

Mémoire présenté le :  
pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA  
et l'admission à l'Institut des Actuaire

Par : **Victor Teissier**

Titre : **Analyse et standardisation de méthodes de lissage pour une application globale à travers tous les risques d'un réassureur vie et santé**

Confidentialité :  NON  (Durée :  1 an  2 ans)

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

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

*Entreprise : SCOR  
Nom : Victor Teissier*

*Signature :*

*Directeur de mémoire en entre-  
prise : Karsten De Braaf  
Nom :*

*Signature :*

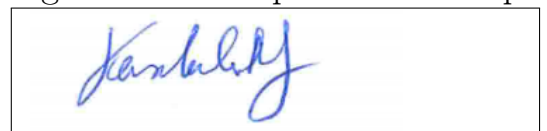
*Membres présents du jury de  
l'ISFA*

*Invité :  
Nom :*

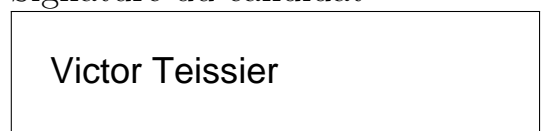
*Signature :* 

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

Signature du responsable entreprise



Signature du candidat



Victor Teissier

MASTER 2 ACTUARIAT

INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

---

**Analyse et standardisation de méthodes de  
lissage pour une application globale à  
travers tous les risques d'un réassureur vie  
et santé**

---

*Auteur :*  
Victor Teissier

*Tuteur entreprise :*  
Karsten De Braaf

*Tuteur académique :*  
Stéphane Loisel

5 juin 2023



---

# Résumé

Dans le domaine de la réassurance, l'asymétrie d'information entre l'assureur et le réassureur est un problème courant qui peut entraîner des tarifs de réassurance injustes. Pour remédier à cela, l'analyse d'expérience est utilisée pour fournir une meilleure compréhension des risques biométriques associés aux portefeuilles réassurés. Cette analyse permet de développer des probabilités ou des taux basés sur des données collectées auprès de diverses sources telles que des compagnies d'assurance, des sociétés de conseil ou des organismes gouvernementaux. Les actuaires peuvent utiliser ces résultats pour élaborer des hypothèses pour la planification financière, l'évaluation, la tarification et la gestion des risques. Cela permet de fournir des *Best Estimate* plus précis. Cependant, cette étude peut être chronophage et nécessiter des hypothèses fortes quant aux modèles d'estimation, ce qui peut impacter directement la tarification des traités de réassurance vie et santé.

L'objectif de ce mémoire est alors de répondre à ces problématiques de modèles et d'estimations en présentant des méthodes d'estimation de taux de mortalité à partir de données d'expérience, puis de proposer des méthodes adaptatives de lissage de taux à partir de modèles simples qu'on confrontera à des modèles plus sophistiqués regroupant pour certains des méthodes dites de *Machine Learning* qui reposent sur des modèles de *boosting* contraints. Les modèles de *boosting* seront utilisés pour obtenir des taux de mortalité plus lisses, en choisissant judicieusement les apprenants faibles. Des schémas « imbriqués » seront également présentés pour optimiser le critère de lissage retenu, en utilisant différents algorithmes d'optimisation en fonction de la méthode de lissage choisie. Une réflexion approfondie sur les avantages et les inconvénients des différentes méthodes au travers de nombreux critères statistiques sera présentée. Tout ce cheminement sera mis en évidence selon une trame d'analyse d'expérience qui tentera, à partir d'un portefeuille étudié, de donner des valeurs plus financières et opérationnelles de notre proposition adaptative.

**Mots Clés :** *Machine Learning*, Méthode de lissage, Réassurance, Mortalité

---

---

# Abstract

In reinsurance, information asymmetry between the insurer and the reinsurer is a common problem that can lead to unfair reinsurance rates. To address this, experience analysis is used to provide a better understanding of the biometric risks associated with reinsured portfolios. This analysis develops probabilities or rates based on data collected from various sources such as insurance companies, consulting firms or government agencies. Actuaries can use these results to develop assumptions for financial planning, valuation, pricing and risk management. This helps to provide more accurate « Best Estimates ». However, this study can be time consuming and require strong assumptions about the estimation models, which can directly impact the pricing of life and health reinsurance treaties.

The objective of this thesis is then to answer these problems of models and estimations by presenting methods of estimation of mortality rates from experience data, then to propose adaptive methods of rate smoothing from simple models which we will confront with more sophisticated models gathering for some of the methods known as Machine Learning which rest on constrained boosting models. Boosting models will be used to obtain smoother mortality rates, by judiciously choosing weak learners. Nested schemes will also be presented to optimise the chosen smoothing criterion, using different optimisation algorithms depending on the chosen smoothing method. An in-depth reflection on the advantages and disadvantages of the different methods through numerous statistical criteria will be presented. The whole process will be highlighted in an experimental analysis framework which will attempt, from a studied portfolio, to give more financial and operational values of our adaptive proposal.

**Key Words :** Machine Learning, Smoothing Method, Reinsurance, Mortality

---



---

# Note de Synthèse

## Contexte

Ce mémoire s'intègre dans un projet plus global de construction de méthode standard d'analyse d'expérience au sein de SCOR qui vise à automatiser l'analyse du portefeuille d'une cédante. On présentera dans les différentes parties du mémoire et le cheminement du projet dans la Figure 1.

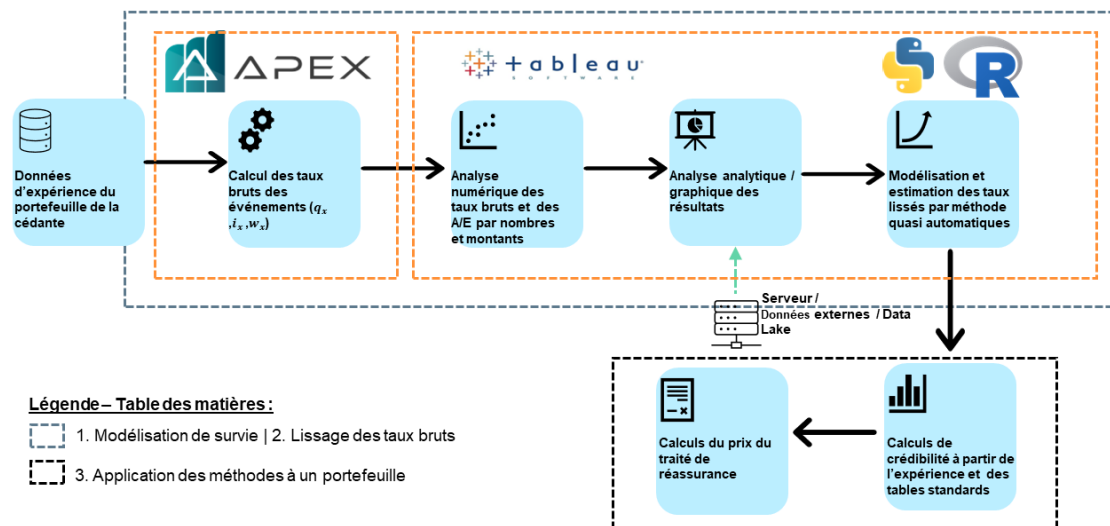


FIGURE 1 – Schéma explicatif du cheminement du mémoire et d'une analyse d'expérience

La finalité est de tenter de proposer des méthodes adaptatives pour de nouveaux utilisateurs. Les modèles seront implémentés sous R et Python et ils seront utilisés au travers de l'application Tableau, application qui permet une analyse complète de résultats analytiques.

## Modèles retenues

On proposera au travers de ce mémoire des modèles pour estimer les taux bruts de mortalité, fondés sur la notion d'exposition au risque. Ensuite, on tentera de répondre à la problématique de création de méthode standardisées et adaptatives. Les modèles testés et retenus seront divers : des modèles classiques comme la méthode de lissage de Whittaker-Henderson ; des modèles plus sophistiqués qui se fondent sur des modèles additifs, de réseaux de neurones ou encore boosting et qui offrent à l'utilisateur la possibilité de considérer plusieurs variables explicatives pour lisser les données, comme la durée (duration) de la police, le montant assuré ou encore le type de produit. On s'intéressera plus particulièrement à tester nos méthodes seulement avec une dimension, l'âge puis avec deux dimensions, la durée de la police d'assurance et l'âge.

---

**TABLE 1** – Résumé des modèles utilisés et des algorithmes d’optimisations utilisés pour le principe « adaptatif » dans le cadre d’une construction de table de mortalité

Modèles	Avantages	Inconvénients
<b>Modélisation des taux lissés   Whittaker-Henderson</b>	<ul style="list-style-type: none"> <li>• Pas de structure de mortalité à définir</li> <li>• Simple à mettre en oeuvre/comprendre</li> <li>• Très interprétable</li> </ul>	<ul style="list-style-type: none"> <li>• Temps de calcul long</li> <li>• Nombre de variables restreints</li> </ul>
<b>B-Spline</b>	<ul style="list-style-type: none"> <li>• Pas de structure de mortalité à définir</li> <li>• Possibilité d’ajout dans un GLM</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration des noeuds</li> <li>• Très sensible</li> </ul>
<b>P-Spline</b>	<ul style="list-style-type: none"> <li>• Ajout d’un critère de régularisation/pénalité</li> <li>• Possibilité d’ajout dans un GLM</li> <li>• Pas de structure à définir</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration des noeuds et de la pénalité</li> <li>• Très sensible</li> </ul>
<b>LOESS</b>	<ul style="list-style-type: none"> <li>• Définition de poids aux voisinages</li> <li>• Optimisation des paramètres simples</li> <li>• Très interprétable</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration du voisinage h</li> </ul>
<b>GLM local</b>	<ul style="list-style-type: none"> <li>• Définition de poids aux voisinages</li> <li>• Choix de la distribution sous-jacente</li> <li>• Interprétable</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration du voisinage h</li> <li>• Optimisation des paramètres</li> </ul>
<b>GAM</b>	<ul style="list-style-type: none"> <li>• Pas de structure à définir</li> <li>• Possibilité de rajouts de variables</li> <li>• Possibilité de rajout d’un critère de pénalité</li> <li>• Interprétable</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration longue des paramètres</li> </ul>
<b>GBM</b>	<ul style="list-style-type: none"> <li>• Pas de structure à définir</li> <li>• Apprend des erreurs précédentes</li> <li>• Peu interprétable</li> </ul>	<ul style="list-style-type: none"> <li>• Très sensible</li> <li>• Calibration longue des paramètres</li> </ul>
<b>RBF Network</b>	<ul style="list-style-type: none"> <li>• Possibilité d’ajout dans un GLM</li> <li>• Peu interprétable</li> </ul>	<ul style="list-style-type: none"> <li>• Problème d’overfitting</li> </ul>
<b>Algorithmes d’optimisations   Scoring de Fisher</b>	<ul style="list-style-type: none"> <li>• Pas de <i>learning rate</i> à définir</li> <li>• Intégration de distribution sous-jacente</li> </ul>	<ul style="list-style-type: none"> <li>• Inversion du Scoring</li> </ul>
<b>Algorithme de Newton</b>	<ul style="list-style-type: none"> <li>• Pas de <i>learning rate</i> à définir</li> <li>• Intégration de distribution sous-jacente</li> </ul>	<ul style="list-style-type: none"> <li>• Inversion de matrice Hessienne</li> </ul>
<b>Algorithme Quasi-Newton</b>	<ul style="list-style-type: none"> <li>• Vitesse d’optimisation</li> </ul>	<ul style="list-style-type: none"> <li>• Choix du <i>learning rate</i></li> </ul>
<b>BFGS (Broyden-Fletcher-Goldfarb-Shanno)</b>	<ul style="list-style-type: none"> <li>• Grande vitesse d’optimisation</li> </ul>	<ul style="list-style-type: none"> <li>• Choix du <i>learning rate</i></li> </ul>

## Méthodologie

L’optimisation des paramètres des différentes méthodes adaptatives est construite autour du schéma imbriqué qui va optimiser le critère GCV. De plus, la méthode GBM Spline repose sur l’ajout de deux contraintes, une contrainte de convexité, mais aussi une contrainte de monotonie qui permet d’empêcher à la méthode du boosting de sûr apprendre. Enfin, les apprenants faibles correspondent à des splines, toujours pour obtenir des résultats de taux de mortalité les plus lisses possibles.



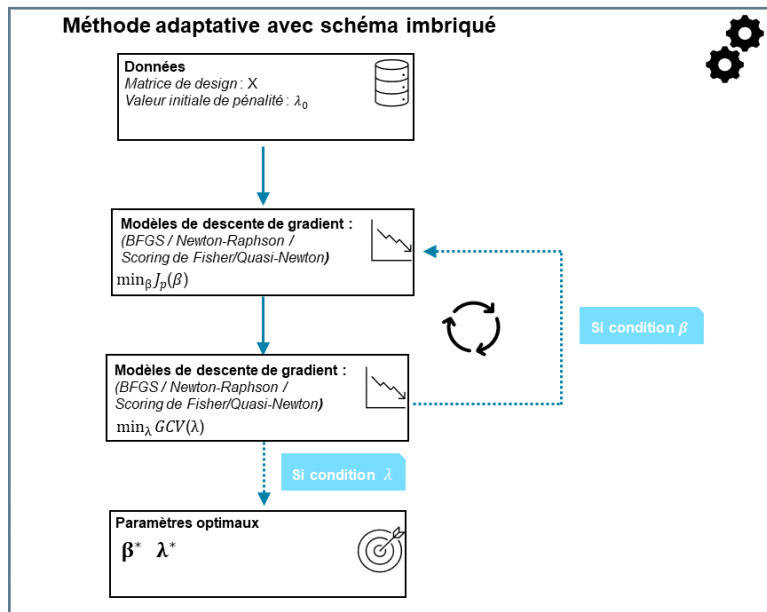


FIGURE 2 – Schéma explicatif de l’algorithme imbriqué

## Présentation des données

L’ensemble des méthodes sera testé dans le cadre d’un exemple appliqué à la construction d’une table de mortalité à partir des données d’une cédante au Portugal. C’est un portefeuille dont les données vont de 1977 à 2013. Cette étude se constitue seulement de 30 % du portefeuille pour tester nos modèles. L’idée était de challenger les méthodes de lissage adaptatives proposées lorsque l’on fait face à une crédibilité plutôt faible. La plus par des méthodes adaptatives fonctionnent bien avec beaucoup de données. Cependant, en 2 dimensions et en particulier avec peu de données, c’est plus difficile de réaliser un bon lissage, c’est une des raisons pour lesquelles nous avons décidé de choisir cette contrainte.

Ce portefeuille est constitué de contrats d’assurance emprunteur, c’est à dire des assurances de crédits divers, certains liés à des hypothèques d’autres liés à des crédits individuels. Ce portefeuille est constitué de 45,1% de Femme et 54,9% d’Homme. On s’intéressera dans ce mémoire à la réalisation d’une analyse d’expérience et à la création d’une table de mortalité du 01/01/2007 au 31/12/2009. Après application du critère de Cochran les âges étudiés seront distribués de 23 à 65 ans.

## Résultats des méthodes adaptatives

TABLE 2 – Résumé des résultats obtenus en 1 dimension pour les méthodes de lissage adaptatives

Critères		GAM	B-Spline	P-Spline	LOESS	WH	GLM local	GBM	RBF
<b>Test du SMR</b>	SMR (%)	99,9%	99,9 %	99,9%	99,7%	98,0%	99,4%	99,9%	99,9%
	$\xi^{SMR}$	1.19	1.20	1.20	1.13	0.53	1.03	1.20	1.19
	p-valeur	0.11	0.11	0.11	0.13	0.29	0.15	0.11	0.12
<b>Déviance</b>		26.13	17.56	20.48	20.47	31.52	28.87	23.34	23.61
<b>MAPE (%)</b>		0.49 %	0.41 %	0.43 %	0.42 %	0.53 %	0.52 %	0.44 %	0.44 %
<b>Résidus standardisés</b>	> 2	0	0	0	0	1	0	0	0
	> 1	10	6	7	7	11	0	8	8
<b>R<sup>2</sup></b>		92.96%	93.97%	92.55%	95.59%	91.89%	91.49%	98.10%	92.64%
<b>Test du run</b>	Nbr runs	23	29	29	29	23	19	19	20
	p_valeur	0.851	0.044	0.040	0.040	0.528	0.292	0.315	0.457

\*Les résultats des méthodes de lissage adaptatives sont des résultats obtenus après application du critère de Cochran présenté Section 2.2

TABLE 3 – Résumé des résultats obtenus en 2 dimensions pour les méthodes de lissage adaptatives

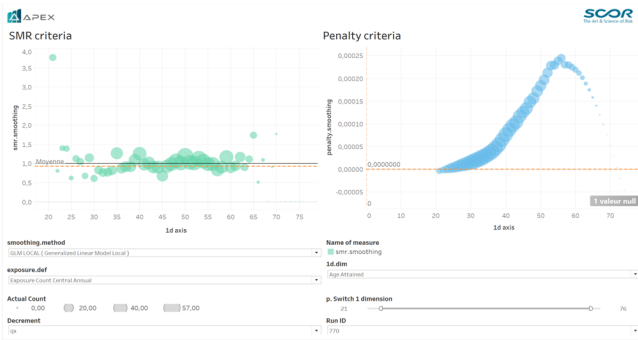
Critères		GAM	LOESS	P-Spline	GBM	GLM local
<b>Test du SMR</b>	SMR (%)	99,9 %	103,2 %	104,5%	99,9%	99,9%
	$\xi^{SMR}$	3.973	-0.001	-0.0012	3.973	3.976
<b>Déviance</b>		106.95	114.38	151.10	86.46	142.62
<b>MAPE (%)</b>		2.58 %	2.74 %	2.89 %	2.29 %	2.95%
<b>Résidus standardisés</b>	> 2	2	0	10	1	7
	> 1	36	0	43	31	44
<b>R<sup>2</sup></b>		53,39%	55,94%	52,11 %	66,09 %	52,07%
<b>Test du run</b>	Nbr runs	72	74	60	80	62
	p_valeur	0.59	0.55	0.07	0.08	0.22

\*Les résultats des méthodes de lissage adaptatives en 2 dimensions sont des résultats obtenus après avoir agrégé les âges pour obtenir une meilleure crédibilité sur le portefeuille étudié

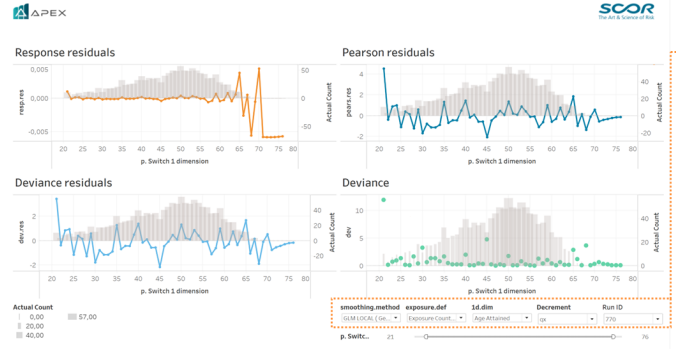
Les résultats concernant les méthodes adaptatives sont plutôt prometteurs. L'intérêt de ces méthodes est surtout d'avoir un processus standardisé pour l'analyse d'expérience, c'est-à-dire qui s'adapte en fonction des données. Ce qui est intéressant à noter, ce sont les résultats d'ajustement de la méthode GBM qui est une méthode de boosting appliquée avec des splines. On note qu'elle a un pouvoir explicatif plus fort que les autres méthodes en 2 dimensions. La difficulté et c'est tout l'intérêt du choix des termes de pénalisation et de contrainte est d'éviter que le modèle sûr ajuste, car le danger est d'avoir un modèle trop précis, avec des taux de mortalité individuels, mais considère mal la structure globale de la mortalité du portefeuille.

## Apport du mémoire et obtention des résultats

Comme présentée lors de la mise en contexte, la finalité du mémoire était de fournir, à des utilisateurs non expérimentés plusieurs méthodes adaptatives de lissage. Mais aussi d'explorer de nouvelles méthodes et de challenger les méthodes existantes. Les *dashboards* analytiques (Figure 3/ Figure 5/ Figure 4) pour les utilisateurs ont été réalisés au travers de l'application Tableau et avec le code implémenté sous R.



tableau



Possibilité de choisir à la fois des méthodes adaptatives et des méthodes calibrées avec nos propres paramètres avec "smoothing method"

tableau

FIGURE 3 – Dashboard du critère de pénalité et du SMR pour les différentes méthodes

FIGURE 4 – Dashboard de l'étude des résidus pour les différentes méthodes

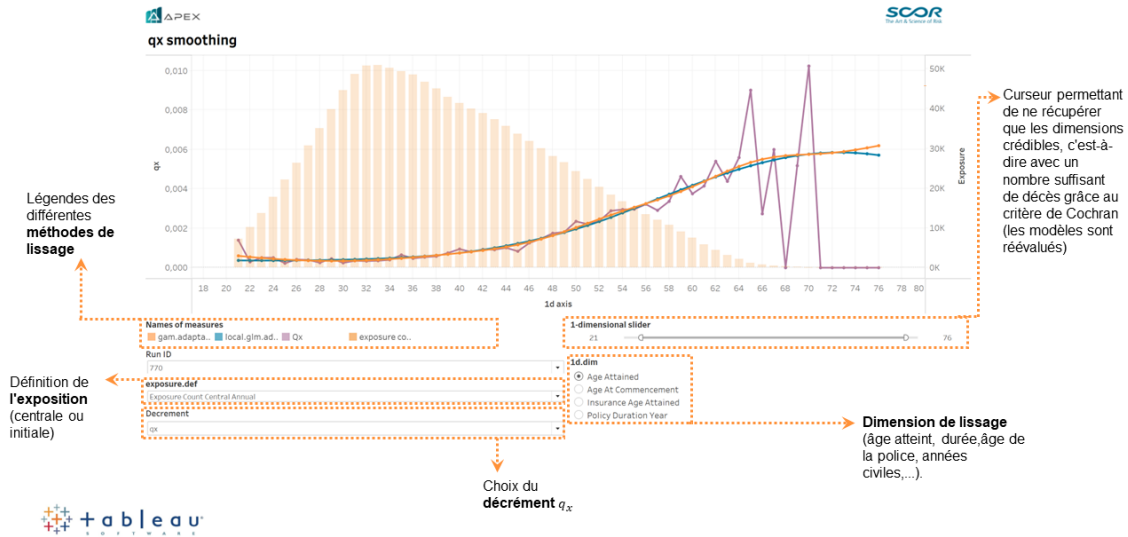


FIGURE 5 – Dashboard permettant de comparer les taux lissés  $\tilde{q}_x$  d'une partie des méthodes adaptatives présentés dans ce mémoire avec les taux bruts  $\hat{q}_x$

En ce qui concerne les résultats sous R, ce mémoire c'est appuyé sur la librairie « *locfit* » pour une partie des résultats qui concernent les méthodes locales. Mais aussi de la librairie « *mgecv* » pour les modèles GAM. La partie sur le modèle GBM Spline a pu être implémentée à l'aide de la librairie « *gamboost* » avec laquelle nous avons modifié la fonction de coût et utilisé les contraintes présentées dans le mémoire. Les méthodes RBF Network, LOESS, B-Spline, P-Spline et Whittaker-Henderson ont été implémentés par nos soins. L'ensemble des résultats de ces librairies étaient ensuite récupérés et les paramètres optimisés à l'aide des méthodes itératives présentées Section 2.10.1. Les graphiques et tests statistiques présentés dans la partie 3 ont été implémentés sans librairie.

En ce qui concerne les résultats sous Python, ils ont été implémentés à l'aide de « *pytorch* », la génération des splines c'est faite à l'aide de la récurrence de De Boor mais à cause de l'instabilité des résultats (surtout dans la partie optimisation) nous avons trouvé préférable d'utiliser les résultats sous R qui étaient plus consistants pour ce mémoire. Les taux bruts  $\hat{q}_x$  par exposition initiale et centrale sont estimés à l'aide de l'application APEX.

## Conclusion

Ainsi, cette étude démontre l'utilité des méthodes de lissage pour obtenir des primes plus cohérentes et des valeurs stables dans la construction de traités de réassurance. Le critère quantitatif GCV permet d'optimiser les paramètres des méthodes de lissage en trouvant un bon compromis entre biais et variance.

Bien que les résultats obtenus soient encourageants en une dimension, les méthodes adaptatives peuvent avoir des difficultés à capturer la structure générale de mortalité en deux dimensions avec peu de données. Néanmoins, la méthode GBM Spline offre de meilleures performances de lissage que les autres méthodes avec un coefficient de détermination ( $R^2$ ) plus élevé.

Les résultats obtenus et illustrés par la Figure 12 soulignent l'importance de choisir avec soin les méthodes de lissage pour obtenir des primes de qualité et des valeurs estimées stables en tarification. En effet, cela permet de mieux gérer le risque total dans le processus de construction d'une table d'expérience.

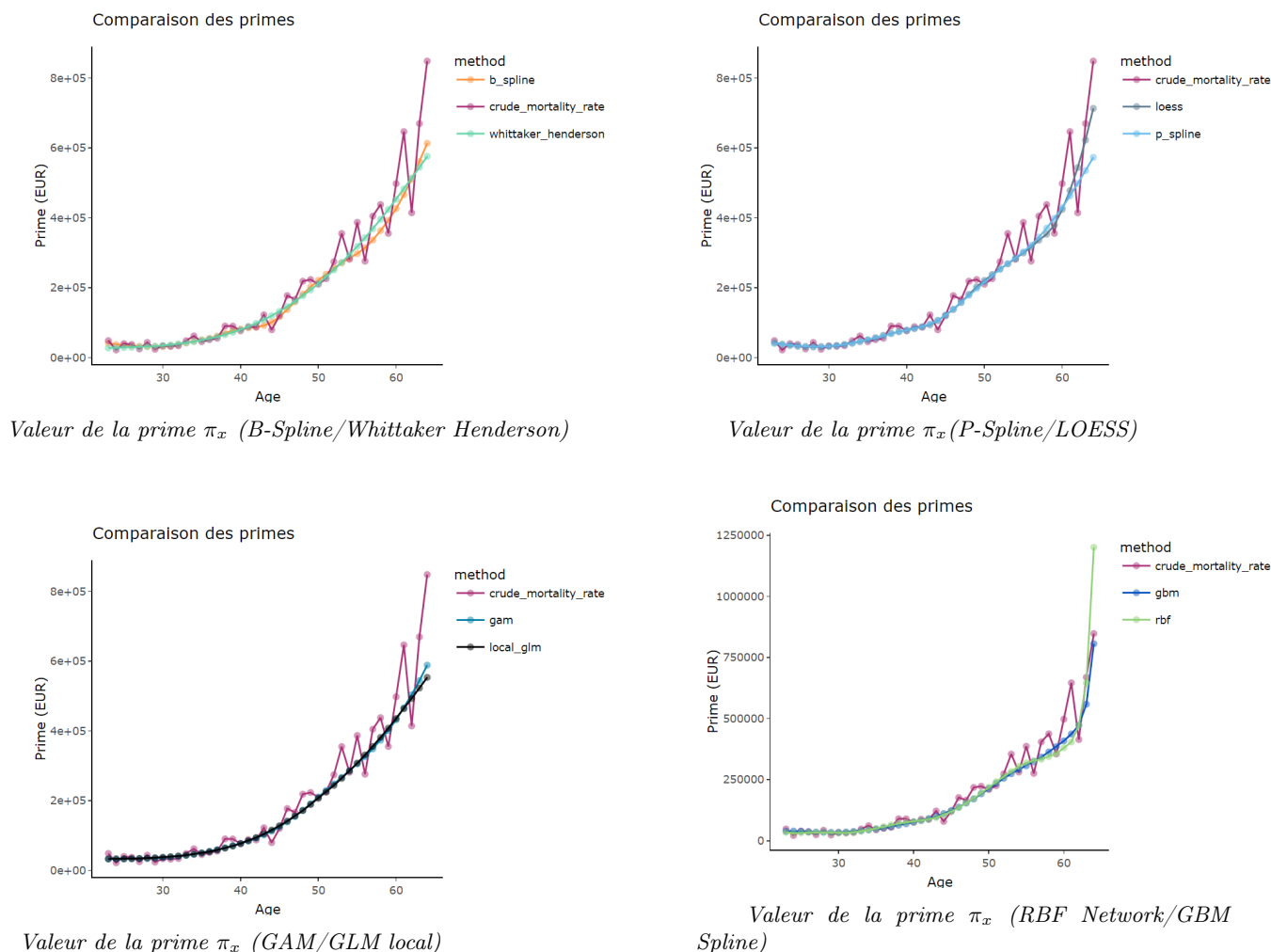


FIGURE 6 – Comparaison des primes  $\pi_x$  pour les différentes méthodes

---

# Executive Summary

## Context

This thesis is part of a more global project to build a standard experience analysis method within SCOR which aims to automate the analysis of a cedant's portfolio. The different parts of the dissertation and the path of the project will be presented in the Figure 7.

