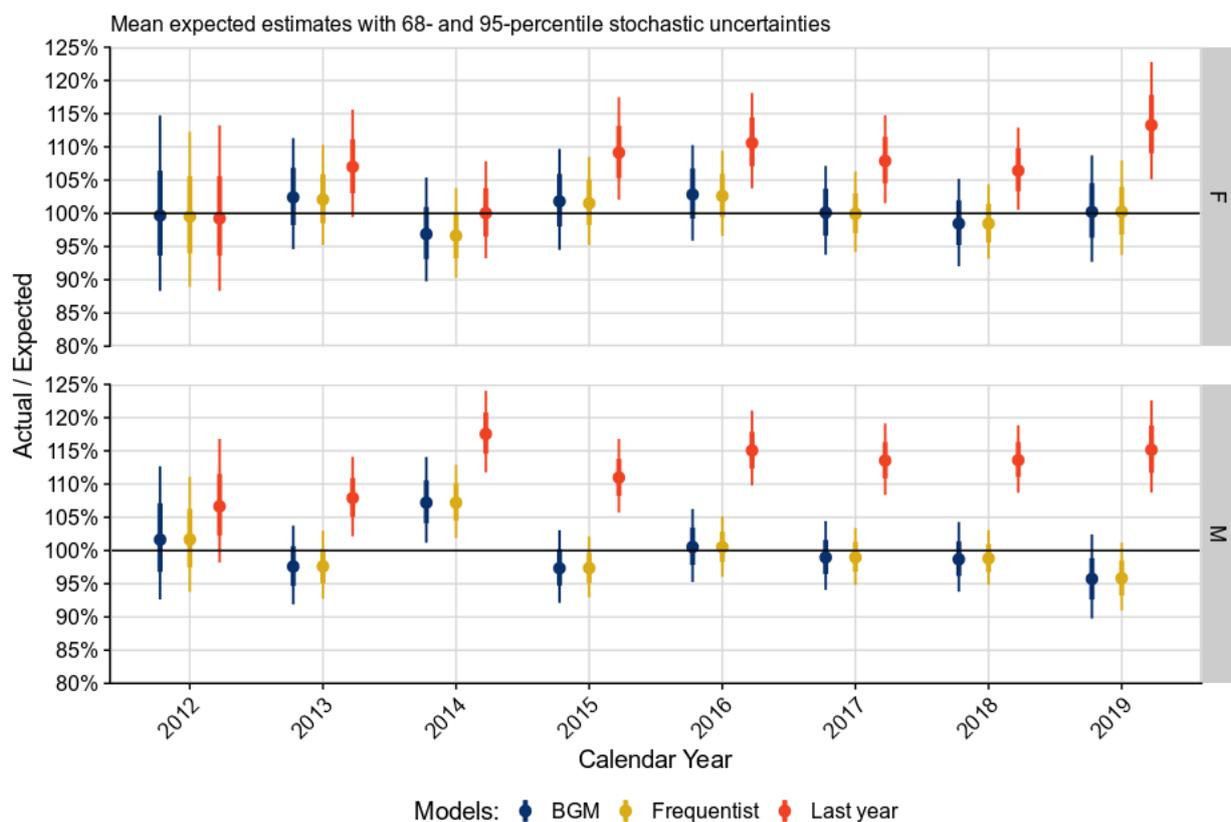


FIG. A.19 : Ratios A/E par année calendaire pour l'intervalle d'âge 80-90

## Annexe A. Presentation de tous les resultats graphiques

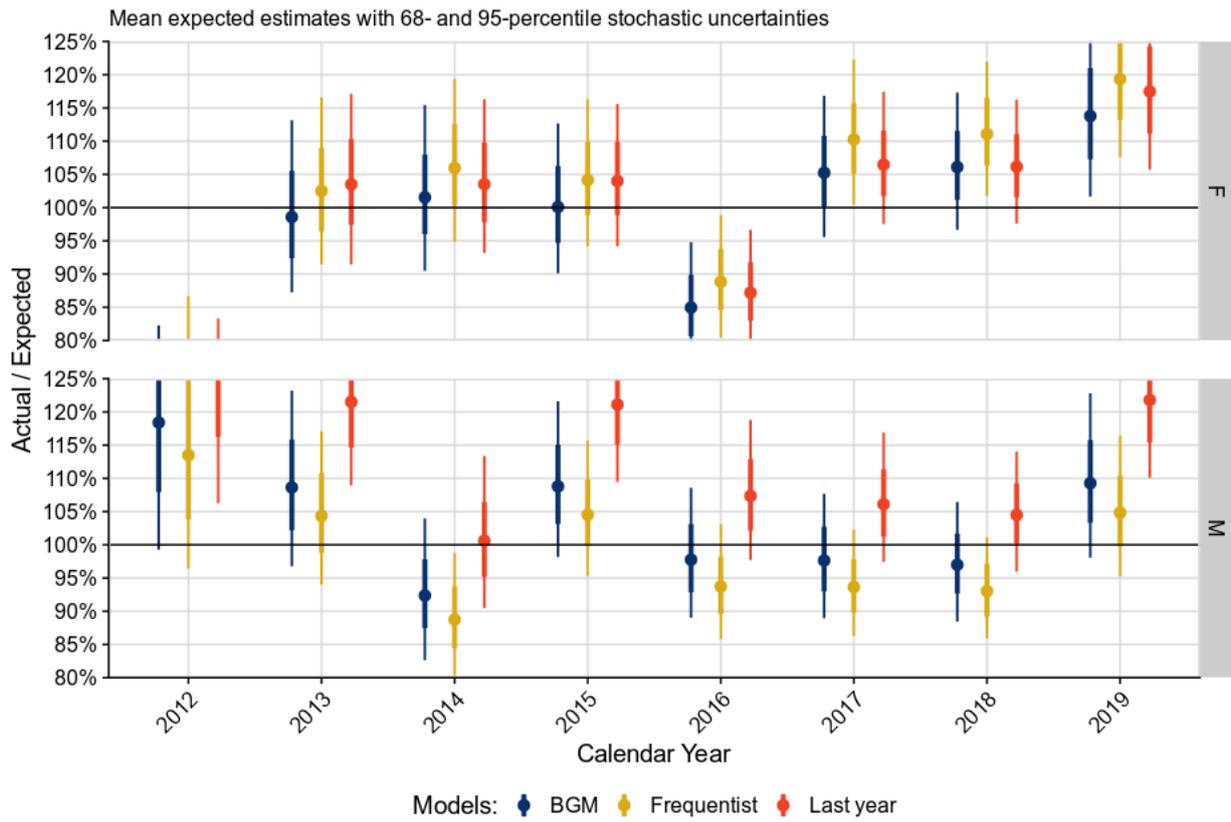


FIG. A.20 : Ratios A/E par année calendaire pour l'intervalle d'âge 90+

# Annexe B

## Code R utiles

### B.1 Code pour produire des B-splines

```
bspline <- function(x,k,i,m=2)
# evaluate ith b-spline basis function of order m at the values
# in x, given knot locations in k
{
  if (m== -1)
  {
    res <- as.numeric(x<k[i+1]&x>=k[i])
  }
  else # construct from call to lower order basis
  {
    z0 <- (x-k[i])/(k[i+m+1]-k[i])
    z1 <- (k[i+m+2]-x)/(k[i+m+2]-k[i+1])
    res <- z0*bspline(x,k,i,m-1)+ z1*bspline(x,k,i+1,m-1)
  }
  return(res)
}
```

### B.2 Code pour créer les matrices de pénalisation différentielle des P-splines

```
k <-6 # example basis dimension
P <- diff(diag(k),differences=1)
S <- t(P)%*%P
```