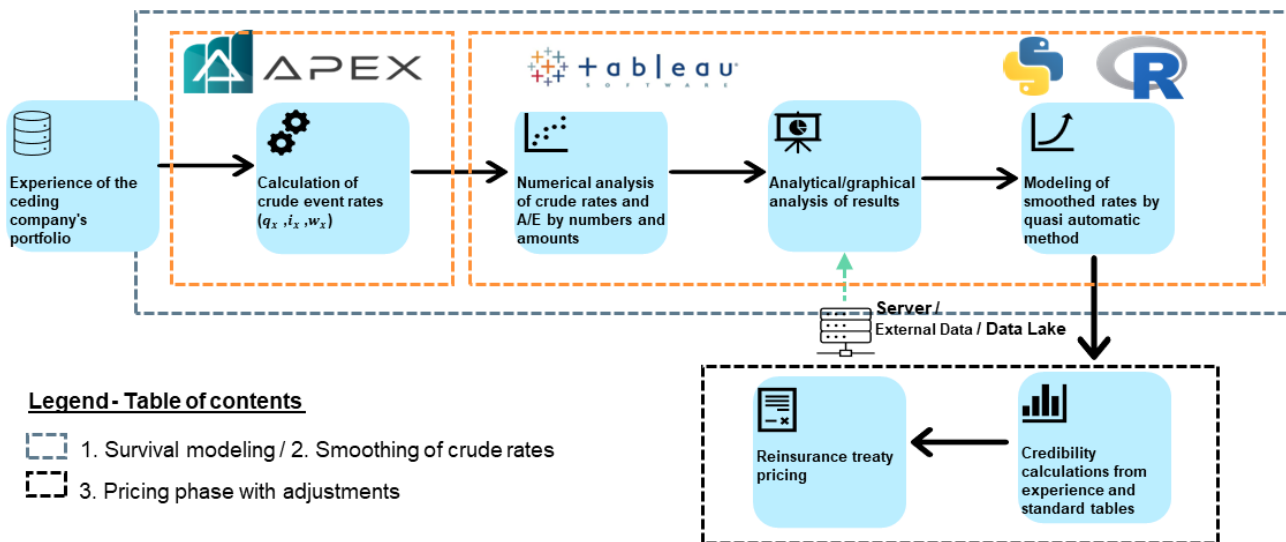


FIGURE 7 – Explanatory diagram of the study and of an experience analysis process

The aim is to try to propose adaptive methods for new users. The models will be implemented in R and Python and will be used through the Tableau application, which allows a complete analysis of analytical results.

## Models retained

This thesis will propose models for estimating crude mortality rates, based on the notion of exposure to risk. Then, we will try to answer the problem of creating standardised and adaptive methods. The models tested and selected will be diverse : classic models such as the Whittaker-Henderson smoothing method ; more sophisticated models based on additive models, neural networks or boosting and which offer the user the possibility of considering several explanatory variables to smooth the data, such as the duration of the policy, the sum insured or the type of product. We are particularly interested in testing our methods only with one dimension, age, and then with two dimensions, policy duration and age.

---

**TABLE 4** – Summary of the models used and the optimisation algorithms used for the « adaptive » principle in the context of a life table construction

Models	Advantages	Drawbacks
<b>Smoothed rate modelling   Whittaker-Henderson</b>	<ul style="list-style-type: none"> <li>• No mortality structure to be defined</li> <li>• Easy to implement/understand</li> <li>• Very interpretable</li> </ul>	<ul style="list-style-type: none"> <li>• Long calculation time</li> <li>• Limited number of variables</li> </ul>
<b>B-Spline</b>	<ul style="list-style-type: none"> <li>• No mortality pattern to be defined</li> <li>• Can be added to a GLM</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration of nodes</li> <li>• Highly sensitive</li> </ul>
<b>P-Spline</b>	<ul style="list-style-type: none"> <li>• Addition of a regularisation/penalty criteria</li> <li>• Can be added to a GLM</li> <li>• No structure to be defined</li> </ul>	<ul style="list-style-type: none"> <li>• Node calibration</li> <li>• Very sensitive</li> </ul>
<b>LOESS</b>	<ul style="list-style-type: none"> <li>• Definition of weights for neighbourhoods</li> <li>• Simplicity in optimising parameters</li> <li>• Very interpretable</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration of h</li> </ul>
<b>GLM local</b>	<ul style="list-style-type: none"> <li>• Definition of weights for neighbourhoods</li> <li>• Choice of the underlying distribution</li> <li>• Fairly interpretable</li> </ul>	<ul style="list-style-type: none"> <li>• Calibration of h</li> <li>• Optimisation of parameters</li> </ul>
<b>GAM</b>	<ul style="list-style-type: none"> <li>• No structure to be defined</li> <li>• Possibility of adding variables</li> <li>• Possibility of adding a regularisation criterion</li> <li>• Fairly interpretable</li> </ul>	<ul style="list-style-type: none"> <li>• Long calibration of parameters</li> </ul>
<b>GBM</b>	<ul style="list-style-type: none"> <li>• No structure to be defined</li> <li>• Learns from previous errors</li> <li>• Poorly interpretable</li> </ul>	<ul style="list-style-type: none"> <li>• Highly sensitive</li> <li>• Long calibration of parameters</li> </ul>
<b>RBF Network</b>	<ul style="list-style-type: none"> <li>• Can be added to a GLM</li> <li>• Poorly interpretable</li> </ul>	<ul style="list-style-type: none"> <li>• Overfitting problem</li> </ul>
<b>Optimisation algorithms   Scoring de Fisher</b>	<ul style="list-style-type: none"> <li>• No learning rate to be defined</li> <li>• Underlying distribution integration</li> </ul>	<ul style="list-style-type: none"> <li>• Scoring reversal</li> </ul>
<b>Algorithm of Newton</b>	<ul style="list-style-type: none"> <li>• No learning rate to be defined</li> <li>• Underlying distribution integration</li> </ul>	<ul style="list-style-type: none"> <li>• Inverse of the Hessian matrix</li> </ul>
<b>Quasi-Newton algorithm</b>	<ul style="list-style-type: none"> <li>• Optimization speed</li> </ul>	<ul style="list-style-type: none"> <li>• Choice of the learning rate</li> </ul>
<b>BFGS (Broyden-Fletcher-Goldfarb-Shanno)</b>	<ul style="list-style-type: none"> <li>• High speed optimisation</li> </ul>	<ul style="list-style-type: none"> <li>• Choice of the learning rate</li> </ul>

## Methodology

The optimization of the parameters of the different adaptive methods is built around the nested scheme that will optimize the GCV criterion. Moreover, the GBM Spline method relies on the addition of two constraints, a convexity constraint, but also a monotonicity constraint which prevents the boosting method from learning safely. Finally, the weak learners correspond to splines, again to obtain the smoothest possible death rate results.

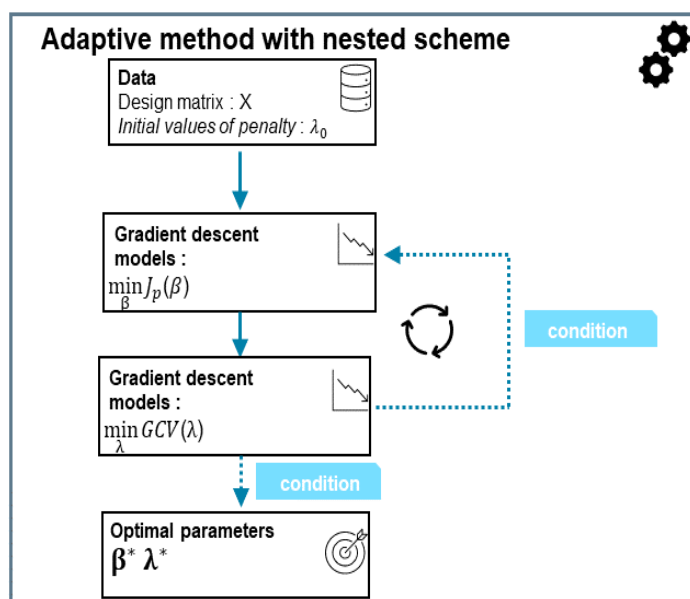


FIGURE 8 – Explanatory diagram of the nested algorithm

## Data presentation

All the methods will be tested in an example applied to the construction of a mortality table from the data of a cedant in Portugal. It is a portfolio with data from 1977 to 2013. This study only consists of 30 % of the portfolio to test our models. The idea was to challenge the proposed adaptive smoothing methods when faced with a rather low credibility. Most adaptive methods work well with a lot of data. However, in 2 dimensions and especially with little data, it is more difficult to achieve good smoothing, which is one of the reasons why we decided to choose this constraint.

This portfolio is made up of loan insurance contracts, i.e. insurance for various loans, some linked to mortgages and others linked to individual loans. This portfolio is made up of 45.1% women and 54.9% men. In this thesis, we will be interested in carrying out an experience analysis and creating a mortality table from 01/01/2007 to 31/12/2009. After application of the Cochran criterion, the ages studied will be distributed from 23 to 65 years.

## Results of adaptive methods

The results concerning adaptive methods are rather promising. The interest of these methods is mainly to have a standardised process for experiment analysis, i.e. one that adapts to the data. What is interesting to note are the fitting results of the GBM method which is a boosting method applied with splines. We note that it has a stronger explanatory power than the other methods in 2 dimensions. The difficulty and it is all the interest of the choice of the terms of penalization and constraint is to avoid that the sure model adjusts, because the danger is to have a too precise model, with individual rates of mortality, but considers badly the global structure of the mortality of the portfolio.

**TABLE 5** – Summary of 1 dimension results for adaptive smoothing methods

Critères		GAM	B-Spline	P-Spline	LOESS	WH	GLM local	GBM	RBF
<b>SMR Test</b>	SMR (%)	99,9%	99,9 %	99,9%	99,7%	98,0%	99,4%	99,9%	99,9%
	$\xi^{SMR}$	1.19	1.20	1.20	1.13	0.53	1.03	1.20	1.19
	p-valeur	0.11	0.11	0.11	0.13	0.29	0.15	0.11	0.12
<b>Deviance</b>		26.13	17.56	20.48	20.47	31.52	28.87	23.34	23.61
<b>MAPE (%)</b>		0.49 %	0.41 %	0.43 %	0.42 %	0.53 %	0.52 %	0.44 %	0.44 %
<b>Standardised residues</b>	> 2	0	0	0	0	1	0	0	0
	> 1	10	6	7	7	11	0	8	8
<b>R<sup>2</sup></b>		92.96%	93.97%	92.55%	95.59%	91.89%	91.49%	98.10%	92.64%
<b>Run test</b>	Nbr runs	23	29	29	29	23	19	19	20
	p_value	0.851	0.044	0.040	0.040	0.528	0.292	0.315	0.457

\*The results of the adaptive smoothing methods are results obtained after applying the Cochran criterion presented in Section 2.2

**TABLE 6** – Summary of the results obtained in 2 dimensions for the adaptive smoothing methods

Critères		GAM	LOESS	P-Spline	GBM	GLM local
<b>SMR Test</b>	SMR (%)	99,9 %	103,2 %	104,5%	99,9%	99,9%
	$\xi^{SMR}$	3.973	-0.001	-0.0012	3.973	3.976
<b>Deviance</b>		106.95	114.38	151.10	86.46	142.62
<b>MAPE (%)</b>		2.58 %	2.74 %	2.89 %	2.29 %	2.95%
<b>Standardised residues</b>	> 2	2	0	10	1	7
	> 1	36	0	43	31	44
<b>R<sup>2</sup></b>		53,39%	55,94%	52,11 %	66,09 %	52,07%
<b>Run Test</b>	Nbr runs	72	74	60	80	62
	p_value	0.59	0.55	0.07	0.08	0.22

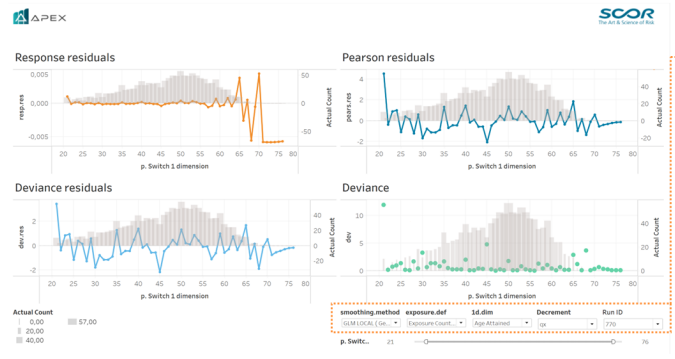
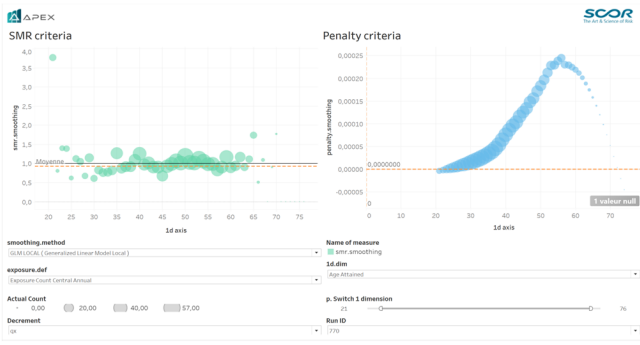
\*The results of the 2-dimensional adaptive smoothing methods are results obtained after aggregating the ages to obtain a better credibility on the studied portfolio

## Submission and results

As presented in the context, the aim of the thesis was to provide inexperienced users with several adaptive smoothing methods. But also to explore new methods and to challenge existing ones. The analytical *dashboards* (Figure 9/ Figure 11/ Figure 10) for the users were carried out through the Tableau application and with the code implemented in R.

As far as the results under R are concerned, this dissertation relied on the « *locfit* » library for part of the results concerning local methods. But also the library « *mgcv* » for the GAM models. The part on the GBM Spline model could be implemented using the library « *gamboost* » with which we modified the cost function and used the constraints presented in the thesis. The RBF Network, LOESS, B-Spline, P-Spline and Whittaker-Henderson methods were implemented by us. All the results of these libraries were then retrieved and the parameters optimised using the iterative methods presented in Section 2.10.1.





Possibility to choose both adaptive methods and methods calibrated with our own parameters with "smoothing method"



FIGURE 9 – Dashboard of penalty criteria and SMR for the different methods

FIGURE 10 – Dashboard of the residue study for the different methods

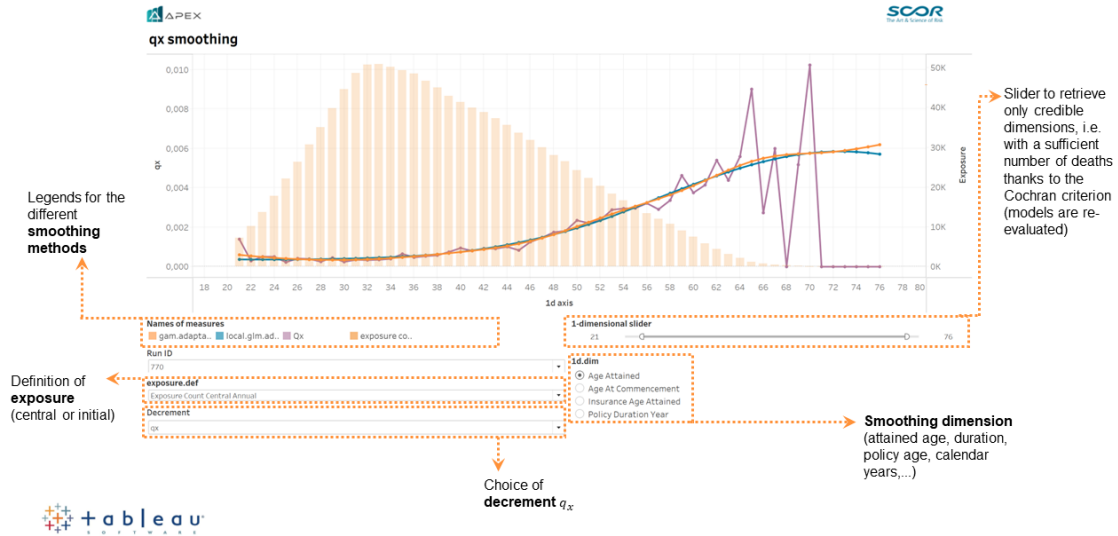


FIGURE 11 – Dashboard allowing to compare the smoothed rates  $\tilde{q}_x$  of a part of the adaptive methods presented in this thesis with the raw rates  $\hat{q}_x$

The graphs and statistical tests presented in Section 3 were implemented without libraries.

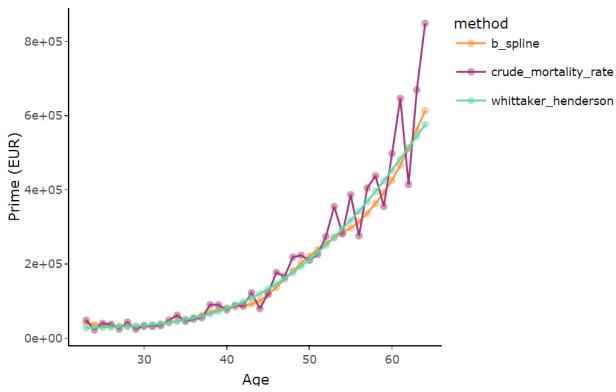
As for the results in Python, they were implemented using « *pytorch* », the generation of the splines was done using De Boor's recurrence but because of the instability of the results (especially in the optimisation part) we found it preferable to use the results in R which were more consistent for this thesis. The gross  $\hat{q}_x$  rates per initial and central exposure are estimated using the APEX application.

## Conclusion

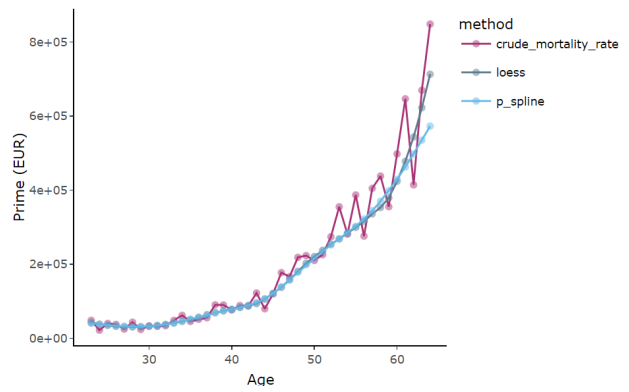
Thus, this study demonstrates the usefulness of smoothing methods to obtain more consistent premiums and stable values in the construction of reinsurance treaties. The quantitative GCV criterion allows to optimise the parameters of the smoothing methods by finding a good compromise between bias and variance.

Although the results obtained are encouraging in one dimension, adaptive methods may have difficulties in capturing the general structure of mortality in two dimensions with little data. Nevertheless, the GBM Spline method offers better smoothing performance than the other methods with a higher coefficient of determination ( $R^2$ ).

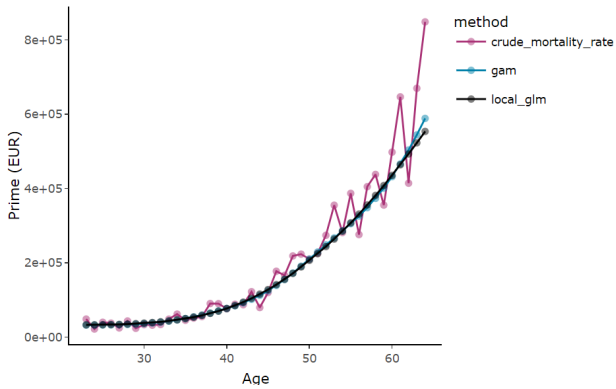
The results obtained and illustrated by Figure 12 underline the importance of carefully choosing the smoothing methods to obtain quality premiums and stable estimated values in pricing. This is because it allows for better management of total risk in the process of building an experience table.



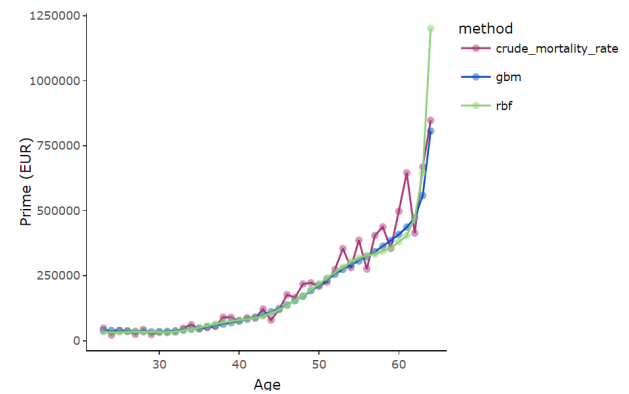
Value of the premium  $\pi_x$  (B-Spline/Whittaker Henderson)



Value of the premium  $\pi_x$  (P-Spline/LOESS)



Value of the premium  $\pi_x$  (GAM/GLM local)



Value of the premium  $\pi_x$  (RBF Network/GBM Spline)

**FIGURE 12** – Comparison of the  $\pi_x$  premiums for the different methods



---

# Remerciements

Tout d'abord, je tiens à exprimer mes sincères remerciements à mon tuteur Karsten de Braaf pour ses conseils, son suivi et son accompagnement tout au long de la rédaction de ce mémoire. Je souhaite également remercier chaleureusement toute l'équipe APEX pour leur accueil bienveillant et leur soutien quotidien, non seulement pour la réalisation de ce mémoire, mais aussi pour les missions que j'ai effectuées dans le cadre de mon alternance. En outre, je souhaiterais adresser mes remerciements à Julien Tomas pour ses explications et ses remarques toujours pertinentes sur les méthodes de lissage.

Dans un second temps, je tiens à remercier Monsieur Stéphane Loisel pour ses remarques et son accompagnement tout au long de ce mémoire, ainsi que Monsieur Frédéric Planchet pour ses apports et ses idées en ce qui concerne les méthodes adaptatives de lissage. Enfin, je suis reconnaissant envers l'ensemble du corps professoral de l'ISFA pour m'avoir permis de découvrir davantage le métier d'actuaire au fil des années.

Pour terminer, je souhaiterais exprimer ma gratitude envers ma famille, en particulier ma grand-mère, pour leur soutien constant dans mes projets et mes études. Je remercie également Marie Péliissier pour sa relecture et son soutien inconditionnel tout au long de la rédaction de ce mémoire.

---

---

# Table des matières

0.1	Contexte . . . . .	17
0.1.1	Les équipes principales en réassurance vie et santé . . . . .	17
0.1.2	Présentation des différents risques . . . . .	18
0.1.3	L'analyse d'expérience au sein de SCOR . . . . .	18
0.1.4	Méthodologie et standardisation proposées par APEX . . . . .	19
0.2	État de l'art en matière de méthode de lissage . . . . .	22
<b>1</b>	<b>Modélisation de survie</b> . . . . .	<b>24</b>
1.1	Censure et troncature . . . . .	24
1.1.1	Définition mathématique . . . . .	26
1.2	Modélisation non-paramétrique de survie . . . . .	27
1.2.1	Estimateur de Kaplan Meier . . . . .	27
1.2.2	Méthodes par discrétisation . . . . .	29
<b>2</b>	<b>Lissage des taux bruts</b> . . . . .	<b>36</b>
2.1	Introduction aux modèles de lissage de taux non-paramétrique . . . . .	37
2.2	Critère de suffisance des données . . . . .	38
2.3	Méthode de Whittaker-Henderson . . . . .	39
2.4	Méthode LOESS . . . . .	41
2.5	Méthode B-Spline / P-Spline . . . . .	43
2.5.1	Définition d'une Spline . . . . .	44
2.5.2	Spline avec base de puissance tronquée . . . . .	44
2.5.3	Récurrence de De Boor . . . . .	45
2.5.4	Méthode par B-Spline . . . . .	46
2.5.5	Méthode par P-Spline . . . . .	47
2.6	Introduction aux modèles de lissage de taux paramétrique . . . . .	48
2.7	Méthode GLM . . . . .	48
2.7.1	Les distributions de lois exponentielles . . . . .	49
2.7.2	La mise en place du GLM . . . . .	49
2.7.3	La dispersion dans le nombre de décès . . . . .	50
2.7.4	Méthode GLM avec lien Poisson . . . . .	52
2.7.5	Méthode GLM avec lien binomial négatif . . . . .	52
2.7.6	Méthode GLM avec quasi-Poisson . . . . .	53
2.7.7	GLM avec caractère local . . . . .	54
2.8	Modèles GAM . . . . .	55
2.8.1	Modèles GAM explication . . . . .	56
2.8.2	Modèle GAM univarié . . . . .	56
2.8.3	Estimation des modèles GAM univariés . . . . .	57
2.8.4	Modèle GAM bivarié et multivarié . . . . .	58
2.9	Modèle avec réseaux de neurones . . . . .	60
2.9.1	Principe des réseaux de neurones . . . . .	60
2.9.2	Les fonctions d'activation . . . . .	62
2.9.3	Modèle RBF Network . . . . .	62
2.9.4	Schéma du réseau du modèle RBF . . . . .	64
2.9.5	Rétropropagation du gradient et algorithmes d'optimisations . . . . .	65
2.10	Modèle GBM : Gradient Boosting Model . . . . .	65
2.10.1	Méthodes d'optimisation des paramètres et construction du caractère « adaptatif » de l'ensemble des techniques de lissage . . . . .	72

---

2.11	Critères d'adéquation et de choix de la méthode . . . . .	82
2.11.1	La déviance $D$ . . . . .	82
2.11.2	Étude des résidus . . . . .	83
2.11.3	La statistique du $R^2$ . . . . .	83
2.11.4	La statistique de $MAPE$ . . . . .	84
2.11.5	Le test du $run$ : . . . . .	84
2.11.6	Le test des signes : . . . . .	85
2.11.7	Le ratio SMR . . . . .	86
2.11.8	Le critère d'adéquation du $\chi^2$ . . . . .	88
2.12	Analyse de sensibilité du portefeuille étudié . . . . .	89
2.13	Obtention des intervalles de confiance des méthodes . . . . .	89
<b>3</b>	<b>Application des méthodes à un portefeuille</b> . . . . .	<b>92</b>
3.0.1	Présentation de la mortalité au Portugal . . . . .	92
3.0.2	Présentation des produits . . . . .	92
3.0.3	Présentation des données . . . . .	93
3.1	Estimation des taux bruts . . . . .	95
3.1.1	Estimation des taux bruts par exposition initiale . . . . .	95
3.1.2	Analyse du niveau de mortalité général de notre portefeuille étudié (LoM) . . . . .	98
3.2	Modélisation des taux lissés . . . . .	99
3.2.1	Analyse de la dispersion dans le nombre de décès . . . . .	99
3.2.2	Analyse graphique des taux lissés avec les méthodes de lissage en 1 dimension . . . . .	99
3.2.3	Analyse du critère de pénalité des méthodes de lissage en 1 dimension . . . . .	101
3.2.4	Analyse des résidus de Pearson pour les méthodes de lissage en 1 dimension . . . . .	102
3.2.5	Analyse des résidus de la réponse pour les méthodes de lissage en 1 dimension . . . . .	103
3.2.6	Analyse des résidus de la déviance pour les méthodes de lissage en 1 dimension . . . . .	104
3.2.7	Analyse du ratio taux bruts et taux lissés (pondéré par l'exposition) par âge des méthodes de lissage . . . . .	105
3.2.8	Analyse des ratios SMR et des valeurs du $\chi^2$ des méthodes de lissage . . . . .	106
3.2.9	Nombre de décès des méthodes de lissage en 1 dimension . . . . .	107
3.2.10	Application des tests et synthèse des résultats pour les méthodes de lissage en 1 dimension . . . . .	108
3.2.11	Résultat des méthodes de lissage adaptatives en 2 dimensions . . . . .	109
3.2.12	Analyse graphique des taux lissés avec les méthodes de lissage en 2 dimension . . . . .	111
3.2.13	Analyse des résidus de Pearson pour les méthodes de lissage en 2 dimensions . . . . .	113
3.2.14	Application des tests et synthèse des résultats pour les méthodes de lissage en 2 dimensions . . . . .	114
3.3	Comparaison qualitative des méthodes de lissage . . . . .	115
3.4	Illustration de l'intérêt des méthodes de lissage sur un exemple : . . . . .	117
3.4.1	Ajustement des niveaux de mortalité grâce à la théorie de la crédibilité . . . . .	117
3.4.2	Application dans le cadre d'un contrat . . . . .	122
	<b>Bibliographie</b> . . . . .	<b>131</b>



---

# Introduction

## 0.1 Contexte

La réassurance est couramment utilisée par les assureurs vie et santé pour gérer leur rentabilité, leur risque et leur capital, et accéder aux services fournis par les réassureurs tiers. L'objectif d'un assureur est d'aider ses assurés à gérer leurs propres risques financiers futurs qui sont incertains. Pour ce faire, il assume ces risques, en les regroupant pour les diversifier et en obtenant un rendement sur le capital requis pour soutenir l'activité. À cette fin, la stratégie et le cadre de réassurance de tout assureur peuvent avoir un impact significatif sur sa performance finale, son profil d'activité et de risque, sa position de capital et sa capacité à rester solvable et à se développer, donc à continuer à répondre aux besoins de ses assurés dans le futur. Des normes comptables et financières régissent la valorisation des produits échangés entre l'assureur et le réassureur.

Cette introduction veillera à présenter les différentes équipes impliquées dans le processus de réassurance (Section 0.1.1) en précisant le fonctionnement au sein du service d'analyse d'expérience, service dans lequel est réalisée cette étude. Il conviendra aussi d'introduire les différentes évolutions dans le cadre des méthodes de lissages, l'historique et l'état actuel en la matière. Enfin, nous nous attacherons à l'étude de la modélisation des taux bruts par exposition lors d'une analyse d'expérience.

### 0.1.1 Les équipes principales en réassurance vie et santé

- **Pricing** : Les équipes pricing s'occupent de mettre en place un prix pour un traité en travaillant en étroite collaboration avec les équipes EA (*Experience Analysis*) et R&D. Ainsi, les équipes EA leur permettent de réajuster leurs hypothèses concernant les prix et les modèles là où l'équipe R&D permet d'avoir des méthodes de calcul de plus en plus performantes en proposant des modèles de prédiction poussés. Enfin, ils collaborent avec les équipes de souscription pour avoir des prix en adéquation avec ceux du marché, correspondant réellement au business.
  - **Reserving** : Les équipes Reserving s'occupent de modéliser et de calculer différentes provisions (définies sous Solvabilité II) qui permettront au réassureur pendant toute la vie du traité de répondre à ses engagements. Les méthodes et hypothèses concernant les provisions sont revues chaque année avec les équipes EA pour éviter un défaut de la part des réassureurs.
  - **R&D** : Les équipes R&D s'occupent de mettre en place des modèles poussés pour appréhender, modéliser et permettre une meilleure gestion du risque. Cela va en créant des scénarios de tendance de 1 à 200 ans.
  - **Equipe de Souscription/Underwriting** : Les équipes de souscription sont directement en lien avec les clients pour proposer des traités adaptés à leurs risques. De plus, elles évaluent la valeur des produits en travaillant de manière étroite avec les équipes Pricing, tout cela dans le but d'éviter une anti-sélection.
  - **Analyse d'expérience (EA)** : Les équipes EA permettent une meilleure compréhension du risque et un ajustement constant des hypothèses des équipes Reserving et Pricing. Elles s'occupent de comparer les prévisions avec les valeurs actuelles en proposant des méthodes qui représentent mieux le risque.
-



### 0.1.2 Présentation des différents risques

Le risque de mortalité (*Mortality*), selon l'article 109-c de Solvabilité II correspond aux obligations d'un réassureur/assureur (assurance-vie temporaire ou vie entière) de verser une série de paiements garantie de manière récurrente ou exceptionnelle en cas de décès de l'assuré pendant la durée de la police.

Le risque de rachat (*Lapse*), dans Solvabilité II (European [17/12/2009]) est présenté de telle manière que cela correspond à l'arrêt prématuré d'un contrat d'assurance par l'assuré. Dans ce cas, l'assureur doit payer à l'assuré ou au bénéficiaire un montant qui correspond soit à la totalité (rachat total) soit à une partie (rachat partiel) de son épargne capitalisée. La police d'assurance à la souscription peut prévoir des pénalités vis à vis de ce rachat. Cette définition est utilisée pour les produits comportant un élément d'épargne ou une valeur de rachat. En pratique, on distingue des rachats les annulations dues au non-paiement de la prime.

Le risque d'invalidité (*Disability*) est un handicap provenant d'une défiance pouvant être physique, cognitive, intellectuelle, mentale, sensorielle ou une combinaison de plusieurs de ces facteurs qui a pour conséquence l'incapacité d'un individu donné de réaliser des tâches considérées comme normales. Les raisons de ce handicap peuvent être diverses : un accident, une maladie par exemple ; les manières de la mesurer le sont aussi, en termes de perte de revenus ou encore de perte de capacité.

Le risque de maladie grave (*Critical Illness*) est un risque présent dans l'assurance-maladie. En effet, les contrats d'assurance-maladie correspondent à des contrats de protection qui emportent le versement d'une somme forfaitaire à la suite d'une maladie grave, comme un cancer ou une crise cardiaque. Cela signifie que, similairement à l'assurance-vie, la prestation est fixe et n'est pas liée à un coût ou à une rente. Le nombre et la gravité des maladies couvertes varient selon les produits et les marchés. Ainsi, certains produits ne couvrent que 5 à 10 pathologies graves alors que d'autres couvrent plusieurs dizaines de maladies. Certaines polices autorisent aussi plusieurs remboursements tout au long de la durée de la police. Ce type de produit est récent, ayant été inventé en Afrique du Sud en 1983. Il est particulièrement populaire dans les pays britanniques. La possibilité de fournir des couvertures longues termes est directement liée à l'appréciation du risque et l'analyse de l'expérience sous-jacente du marché et plus particulièrement du portefeuille dans ce type de risque.

### 0.1.3 L'analyse d'expérience au sein de SCOR

L'analyse d'expérience (*Experience Analysis*) est un processus qui assiste et appuie différentes équipes tout le long de la durée d'un traité. L'offre de la cédante (**Phase 1**) permet une revue et le calcul de différentes valeurs ( $q_x$ ,  $i_x$ ,  $w_x$ ) (taux de mortalité, taux d'incidences et taux rachats) pour les équipes **Pricing** qui, ensuite, déterminent les primes associées (pure, technique, ...) et renvoient leurs informations aux équipes de souscriptions, directement en lien avec les courtiers voire directement avec les assureurs, et négocient les traités. Dans la mesure où il y a une signature du contrat, il y a un calcul des provisions associées à ce traité. L'analyse d'expérience joue encore un rôle important dans cette phase (**Phase 2**) car elle permet de comparer les  $q_x$  estimés par les équipes Pricing lors de la signature du contrat à la réalité et d'ajuster en fonction les réserves (en analysant les *Actual/Expected*, c'est-à-dire la valeur actuelle de l'expérience sur les projections de mortalité). Ce processus est ensuite renouvelé jusqu'à la fin du contrat en ajustant chaque année l'expérience du portefeuille.

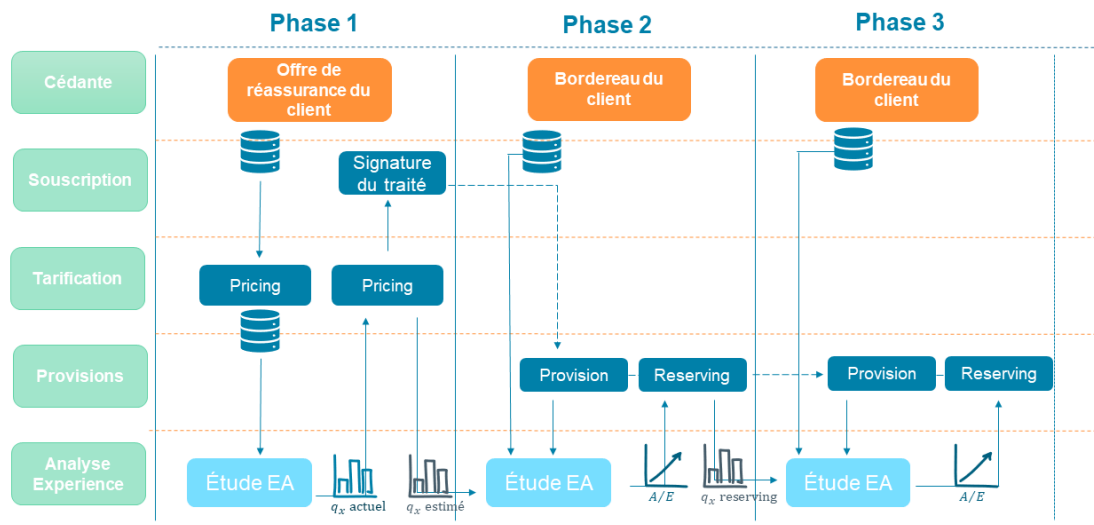


FIGURE 13 – Analyse d'expérience au sein de SCOR

Ainsi, plus généralement, l'analyse d'expérience et la construction de différentes tables (mortalités, rachats, incapacités, invalidités, ...) viennent répondre à différents besoins chez un réassureur/assureur. De nombreux assureurs vie réalisent des études de mortalité à partir de leurs propres données, dont les résultats sont souvent comparés aux références du marché. Les réassureurs/assureurs effectuent des études annuelles sur l'utilisation des sinistres, tant sur leur fréquence que sur leur gravité. Ces études annuelles sont nécessaires au suivi de l'évolution rapide de l'utilisation et du coût des services médicaux. Des études distinctes sont souvent réalisées pour les grands employeurs.

De plus, en raison du volume relativement élevé de rachats et des variations importantes d'une compagnie à l'autre, la plupart des réassureurs/assureurs vie effectuent leurs propres études de rachats de leur portefeuille plutôt que de se fier aux études de l'industrie, qui sont produites moins fréquemment et représentent de manière moins fidèle la réalité du portefeuille.

Les analyses d'expériences dans les régimes de retraite sont, quant à elles, effectuées pour fournir des renseignements à jour sur l'ensemble des employeurs afin d'établir les hypothèses d'évaluation des régimes de retraite. En raison d'une certaine difficulté à recueillir des données auprès des employeurs, ces études ont historiquement été réalisées moins souvent que les études de mortalité en assurance-vie et c'est aussi pour cela que la plupart des réassureurs réalisent leurs propres analyses d'expérience pour avoir des données plus fiables et plus récentes.

Les produits dont la part de marché est plus faible et qui comptent moins d'intervenants, comme les produits de soins de longue durée (*LTC*) ou les produits individuels d'invalidité (*DI*), ont tendance à faire l'objet d'études moins fréquentes. De plus, ils présentent des complexités que l'on ne retrouve pas dans l'assurance-vie. Il est donc préférable là aussi de construire sa propre table d'expérience en prenant des hypothèses précises qui répondent au mieux à la complexité des produits.

#### 0.1.4 Méthodologie et standardisation proposées par APEX

Pour analyser, modéliser et produire une analyse d'expérience standard à travers la majorité des risques auxquels est confronté un réassureur, mais aussi un assureur, la SCOR a mis en place un outil standard de modélisation de différents taux (décréments) à partir de l'expérience du portefeuille étudié par type de risques, APEX.

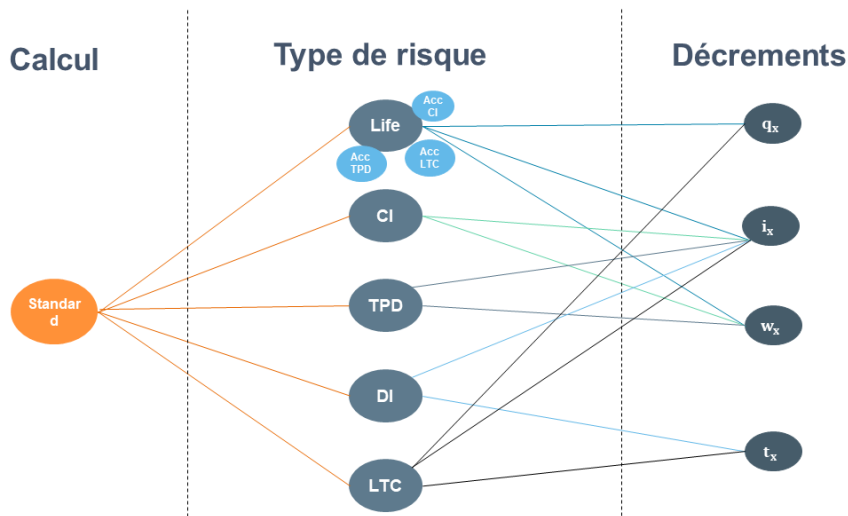


FIGURE 14 – Décomposition par type de risques sous **APEX**

Les notations sur la Figure 14 correspondent à :

- Mortalité (**Life**) : Correspond aux risques liés à la mortalité.
- Critical Illness (**CI**) : Correspond aux risques de maladie grave.
- Total and Permanent Disability (**TPD**) : Correspond aux risques liés à une blessure ou à une maladie qui vous rend incapable de reprendre un travail rémunéré.
- Disability (**DI**) : Correspond aux risques liés à une incapacité.
- Long-Term-Care (**LTC**) : Correspond aux risques liés à une dépendance.

Ce moteur de calcul permet de retourner aux différentes équipes impliquées lors de l'évolution d'un traité les taux bruts ( $q_x, i_x, w_x, t_x$ ), qui correspondent respectivement aux taux de mortalité, taux d'incidences, taux de rachats et taux de cessation de la demande issus de leur expérience. Cela permet aussi de proposer aux cédantes et aux différents clients une analyse complète de leur portefeuille et d'avoir une confiance accrue. Les motivations de cette étude et de ce mémoire et de façon plus générale pour le projet **APEX** sont :

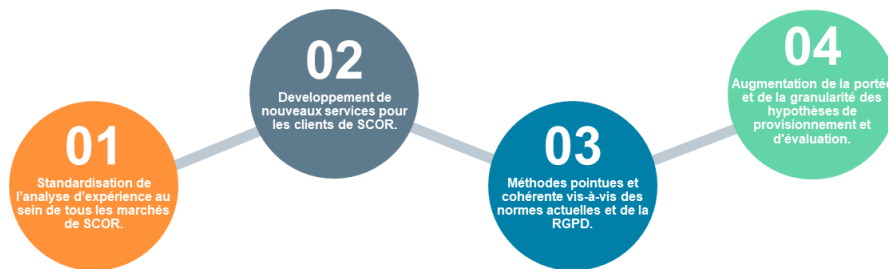


FIGURE 15 – Motivation et points clés derrière cette étude et **APEX**

La Section 1.1.1 présentera la manière dont sont estimés les taux étudiés, à savoir les taux de décès,  $q_x$ , reliés au risque de mortalité. Cependant, la majeure partie de ce mémoire sera consacrée à la présentation et à la comparaison des différentes méthodes pour lisser ces taux de la manière la plus judicieuse en fonction des critères étudiés, mais aussi de nos données (variables explicatives, exposition, ...). On proposera aussi des méthodes automatiques de choix de paramètres dans le cadre d'utilisation par des nouveaux utilisateurs, tout en les confrontant les unes aux autres.

Plus généralement, l'analyse d'expérience et la construction de tables d'expérience/mortalité dans le cadre d'un processus de réassurance visent à réduire le risque total qui peut se décomposer de la manière suivante en considérant plusieurs facteurs qui influent sur le risque total et dont les équipes doivent tenter de minimiser par le biais de plusieurs actions, cette notion de risque dans la construction de tables de mortalité est présentée plus amplement dans la thèse de Kamega [2011] :

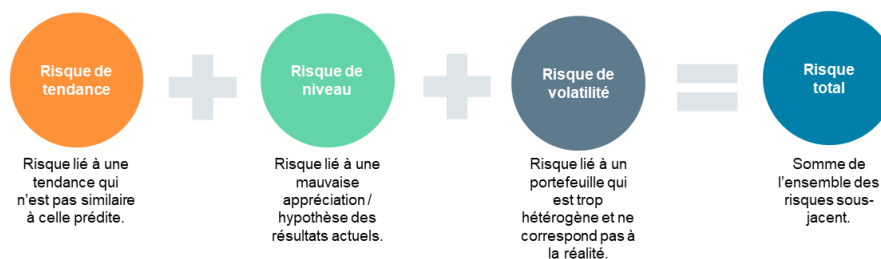


FIGURE 16 – Décomposition du risque

La construction de sa propre table d'expérience/mortalité permet à la fois de diminuer la volatilité, ainsi que le risque de niveau en proposant des taux plus proches de la réalité du marché et de notre portefeuille. Elles influent aussi indirectement sur le risque de tendance, car c'est avec les nouvelles hypothèses que seront estimés les taux projetés avec des ajustements en fonction du lissage réalisé ou de tables standards. Cela nous permet de réduire le risque total de modélisation et d'avoir une prime pure relativement plus proche de la réalité critère essentiel dans le cadre du marché compétitif de la réassurance.

Il est primordial d'avoir certaines données pour construire une table d'expérience. Les données nécessaires varient en fonction du problème et du risque étudiés. En théorie, les données à analyser pour les études d'expérience doivent être homogènes. Dans la réalité, c'est très difficile. En pratique, les données doivent être subdivisées en fonction des facteurs qui ont, par expérience, un effet significatif sur la mortalité. Cette approche est réalisable si les informations appropriées sont disponibles et si nous disposons de suffisamment de données pour rendre possible une analyse aussi détaillée. Dans le cadre de ce mémoire, nous nous contenterons d'étudier des modèles à 1 ou 2 dimensions, âge et durée ou duration de la police. Les statistiques de mortalité/morbidité de l'assurance vie qui sont prises en comptes dans le cadre de l'application proposée et qui sont généralement étudiées sont subdivisées par :

- Sexe (il n'est pas pris en compte dans le cadre de la tarification)
- Age
- Type de produit
- Statut de fumeur/non fumeur
- Médical/non médical (basé sur le niveau de souscription)
- Canal de distribution
- Profession de l'assuré
- Somme assurée

Une fois le rapport d'analyse de l'expérience achevé et la table d'expérience construite, les résultats doivent être pris en compte pour produire une base de tarification affinée. Un élément majeur de ce processus consiste à déterminer le niveau de crédibilité à attribuer à l'expérience. On définit la « crédibilité » comme une mesure de la valeur prédictive pour une application précisée que l'actuaire attache à un ensemble particulier de données (le terme prédictif est utilisé ici dans le sens statistique du terme et non dans le sens de prédire l'avenir). La crédibilité totale est le niveau auquel l'expérience du sujet se voit attribuer une valeur prédictive totale sur la base d'un intervalle de confiance sélectionné. L'application la plus courante de la théorie de la crédibilité consiste à évaluer comment l'expérience passée peut être combinée aux attentes actuelles en matière de mortalité pour obtenir une hypothèse de mortalité actualisée et un contrat de réassurance tarifé de manière plus précise. La Partie 3 de cette étude permettra, au travers d'un exemple, plus en détail d'expliquer le fonctionnement de ce processus.

## 0.2 État de l'art en matière de méthode de lissage

Certaines techniques élémentaires de lissage des données sont apparues au cours du XVIII<sup>e</sup> siècle. À cette époque, les techniques de lissage consistaient en une simple interpolation des données. Elles ont finalement évolué vers des méthodes modernes plus complexes. La réflexion actuarielle sur le lissage, ou la graduation, des estimations des taux de mortalité en fonction de l'âge a commencé à se développer dans les années 1820. Concrètement, deux types d'approches ont vu le jour : une approche de lissage non paramétrique qui a défini le taux de mortalité d'un âge donné comme une moyenne pondérée des taux de mortalité « bruts » estimés sur une plage d'âges centrée sur l'âge donné (méthode par moyenne mobile); et une approche paramétrique qui a spécifié une forme fonctionnelle pour la variation des taux de mortalité en fonction de l'âge (et potentiellement de la durée de la police), puis a ajusté les paramètres de cette fonction aux taux de mortalité observés. Ces deux approches visaient le même objectif, mais étaient philosophiquement différentes. La dernière approche présentée visait à fournir une « loi de la mortalité » explicative qui soit cohérente avec les données, tandis que la première tentait simplement d'éliminer le « bruit » ou encore les

« résidus » dans les échantillons statistiques des taux de mortalité.

Dans un contexte plus large que le simple contexte actuariel, c'était une ère d'exploration dans l'application des approches statistiques aux sciences sociales et comportementales. Ensuite on a vu l'émergence des modèles linéaires généralisée en 1989 McCullagh et Nelder [1989]. Cela a alors permis d'étendre les distributions de lois de probabilités prises en considérations pour nos modèles, autres que la loi normale dans le cadre des modèles linéaires. On observe les prémices de modèles linéaires généralisés publiés par Hastie et Tibshirani [1986], l'idée sous-jacente consistant à étudier des relations qui ne sont pas seulement linéaires entre nos variables explicatives et notre variable de réponse.

Plus tard, Wood [2006] publie la première édition de son ouvrage qui s'est imposée comme l'une des principales références en matière de modèles GAM (*Generalized Additive Model*) ou Modèles Additifs généralisés en proposant l'utilisation de splines pénalisées pour construire cette fonction non linéaire. Nous nous attacherons ainsi, lors de cette étude, à présenter à la fois des méthodes de lissage paramétrique et non paramétrique, tout en opposant des méthodes plus récentes qui font intervenir des notions de boosting et de réseaux de neurones à celles plus anciennes présentées précédemment dont on connaît la fiabilité et la robustesse. Ces méthodes de lissage visant à être standardisées et quasi automatisées pour des nouveaux utilisateurs, nous avancerons des méthodes d'optimisation de paramètres tout en proposant des critères pour leurs choix.

---

# Chapitre 1

## Modélisation de survie

La réassurance comme l'assurance cherchent toutes deux à garantir à leurs bénéficiaires un montant forfaitaire lors de la survenance d'un événement. En réassurance dans le cadre de traité réalisé avec les assureurs et en assurance dans le cadre d'un assureur avec ses assurés. Les normes comptables et réglementaires internationales accordent beaucoup d'importance dans le calcul du bilan à la durée de survie. Il faut donc, pour ces deux entités, (assureur et réassureur) modéliser le temps de survie avant l'arrivée d'un sinistre, pour calculer au mieux la valeur de la prime payée par l'assureur en cas de réassurance comme pour estimer les provisions nécessaires lors de la survenance de celui-ci.

Les acteurs du domaine doivent donc composer avec des données de durée. Ils collectent ces données en continu et les étudient sur une période déterminée. Nous présenterons au travers de cette partie les modèles et phénomènes intrinsèques reliés à l'étude et à la modélisation de durée.

### 1.1 Censure et troncature

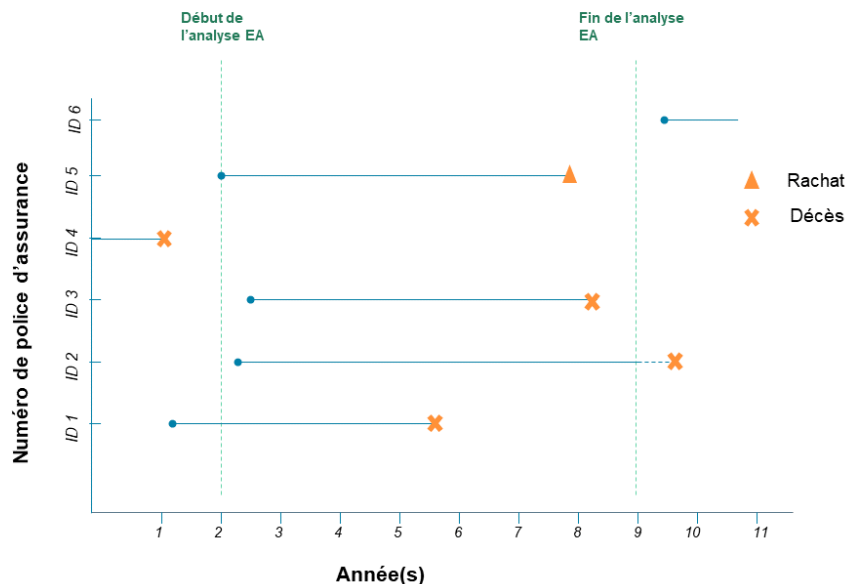


FIGURE 1.1 – Schéma explicatif de la notion de censure et de troncature

Une étude est réalisée pendant une durée, la durée d'observation, qui a une date de début et une date de fin. L'idée lors de cette étude est d'observer la survenance d'un événement particulier

---

et d'étudier la manière dont réagit notre portefeuille. Cependant entre cette date de début et de fin nous avons seulement une certaine période de temps, durant laquelle les assurés peuvent régir de plusieurs manières : certains ne sont pas observés pendant la période d'étude, certains souscrivent une police d'assurance pendant cette durée, certains rachètent une partie de leur contrat d'assurance avant la fin l'étude, d'autres décèdent ou tombent malades pendant la durée d'observation là où certains n'ont aucun événement majeur.

Cette liste non exhaustive d'événements nous montre que d'une police d'assurance à une autre nous observons des niveaux « d'information » différents et des données qui sont incomplètes. Nous ne pouvons donc pas modéliser et étudier notre expérience avec le même modèle pour toutes les polices, nous devons prendre un compte et adapter nos modèles en fonction de ces données incomplètes. On définit ces phénomènes de données incomplètes ou inobservées par **censure** et **troncature**.

#### **La notion de censure :**

Dans le cadre d'une analyse d'expérience liée à la survenance d'un événement de mortalité, on dira qu'il y a **censure à droite** lorsqu'on observe toute la police d'assurance, mais que l'événement de décès ne survient pas ou survient après la date de fin d'observation. Ce cas correspond sur notre Figure 1.1 aux polices ID 5 et ID 2 car pour la police ID 5 nous observons que l'individu quitte la police, mais ne décède pas. L'événement correspondant est donc un rachat. Pour ce qui est de la police ID 2 l'événement correspondant est un décès, mais qui survient après la fin de l'observation : nous ne connaissons alors pas la durée exacte jusqu'à la survenance du décès.

La **censure à gauche** correspond au fait qu'on observe pas toute la police d'assurance pendant la durée de l'étude. En effet on observe que la police ID 1 est présente avant l'analyse d'expérience quand bien même l'événement de décès se produit pendant la durée d'observation, nous n'avons là pas, là non plus, la durée exacte jusqu'à la survenance du décès car avant même le début de l'étude EA (*Experience Analysis*) l'individu est d'ores et déjà présent dans le portefeuille.

#### **La notion de troncature :**

La troncature est différente de la censure. Ainsi, dans le cadre de la censure, nous avons une information dite incomplète, alors que dans le cadre de la troncature nous n'avons tout simplement pas d'information concernant la police d'assurance.

La **troncature à droite** correspond à la police d'assurance ID 6 : l'événement de départ qui est censé l'inclure dans l'étude EA se produit après la fin de l'observation. Dans le cadre de notre exemple cela peut-être car cet individu a souscrit à la police d'assurance après la fin de l'étude.

La **troncature à gauche** en est l'opposé, ce qui signifie que l'événement étudié s'est déjà produit avant même le début de l'étude EA. Cela correspond dans notre exemple à la police d'assurance ID 4. Les individus sont d'ores et déjà décédés avant même le début de l'observation et on dira que l'information les concernant est tronquée à gauche.

Les deux phénomènes les plus répandus en assurance sont la censure à droite et la troncature à gauche.



### 1.1.1 Définition mathématique

#### Censure : formalisme mathématique

Dans le cadre de notre étude et de l'estimation au sein d'une analyse d'expérience dans le but de fournir des tables prospectives, nous prendrons en compte seulement la censure à droite. Pour la modéliser, la convention est d'utiliser une variable  $T$  qui correspond à une durée de survie observée sur notre Figure 1.1. Cela correspond aux différents traits. Nous introduisons aussi  $C$  qui correspond à l'instant de censure, instant qui peut être **aléatoire** ou **fixe**. Dans le cas où  $C$  est **aléatoire**, l'instant de censure sera une variable aléatoire. Dans le cas de notre étude EA, comme nous étudions les différents risques (rachat, maladie grave, mortalité) séparément puis recroisons nos taux à la fin, nous considérerons que la censure est fixe dans le cadre de l'étude d'un risque précis.

On pose ainsi :

$$\delta_i = \mathbb{1}_{T_i \leq C} \quad Y_i = \min(T_i, C)$$

Avec,  $\delta_i$  qui vaut 1 dans le cas où on observe l'événement étudié avant la censure, on a une **information complète**. Dans le cas où sa valeur serait égale à 0, on observe l'événement étudié après la censure, on a une **information incomplète**.

#### La force de mortalité

**La force de mortalité** dans le cadre de l'étude de la mortalité ou plus généralement dans le cas d'une étude de durée qu'on appelle **la fonction de hasard**, correspond à l'effet instantané de la survenance du risque étudié (rachat, mortalité, incidence) à un certain âge ou à un certain moment donné.

Instinctivement, on définira cette valeur en termes de variation du nombre « survivant » lié à un événement donné par :

$$\mu(t) = \frac{1}{E_t} \frac{dE_t}{dt} \quad (1.1)$$

Avec  $E_t$  qui correspond au **nombre de survivants à l'instant  $t$**  et  $\frac{dE_t}{dt}$  qui correspond à la variation du nombre de survivants par rapport au temps  $t$ , c'est taux de diminution de la survie par rapport à un événement donné sur un incrément de temps infiniment petit.

Une autre définition possible revient à utiliser une fonction, **la fonction de survie**. Elle traduit la probabilité que la durée de survie dépasse l'instant  $t$ . Cette fonction est définie de la manière suivante avec  $F$  la fonction de répartition liée à la variable aléatoire  $T$  :

$$S(t) = (1 - F(t)) = \mathbb{P}(T > t) \quad (1.2)$$

Il est alors possible de décrire la force de mortalité en utilisant la fonction de survie :

$$\mu(t) = \lim_{\Delta t \rightarrow 0} \mathbb{P}(T \leq t + \Delta t | T > t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{S(t)} = \frac{-S'(t)}{S(t)} \quad (1.3)$$

On interprétera cette valeur comme la probabilité que l'événement étudié dans notre analyse se produise dans un petit intervalle de temps après  $t$  sachant qu'il ne s'est pas produit avant  $t$ . On déduit ensuite de l'équation 1.3 une relation entre **la fonction de survie** et **le taux de hasard** :

$$S(t) = \exp\left(-\int_0^t \mu(u) du\right) = \exp(-\Lambda(t)) \quad (1.4)$$

Avec,  $\Lambda(t)$  qui correspond à la fonction de risque de hasard cumulé

### Vraisemblance en présence de censure à droite

En présence de censure à droite on obtient la vraisemblance suivante :

$$L = \prod_i (\mu(t_i) S(t_i))^{\delta_i} \cdot (S(t_i))^{1-\delta_i} = \prod_i \mu(t_i)^{\delta_i} S(t_i) \quad (1.5)$$

- $\mu(t_i) S(t_i)$  : Correspond à la **densité des individus dont on a observé l'événement à la date  $t_i$**  (décès, rachat, maladie), c'est à dire les individus dont on a totalement observé l'information. On sait que l'événement se produit à  $t_i$ . Ce type de cas correspond sur la Figure 1.1 à **ID 1** et **ID 3**.
- $S(t_i)$  : Correspond aux individus censurés, c'est à dire dont la seule connaissance qu'on en a est que l'événement que l'on souhaite observer ne s'est pas produit et qu'il s'est produit après  $t_i$ . On ne sait finalement que le fait que l'événement se produit donc après  $t_i$ . Ce type de cas correspond sur la Figure 1.1 à **ID 5** et **ID 2**.

## 1.2 Modélisation non-paramétrique de survie

### 1.2.1 Estimateur de Kaplan Meier

Pour estimer la **fonction de survie** présentée dans la partie précédente on pourrait calculer cette valeur de manière empirique :

$$\hat{S}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{t_i > t} \quad (1.6)$$

Cependant nous ne prenons pas en compte la censure présente dans notre étude, car en effet dans le cas de l'estimation empirique on sur estimera notre fonction de survie. Par exemple on n'observera pas les individus correspondants à la police **ID 5** sur la Figure 1.1. On observe avec cette méthode empirique la variable aléatoire  $T_i$  et non  $Y_i$ . Une proposition d'estimateur en présence de données incomplètes est alors proposé par Kaplan et Meier [1958]. Ce modèle s'appuie sur la théorie Bayésienne, c'est à dire que l'on considère les observations passés pour estimer la valeur présente.

On part du postulat que survivre après un temps  $t$  c'est ne pas mourir avant  $t$  et être en vie juste avant, de cette manière on remonte en conditionnant par les données passées jusqu'à obtenir un estimateur de la **fonction de survie**. En effet on a en introduisant un temps infinitésimal qui sera caractérisé ici par  $\epsilon$  :

$$\begin{aligned} \forall \epsilon > 0, \quad S(t + \epsilon) &= \mathbb{P}(T > t + \epsilon) \\ &= \mathbb{P}(T > t + \epsilon | T > t) \mathbb{P}(T > t) \\ &= (1 - \mathbb{P}(T < t + \epsilon | T > t)) S(t) \end{aligned} \quad (1.7)$$

En faisant tendre  $\epsilon \rightarrow 0$  on obtient :

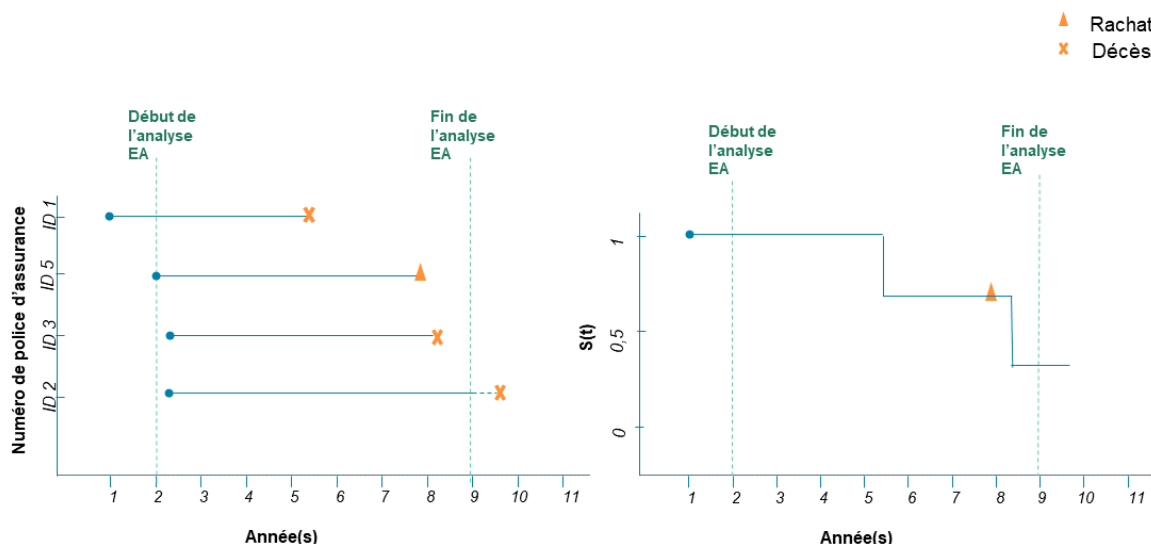
$$S(t^+) = (1 - \mathbb{P}(T < t^+ | T > t^-)) S(t^-) \quad (1.8)$$

Ceci nous permet d'obtenir l'estimateur de la fonction de survie, de telle manière que :

$$\begin{aligned}
 \mathbb{P}(T < t^+ | T > t^-) &= \frac{S(t^-) - S(t^+)}{S(t^-)} \\
 &= \frac{\text{Nombre d'évènements observés au temps } t}{\text{Nombre de personne observées au temps } t^-} \\
 &= \frac{d_t}{E_{t^-}}
 \end{aligned}
 \tag{1.9}$$

Ce résultat nous permet, en posant une série temps consécutifs croissants d'évènements  $t_i$ , ( $\forall i \in \{1, \dots, n\}$ ) (décès, rachats, censures), d'obtenir la fonction de survie de Kaplan Meier :

$$\begin{aligned}
 \hat{S}(t) &= \prod_{t_i < t} \mathbb{P}(T > t_{i+1} | T > t_i) \\
 &= \prod_{t_i < t} \left(1 - \frac{d_{t_i}}{E_{t_i}}\right)
 \end{aligned}
 \tag{1.10}$$



**FIGURE 1.2** – Représentation de l'estimateur de la fonction de survie de Kaplan Meier avec les données de survie de la Figure 1.1

L'estimateur de Kaplan Meier nécessite un nombre de données qui est large. La qualité de l'approximation dépend fortement du nombre d'observations (correct si  $n \gg 1000$ ). De plus, les observations censurées fournissent beaucoup moins d'informations que les temps d'incidences observés. Ainsi, dans le cadre d'une analyse d'expérience, cela peut poser problème. En effet certains groupes de portefeuilles sont découpés en fonction de leurs caractéristiques ce qui peut réduire le nombre de données, ce qui emporte la réduction de la qualité de prédiction du modèle, car nous devrions l'utiliser pour chaque groupe de portefeuilles.

La mise en place de cette méthode peut parfois être difficile sur le plan opérationnel, car elle

requiert une bonne connaissance de la chronologie des événements et des censures. Toutefois, elle présente de nombreux avantages, étant convergente, sans biais, cohérente et facilement interprétable.

Enfin, cela permet de prendre en compte la censure dans le cadre d'une analyse d'expérience, évitant ainsi une sous-estimation de la durée de vie individuelle, ce qui est un désavantage de l'estimation empirique.

### 1.2.2 Méthodes par discrétisation

Une méthode alternative et plus opérationnelle en comparaison à la méthode présentée précédemment et aussi à celle utilisée par APEX consiste à discrétiser les temps et à considérer des expositions différentes en cas d'événement. La principale différence entre les deux méthodologies est l'hypothèse sur la distribution des événements dans l'intervalle d'étude  $j$  de l'analyse d'expérience noté  $I_j = [\tau_1; \tau_j]$

Prenons par exemple un individu avec les caractéristiques suivantes :

Date de naissance	Date de début de condition	Date de fin de condition	Cause
13/04/1961	15/06/2012	15/07/2017	Décès

En considérant que nous réalisons une analyse d'expérience de notre portefeuille entre les temps  $\tau_j = 01/01/2015$  (date de début d'observation) et  $\tau_{j+1} = 31/12/2018$  (date de fin d'observation).

On a donc le schéma suivant :

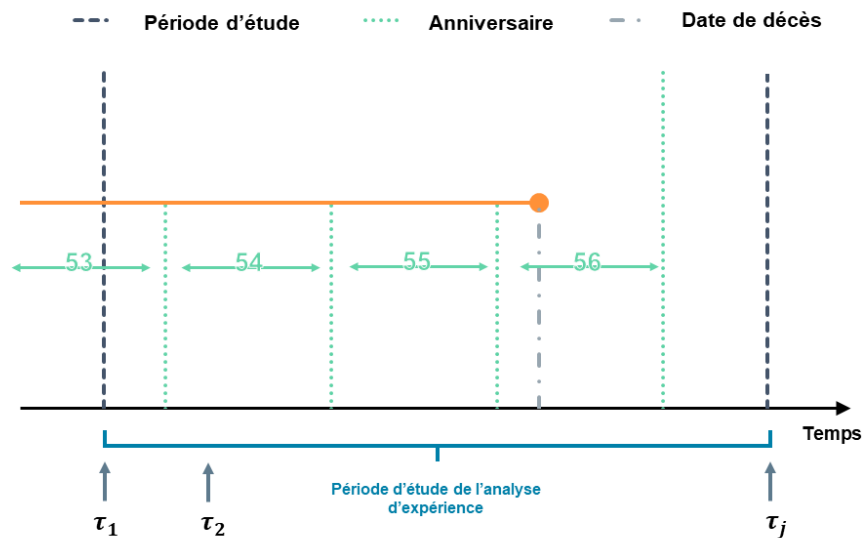


FIGURE 1.3 – Schéma de l'exposition dans le cadre de l'exemple introductif

Un estimateur instinctif pour calculer le taux de mortalité brut à un âge  $x$  pour cette étude

serait de considérer la valeur suivante :

$$\hat{q}_x = \frac{d_x}{N_x} = \frac{N_x - N_{x+1}}{N_x} \quad (1.11)$$

Avec,  $d_x$  le nombre de décès à l'âge  $x$  et  $N_x$  le nombre d'individus vivants à l'âge  $x$ . Cependant cette méthode omet les décès de notre portefeuille liés à la censure mais aussi à la troncature. En effet dans l'exemple présenté précédemment notre individu aurait pu résilier son contrat le 20/05/2017. Cependant dans ce cas de censure nous n'aurions pas pris en compte son décès dans notre estimation alors qu'il décède le 15/07/2017 à l'intérieur de la période d'étude. L'estimateur instinctif est donc biaisé à cause du phénomène de censure et troncature.

On voit bien les limites de cette méthode, une alternative consiste donc à considérer l'exposition au risque de l'individu. On va pondérer les individus censurés par leur durée de présence dans le portefeuille. On considérera aussi différentes hypothèses quant à la distribution de la mortalité sur l'année. Cela signifiera que notre force de mortalité sur un intervalle de temps (annuel, mensuel, journalier) sera différente. On traitera dans ce mémoire de deux méthodes : l'exposition centrale et initiale. On proposera aussi des intervalles de confiance suivant les recommandations de la ligne directrice de mortalité.

### Exposition initiale et hypothèse de Balducci

En reprenant les notations précédentes, c'est-à-dire dans le cas où nous souhaitons réaliser une étude d'analyse d'expérience qui débute à  $\tau_1$  et se termine à  $\tau_2$ . On étudiera notre individu d'âge  $x + \tau_1$  au début de l'étude (dans notre exemple Section 1.2.2 d'âge  $53 + \tau_1$ ) et à la fin de notre analyse d'expérience d'âge  $x + \tau_2$  (dans notre exemple Section 1.2.2 d'âge  $53 + \tau_2$ ). On s'intéressera alors à étudier notre individu dans l'intervalle d'observation suivant :  $[x + \tau_1; x + \tau_2]$ .

On observe donc que pour cet individu on a une exposition partielle pour les âges de 53 et 56 ans, une exposition totale pour les âges de 54 et 55 ans et un décès à l'âge de 56 ans. En reprenant les indications fournies par la Commission d'Agrément [2006] et en ignorant la présence de troncature, on pose les notations suivantes :

- $c_i$  : L'instant de rachat (censure) pour l'individu  $i$  dans l'intervalle d'observation
- $t_i$  : L'instant de décès pour l'individu  $i$  lors de l'étude dans l'intervalle d'observation
- $W_x$  : Le nombre de rachats ou d'individus censurés à l'âge  $x$  dans l'intervalle d'observation
- $N_x$  : Le nombre d'individus vivants à l'âge  $x$ , cela correspond à la taille de la population à l'âge  $x$  dans l'intervalle d'étude
- $ec_x$  : L'exposition centrale à l'âge  $x$
- $ei_x$  : L'exposition initiale à l'âge  $x$

Cette étude se schématisera Figure 1.4 avec les notations précédentes.

On a donc que  $W_x$  correspond à une somme de fonctions indicatrices, qui correspondent à un rachat ou non des individus étudiés. On comptera alors le nombre de personnes dans notre portefeuille grâce à cette indicatrice qui rachètent leurs contrats.

$$W_x = \text{Card}(\{c_i < \tau_2\} \quad , \forall i \in \{1, \dots, N_x\}) \quad (1.12)$$

Grâce à ces notations, on peut estimer et approximer le nombre de décès dans l'intervalle d'observation  $j$  à l'âge  $x$  c'est-à-dire dans l'intervalle  $[x + \tau_1; x + \tau_2]$ , en obtenant ainsi le nombre de décès total avec prise en compte de censure à droite dans notre portefeuille étudié :

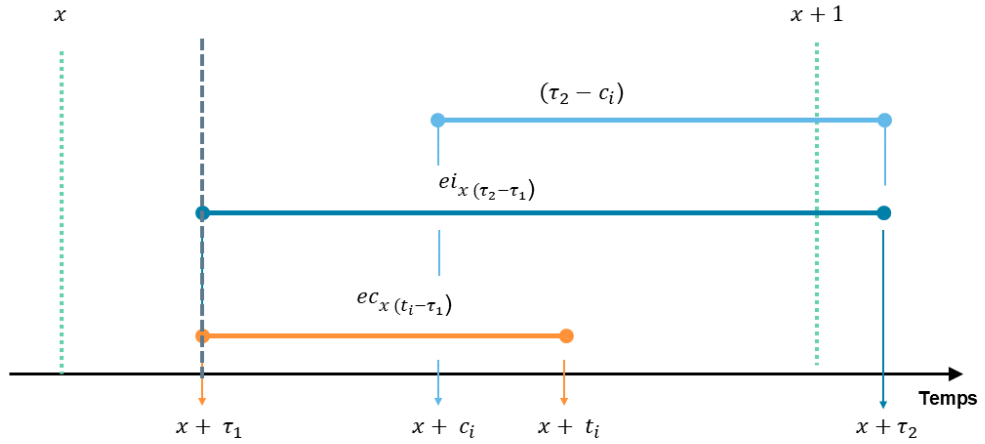


FIGURE 1.4 – Schéma explicatif de l'étude pour l'âge partielle 53 ans avec l'exposition centrale et initiale

$$D_x = \underbrace{\sum_{i=1}^{N_x} \mathbb{P}(T_{i,x} \leq \tau_2 | T_{i,x} > \tau_1)}_{\text{Nombre de décès réellement observé}} - \underbrace{\sum_{i=1}^{W_x} \mathbb{P}(T_{i,x} \leq \tau_2 | T_{i,x} > \tau_1 + c_i)}_{\text{Nombre de décès inobservé à cause de la censure à droite}} \quad (1.13)$$

Pour des raisons de lisibilité, on exprime la variable aléatoire  $T_{i,x}$  de la manière suivante :  $T_{i,x} = T_i - x$ , variable aléatoire qui représente le temps de sortie de l'individu  $i$  d'âge  $x$ . L'équation 1.16 porte le nom de **relation de cohérence**. On l'écrira dans la suite du mémoire plutôt de la manière suivante :

$$D_x = \sum_{i=1}^{N_x} \tau_2 - \tau_1 q_{i,x+\tau_1} - \sum_{i=1}^{W_x} \tau_2 - c_i q_{i,x+c_i} \quad (1.14)$$

Pour pouvoir résoudre l'équation de cohérence on doit réaliser une hypothèse sur la distribution sous-jacente de nos taux de mortalité. Dans le cadre de la méthode par exposition initiale on réalise l'hypothèse dite de Balducci. C'est à dire que le taux de mortalité diminue dans l'intervalle d'étude de la manière suivante :

$$\begin{aligned} \tau_2 - c_i q_{x+c_i} &= \mathbb{P}(T_x \leq \tau_2 | T_x > \tau_1 + c_i) \\ &= (\tau_2 - c_i) \mathbb{P}(T_x \leq \tau_2 | T_x > \tau_1) = (\tau_2 - c_i) q_x \end{aligned} \quad (1.15)$$

On estime alors grâce à l'équation de cohérence le taux de mortalité à l'âge  $x$  pour l'intervalle d'observation :

$$\hat{q}_x = \frac{D_x}{\underbrace{\sum_{i=1}^{N_x} (\tau_2 - \tau_1) + \sum_{i=1}^{W_x} (c_i - \tau_2)}_{\text{Exposition Initiale}}} \quad (1.16)$$

Une autre expression de l'équation de cohérence obtenue consiste à poser la variable indicatrice de décès de l'individu  $i$  suivante :

$$\forall i \in \{1, \dots, N_x\} \quad D_i = \begin{cases} 0 & \text{S'il y a sortie avec censure} \\ 1 & \text{Sinon} \end{cases} \quad (1.17)$$

Alors on a le taux de mortalité suivant :

$$\hat{q}_x = \frac{D_x}{\underbrace{\sum_{i=1}^{N_x} (\tau_2 - \tau_1) + \sum_{i=1}^{N_x} (1 - D_i)(c_i - \tau_2)}_{\text{Exposition Initiale noté } e_i}} \quad (1.18)$$

On obtient ainsi notre estimation du taux de mortalité corrigé des possibles censures qui peuvent être présentes dans notre portefeuille. On estime ensuite notre taux en réalisant l'hypothèse de Balducci. Cependant, cette hypothèse n'est plus réellement valable pour les âges élevés. En effet, la mortalité a plutôt tendance à augmenter avec le temps à des âges élevés, ce qui réfute notre hypothèse de diminution de la mortalité sur notre intervalle d'étude.

En utilisant cette méthode, nous obtenons une estimation améliorée de nos  $q_x$ , car l'exposition réelle de notre portefeuille dans l'intervalle d'observation est mieux estimée. En outre, il convient de souligner que l'hypothèse de Balducci implique que la probabilité de décès de l'individu diminue sur la plage d'observation.

### Exposition centrale et constance de la force de mortalité

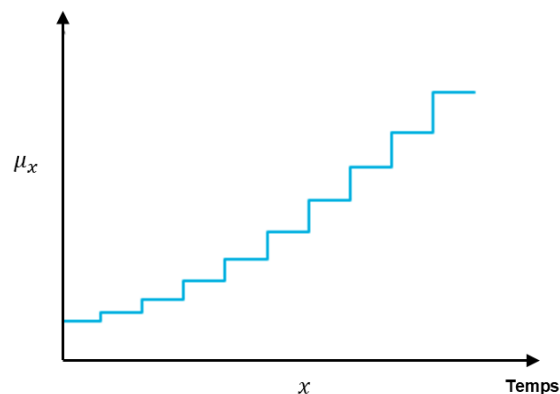
Dans certaines situations, en fonction du nombre de décès et du portefeuille, il peut être préférable de supposer que la force de mortalité reste constante entre les différents âges. Cette hypothèse implique que l'exposition au risque doit être ajustée. Nous opterons pour une exposition au risque appelée « centrale ». La différence avec l'exposition au risque initiale est que seuls les individus qui ont survécu pendant toute la durée d'observation sont considérés comme étant exposés au risque. En d'autres termes, seuls ces individus seront affectés à 1.

Dans cette section, on pose la notation suivante qui correspond à la longueur de l'intervalle d'observation :  $\tau_2 - \tau_1 = d_\tau$ . Cette notation nous permet d'exprimer notre hypothèse de constance de la force de mortalité dans l'intervalle considéré.

L'hypothèse de constance de la force de mortalité se traduit de la manière suivante avec  $\epsilon \in [0, d_\tau]$  :

$$\mu(x + \tau_1 + \epsilon) = \mu(x + \tau_1) \quad (1.19)$$

En conséquence, il est essentiel de sélectionner un intervalle d'observation suffisamment restreint pour que cette « approximation » soit la plus précise possible. En effet, plus la distance  $d_\tau$  est petite, plus l'approximation est fiable, car elle peut ne pas être vérifiée pour des plages d'observation larges.



**FIGURE 1.5** – Schéma explicatif de l'hypothèse de constance de la force de mortalité

On peut ensuite déterminer la fonction de survie à partir de la relation équation 1.19,

$$\begin{aligned}
 {}_{t_i - \tau_1}p_{x + \tau_1} &= \exp \left( - \int_0^{t_i - \tau_1} \mu(x + \tau_1 + s) ds \right) \\
 &= \exp \left( - \int_{\tau_1}^{t_i} \mu(x + u) du \right) \\
 &= \exp \left( - (t_i - \tau_1) \mu \right)
 \end{aligned} \tag{1.20}$$

Nous écrivons, par abus de notation  $\mu$  car dans l'intervalle  $x + \tau_1$  et  $x + t_i$  avec nos hypothèses,  $\mu$  est une constante. On réutilise ensuite la variable indicatrice présentée équation 1.17 qui est égale à 1 si un décès est observé dans l'intervalle  $[x + \tau_1, x + \tau_2]$  pour l'individu  $i$  et à 0 sinon. On a que la vraisemblance des individus qui sont vivants à la fin de notre analyse d'expérience n'est pas modifiée. En effet, seulement les individus qui sont décédés ont leur densité de modifiée, elle ne correspond plus à une variable de Bernoulli avec  $x + T_{i,x}$  l'âge au décès pour l'individu  $i$  d'âge  $x$  mais la variable aléatoire conditionnée par rapport au fait que le temps de décès est supérieur à  $\tau_1$ , on a étudié alors  $x + T_{i,x} | T_{i,x} > \tau_1$ . On peut ainsi exprimer la vraisemblance,

$$L = \prod_{i=1}^{N_x} {}_{t_i - \tau_1}p_{x + \tau_1} \mu(x + t_i)^{D_i} \tag{1.21}$$

Grâce à l'expression de la vraisemblance et notre hypothèse de distribution constante de force de mortalité dans l'intervalle d'analyse, on obtient le résultat suivant :

$$\log(L) = \sum_{i=1}^{N_x} \left( (t_i - \tau_1) \mu + D_i \log(\mu) \right) \tag{1.22}$$



On obtient ensuite l'estimateur  $\hat{\mu}$  du maximum de vraisemblance en dérivant l'équation 1.22 (noté *ec*). On obtient avec  $\frac{\partial}{\partial \mu} \log(L) = 0$ . Le résultat correspond au ratio entre le nombre total de décès observés sur **l'exposition centrale** définie précédemment.

$$\hat{\mu} = \frac{\sum_{i=1}^{N_x} D_i}{\sum_{i=1}^{N_x} (t_i - \tau_1)} \quad (1.23)$$

On peut ensuite exprimer notre probabilité de décès à l'âge  $x$  :

$$\hat{q}_x = 1 - \exp \left( - \frac{\sum_{i=1}^{N_x} D_i}{\underbrace{\sum_{i=1}^{N_x} (t_i - \tau_1)}_{\text{Exposition centrale noté ec}}} \right) \quad (1.24)$$

Cette approximation est encore discutable aux bas âges et aux âges élevés. En effet après 95 ans, la force de mortalité n'est pas constante dans un mois donné. De même pour les nouveaux nés : la mortalité infantile chez les moins de 5 ans est très élevée. De plus, elle n'est ni constante ni tout le temps décroissante. Il faudra donc, lorsqu'on applique ces hypothèses, se pencher à la fois l'âge du portefeuille, sa crédibilité, mais aussi le pays étudié. En effet, certains pays ont des tendances de mortalité qui diffèrent. De plus l'ensemble de ces précautions sont encore plus importantes dans le cas d'autres risques pour estimer par exemple des taux d'incapacité ou encore de maladie grave. Nous comparerons et mettrons en évidence les limites et avantages des méthodes présentées précédemment dans la Partie 3, qui vise à réaliser une étude complète d'analyse d'expérience sur un portefeuille.

### Implémentation des méthodes par discrétisation

Pour appliquer les méthodes par discrétisation, on doit transformer nos données. En effet, les modèles présentés précédemment s'appliquent sur des « *pseudo données* ». Ces pseudos données constituent une structure qui nous permet de déterminer nos taux à partir des indicateurs précédents (exposition centrale, initiale et variable indicatrice de l'événement). Cependant, un désavantage est que la taille des données dépendra de la granularité que l'on souhaite obtenir. C'est-à-dire qu'il y aura autant de lignes que d'intervalles de temps. Pour chaque intervalle de temps on déterminera l'exposition à la fois centrale et initiale de l'individu considéré.

Dans le cadre d'une analyse d'expérience, les intervalles de temps sont variés et dépendent du risque étudié. Par exemple pour certains risques, comme le rachat, on se concentrera plutôt sur la durée de la police. Pour une étude de mortalité, on pourra étudier la durée, mais aussi l'âge de l'individu. Les intervalles de temps peuvent ensuite être annuels, mensuels ou encore trimestriels. La granularité apportée à l'étude dépend de la précision du résultat souhaité, mais aussi de sa finalité.

x	$X_{i,1}$	...	$X_{i,p}$	ec	ei	$D_x$
53	$x_{i,1,1}$	...	$x_{i,1,p}$	0.28	0.28	0
54	$x_{i,2,1}$	...	$x_{i,2,p}$	1	1	0
55	$x_{i,3,1}$	...	$x_{i,3,p}$	1	1	0
56	$x_{i,4,1}$	...	$x_{i,4,p}$	0.25	1	1

**TABLE 1.2** – Pseudo donnée dans le cadre de notre exemple

Ainsi, si l'on reprend notre exemple présenté Section 1.2.2 c'est-à-dire dans l'intervalle  $I_j = [\tau_1; \tau_j]$  constitué de plusieurs sous-intervalles on a des expositions centrales et initiales qui sont différentes. Avec l'indicatrice  $D_x$  qui représente le moment de sortie, censure ou décès c'est-à-dire qui vaut toujours 0 sauf pour le dernier intervalle de temps observé. De la même manière l'exposition est toujours égale à 1 sauf pour le dernier intervalle et le premier intervalle de temps, car notre observation au début de l'étude n'est pas complète : elle représente la quantité de temps observée dans tous les cas lorsque nous calculons l'exposition centrale ou seulement en cas de censure lorsque nous calculons l'exposition initiale. Ainsi, nos *pseudos données* sont enregistrées de la manière présentée Table 1.2.

### Intervalles de confiance des taux bruts

Pour obtenir les intervalles de confiance pour les deux méthodes proposés précédemment la Commission d'Agrément [2006] propose une solution qui vient palier le fait d'avoir un dénominateur aléatoire, qui peut rendre les calculs de variance ardu. La proposition est alors en d'utiliser la méthode par exposition centrale et de réaliser un développement limité de notre expression pour obtenir,

$$\hat{q}_x = 1 - \exp\left(-\frac{\sum_{i=1}^{N_x} D_i}{ec_x}\right) \simeq \frac{\sum_{i=1}^{N_x} D_i}{ec_x} \quad (1.25)$$

On obtient ainsi,

$$\mathbb{E}(\hat{q}_x) = q_x \quad Var(\hat{q}_x) = \frac{q_x(1-q_x)}{ec_x}$$

On obtient alors en considérant que nous avons un nombre suffisant d'individus à l'âge x  $N_x$  le caractère gaussien de  $\hat{q}_x$  qui nous permet d'obtenir,

$$IC_{(1-\alpha)}(\hat{q}_x) = \hat{q}_x \pm z_{1-\alpha/2} \sqrt{\frac{q_x(1-q_x)}{ec_x}} \quad (1.26)$$

Avec,  $z_{1-\alpha/2}$  le quantile d'ordre  $1 - \alpha/2$  d'une loi normale centrée réduite.

---

## Chapitre 2

# Lissage des taux bruts

Après avoir modélisé les taux bruts dans le cadre d'une analyse d'expérience, il est généralement recommandé de lisser les données pour éviter toute incohérence ou variation trop importante dans le calcul de la prime. En effet, il peut arriver que pour certains jeux de données, la prime pure pour un individu de 50 ans soit plus élevée que celle pour un individu de 52 ans, ce qui est paradoxal. Cette situation s'explique par la variabilité des données et la diminution de l'exposition de notre portefeuille au fil du temps. Par conséquent, certains portefeuilles avec peu de données peuvent présenter de nombreuses irrégularités dans les estimations des taux bruts. Bien que la solution la plus simple aurait été de lisser directement les primes, cela prendrait plus de temps, car une table d'expérience peut être utilisée pour plusieurs produits et risques différents. De plus, chaque pays a ses propres méthodes de crédibilité, ce qui aurait pu aggraver la variabilité déjà existante et entraîner des erreurs ou des incohérences dans l'estimation de la **prime pure**.

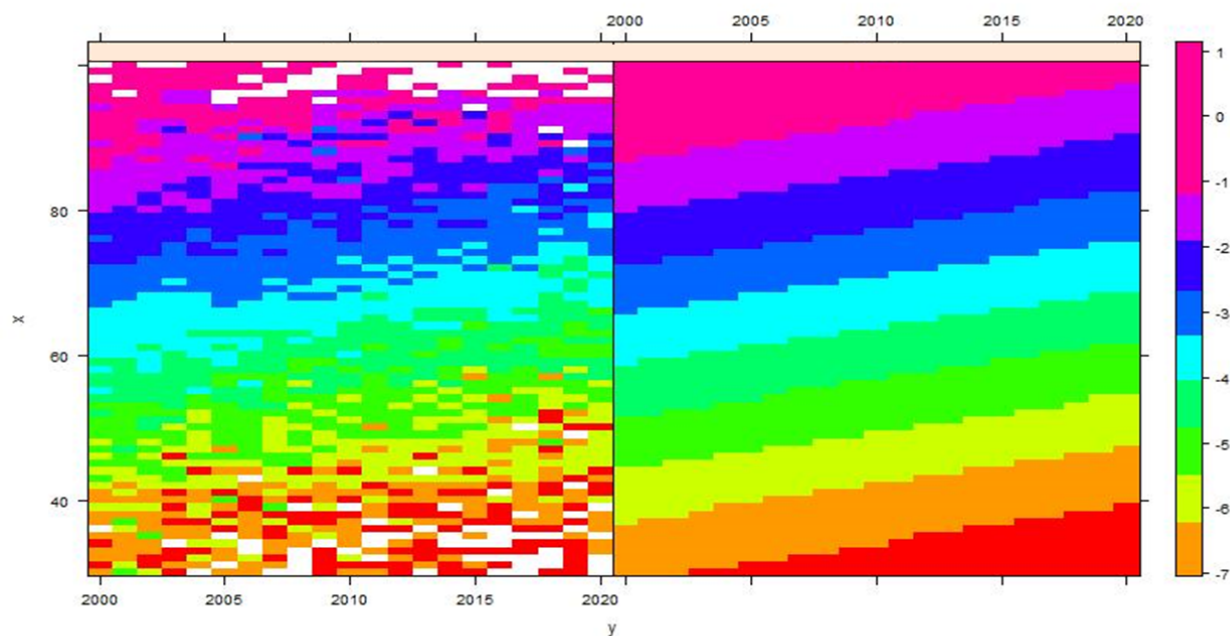


FIGURE 2.1 – Comparaison du log des taux bruts (à gauche) vs taux lissés (à droite)

Par exemple sur la Figure 2.1 ci-jointe on observe que la représentation du log des taux bruts présente des valeurs très variables et beaucoup d'erreurs. À terme, si on conservait ces taux bruts, cela mènerait à des incohérences de valeurs de primes entre les individus. On va donc chercher à lisser pour obtenir des taux comme la représentation à droite avec des taux qui varient moins et

---

apportent des résultats plus consistants.

## 2.1 Introduction aux modèles de lissage de taux non-paramétrique

Dans cette section, nous allons examiner les techniques de lissage non-paramétriques. C'est-à-dire qu'elles ne supposent pas de distribution particulière pour nos taux lissés. On va donc estimer nos taux lissés et la distribution de nos données grâce à un **apprentissage supervisé**. En notant  $\mathcal{X}$  : l'ensemble de nos données explicatives et  $\mathcal{Y}$  : l'ensemble des variables à expliquer.

On va introduire une quantité,  $l : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$  appelée **fonction de perte** qui va caractériser l'éloignement ou la perte lorsque la sortie réelle est  $y$  et la sortie prédite est  $\hat{y}$ . Plus la fonction de perte est faible plus notre prédiction est proche de la valeur réelle et plus les données prédites sont en adéquation avec nos données. Cette fonction de perte va être spécifique à chaque modèle, quand on imposera une distribution sous-jacente aux nombres de décès, on utilisera la différence de la log-vraisemblance.

On va aussi introduire une fonction  $f : \mathcal{X} \rightarrow \mathbb{R}$  qui peut être une combinaison linéaire, combinaison locale, produit de variables explicatives, fonctions d'activations, splines. Et qui dans le cadre du mémoire prendra différentes formes. C'est une fonction qu'on s'attachera à estimer de telle manière qu'elle approche les données réelles de la manière la plus fidèle possible.

Dans le cadre de la régression et du lissage, la fonction de perte utilisée est la fonction :

$$l_p(y, \hat{y}) = |y - \hat{y}|^p \quad (2.1)$$

On parle alors de régression  $L^p$  et dans le cas où  $p = 2$  on parlera de régression aux moindres carrés.

Avec toutes ces quantités, nous définissons le **risque empirique** qui correspond à la fonction suivante avec  $\pi$  la distribution de probabilité associée à nos données :

$$R_n(f) = \mathbb{E}_\pi(l_p(Y, f(X))) = \frac{1}{n} \sum_{i=1}^n l_p(y_i, f(x_i))$$

Ce risque correspond à l'erreur moyenne de nos prédictions. Nous allons donc chercher à minimiser cette quantité tout au long du mémoire pour que nos estimation soient le plus proche possible des valeurs réelles. Plus le risque d'estimation est faible plus nos prédictions sont proche de nos valeurs à prédire. En effet nous chercherons donc à estimer :

$$\hat{\alpha} \in \arg \min_{\alpha \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^n l_p(y_i, f(x_i, \alpha)) \right\} \quad (2.2)$$

Cependant au lieu de minimiser cette quantité nous allons plutôt minimiser celle-ci :

$$\begin{aligned} \hat{\alpha} &\in \arg \min_{\alpha \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^n l_p(y_i, f(x_i, \alpha)) + \lambda \text{pen}(\alpha) \right\} \\ \Leftrightarrow \hat{\alpha} &\in \arg \min_{\alpha \in \mathbb{R}^d} \left\{ \sum_{i=1}^n l_p(y_i, f(x_i, \alpha)) + \lambda \text{pen}(\alpha) \right\} \end{aligned} \quad (2.3)$$

Cela permet à notre estimation de garder une forme lisse et de ne pas avoir un modèle qui s'ajuste de manière trop proche avec nos données. Ce terme permet de balancer notre modèle entre **biais**

et **variance**. Plus le lambda est élevé plus le modèle s'éloigne des données. Nous réutiliserons cette spécification dans la plupart des modèles présentés par la suite qui sont basés sur ces principes.

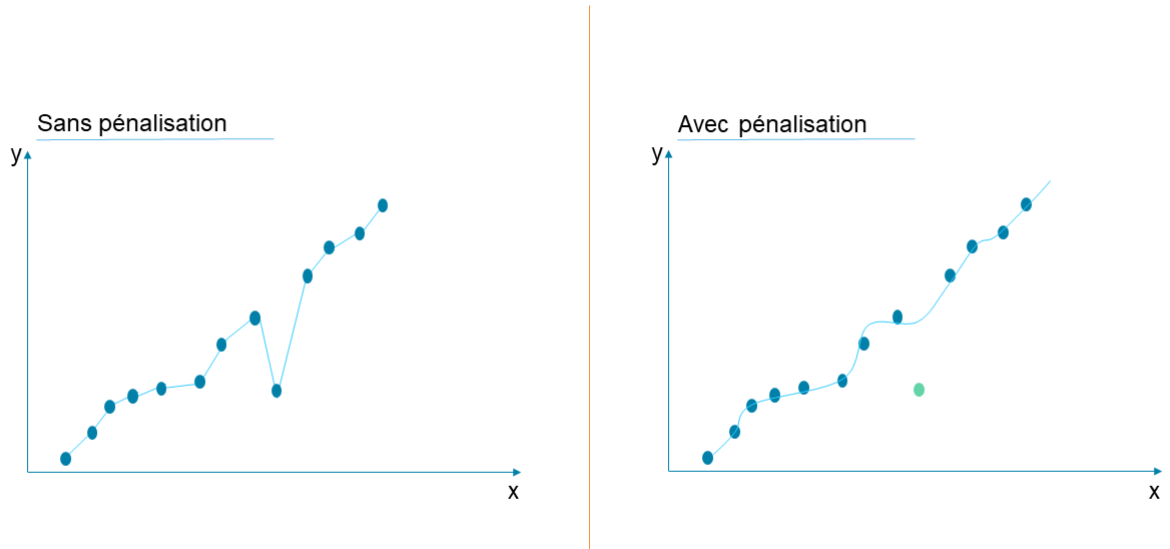


FIGURE 2.2 – Comparaison des méthodes lorsque  $\lambda = 0$  et  $\lambda \neq 0$

Nous constatons que lorsque  $\lambda$  est égal à 0, l'algorithme supervisé se contentera simplement de reproduire les données à prédire. Toutefois, si nous avons peu de données pour construire ce point, cela peut être considéré comme un point aberrant, ou du moins un point qui ne suit pas la tendance générale de nos données. Cependant, selon le type de produit, de risque ou encore la période étudiée, ce point peut avoir une signification particulière.

Par exemple, si notre abscisse  $x$  représente une période temporelle, comme une année, et que notre ordonnée  $y$  correspond à un taux de rachat, cela peut indiquer un événement économique particulier ayant entraîné une baisse significative des taux de rachat cette année-là, de sorte que ce point ne serait pas considéré comme « aberrant » au sens propre du terme.

Dans le deuxième graphique avec une pénalisation  $\lambda \neq 0$ , l'algorithme cherche à suivre la courbe générale des points tout en mettant l'accent sur la tendance générale en pénalisant les points qui s'éloignent trop de cette tendance. Ce compromis, appelé **le compromis biais-variance**, sera étudié en détail tout au long de cette partie consacrée aux méthodes de lissage supervisé.

## 2.2 Critère de suffisance des données

Dans le cadre du risque de décès, pour s'assurer que nous avons suffisamment de données et d'exposition dans notre portefeuille nous utilisons un critère, **le critère de Cochran**. Ce critère nous permettra de classifier nos bandes d'âges par quantité de données de sinistres et d'exposition. En fonction de la classification de la bande d'âges, nous ajusterons différents modèles, adaptés à la quantité de données et l'exposition sous-jacente. Ce critère assure un niveau de crédibilité quant à notre lissage.

**Critère de Cochran :**

**Critère de sinistralité :**  $E_x \cdot \hat{q}_x \geq 5$

**Critère d'exposition :**  $E_x \cdot (1 - \hat{q}_x) \geq 5$

Avec,  $\hat{q}_x$  le taux de mortalité estimé dans la Partie 1, convergent et non biaisé.

Si le critère de **sinistralité** et **d'exposition** est respecté pour la bande d'âge étudiée alors dans le cadre de cette étude on considérera que les données sont suffisantes et on appliquera des modèles qui s'ajustent avec les données observées. Si le critère de **sinistralité** n'est pas respecté, mais que celui de **l'exposition** l'est alors généralement, on adaptera nos modèles en ajustant nos valeurs avec des **tables de référence**, nous n'étudierons pas ce cas dans le cadre de cette étude.

## 2.3 Méthode de Whittaker-Henderson

Pour le lissage de nos taux bruts l'une des méthodes les plus connues et anciennes est la méthode de Whittaker-Henderson (Henderson [1924]), méthode qualifiée comme étant « Une nouvelle méthode de graduation » en 1924, dans Transactions XXV de l'Actuarial Society of America. C'est une méthode qui consiste à ajuster nos taux lissés avec nos taux bruts en supposant que nos taux lissés n'ont pas de forme particulière. A cet ajustement avec nos taux bruts on ajoute un terme de pénalité noté P, qui va pénaliser les écarts trop importants entre les taux. Plus les taux consécutifs sont différents ou éloignés plus la pénalisation sera importante. Ceci permet de compenser le terme d'adéquation et de conserver un caractère lisse dans nos données. Le modèle consiste donc à estimer le programme suivant :

$$\underset{\tilde{q}_{x_{min}} \dots \tilde{q}_{x_{max}}}{arg\ min} \{(F + \lambda P)\} \quad (2.4)$$

Avec,

**La mesure de pénalité :**

$$P = \sum_{x=x_{min}}^{x_{max}} (\Delta^z \tilde{q}_x)^2$$

**La mesure de fidélité :**

$$F = \sum_{x=x_{min}}^{x_{max}-z} w_x (\hat{q}_x - \tilde{q}_x)^2$$

Dans le paramètre d'adéquation F on pondère nos données à l'aide de  $w_x$  qui correspond à l'exposition au risque à l'âge x. Concernant le terme P on a la relation suivante pour calculer **le terme de différenciation** :

$$(\Delta^2 \tilde{q}_x) = (\tilde{q}_x - \tilde{q}_{x-1}) - (\tilde{q}_{x-1} - \tilde{q}_{x-2})$$

On peut généraliser ce terme et obtenir une relation pour estimer ce paramètre de pénalité noté P :

$$(\Delta^z \tilde{q}_x) = \sum_{k=0}^z \binom{z}{k} (-1)^{z-k} \tilde{q}_{x-k} \quad (2.5)$$

On remarque que le paramètre de différenciation est important : **plus le paramètre z est élevé**

**plus on pénalise par rapport à l'expérience passée** car on pénalise notre erreur actuelle par une erreur de plus en plus éloignée. La relation de différenciation est importante, car nous réutiliserions cette relation pour d'autres méthodes qui nécessitent une pénalisation. En effet, nous poserons **la matrice de différenciation**, notée  $D_z$  qui est défini par les valeurs des coefficients binomiaux. Par exemple, pour  $z = 3$  on obtient la valeur suivante :

$$D_3 = \begin{bmatrix} -1 & 3 & -3 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & \dots & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & -1 & 3 & -3 & 1 \end{bmatrix}$$

On construit les éléments de cette matrice qu'on notera  $D_z = (d_{i,j}^z)$  avec la relation suivante pour un pas de différenciation  $z$  donné, relation de différenciation avec,  $n_{age} = x_{max} - x_{min}$  qui correspond au nombre d'âges,

$$\forall (i, j) \in \llbracket 0, n_{age} - z \rrbracket \times \llbracket 0, n_{age} \rrbracket, \quad d_{i,j}^z = \begin{cases} \binom{z}{j-i} (-1)^{z-(j-i)} & \text{si } 0 \leq (j-i) \leq z \\ 0 & \text{sinon} \end{cases}$$

On peut ainsi résoudre matriciellement le programme d'optimisation de l'équation 2.4 et obtenir le programme suivant :

$$\underset{\tilde{q}}{\text{arg min}} \{ (\hat{q} - \tilde{q})^T W (\hat{q} - \tilde{q}) + \lambda (D_z \tilde{q})^T (D_z \tilde{q}) \} \quad (2.6)$$

On obtient la solution suivante pour nos **taux lissés** en dérivant le terme de l'équation 2.6 :

$$\tilde{q} = (W + \lambda D_z^T D_z)^{-1} W \hat{q} \quad (2.7)$$

Les résultats présentés dans la Partie 3 sont sensibles à la valeur de  $\lambda$ , qui peut prendre des valeurs comprises entre 1 et 1000. Cela rend son calibrage difficile, d'où la nécessité de proposer des méthodes semi-adaptatives avec des critères de régularité et de biais pour optimiser ce paramètre. En revanche, la méthode de Whittaker-Henderson ne présente que deux paramètres, ce qui la rend facile à mettre en place et à standardiser/automatiser dans le cadre d'une analyse d'expérience, qui est l'objet de notre étude. Il est possible de proposer une méthode adaptative et automatique pour minimiser les critères présentés dans la Section 2.11, à condition de respecter certaines conditions de résolution.

Cependant elle présente certains désavantages numériques : le lissage avec la méthode Whittaker-Henderson nécessite un espacement des nœuds d'une année, ce qui nécessite opérationnellement une bonne gestion et des données bien renseignées. De plus le recours à des données bidimensionnelles, entraîne des difficultés, par exemple, si nous avons les âges de 50 à 90 ans et les années 1960 à 2020, l'équation 2.7 nécessite le stockage et l'inversion d'une matrice  $2501 \times 2501$ . Cette méthode sera une référence pour comparer les résultats des méthodes plus poussées et la standardisation de celle-ci présentée dans l'exemple Partie 3.

## 2.4 Méthode LOESS

On observe qu'un des inconvénients de la méthode de Whittaker-Henderson est qu'elle ne considère pas le caractère local des taux. C'est à dire que l'on souhaite juste ajuster nos taux lissés le plus proche possible de nos taux bruts pour un âge donné tout en ajoutant un terme de pénalisation correspondant à la variation d'un âge à l'autre. Cependant on ne prend pas en compte la « tendance » de nos taux sur plusieurs âges. Cleveland [1979] propose ainsi une méthode locale, la méthode LOESS (*LOcally Estimated Scatterplot Smoothing*). Elle réalise une régression locale avec une fenêtre d'observation donnée. Cette fenêtre d'observation notée  $h(x)$  permet de capter les interactions qui peuvent exister entre les taux bruts à l'âge  $x$  et les âges passés et futurs.

Pour cela on pose la **régression au voisinage** de  $x_0$  :

$$\begin{aligned} \ln(\tilde{q}_x) &= \langle \beta_x, X_{x,local} \rangle \\ &= \beta_{x,0} + \beta_{x,1}(x-x_0) + \beta_{x,2}(x-x_0)^2 + \dots + \beta_{x,n}(x-x_0)^n \quad \text{si } |x-x_0| < h(x) \end{aligned} \quad (2.8)$$

Ainsi nous avons deux matrices, celles des coefficients  $\alpha$  et celle des données d'âges locales  $X_{locale}$  :

$$\beta_x = \begin{bmatrix} \beta_{x,0} \\ \beta_{x,1} \\ \cdot \\ \cdot \\ \beta_{x,n} \end{bmatrix} \quad X_{x,local} = \begin{bmatrix} 1 \\ (x-x_0) \\ \cdot \\ \cdot \\ (x-x_0)^n \end{bmatrix}$$

En plus de l'ajustement local, des poids sont ajoutés pour refléter l'importance relative de certains âges par rapport à d'autres. Ces poids permettent également de satisfaire la condition  $|x-x_0| < h$  directement dans la fonction de poids. Une possibilité est d'accorder plus d'importance aux âges récents et moins d'importance aux âges les plus éloignés de notre fenêtre d'observation, en utilisant une **fonction de poids** telle que :

$$\omega_x = \begin{cases} W(z) & \text{Si } z \leq 1 \\ 0 & \text{Sinon} \end{cases}$$

Il existe de nombreuses fonctions de poids, voici une liste non exhaustive des plus utilisées. Dans le cadre de ce mémoire, nous avons testé et utilisé les fonctions de poids de **Epanechnikov** et **Biweight** :

Nom	Fonction de poids $W(z)$
Uniforme	$\frac{1}{2} \mathbb{1}_{ z  \leq 1}$
Triangulaire	$(1 -  z ) \mathbb{1}_{ z  \leq 1}$
Tricube	$(1 -  z ^3)^3 \mathbb{1}_{ z  \leq 1}$
Gaussienne	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$
Quartic (Biweight)	$\frac{15}{16}(1 - z^2)^2 \mathbb{1}_{ z  \leq 1}$
Epanechnikov	$\frac{3}{4}(1 - z^2) \mathbb{1}_{ z  \leq 1}$
Exponential	$\exp\left(-\frac{(-2.5z)^2}{2}\right) \mathbb{1}_{ z  \leq 1}$
McLain	$\frac{1}{(z+\epsilon)^2} \mathbb{1}_{ z  \leq 1}$
Bicarré	$(1 - z^2)^2 \mathbb{1}_{ z  \leq 1}$

La valeur  $z$  dépend de la dimension utilisée pour modéliser nos taux lissés.

Si l'on souhaite modéliser les taux de mortalité  $\tilde{q}_x$  en utilisant uniquement l'âge comme variable explicative, on peut obtenir la différence en **valeur absolue** entre les taux de mortalité à différents



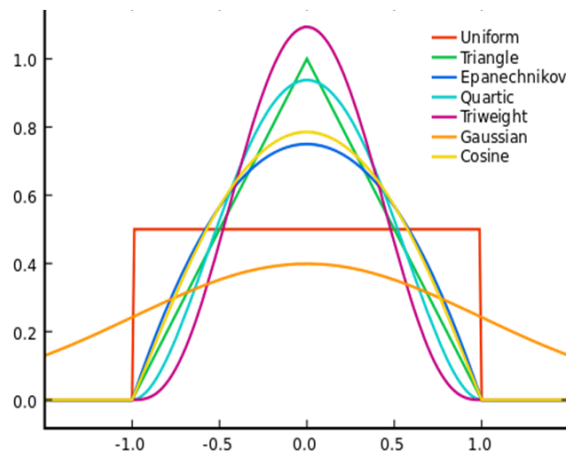


FIGURE 2.3 – Représentation des différentes fonctions de poids  $W(\cdot)$

âges. En revanche, si l'on souhaite modéliser les taux de mortalité  $\tilde{q}_x(t)$  lissés, il est nécessaire de considérer deux dimensions, ce qui implique d'adapter la notion de distance entre différents points dans un espace à deux dimensions. Dans ce cas, on utilise la **norme euclidienne** pour calculer la distance entre ces points.

$$z = \begin{cases} \frac{(x-x_i)}{h(x)} & \text{Si } 1d \\ \frac{\sqrt{(x-x_i)^2+(t-t_i)^2}}{h(x,t)} & \text{Si } 2d \end{cases}$$

Pour l'étude de la mortalité nous ajusterons notre fonction de poids en fonction de l'exposition de notre portefeuille, en ajustant la fonction de la manière suivante :

**En 1 dimension,**

$$\omega_x^{adjust} = \frac{\omega_x E_x}{\sum_x E_x}$$

**En 2 dimensions,**

$$\omega_{x,t}^{adjust} = \frac{\omega_{x,t} E_{x,t}}{\sum_t \sum_x E_{x,t}}$$

L'application uniforme de la fonction de pondération à chaque tranche d'âge n'est pas optimale. En effet, certains âges avancés ont une exposition faible et ne devraient donc pas être considérés avec autant de poids. Pour prendre en compte ce manque de données à certains âges, nous ajoutons une pondération liée à l'exposition à un âge donné.

Dans notre application nous considérerons une fenêtre d'observation noté  $h$  qui est constante c'est à dire que  $h(x) = h$  et  $h(x,t) = h$ . Celle-ci ne dépendra pas des âges et de la durée de la police d'assurance elle sera commune à tous les points à lisser.

Dans la suite du mémoire nous proposerons des méthodes adaptatives pour le choix de la fenêtre d'observation en fonction de l'âge et de l'exposition associée. Concernant le degré du polynôme nous choisirons un degré égale à  $n = 2$ .

Avec l'ensemble de ces hypothèses on a donc le programme d'optimisation suivant sous forme linéaire et matricielle en 1 dimension au voisinage de l'âge  $x_i$  :

$$\begin{aligned} \arg \min_{\beta} \left\{ \sum_{x=x_{min}}^{x_{max}} w_x^{adjust} (\ln(\hat{q}_x) - (\beta_{x,0} + \beta_{x,1}(x - x_i) + \beta_{x,2}(x - x_i)^2))^2 \right\} \\ \Leftrightarrow \arg \min_{\beta} \left\{ (\hat{q} - X_{local}\beta)^T W^{adjust} (\hat{q} - X_{local}\beta) \right\} \end{aligned} \quad (2.9)$$

Avec,

$$\beta = \begin{bmatrix} \beta_{x_{min},0} & \beta_{x_{min},1} & \beta_{x_{min},2} \\ \vdots & \vdots & \vdots \\ \beta_{x_{max},0} & \beta_{x_{max},1} & \beta_{x_{max},2} \end{bmatrix}^T \quad X_{local} = \begin{bmatrix} 1 & (x_{min} - x_i) & (x_{min} - x_i)^2 \\ \vdots & \vdots & \vdots \\ 1 & (x_{max} - x_i) & (x_{max} - x_i)^2 \end{bmatrix}$$

Cette méthode nous permet d'obtenir différents résultats en fonction de la valeur de  $h$ . On peut observer que le choix de la fenêtre d'observation impacte beaucoup nos taux lissés. Plus elle est élevée plus la forme de la courbe est lisse car on agrandit le voisinage d'âges considérés. Plus elle est proche, plus elle s'ajuste avec les taux bruts. Nous avons donc un compromis à réaliser entre une fenêtre qui a un biais qui est faible mais aussi un caractère lisse et une tendance pas trop proche des données (voir *Local Likelihood Estimation* [1987]). Cela revient encore au compromis biais-variance dans nos modèles.

Pour les âges élevés, nous disposons de moins de données pour effectuer la régression, ce qui peut poser problème. De plus, la régression présente un autre défi, à savoir sa nature linéaire ou quadratique. En effet, nos taux lissés peuvent avoir des relations non-linéaires avec l'âge des individus qui ne sont pas captées par ce modèle. En outre, estimer le paramètre de la fenêtre  $h$  est difficile, car il faudrait effectuer plusieurs régressions pour déterminer la valeur la plus adaptée à nos données.

## 2.5 Méthode B-Spline / P-Spline

Dans la section précédente, nous avons interpolé nos taux de manière linéaire puis de manière locale grâce à la méthode LOESS. Cependant cette méthode restreint la forme de la relation entre la variable explicative (l'âge, la durée ou durée du contrat, l'année calendaire,...) et le taux dont nous souhaitons réaliser le lissage. Historiquement, les splines sont intervenues suite au **phénomène de Runge** lors d'une interpolation polynomiale, l'interpolation de lagrange par exemple. Effectivement, on observe qu'avec un polynôme qui interpole plusieurs points, lorsqu'on augmente le nombre de points, le polynôme se met à osciller fortement entre les points avec une amplitude de plus en plus grande. Ainsi, malgré leurs propriétés recherchées telles des dérivées et des intégrales facilement évaluables, les polynômes présentent un manque de flexibilité.

En effet, un polynôme peut bien suivre une courbe sur un intervalle, mais être très variables dans certaines zones de points, l'ajustement est valable seulement sur un petit intervalle. Afin d'obtenir une classe de fonctions plus flexibles, des polynômes de bas ordre sont nécessaires, et ce, sur de petits intervalles. Ces problèmes de stabilité nous amènent alors à considérer une classe de polynôme qui interpole plusieurs sous-intervalles tout en étant continue par morceaux, les splines.

### 2.5.1 Définition d'une Spline

La théorie des splines s'est vue grandir au début des années 1960. Les splines étaient alors considérées comme un outil d'analyse numérique pour modéliser des fonctions. L'utilisation des splines comme méthode de régression non paramétrique est attribuée à Wahba [1975], qui a démontré leurs propriétés statistiques. Elles permettent de capturer les relations non linéaires entre les variables explicatives et la variable réponse, tout en évitant le phénomène de Runge observé dans les interpolations polynomiales. L'idée est de diviser l'ensemble des points à ajuster en sous-intervalles (nœuds) et d'ajuster une fonction simple à chaque sous-intervalle. Cependant, nous devons, pour chaque fonction, respecter des relations de raccordement aux points de jonction. Dans les splines les fonctions simples utilisées sont des polynômes. On impose à ces polynômes des conditions d'égalité, des polynômes, des pentes et de courbures aux différents points de jonction.

$$\forall k_i \in \llbracket x_{min}, x_{max} \rrbracket, \quad \begin{cases} B_{j,k}(k_i) = B_{j+1,k}(k_i) & \text{Condition de continuité} \\ B'_{j,k}(k_i) = B'_{j+1,k}(k_i) & \text{Condition de conservation de la pente} \\ B''_{j,k}(k_i) = B''_{j+1,k}(k_i) & \text{Condition de conservation de la courbure} \end{cases} \quad (2.10)$$

Avec :

- $x_{min}$  : L'âge minimum du jeu de données
- $x_{max}$  : L'âge maximum du jeu de données

Nous introduirons les notations suivantes pour la suite de l'étude,  $\mathcal{U} = \{0 \leq x_{min} = u_0, \dots, u_N = x_{max}\}$  qui est une suite croissante de nombres réels sur l'intervalle  $\llbracket x_{min}, x_{max} \rrbracket$  constitué des âges du portefeuille étudié. Cette suite décroissante est appelée **nœuds**. Le vecteur  $\mathcal{U}$ , appelé le vecteur nœud, il est utilisé pour construire les splines. Les points de contrôle sont des poids attribués à chaque spline, ce qui permet de la modifier et de lui accorder plus ou moins d'importance par rapport aux autres. En sommant ces poids pour toutes les splines estimées, nous obtenons notre prédiction finale.

**Notations :**

- $\mathcal{U}$  : Le vecteur des noeuds
- $N = \#\mathcal{U}$  : Le nombre de noeuds de notre fonction spline
- $k$  : L'ordre de la fonction spline
- $n = N-k-1$  : Le nombre de points de contrôles
- $\forall i \in \{1, \dots, n\}$ ,  $\beta_i$  : Les points de contrôles

### 2.5.2 Spline avec base de puissance tronquée

Historiquement, les splines étaient construites à l'aide de puissance tronquée qui étaient définies de la manière suivante :

En réutilisant les notations définies Section 2.5.1 :  $\mathcal{U} = \{x_{min} \leq u_1, \dots, u_N \leq x_{max}\}$  un ensemble de noeuds qui sont compris entre  $\llbracket x_{min}, x_{max} \rrbracket$  avec les noeuds (dans notre cas certains âges sélectionnés) qui sont regroupés de la manière suivante  $u_1 \leq u_2 \leq \dots \leq u_{N-1} \leq u_N$ . Avec,  $f$ , une fonction polynomiale d'ordre  $k$  qui respecte les conditions de l'équation 2.10.

Nous avons alors avec les puissances tronquées  $\forall x \in \llbracket x_{min}, x_{max} \rrbracket$  :

$$\begin{aligned} f(x) &= \beta_1 + \sum_{j=1}^k \beta_{j+1} x^j + \sum_{j=k+2}^{m+k-1} \beta_j (x - u_j)_+^k \\ &= B\beta \end{aligned} \quad (2.11)$$

Il est important de noter que pour construire notre fonction spline, nous devons définir nous-mêmes un certain nombre de noeuds. De plus, une observation importante est que pour les bas âges, lorsque  $x < u_{k+2}$ , notre lissage n'a qu'un terme polynomiale et ne présente pas de caractère « local », ce qui peut affecter la qualité de notre prédiction. Il convient donc de faire attention au nombre de noeuds et à leur répartition dans l'espace des points.

De plus elle est exigeante en termes de calcul numérique, comme présenté par de Boor [1978], cette méthode demande le calcul de  $k + j$  terme pour calculer un âge  $x$  tel que  $x > u_j$ . De Boor proposera alors une méthode alternative, plus opérationnelle dans le cas d'estimation par spline.

### 2.5.3 Récurrence de De Boor

Pour construire ce type de polynômes, de Boor [1978] propose une relation de récurrence que l'on réutilisera pour construire nos polynômes sous **R** et **Python**. Il propose une relation de récurrence qui permet d'itérer des polynômes de ce type à un certain degré  $k$ , tout en respectant les conditions présentées à l'équation 2.10 et en conservant toujours  $\mathcal{U} = \{u_1, \dots, u_N\}$ , c'est-à-dire un ensemble de noeuds compris entre  $\llbracket x_{min}, x_{max} \rrbracket$ .

On définit cette relation de la manière suivante :

$$B_{i,0}(x) = \begin{cases} 1 & \text{si } u_i \leq x < u_{i+1} \\ 0 & \text{sinon} \end{cases}$$

Les B-spline d'ordres supérieurs sont obtenues par récurrence de la manière suivante :

$$B_{i,k}(x) = \omega_{i,k} B_{i,k-1}(x) + (1 - \omega_{i+1,k}) B_{i+1,k-1}(x)$$

A noter :

- **k** : Représente l'ordre du polynôme, c'est-à-dire qu'ici on a un polynôme d'ordre  $k$  c'est-à-dire qu'il est de degré  $k - 1$  (car il y a un terme constant)
- **i** : Représente le polynôme du noeud  $i$ , c'est-à-dire le polynôme entre les valeurs  $x_i$  et  $x_{i+1}$

Avec le coefficient :

$$\omega_{i,k}(x) = \begin{cases} \frac{x - u_i}{u_{i+k} - u_i} & \text{si } u_i \neq u_{i+k} \\ 0 & \text{sinon} \end{cases}$$

Ce coefficient est important car il permet d'avoir seulement  $k$  valeurs non nulles à déterminer

pour une spline d'ordre  $k$ .

### 2.5.4 Méthode par B-Spline

Cela nous permet ensuite de construire une fonction, qui est une combinaison linéaire des différentes splines. Et nous définirons à l'aide des **points de contrôle** l'importance associée à une spline «  $i$  » donnée.

$$\forall x \in \llbracket x_{min}, x_{max} \rrbracket, \quad f(x) = \sum_{j=1}^n \beta_j B_{j,k}(x) \quad (2.12)$$

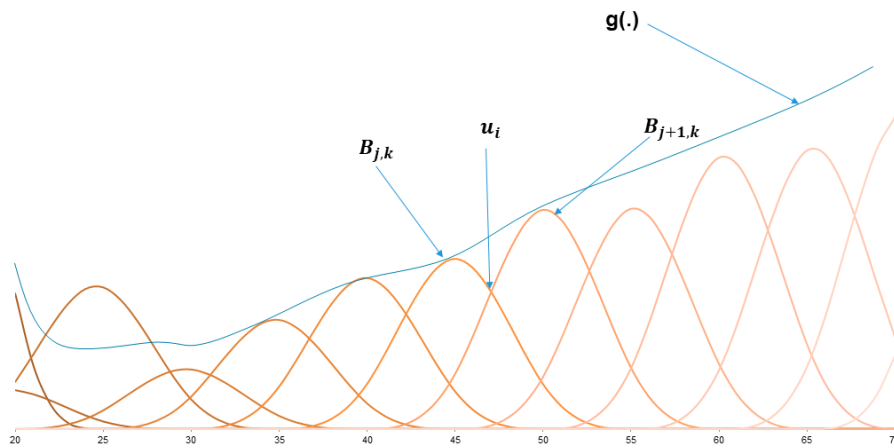


FIGURE 2.4 – Représentation de la fonction  $f(\cdot)$

On observe ici que notre fonction  $f(\cdot)$  est une fonction qui présente  $n$  **points de contrôle** par une **combinaison linéaire de polynômes de degré  $k-1$**  (car notre spline est d'ordre  $k$ ) qui présentent les caractéristiques des splines. Nous allons chercher à estimer les points de contrôle grâce à nos données de mortalités bruts. L'idée est d'accorder les poids aux points de contrôle de manière que notre fonction  $f(\cdot)$  s'ajuste au mieux avec les taux de mortalité bruts. Plus le  $\alpha_i$  est important plus on accorde d'importance à la spline  $B_{i,k}$  et donc plus notre taux lissé prendra la courbure de cette spline. Cette méthode s'applique aussi avec d'autres taux. On obtient ainsi le programme d'optimisation suivant dans le cadre d'une régression :

$$\begin{aligned} & \underset{\beta}{\operatorname{arg\,min}} \left\{ \sum_{x=x_{min}}^{x_{max}} (\hat{q}_x - f(x))^2 \right\} \\ \Leftrightarrow & \underset{\beta}{\operatorname{arg\,min}} \left\{ (\hat{q} - B\beta)^T (\hat{q} - B\beta) \right\} \end{aligned} \quad (2.13)$$

Avec,

$$B = \begin{bmatrix} B_{1,k}(x_{min}) & B_{2,k}(x_{min}) & B_{3,k}(x_{min}) & \dots & B_{n,k}(x_{min}) \\ \vdots & \vdots & \vdots & & \vdots \\ B_{1,k}(x_{max}) & B_{2,k}(x_{max}) & B_{3,k}(x_{max}) & \dots & B_{n,k}(x_{max}) \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_{1,x_{min}} & \cdot & \cdot & \beta_{n,x_{min}} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{1,x_{max}} & \cdot & \cdot & \beta_{n,x_{max}} \end{bmatrix}^T$$

La prédiction des B-Spline est très dépendante des noeuds. En effet plus il y a de noeuds plus les B-Spline vont envelopper nos points de la manière la plus fidèle possible en conservant l'allure générale des points. Cependant pour les âges élevés, ou pour certaines expositions faibles ce n'est pas ce qu'on veut car on manque de crédibilité au travers de nos données. On va donc chercher à pénaliser nos coefficients qui ont une grande variabilité c'est à dire ceux pour lesquels la différence est élevée.

Dans le **cas extrême**, soit celui où le nombre de noeuds correspond exactement au nombre de points à estimer c'est à dire avec :

$$\mathcal{U} = \{x_{min}, x_{min-1}, \dots, x_{max-1}, x_{max}\}$$

Dans ce cas extrême nos splines enveloppent parfaitement la courbure de nos points car ils ont exactement le même nombre de noeuds à interférer que de points. Le désavantage de cette méthode c'est qu'elle n'a qu'un terme de fidélité et d'adéquation des données lissées. Elle ne pénalise pas les coefficients. Une méthode alternative proposée est la méthode P-Spline qui, quant-à-elle prend en compte un terme de pénalité. Dans le cadre de la méthode B-Spline nos méthodes adaptatives chercheront plutôt à définir un nombre optimal de noeuds là où les méthodes adaptatives des P-Splines chercheront à déterminer un terme optimal de pénalité.

### 2.5.5 Méthode par P-Spline

Dans la méthode B-Spline les taux lissés ont tendance à avoir un biais faible mais une forte variance ce qui n'est pas recherché en pratique dans le cas du lissage : on cherche à éliminer l'erreur de modèle de nos estimations bruts pas à ajuster parfaitement avec nos taux bruts. On pénalise donc notre programme d'optimisation par une valeur P comme pour la méthode de Whittaker-Henderson de telle manière que les données ne s'ajustent pas trop à nos taux bruts et qu'on conserve une courbure polynomiale. On a donc la formule suivante :

$$\arg \min_{\beta} \left\{ \sum_{x=x_{min}}^{x_{max}} (\hat{q}_x - f(x))^2 + \lambda \left( \sum_{x=x_{min}}^{x_{max}-z} (\Delta^z \beta)^2 \right) \right\} \quad (2.14)$$

En utilisant les notations  $B$  et  $\beta$  présentées Section 2.5.4 et la matrice de différenciation présentée Section 2.3 on obtient le programme d'optimisation matriciel suivant :

$$\arg \min_{\alpha} \left\{ (\hat{q} - B\beta)^T (\hat{q} - B\beta) + \lambda (D_z \beta)^T (D_z \beta) \right\} \quad (2.15)$$

On obtient ainsi les coefficients optimaux, appelés historiquement **les points de contrôle** optimaux :

$$\hat{\beta} = (B^t B + \lambda D_z^T D_z)^{-1} B^T \hat{q} \quad (2.16)$$

La méthode par P-Spline à plusieurs avantages en comparaison des méthodes présentées précédemment. Tout comme la méthode par B-Spline elle permet aux utilisateurs ou praticiens d'utiliser cette méthode avec des données avec des noeuds (âges, temps, ...) espacés de plus de 1 ans il est par exemple possible de l'utiliser pour des noeuds espacé de 5 ans. Ainsi, opérationnellement, cela permet d'agrèger nos données lorsque nous n'avons pas assez d'incidents pour un noeud donné. Ce qui est souvent le cas lorsqu'on agrège nos données par types de risques ou par types de produits. Cette méthode, tout comme la méthode par B-Spline, permet aussi à l'utilisateur de proposer une distribution sous-jacente à nos incidents dans le cadre d'un GLM (voir Section 2.7), par exemple une distribution de Poisson ou binomiale négative.

Enfin d'un point de vu informatique et calculatoire, reprenons l'exemple d'un portefeuille d'individus de 50 à 90 ans qui évoluent dans les années 1960 à 2020. Avec des noeuds espacé de 5 ans l'équation 2.16 nécessite seulement l'inversion d'une matrice  $165 \times 165$ . La prochaine partie propose un prolongement des méthodes proposées jusqu'alors en introduisant une distribution sous-jacente à nos incidents.

De plus, le choix des noeuds est un choix très délicat (voir Li et Cao [2022]). Plusieurs procédures existent dans la littérature afin de déterminer l'emplacement et le nombre de noeuds (voir Goepp et al. [2018] / Huang [2021]). Une méthode simple consiste à choisir des noeuds équidistants pour un nombre  $z$  fixé d'avance. Cependant, cette technique n'offre pas un compromis optimal entre une grande flexibilité (donc un bon ajustement) et un faible nombre de noeuds. Plusieurs auteurs, ont avancé une méthode pas à pas descendante (*backward selection*) pour sélectionner les noeuds. Selon ces derniers, à partir d'un ensemble de noeuds équidistants, aussi appelés noeuds uniformes, l'approche consiste à enlever un noeud à la fois jusqu'à l'obtention d'un emplacement optimal.

C'est un critère important, car dans le cadre de notre étude, les taux bruts présentent dans certains cas et pour certains âges des changements brusques dans leur dérivée avec des pentes positifs ou négatives élevées. De plus, cette variabilité est généralement exacerbée lorsqu'on ajoute la durée ou duration de notre police d'assurance avec les âges pour voir une surface de mortalité. Il est donc essentiel dans le cadre d'une standardisation de cette méthode de trouver un compromis concernant la disposition des noeuds pour obtenir un modèle parcimonieux. Nous proposerons des méthodes d'optimisations visant à répondre à ce besoin dans la Section 2.10.1 et les appliquerons ensuite dans la Partie 3.

## 2.6 Introduction aux modèles de lissage de taux paramétrique

Dans cette section nous présenterons des méthodes de lissage dite **paramétriques** ou **semi-paramétriques** c'est à dire que nous supposerons une forme générale, une distribution de probabilité pour notre variable de comptage  $Y$ , autrement dit  $Y \sim \pi(\cdot)$ . Dans toute cette section  $Y$  pourra représenter une variable qui indique un nombre de décès, d'incidences ou encore de rachats.

## 2.7 Méthode GLM

Ce chapitre introduit la notion de GLM (*Generalized Linear Model*) présenté par Nelder et McCullagh (1989) et Dobson et Barnett (2008). Il présente les méthodes de lissage avec GLM, les différentes lois associées aux GLM et ses résultats. Nous expliquerons aussi comment déterminer les paramètres de notre GLM et ses limites tout en proposant une version raffinée en prenant en compte l'hétérogénéité de notre portefeuille et un caractère local en réutilisant les résultats de la méthode LOESS (Section 2.4). Cette partie s'attardera tout d'abord brièvement à expliquer la forme générale d'un modèle GLM, son intérêt dans le cadre de méthodes de lissage, mais aussi ses modifications pour arriver à une version locale.

### 2.7.1 Les distributions de lois exponentielles

La famille de lois exponentielles regroupe les lois qui ont une fonction de densité de la forme suivante :

Soit,  $Y \sim \pi$  et  $\pi$  une distribution de la famille exponentielle, on a :

$$f(y) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (2.17)$$

Avec le paramètre  $\theta$  qui est le **paramètre canonique**,  $\phi$  qui est le **paramètre de dispersion**. En général, pour ce paramètre on prend en considération l'exposition au risque comme dispersion. Les fonctions  $a(\cdot)$ ,  $b(\cdot)$  et  $c(y, \cdot)$  sont connues et déterminées à partir de la fonction de densité. On obtient l'espérance et la variance (voir McCullagh 1989) suivantes :

$$\mu = \mathbb{E}(Y) = b'(\theta) \quad \text{Var}(Y) = a(\phi)b''(\theta) \quad (2.18)$$

Avec  $b'(\cdot)$  et  $b''(\cdot)$  respectivement la dérivé première et seconde de  $b$ . On utilise la distribution exponentielle pour modéliser la distribution de la **variable réponse**  $Y$  dans le cadre du GLM.

### 2.7.2 La mise en place du GLM

Les modèles linéaires généralisés forment une catégorie de modèle très utilisée en actuariat, pour modéliser des montants de sinistres, comme pour modéliser des nombres de sinistres. Le GLM est une extension des modèles linéaires, extension dans laquelle on prend en compte une distribution sous jacente pour la loi de  $Y$ . On suppose que notre variable aléatoire  $Y$  va suivre une loi de la famille exponentielle. On choisit la loi utilisée en fonction de ce que l'on veut modéliser.

Dans le cadre du GLM il y a des hypothèses fortes concernant les variables, on a que les variables  $Y_1, \dots, Y_n$  sont mutuellement indépendantes. De plus on considère que  $Y$  doit suivre une loi de la famille exponentielle.

Il y a plusieurs composantes lors de la construction d'un GLM,

- **Une composante aléatoire** pour la variable réponse  $Y_i$  qui est censée suivre une loi de la famille exponentielle, c'est-à-dire que :

$$f(y) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}$$

Avec,  $\theta$  : le paramètre canonique et  $\phi$  le paramètre de dispersion

De manière générale pour toutes les fonctions de la famille exponentielle, la log-vraisemblance vaut :

$$\ln(\mathcal{L}(\alpha|y, \phi)) = \sum_i \left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right\} \quad (2.19)$$

- **Le prédicteur linéaire** ou variable réponse noté  $\eta$  qui est défini de la manière suivante :

$$\eta_i = X_i\alpha^t = \alpha_0 + \alpha_1x_{i,1} + \dots + \alpha_nx_{i,n}$$



- **Une fonction de lien**, notée  $g(\cdot)$  qui va relier le prédicteur linéaire à l'espérance conditionnelle de  $Y$ . De telle manière que :

$$\eta_i = g(\mathbb{E}(Y|X = x_i)) = X_i \alpha^T \quad (2.20)$$

Pour chaque distribution exponentielle, il existe une fonction de lien qui lui est propre. Cette fonction de lien  $g(\cdot)$  est une fonction monotone et différentiable.

Dans le cas de notre étude pour le lissage de nos données de nos taux de mortalité, on a ici qu'une seule variable explicative, l'âge ou la durée. Le modèle GLM est donc le suivant dans le cadre simple où on considère seulement l'âge :

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket, \quad \eta_i = g(\mathbb{E}(Y|X = x_i)) = \alpha_0 + \alpha_1 x_i \quad (2.21)$$

Comme nous souhaitons modéliser une fréquence, un taux d'incidence pour un âge  $x$  nous devons introduire l'exposition au risque pour l'âge  $x$  qui correspondra dans notre modèle GLM à la variable offset. De telle manière que :

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket, \quad \eta_i = g(\mathbb{E}(Y|X = x_i)) = g(\mu_i) = g\left(\frac{D_i}{E_i}\right) = \alpha_0 + \alpha_1 x_i \quad (2.22)$$

Nous devons définir une loi de comptage pour nos incidences c'est à dire pour notre variable réponse  $Y_i$ . Nous proposerons dans la suite du mémoire différentes lois de comptages en fonction de notre portefeuille.

### 2.7.3 La dispersion dans le nombre de décès

La dispersion dans le portefeuille étudié est un critère important. C'est un terme statistique qui caractérise la taille de la distribution de valeurs. Il existe deux grands cas : la sur-dispersion et la sous-dispersion. Lorsque la variance du nombre de décès est supérieure à la moyenne du nombre de décès on dit que le nombre de décès du portefeuille est sur-dispersé. Lorsque c'est l'opposé on dit que le nombre de décès du portefeuille est sous-dispersé. Une sur-dispersion de notre jeu de données présente de nombreux problèmes.

Les raisons d'une sur-dispersion sont diverses, le portefeuille étudié peut présenter des erreurs, il peut y avoir **une corrélation positive entre les âges** dans le cadre d'une étude de la mortalité ou encore le fait qu'on prend en compte une relation linéaire entre la variable réponse et les variables explicatives. Il y a donc plusieurs raisons pour avoir une sur-dispersion de nos sinistres dans notre portefeuille étudié.

Cela peut avoir plusieurs impacts : certaines variables non-significatives peuvent le devenir dans notre modèle à cause de la statistique de test qui dépend directement de la variance. En effet, une sur-dispersion implique une sous-estimation des résidus et ainsi une surestimation de la statistique du  $\chi^2$  qui a pour impact une significativité sur-élevée de nos variables. Ce problème est présenté dans Denuit et al. [2007]. De plus dans notre cas pour le lissage des âges élevés cela peut avoir une forte conséquence, car les décès présentent une grande variabilité. Nous pouvons estimer cette statistique de dispersion à l'aide de la **déviante** avec la formule suivante :

$$\phi_D = \frac{D(y, \hat{\mu}_i)}{n - p} \quad (2.23)$$

Avec,

- $n$  : Le nombre d'individus, dans notre cas le nombre d'âges
- $p$  : Le nombre de paramètres  $\beta$

Mais aussi à l'aide de la statistique du  $\chi^2$  de Pearson exprimé à l'aide des observations et de  $\mu_i$  définie Section 2.7.2 :

$$X^2 = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{Var(\hat{\mu}_i)} \quad (2.24)$$

À l'aide de cet estimateur on peut déterminer la dispersion de nos données de la manière suivante :

$$\phi_{X^2} = \frac{X^2}{n - p} \quad (2.25)$$

Nous pouvons aussi analyser la dispersion de nos données en traçant le graphique suivant en notant **N la variable aléatoire qui traduit le nombre de sinistres** (décès, incidences, rachats) ( $Var(N), E(N)$ ) qu'on estime de manière empirique :

$$E(N) = m_N = \frac{\sum_{i=1}^n N_i}{\sum_{i=1}^n E_i}, \quad Var(N) = S_N^2 = \frac{\sum_{i=1}^n (N_i - m_N E_i)^2}{\sum_{i=1}^n E_i}$$

Dans le cadre du **risque de mortalité** ces valeurs deviendront, avec  $E_i$  qui correspond à l'exposition pour l'âge  $i$  :

$$E(D_x) = m_{D_x} = \frac{\sum_{i=x_{min}}^{x_{max}} D_i}{\sum_{i=x_{min}}^{x_{max}} E_i}, \quad Var(D_x) = S_{D_x}^2 = \frac{\sum_{i=x_{min}}^{x_{max}} (D_i - m_{D_x} E_i)^2}{\sum_{i=x_{min}}^{x_{max}} E_i}$$

On peut aussi **conditionner** nos valeurs par **classe**, par exemple en créant des **tranches de sommes assurés** ou de **types de produits**, on introduira donc  $C_x = c$ , qui signifiera que l'individu d'âge  $x$  qui appartient à la classe  $c$  :

$$E(D_x | C_x = c) = m_{D_x, c} = \frac{\sum_{i, C_x = c} D_{i, c}}{\sum_{i, C_x = c} E_{i, c}}, \quad Var(D_x | C_x = c) = S_{D_x, c}^2 = \frac{\sum_{i, C_x = c} (D_{i, c} - m_{D_x, c} E_{i, c})^2}{\sum_{i, C_x = c} E_{i, c}}$$

Pour étudier la dispersion nous traçons les valeurs empiriques ( $S_{D_x}^2, m_{D_x}$ ) si celles-ci présentent une droite telle que  $E(D_x) = Var(D_x)$ , cela voudra dire que **nous ne sommes pas en présence de sous ou sur-dispersion**. Si les valeurs sont en dessous de cette droite de régression cela signifiera que nous sommes en présence de **sur-dispersion** et si les valeurs sont au dessus de cette droite de régression cela signifiera que nous sommes en présence de **sous-dispersion**.

### 2.7.4 Méthode GLM avec lien Poisson

Tout d'abord nous pouvons considérer que nos **nombres de décès** suivent une **loi de Poisson**. Ainsi,  $D_x \sim \mathcal{P}(\lambda_x)$  ce qui implique que pour un individu  $i$ ,  $\theta = \ln(\lambda_i)$  par identification avec la **loi de Poisson** et une loi de la famille exponentielle et que la fonction de lien canonique pour une loi de Poisson est la fonction  $g(\cdot) = \ln(\cdot)$ . Nous calculons alors la log-vraisemblance de cette loi pour un individu d'âge  $i$ ,

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$\begin{aligned} \ln(f(d_i)) &= \ln(\mathbb{P}(D_i = d_i)) \\ &= \ln\left(e^{-\lambda_i} \frac{\lambda_i^{d_i}}{d_i!}\right) \\ &= -\lambda_i + d_i \ln(\lambda_i) - \ln(d_i!) \\ &= -e^\theta + d_i \theta - \ln(d_i!) \end{aligned} \quad (2.26)$$

Ainsi, on obtient pour la distribution de Poisson la **log-vraisemblance** suivante :

$$l(\lambda) = \ln(\mathcal{L}(\lambda|y, \phi)) = \sum_{i=x_{min}}^{x_{max}} \{-e^\theta + y_i \theta - \ln(y_i!)\} \quad (2.27)$$

Un problème qui peut intervenir est une **sur-dispersion** de nos données. En effet en considérant un GLM avec lien Poisson on considère que notre distribution du nombre de décès est **équidispersée** entre les âges et comme présenté Section 2.7.3. Cette propriété n'est pas présente dans tous les portefeuilles et il convient donc d'y attarder une importance. Nous proposerons dans la prochaine partie une méthode qui apporte une prise en compte de l'hétérogénéité du portefeuille.

### 2.7.5 Méthode GLM avec lien binomial négatif

Nous avons observé qu'il y avait des insuffisances dans le modèle avec lien de Poisson. En effet, ce modèle présuppose une équidispersion de notre mortalité au fil des âges pour aller au-delà de cette supposition ou dans la situation où ce n'est pas le cas, par exemple à cause d'une corrélation positive entre la mortalité de certains âges. Ainsi, dans le cadre de l'étude nous avons étudié une alternative à la modélisation par une distribution de Poisson comme variable de comptage. Le choix se tournera alors vers une loi binomiale négative lorsqu'on rejettera  $H_0$  lors du test de dispersion.

On supposera, que  $D_x \sim \mathcal{P}(E_x \Lambda_x)$  et  $\Lambda_x \sim G(r, \frac{p}{1-p})$  avec  $r \in \mathbb{N}$  et  $p \in [0, 1]$ . Cela signifie qu'on considère une sur-dispersion de notre jeu de données, c'est-à-dire que  $\mathbb{E}(D_x) \neq \text{Var}(D_x)$ .

On obtient la **densité de probabilité conditionnelle** suivante pour le nombre de décès à l'âge  $x$  noté  $D_x$  :

$$f_{D_x}(d) = \mathbb{P}(D_x = d | \Lambda_x = \lambda) = e^{-\lambda} \frac{\lambda^d}{d!} \quad (2.28)$$

On obtient ensuite de cette manière la **distribution marginale de  $D_x$**  :

$$\begin{aligned}
\mathbb{P}(D_x = d) &= \int_0^{+\infty} \mathbb{P}(D_x = d | \Lambda_x = \lambda) \mathbb{P}(\Lambda = \lambda) d\lambda \\
&= \frac{\Gamma(r+d)}{d! \Gamma(r)} p^r (1-p)^d \\
&= \frac{\Gamma(d + \frac{1}{\phi})}{d! \Gamma(\frac{1}{\phi})} \left( \frac{1}{1 + \phi\mu} \right)^{\frac{1}{\phi}} \left( \frac{\phi\mu}{1 + \mu\phi} \right)^d
\end{aligned} \tag{2.29}$$

On a la valeur de l'espérance et de la variance :

$$\mathbb{E}(D_x) = \mu \quad \text{et} \quad \text{Var}(D_x) = \mu(1 + \phi\mu)$$

De plus, la loi binomiale négative appartient à la famille des lois exponentielles. En effet :

$$f_{D_x}(d) = \exp \left\{ r \ln(p) + d \ln(1-p) + \ln \binom{d}{d+r-1} \right\} \tag{2.30}$$

Avec les paramètres :

$$\theta = \ln(1-p) = \ln \left( \frac{\phi\mu}{1 + \phi\mu} \right), \quad b(\theta) = -r \ln(p) = -r \ln \left( \frac{1}{1 + \mu\phi} \right), \quad a(\varphi) = 1$$

On a donc le lien canonique suivant pour **la loi binomiale négative**  $g(\mu) = \ln \left( \frac{\phi\mu}{1 + \phi\mu} \right)$ .

Ainsi, on obtient aussi avec la distribution binomiale négative la log-vraisemblance :

$$l(r, p) = \ln(\mathcal{L}(r, p)) = \sum_{i=x_{min}}^{x_{max}} \{ \ln(\Gamma(r + y_i)) - \ln(y_i!) - \ln(\Gamma(r)) + r \ln(p) + y_i \ln(1-p) \} \tag{2.31}$$

### 2.7.6 Méthode GLM avec quasi-Poisson

La méthode GLM avec quasi-Poisson est une méthode qui est fondée sur la notion de quasi-vraisemblance elle est présentée dans le cas de variables de comptages par Wooldridge [1997]. De plus dans le cas de la méthode par quasi-poisson on modifiera notre variance de telle manière qu'on prenne en considération le phénomène de dispersion présenté précédemment. On définira une variance « modifiée, » de la façon suivante,

$$\text{Var}(D_x | X) = \phi \lambda_x \tag{2.32}$$

On estimera le paramètre  $\phi$  grâce à l'estimation  $\phi_{X^2}$  présentée précédemment (cf Section 2.7.3). Cela signifie que le modèle de quasi-Poisson est équivalent à un modèle de Poisson, avec les erreurs standards des  $\hat{\beta}$  multipliées par  $\sqrt{\phi_{X^2}}$ . Mais ce n'est pas exactement un modèle de Poisson car on n'a pas la propriété  $\mathbb{E}(D_x | X) = \text{Var}(D_x | X)$ . De plus la quasi-vraisemblance dans le cadre du modèle de Poisson n'est pas modifiée.

L'avantage de cette approche est de conserver les propriétés numériques de la distribution de

Poisson et de ne pas avoir à estimer plusieurs paramètres dont les paramètres de la distribution conditionnelle comme ce qui peut être le cas dans le cadre de la méthode avec la distribution binomiale négative. Ainsi avec cette méthode on remplacera la variance  $Var(D_x|X)$  du modèle original à l'âge  $x$  par  $\phi Var(D_x|X)$  avec  $\phi > 1$  utilisé pour tenir compte de la surdispersion par rapport au modèle original. On aura les valeurs suivantes en fonction de nos résultats et de nos données,

- $\phi < 1$  Qui est utilisé pour tenir compte de la sous-dispersion par rapport au modèle original.
- $\phi > 1$  Qui est utilisé pour tenir compte de la sur-dispersion par rapport au modèle original.

Le paramètre  $\phi_{X^2}$  est généralement estimé après l'estimation des paramètres du modèle  $\hat{\beta}$ .

On pourra résumer le choix de la loi de comptage la plus adaptée en fonction des données par le schéma suivant :

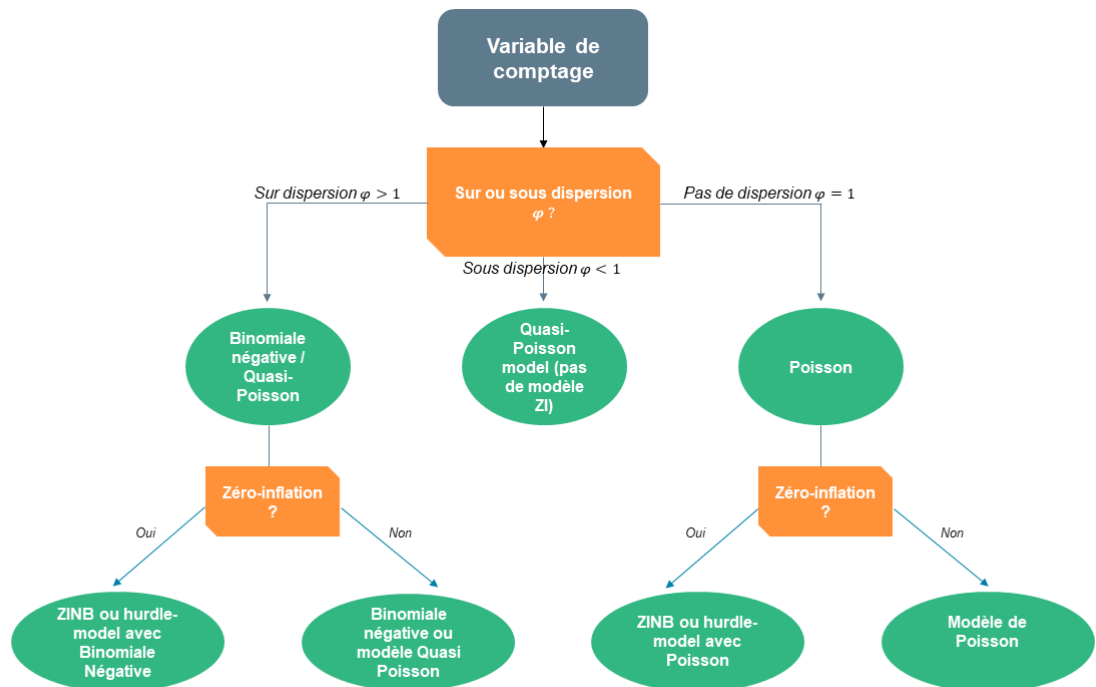


FIGURE 2.5 – Résumé du choix des lois en fonction de l'hétérogénéité et de la distribution du portefeuille pour le nombre de sinistres

### 2.7.7 GLM avec caractère local

Le modèle GLM local (mis en évidence de manière pratique par Tomas et Planchet [2013]) récupère les avantages de la méthode LOESS en supposant une distribution inconnue pour nos taux. Cela permet de considérer des relations non linéaire entre nos taux lissés et les âges tout en rajoutant un caractère local, avantage de la méthode LOESS (cf Section 2.4). Le modèle GLM prend donc la forme suivante au voisinage de  $x_0$  avec la fonction de lien  $g(\cdot)$  propre à la distribution choisie judicieusement par l'étude du portefeuille :

$$\forall x_i \in [x_{min}, x_{max}],$$

$$\begin{aligned} \eta_i &= g(E(Y|X = x_i)) = \beta_{i,0} + \beta_{i,1}(x_i - x_0) + \beta_{i,2}(x_i - x_0)^2 \\ &= X_i \beta_i \end{aligned} \tag{2.33}$$

Avec,

$$X_i = [1, (x_i - x_0), (x_i - x_0)^2], \quad \beta_i = [\beta_{i,0}, \beta_{i,1}, \beta_{i,2}]^T$$

Cette relation « locale » entre la force de mortalité est mis en évidence par Gavin et al. [1994], par la conviction que les forces de mortalité progressent de façon régulière d'une observation à l'autre. Il s'ensuit que les données de plusieurs observations de part et d'autre d'un point  $x_i$  peuvent être utilisées pour augmenter l'information de base dont nous disposons à  $x_i$ , en permettant d'avoir une meilleure estimation et un lissage plus performant. On prendra alors en compte une tendance générale de mortalité à caractère non-linéaire et local.

Cette méthode est particulièrement intéressante pour des portefeuilles avec peu de données de mortalité. La méthode d'estimation de cette méthode diffère des GLM « classiques », car on devra modifier les poids de notre méthode d'estimation des paramètres. À partir des paramètres estimés  $\hat{\beta}_i$  on sera en capacité de déterminer  $\hat{\eta}_i$  puis d'estimer  $\hat{\mu}_i$ .

$$\hat{\mu}_i = g^{-1}(\hat{\eta}_i) \tag{2.34}$$

On présentera la méthode d'estimation des  $\hat{\beta}_i$  (à partir des travaux de Tomas [2015]) dans la Section 2.10.1, qui fait intervenir les poids réitérés notés  $\gamma_{i,i}$  qui sont définis par :

$$\gamma_{i,i} = \frac{1}{b''(\theta_i) \left(\frac{d\eta_i}{d\mu_i}\right)^2} = \frac{1}{\mathbb{V}(\mu_i)} \left(\frac{d\mu_i}{d\eta_i}\right)^2$$

Cette valeur de poids est une résultante de la méthode de Scoring de Fisher et de l'algorithme de Newton-Raphson pour la résolution des équations de vraisemblance. Ce poids contient les informations apportées par les paramètres vis à vis de la courbure. (on note ici  $\gamma$  **le poids réitéré** pour ne pas le confondre avec les poids  $\omega$  **de la vraisemblance locale**).

## 2.8 Modèles GAM

Lors de l'étude des modèles GLM, nous avons constaté que ces modèles étaient efficaces pour capturer les relations locales entre les variables, tout en considérant une distribution antérieure pour notre variable du nombre de décès  $D_x$ . Cependant, nous avons également observé que notre variable réponse, notée  $\eta_i$  avec  $x_i \in \llbracket x_{min}, x_{max} \rrbracket$ , présentait une relation linéaire ou quadratique avec les différents âges de notre portefeuille.

Cela peut poser problème, car la mortalité suit souvent une trajectoire non linéaire, avec une augmentation à la naissance due à **la mortalité infantile**, puis une diminution progressive jusqu'à l'âge de 5 ans, suivie d'une diminution plus marquée jusqu'à environ 10 ans, puis une remontée progressive jusqu'à l'âge de 19-20 ans avant de se stabiliser et de continuer à augmenter. Il est donc nécessaire de considérer d'autres relations ou fonctions pour estimer la mortalité.

De plus dans le GLM, nous n'avons pas étudié les relations entre les covariables et les relations entre les différents âges, ou avec les durées de police d'assurance. Par exemple, on peut imaginer que pour une année de pandémie, on aura une corrélation assez forte entre l'année et la mortalité à un âge donné. Les désavantages du GLM sont donc qu'on impose une relation linéaire entre  $\eta_i$  et les âges et qu'on ne prend pas en compte des corrélations, relations entre les différents âges, années calendaires ou plus généralement les différentes variables explicatives de notre modèle.

### 2.8.1 Modèles GAM explication

Dans le **modèle GAM** nous avons une mise en place qui est similaire au GLM, c'est-à-dire que nous avons toujours une composante aléatoire et un prédicteur linéaire ou une variable réponse comme présentée Section 2.7.2. Nous devons aussi choisir une distribution pour notre variable  $Y$  qui modélise le nombre de décès noté  $D_x$ . On effectuera ce choix en fonction de critères de dispersion de nos nombres de sinistres (décès dans notre cas) présentés Section 2.7.3. Le modèle se présente de la manière suivante, avec  $\mathbf{h}$  : **une fenêtre d'observation** qui caractérise l'importance locale que l'on apporte à certaines variables :

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$\begin{aligned} \eta_i = g(\mathbb{E}(Y|X = x_i)) &= \underbrace{\sum_{u=i-h}^{i+h} f_{u,1}(x_u)}_{\text{Effet individuel}} + \underbrace{\sum_{u=i-h}^{i+h-1} f_{u,2}(x_u, x_{u+1})}_{\text{Effet croisé}} + \underbrace{f_3(x_{i-h}, \dots, x_{i+h})}_{\text{Effet collectif}} + X\alpha \\ &= X\alpha + Z\gamma \end{aligned} \quad (2.35)$$

Avec,

$$Z = \begin{bmatrix} f_{i-h,1}(x_{i-h}) & f_{i-h+1,1}(x_{i-h+1}) & \dots & f_{i+h-1,1}(x_{i+h-1}) & f_{i+h,1}(x_{i+h}) \\ f_{i-h,2}(x_{i-h}, x_{i-h+1}) & f_{i-h+1,2}(x_{i-h+1}, x_{i-h+2}) & \dots & f_{i+h,2}(x_{i+h-1}, x_{i+h}) & 0 \\ f_3(x_{i-h}, \dots, x_{i+h}) & 0 & 0 & \dots & 0 \end{bmatrix} \quad \text{et} \quad \gamma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$X$  et  $\alpha$  correspondent quant-à-eux respectivement aux **variables explicatives locales** qui capturent les relations linéaires et  $\alpha$  la **matrice des coefficients** associée à ces variables. On supposera comme avec les modèles GLM que notre variable aléatoire,  $Y \sim \pi(\cdot)$  avec  $\pi(\cdot)$  est une loi de comptage adaptée à notre portefeuille/expérience et notre risque étudié. Pour notre portefeuille et dans le cadre de la mortalité nous étudierons le risque de mortalité.

### 2.8.2 Modèle GAM univarié

Dans le contexte de l'étude de mortalité, nous allons commencer par créer un modèle GAM simplifié afin d'examiner et de comprendre les effets impliqués dans la mise en place d'un modèle GAM. Ce modèle ne tiendra compte que de la variable explicative de l'âge. Cependant, nous ne ferons pas d'hypothèses sur la forme de la relation entre l'âge et  $\eta_i$  :

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$\begin{aligned} \eta_i &= g(\mu_i) \\ &= g\left(\frac{D_i}{E_i}\right) \\ &= \beta_0 + f_1(x_i) \end{aligned} \quad (2.36)$$

Nous pourrions également prendre en compte que notre taux dépend de l'expérience passée et des âges antérieurs, auquel cas nous utiliserons le modèle suivant :

$$\begin{aligned}
\eta_i &= g(\mu_i) \\
&= g\left(\frac{D_i}{E_i}\right) \\
&= \beta_0 + f_1(x_i) + f_2(x_{i+1}) + f_3(x_{i-1})
\end{aligned} \tag{2.37}$$

L'ensemble des fonctions  $f_1$ ,  $f_2$  et  $f_3$  sont des fonctions univariées. Nous présenterons dans la prochaine partie comment les estimer.

### 2.8.3 Estimation des modèles GAM univariés

Pour estimer la fonction individuelle qui correspond à un modèle GAM avec une interaction univariée présentée dans l'équation 2.36, on utilise les notions présentées dans la Section 2.5.5 sur les splines en approximant la fonction  $f_1$  grâce à une combinaison de fonctions de base de la manière suivante :

$$f_u(x_i) = \sum_{j=1}^{m+k-1} \alpha_{u,j} B_{j,k}^u(x_i) \tag{2.38}$$

Les notations sont les mêmes que dans la Section 2.5.5. Néanmoins, nous indexons nos points de contrôles  $\alpha$  avec un  $u$  pour faire référence à la fonction approximée, de même pour les fonctions de base  $B_{j,k}$ .

Cependant dans le modèle présenté à l'équation 2.36 on observe qu'en sommant deux composantes non paramétriques nous aurons des problèmes d'identifiabilité. Les sommes des  $f_u$  ne sont identifiables qu'à une constante près, constituant un problème que l'on retrouve en additionnant des fonctions non-paramétriques et non restreintes. Nous souhaitons que nos termes soient orthogonaux à l'*intercept* (ordonnée à l'origine). Pour réaliser cela on impose que :

$$\forall u, \quad \sum_i f_u(x_i) = 0 \tag{2.39}$$

Dans notre exemple présenté Section 2.8.2 cela impose que :

$$\forall u \in \{1, 2, 3\} : \quad f_u(x_i) + f_u(x_{i+1}) + f_u(x_{i-1}) = 0 \tag{2.40}$$

On peut réécrire cette équation matriciellement de la manière suivante :

$$\mathbb{1}^T \tilde{B}_u \alpha_u = 0 \tag{2.41}$$

Avec,

$$\tilde{B}_u = \begin{bmatrix} B_{1,k}^u(x_i) & \cdot & \cdot & B_{n,k}^u(x_i) \\ B_{1,k}^u(x_{i+1}) & \cdot & \cdot & B_{n,k}^u(x_{i+1}) \\ B_{1,k}^u(x_{i-1}) & \cdot & \cdot & B_{n,k}^u(x_{i-1}) \end{bmatrix} \quad \alpha_u = [\alpha_{u,1}, \dots, \alpha_{u,n}]^T$$

Pour re paramétrer nos coefficients  $\alpha$  avec cette contrainte nous avons recours à une décomposition utilisée en optimisation, la décomposition . Elle consiste à réécrire la matrice  $\mathbb{1}^T B_u$  de la manière suivante :

$$(\mathbb{1}^T \tilde{B}_u)^T = QR \tag{2.42}$$



Avec  $Q$  une matrice orthogonale de taille  $(m+k-1) \times (m+k-1)$  et  $R$  une matrice triangulaire supérieure de taille  $1 \times (m+k-1)$ . En utilisant la décomposition dite de **Householder** pour construire la matrice  $Q = (D : Z)$  avec  $Z$  une matrice orthogonale de dimension  $(m+k-2) \times (m+k-1)$ , on démontre qu'en posant :

$$\tilde{\alpha} = Z\alpha \quad (2.43)$$

**La contrainte d'identifiabilité** est respectée et nous permet, en redéfinissant notre matrice  $\tilde{B}_u$ , de respecter cette contrainte en posant  $B_u = \tilde{B}_u Z$ . On obtient alors le modèle GAM suivant :

$$\begin{aligned} \eta_i &= g(\mu_i) \\ &= g\left(\frac{D_i}{E_i}\right) \\ &= \alpha_0 + X\alpha \end{aligned} \quad (2.44)$$

Avec,

$$X = [B_1, B_2, B_3] \quad \alpha = [\alpha_1, \alpha_2, \alpha_3]^T$$

On pourra généraliser cette méthode avec plusieurs effets bivariés en posant le GLM suivant qui regroupe les effets d'une fenêtre d'observation  $h$  :

$$\begin{aligned} \eta_i &= g(\mathbb{E}(Y_i | (X_{i-h}, \dots, X_{i+h}) = (x_{i-h}, \dots, x_{i+h}))) \\ &= \alpha_0 + \sum_{u=i-h}^{i+h} \sum_{j=1}^{m+k-1} \alpha_{u,j} B_{j,k}^u(x_u) \end{aligned} \quad (2.45)$$

### 2.8.4 Modèle GAM bivarié et multivarié

Dans le modèle GAM uni varié présenté Section 2.8.2 on additionne plusieurs effets non-linéaires des âges. Cependant, on ne prend pas en compte les effets croisés entre les variables, les effets bivariés entre deux âges et multivariés entre plusieurs âges. Dans les sections précédentes on omettait ces interactions. Bien sûr certaines interactions ne sont pas nécessaires à l'amélioration de la qualité de prédiction et additionner trop d'interactions peut mener à un sur-ajustement. Pour construire cette fonction multivariée nous utilisons le produit tensoriel. Il se construit de la manière suivante, en définissant deux fonctions bivariées liées aux âges des individus et la durée de la police d'assurance :

$$\forall (x_i, t_z) \in \llbracket x_{min}, x_{max} \rrbracket \times \llbracket t_{min}, t_{max} \rrbracket,$$

$$f_i(x_i) = \sum_{j=1}^n \alpha_{i,j} B_{j,k}^i(x_i) \quad f_z(t_z) = \sum_{j=1}^n \alpha_{z,j} B_{j,k}^z(t_z)$$

Avec  $\alpha_i$  et  $\alpha_z$  les paramètres reliés aux fonctions de bases des âges  $x_i$  et de la durée de la police  $t_z$  et  $B_i, B_z$  correspondent aux fonctions de base approximées à l'aide de splines.

Pour construire une fonction  $f_{i,z}$ , on utilise d'abord la fonction  $f_i$  puis on fait varier la valeur des  $\alpha_i$  en fonction des valeurs prises par la fonction  $f_z$  de telle manière que :

$$\forall l \in \llbracket 1, n \rrbracket,$$

$$\alpha_{i,l}(z) = \sum_{j=1}^n \beta_{z,j} B_{j,k}^z(t_z) \quad (2.46)$$

On a donc la fonction  $f_{i,z}$  suivante :

$$f_{i,z}(x_i, t_z) = \sum_{l=1}^n \sum_{j=1}^n \beta_{l,j} B_{l,k}^i(x_i) B_{j,k}^z(t_z) \quad (2.47)$$

On notera aussi matriciellement :

$$f_{i,z}(x_i, t_z) = X^T \beta \quad (2.48)$$

Avec,  $X$  qui est défini de la manière suivante en utilisant l'opérateur matriciel produit tensoriel  $\otimes$ ,

$$X = X_i \otimes X_z \quad (2.49)$$

Et les vecteurs composaient des splines,

$$X_i = [B_{1,k}^i, \dots, B_{n,k}^i] \quad X_z = [B_{1,k}^z, \dots, B_{n,k}^z]$$

On pourra ré-itérer le processus pour plusieurs dimensions. L'intérêt de cette fonction est de capturer à la fois des effets liés à la durée de la police d'assurance mais aussi à l'âge atteint de notre assuré. Cependant plus nous souhaitons ajouter de dimensions et de variables explicatives, plus le nombre de paramètres à estimer est élevé et le temps de calcul augmente. On s'intéressera alors dans la section suivante à comparer les algorithmes d'optimisation des paramètres du modèle GAM. Nous présenterons aussi celui retenu dans le cadre de l'application à un portefeuille Section 3.

En reprenant les notations Section 2.8.4 qui concernent le modèle GAM bivarié avec la durée et l'âge de la police d'assurance, on pourra écrire la vraisemblance dite pénalisée de la manière suivante toujours en supposant que le nombre de décès à l'âge  $x_i$  et à la durée  $t_z$  suit une loi de Poisson,  $Y_{i,z} \sim \mathcal{P}(\mu_{i,z})$  :

$$\eta_{i,z} = \log(\mathbb{E}(Y_{i,z}|X)) = \log(\mu_{i,z}) = X^T \beta \quad (2.50)$$

A partir de l'hypothèse d'une distribution de Poisson pour nos nombres de sinistres on peut obtenir la log-vraisemblance du modèle. Cependant concernant les termes de pénalités, on doit réaliser des modifications en l'état. En effet on souhaite pénaliser suivant les âges mais aussi suivant la durée de la police d'assurance. De plus les coefficients de nos fonctions bivariées sont soumis aux produit de kroneckers. On doit alors pénaliser nos coefficients en prenant cela en compte de telle manière que la matrice de différenciation  $D_i$  interviendra sur les colonnes de notre matrice de coefficients alors que la matrice de différenciation  $D_z$  interviendra sur les lignes de la matrice  $\beta$ .

Ainsi on obtient l'expression de la log-vraisemblance pénalisée notée  $l_p$  suivante :

$$\begin{aligned} l_p(\beta; X, y, \phi) &= l(\beta; X, y, \phi) - \frac{1}{2} (\lambda_i \beta^T (I \otimes D_i^T D_i) \beta + \beta^T \lambda_z (D_z^T D_z \otimes I) \beta) \\ &= l(\beta; X, y, \phi) - \frac{1}{2} \beta^T P \beta \end{aligned} \quad (2.51)$$

Avec,

- I : La matrice identité  $n \times n$
- P : La matrice de pénalisation  $P = \lambda_i (I \otimes D_i^T D_i) + \lambda_z (D_z^T D_z \otimes I)$

En posant la fonction de coût pénalisée  $J_p = -l_p$ , on pourra déterminer les coefficients  $\beta$  avec les algorithmes d'optimisation présentés Section 2.10.1.

## 2.9 Modèle avec réseaux de neurones

Dans cette partie nous présenterons des modèles adaptatifs qui se fondent sur le principe des « réseaux de neurones ». Les réseaux de neurones incarnent la représentation habituelle que nous faisons du cerveau, c'est à dire qu'une information simple transite dans un grand nombre d'autres informations jusqu'à créer l'information finale. On insère donc en *input* de l'information concernant une variable puis après différents calculs on a une nouvelle information sur une variable puis on ré-itére le processus jusqu'à avoir l'information finale.

Les précurseurs du principe sous-jacent de réseaux de neurones ont été Yann Le Cun et Yoshua Bengio avec des publications sur le « Perceptron » et les « Réseaux de neurones convolutives ». On a pu observer que dans les modèles additifs, on approxime les fonctions du modèle par des splines lisses, qui réduisent l'overfitting et sont ajustés analytiquement. Les méthodes présentées ci-dessous proposent d'utiliser des réseaux de neurones multicouches pour approximer les fonctions de lissage. On s'attardera à choisir judicieusement les fonctions d'activations pour répondre au mieux à nos exigences de régularité et de proximité avec les taux bruts.

### 2.9.1 Principe des réseaux de neurones

Les réseaux *feedforward* profonds, également appelés réseaux de neurones *feedforward* ou perceptrons multicouches, sont des modèles d'apprentissage profonds. L'intérêt des réseaux de neurones *feedforwards* est de permettre au modèle de modéliser des fonctions avec plusieurs variables explicatives. En effet dans les modèles linéaires généralisés on considère des interactions linéaires entre nos variables. Les réseaux de neurones permettent de comprendre des interactions entre 2 ou plusieurs variables. Les réseaux de neurones sont inspirés de la neuroscience et du cerveau, c'est à dire qu'à une série d'informations d'entrées  $[x_1, \dots, x_n]$ , le modèle va associer des poids  $[w_1, \dots, w_n]$ . On va ensuite sommer l'ensemble des informations pondérées par leur poids respectif auquel on va ajouter un biais  $b$  propre à cette estimation. De telle manière que à chaque neurone on calcul :

$$\begin{aligned} f(x; w, b) &= x^T w + b \\ &= \sum_{i=1}^n w_i x_i + b \end{aligned} \tag{2.52}$$

Jusque là notre modèle revient à réaliser une régression linéaire. Pour cela nous allons appliquer à cette sortie une fonction appelée fonction d'activation. Celle-ci décide de si un neurone doit être activé ou non. Le but de la fonction d'activation est d'introduire une non-linéarité dans la sortie d'un neurone. Dans le cas du lissage cela permet de considérer une prise en compte de non linéarité entre nos variables mais aussi de permettre d'activer seulement les variables explicatives qui présentent un intérêt pour le lissage de nos taux, par exemple, certains âges qui sont trop éloignés

par rapport à celui que l'on veut déterminer ne seront pas pris en compte. On obtient alors un nouvel input transmis à un nouveau neurone de telle manière que avec  $\sigma$  la fonction d'activation (*activation function*), la couche cachée (*hidden layer*)  $k$  et à la  $j$ -ème unité de couche (*unit layer*) on a la valeur suivante :

$$\begin{aligned} h_j &= \sigma \left( x^T w_j^{[k]} + b_j^{[k]} \right) \\ &= \sigma \left( \sum_{i=1}^n w_{i,j}^{[k]} x_i + b_j^{[k]} \right) \end{aligned} \tag{2.53}$$

Dans un réseau neuronal, nous mettons à jour les poids/coefficients et les biais des neurones sur la base de l'erreur à la sortie. Ce processus est connu sous le nom de rétropropagation. Nous présenterons ce processus, qui nous permet de déterminer les poids optimaux dans la suite du mémoire.

Nous allons illustrer le mode de fonctionnement et le vocabulaire pour décrire un réseau de neurones, avec un réseau de neurones simple, considérons un réseau avec 3 variables explicatives seulement de telle manière que :

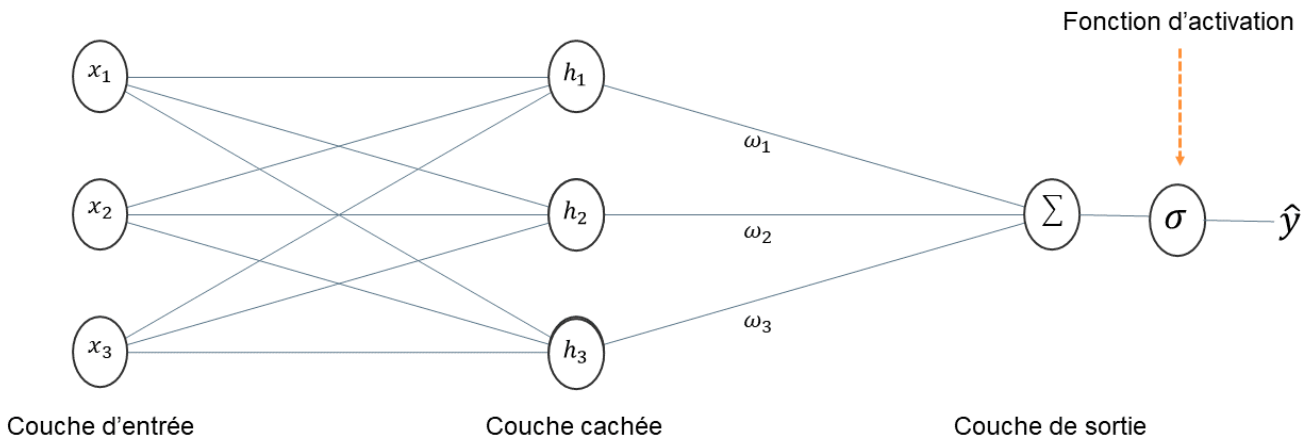


FIGURE 2.6 – Schéma explicatif d'un réseau de neurones

Dans ce cas de figure on a que :

$$\forall i, \quad h_i = \sigma \left( \sum_{j=1}^3 w_{i,j}^{[1]} x_j + b_i^{[1]} \right) \tag{2.54}$$

Et la valeur finale prédite :

$$\hat{y} = \sigma \left( \sum_{k=1}^3 w_k^{[2]} h_k + b^{[2]} \right) \quad (2.55)$$

On observe que même avec un réseau de neurones simple, cela fait intervenir beaucoup de paramètres. De plus le choix de la fonction d'activation a un impact important sur la valeur prédite résultante, car c'est elle qui va transformer les valeurs de sortie des couches cachées. Enfin on remarque que l'optimisation des paramètres semble ne pas être aisée : il faudra utiliser un principe présenté plus tard, la rétropropagation du gradient (*back propagation*). Cette méthode utilise une astuce calculatoire pour calculer la dérivée d'une fonction composée d'autres fonctions dont la dérivée est connue : c'est la règle de dérivation en chaîne (*chain rule*).

### 2.9.2 Les fonctions d'activation

Comme nous l'avons mentionné précédemment, pour faire fonctionner les réseaux de neurones il est nécessaire d'avoir des fonctions dites d'activation. Une fonction d'activation dans un réseau neuronal définit comment la somme pondérée de l'entrée est transformée en une sortie à partir d'un ou plusieurs nœuds dans une couche du réseau.

Nom	Fonction	Plage de valeurs
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$	$[0,1]$
tanh	$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$[-1,1]$
ReLU	$\sigma(x) \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0. \end{cases}$	$]0,\infty[$
Radiale gaussienne	$\sigma(x) = \exp\left(\frac{-\ x-c\ ^2}{2a^2}\right)$	$]0,\infty[$
SoftPlus	$\sigma(x) = \ln(1 + e^x)$	$]0,\infty[$
Radiale multiquadratique	$\sigma(x) = \sqrt{\ x - c\ ^2 + a^2}$	$]0,+\infty[$

TABLE 2.2 – Exemple de fonction d'activations

Le choix de la fonction d'activation a un impact important sur la capacité et les performances du réseau neuronal et différentes fonctions d'activation peuvent être utilisées dans différentes parties du modèle. Par exemple, dans le cadre de notre étude, nous retiendrons des fonctions d'activations qui fournissent des courbes lisses en sortie et cela dans le cadre d'un problème de régression. Dans le cadre du modèle RBF (Section 2.9.3) on fait le choix d'une fonction radiale aux différentes couches, qui apporte plus de poids aux données proches du point à estimer et renvoie une fonction lisse. Puis on applique une fonction d'activation linéaire en sortie du réseau, qui a pour intérêt de sommer l'ensemble des fonctions radiales estimés avec leurs poids associés. Par exemple, dans le cadre d'un problème de classification le choix de la fonction d'activation de sortie aurait été différent car la réponse de sortie aurait été binaire.

### 2.9.3 Modèle RBF Network

Le modèle RBF Network (*Radial Basis Function Network*) (voir Grievink et van Schijndel [2002] / Baughman et Liu [1995]) est un modèle qui combine deux notions, les fonctions radiales de base et les réseaux de neurones. Les fonctions radiales de base sont des fonctions que l'on définit

de la manière suivante : une fonction de base radiale,  $\phi(x)$  est une fonction par rapport à l'origine ou à un point noté  $c$  (*centroïde* par exemple), c'est-à-dire que  $\phi(x) = f(\|x - c\|)$  où la norme est généralement la norme euclidienne mais peut être une autre norme. On utilise cette méthode dans le cadre du lissage car elle permet à la fois de capturer des relations non linéaires mais aussi de seulement considérer des points/âges proches de ceux qu'on veut estimer. L'idée sous-jacente derrière cette méthode revient en plusieurs fonctions radiales de base en leur associant des poids pour obtenir une fonction lisse.

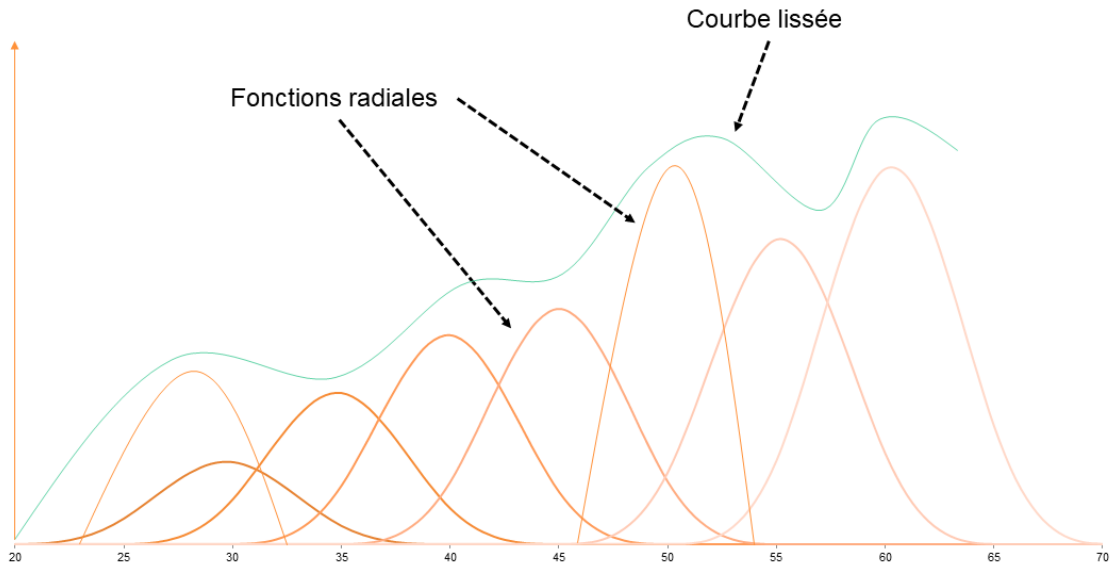


FIGURE 2.7 – Schéma de la fonction lissée avec le modèle RBF

Dans le cas proposé via cet exemple, on observe que les poids associés aux fonctions radiales sont différents : plus le poids d'une fonction radiale est élevé plus la courbe lissée sera proche de la fonction radiale. Dans le cadre du modèle que nous présenterons, nos fonctions radiales qui seront considérées dans le cadre du modèle comme nos fonction d'activation  $\sigma$  (par soucis de notation nous noterons la fonction d'activation radiale  $\phi$ ), seront définies de la manière suivante :

$$\phi_i = \exp\left(\frac{-\|x - c_i\|^2}{2\sigma_i^2}\right) \quad (2.56)$$

Plus le point  $x$  est proche du centroïde  $c_i$  plus la valeur de la fonction d'activation est élevée. Dans le cadre du lissage, cela signifie que la fonction radiale accordera plus d'importance aux points récents (âges, temps, ...). Cependant cette fonction peut être difficile à optimiser.

Pour déterminer le centroïde  $c_j$  il y a plusieurs méthodes. Dans le cadre de ce mémoire, comme nous souhaitons mettre en place des méthodes de façon standardisée et automatisée, nous choisissons de déterminer le centroïde par la méthode des *K-means* (initialisé avec `kmeans++` sous R).

Le schéma sera le suivant pour estimer le taux lissé à un âge  $i$  :

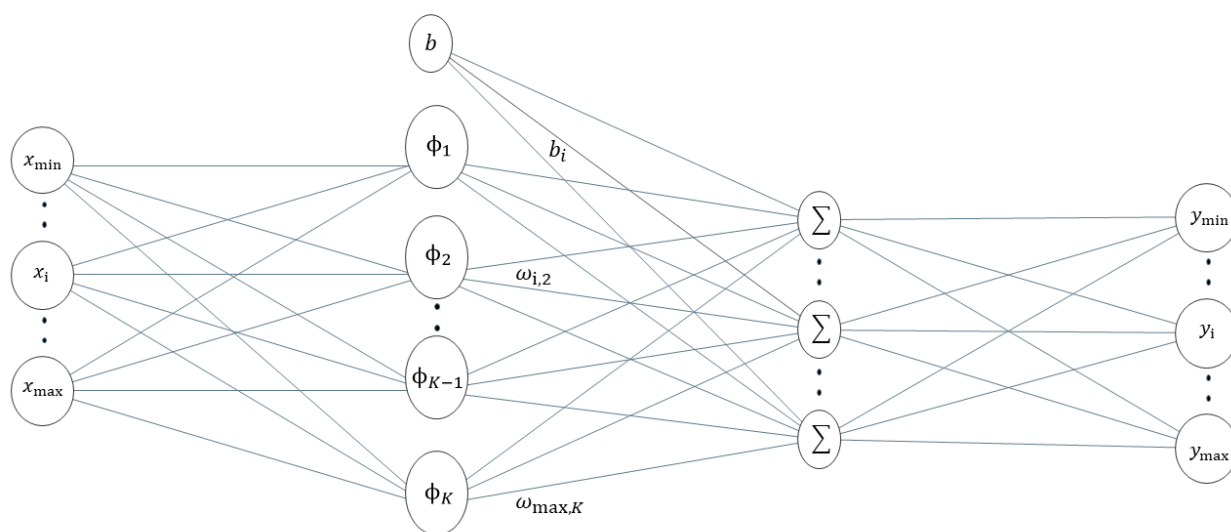


FIGURE 2.8 – Schéma du réseau de neurones RBF

Et la valeur prédite finale vaudra en utilisant les notations précédentes avec  $K$  le nombre de fonctions radiales choisies grâce aux nombres de clusters dans nos données. **Plus nos données sont volatiles et variables plus il y aura des clusters et donc  $K$  sera élevé :**

$$g_i(x; w, b) = \sum_{k=1}^K w_{i,k} \phi_{i,k}(x, c_k) + b_i^{[1]} \quad (2.57)$$

### 2.9.4 Schéma du réseau du modèle RBF

On estime ensuite les poids de nos différents noeuds, pour cela on résout le programme d'optimisation suivant avec le vecteur des âges  $x = [x_{min}, \dots, x_{max}]^T$ ,

$$\arg \min_w \left\{ \frac{1}{2} \sum_{i=x_{min}}^{x_{max}} (\hat{q}_i - g_i(x; w, b))^2 \right\} \quad (2.58)$$

On obtient matriciellement le programme d'optimisation suivant que l'on va résoudre :

$$\arg \min_w \{ (\hat{q} - g)^t (\hat{q} - g) \} \quad (2.59)$$

Pour optimiser les paramètres de poids on utilise la rétropropagation du gradient présenté dans la section 2.9.5.

### 2.9.5 Rétropropagation du gradient et algorithmes d'optimisations

Pour optimiser les paramètres de notre modèle RBF on utilise la règle des dérivées à la chaîne (*chain rule*) qui nous permet de trouver la dérivée de fonctions composées. Elle est calculée de manière extensive par l'algorithme de rétropropagation, afin d'entraîner des réseaux de neurones. Pour cela on a  $\forall (i, j) \in \llbracket x_{min}, x_{max} \rrbracket \times \llbracket 1, K \rrbracket$  la dérivée correspondant à nos poids  $w_{i,j}$ ,

$$\begin{aligned} \frac{\partial J}{\partial w_{i,j}} &= \frac{\partial J}{\partial g_i} \cdot \frac{\partial g_i}{w_{i,j}} \\ &= \frac{\partial}{\partial g_i} \left( \frac{1}{2} \sum_{i=x_{min}}^{x_{max}} (\hat{q}_i - g_i(x, w, b))^2 \right) \cdot \frac{\partial}{\partial w_{i,j}} \left( \sum_{k=1}^K w_{i,k} \phi_{i,k}(x, c_k) + b_i^{[1]} \right) \\ &= (\hat{q}_i - g_i(x, w, b)) \cdot \phi_{i,j}(x, c_j) \end{aligned} \quad (2.60)$$

De même on obtient la dérivée par rapport aux coefficient  $b_i$  de la même manière,

$$\begin{aligned} \frac{\partial J}{\partial b_i} &= \frac{\partial J}{\partial g_i} \cdot \frac{\partial g_i}{b_i} \\ &= \frac{\partial}{\partial g_i} \left( \frac{1}{2} \sum_{i=x_{min}}^{x_{max}} (\hat{q}_i - g_i(x, w, b))^2 \right) \cdot \frac{\partial}{\partial b_i} \left( \sum_{k=1}^K w_{i,k} \phi_{i,k}(x, c_k) + b_i^{[1]} \right) \\ &= (\hat{q}_i - g_i(x, w, b)) \end{aligned} \quad (2.61)$$

A partir de cela on peut construire notre algorithme d'optimisation.

## 2.10 Modèle GBM : Gradient Boosting Model

Nous avons pu observer dans nos précédents modèles que le choix d'un modèle à la fois stable, parcimonieux et qui gomme la variabilité de notre estimation des taux bruts est difficile. Dans cette partie nous étudierons un modèle qui combine deux notions, le boosting et la descente de gradient pour lisser nos données. En effet une tâche commune qui apparaît dans différentes applications d'apprentissage automatique est de construire un modèle de régression non paramétrique à partir des données clients : dans notre cas de la cédante.

Lors de la conception d'un modèle, une stratégie consiste à construire un modèle à partir de la théorie et à ajuster ses paramètres en fonction des données et de l'expérience du portefeuille étudié. Malheureusement, dans la plupart des situations, de tels modèles ne sont pas disponibles. On cherche donc à estimer les relations potentielles entre les variables d'entrée. Le principe du boosting répond à cette problématique en créant un modèle à partir des erreurs des précédents.

En effet à chaque itération, un nouveau modèle faible, d'apprentissage de base, est formé en fonction de l'erreur du modèle appris jusqu'à présent. On estime ensuite les paramètres à l'aide d'une méthode de descente de gradient. D'où le sigle *Gradient Boosting Model* (GBM).

### Le boosting

Nous nous sommes tournés vers le modèle GBM dans le cadre de notre projet d'automatisation ou quasi-automatisation des méthodes de lissage car il présente certains atouts. Ainsi, il permet une grande flexibilité, il est hautement personnalisable. De plus, il introduit une grande liberté dans la conception du modèle : nous pouvons par exemple définir une fonction dite de perte (*loss*)



function) et des fonctions d'apprentissages de base (*weak learner/base learner*). L'idée est de complexifier un modèle en ajoutant successivement plusieurs modèles simples, appelés « **apprenants faibles** ». On obtient ainsi un modèle consistant et robuste.

$$G^{(m)}(x) = g^{(1)}(x) + g^2(x) + \dots + g^{m-1}(x) \tag{2.62}$$

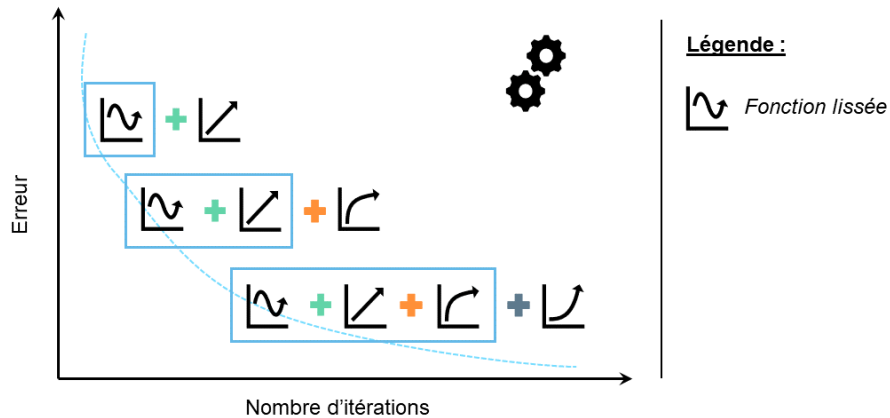


FIGURE 2.9 – Schéma représentatif du principe de boosting

De fait, comme l'illustre l'équation 2.62 notre modèle final n'est rien d'autre qu'un méta modèle constitué des différentes fonctions estimées. Dans notre cas, les apprenants faibles seront des fonctions splines, car de par leur construction ce sont des fonctions qui ont des bonnes qualités de lissage. Nous pourrions également légitimement utiliser des fonctions radiales. Cependant, de par leurs constructions, les splines cubiques sont les fonctions les plus lisses pour approximer une série de points. En effet on a la proposition suivante,

**Proposition :** La spline  $B(x)$  construite grâce à la formule de récurrence de De Boor est l'interpolation la plus « lisse » dans le sens de la minimisation du critère de régularité/pénalité :

$$P(f) = \int_{x_{min}}^{x_{max}} (f''(x))^2 dx \tag{2.63}$$

Pour toute fonction  $f$  continue dans l'intervalle d'âges  $[x_{min} - x_{max}]$ . Et dont la dérivée première est absolument continue et qui interpole  $\{x_i, y_i\} \forall x_i \in \llbracket x_{min}, x_{max} \rrbracket$ .

**Preuve :** On suppose qu'il existe  $g(\cdot)$  une fonction qui est plus lisse que  $B(\cdot)$  mais qui coïncide en les mêmes nœuds, on pose  $h(x) = g(x) - B(x)$

$$\begin{aligned} \int_{x_{min}}^{x_{max}} (g''(x))^2 dx &= \int_{x_{min}}^{x_{max}} (h''(x) + B''(x))^2 dx \\ &= \int_{x_{min}}^{x_{max}} (B''(x))^2 dx + 2 \int_{x_{min}}^{x_{max}} B''(x)h''(x) dx + \int_{x_{min}}^{x_{max}} (h''(x))^2 dx \end{aligned}$$

On obtient par intégration par parties et propriété des splines construites à l'aide de la récurrence de De Boor,

$$\begin{aligned} \int_{x_{min}}^{x_{max}} B''(x)h''(x) dx &= \underbrace{B''(x_{max})h'(x_{max})}_{=0} - \underbrace{B''(x_{min})h'(x_{min})}_{=0} - \int_{x_{min}}^{x_{max}} B'''(x)h'(x) dx \\ &= - \sum_{i=min}^{max-1} B'''(x_i^+) \int_{x_i}^{x_{i+1}} h'(x) dx \\ &= - \sum_{i=min}^{max-1} B'''(x_i^+) (h(x_{i+1}) - h(x_i)) \\ &= 0 \end{aligned} \tag{2.64}$$

On a que  $B'''(\cdot)$  est constante sur l'intervalle  $[x_i, x_{i+1}]$  car  $B(\cdot)$  est une spline cubique entre ces valeurs. De plus, on a que  $B(\cdot)$  et  $g(\cdot)$  coïncident aux nœuds, c'est-à-dire que  $\forall x_i \in \{x_{min}, \dots, x_{max}\}$  on a  $h(x_i) = 0$ . Ainsi, on obtient,

$$\int_{x_{min}}^{x_{max}} (g''(x))^2 dx = \underbrace{\int_{x_{min}}^{x_{max}} (B''(x))^2 dx}_{\geq 0} + \int_{x_{min}}^{x_{max}} (h''(x))^2 dx \geq \int_{x_{min}}^{x_{max}} (B''(x))^2 dx \tag{2.65}$$

On conclut que la spline cubique est la fonction la plus « lisse » dans le sens du critère de régularité comme apprenant faible pour notre modèle de boosting.

Le boosting avec des apprenants de base (dans notre cas les splines cubiques) implique généralement les paramètres suivants :

- L'itération d'arrêt  $m_{stop}$ , c'est à dire à partir de quel moment on stop l'apprentissage du modèle à partir des erreurs.
- Le pas d'apprentissage de la descente ou paramètre de rétrécissement (*learning rate* ou *shrinkage parameter*) noté  $\nu$
- Les degrés de liberté  $df$  (qui sont égaux à la trace de la matrice  $\hat{H}$  pour une P-Spline).

Dans le cadre de cette étude on définira les fonctions de coûts pour l'apprentissage du modèle de boosting de la manière suivante :

- Si on suppose une distribution sous-jacente pour nos décès :

**La log vraisemblance négative pénalisée :**  $J_p = -l_p$ .

- Si on ne suppose pas de distribution :

**En 1 dimension :**  $J(q, g, \beta) = \frac{1}{2} \sum_{x=x_{min}}^{x_{max}} (\hat{q}_x - g(x, \beta))^2$

**En 2 dimensions :**  $J(q, g, \beta) = \frac{1}{2} \sum_{t=t_{min}}^{t_{max}} \sum_{x=x_{min}}^{x_{max}} (\hat{q}_{x,t} - g(x, t, \beta))^2$

L'algorithme de boosting en 1 dimension appliqué à notre base de spline se présente de la manière suivante :

**Algorithm 1** – *Algorithme GBM Spline avec la fonction de coût  $J_p(\beta)$* **Require:** Valeurs initiales :  $F_0 = \beta_0 = \underset{\beta}{\operatorname{argmin}} J_p(Y, \beta)$ **Require:** Matrice de design :  $X$ 2: **for**  $k=1, \dots, m$  **do**

$$z^{(k-1)} \leftarrow \left( -\nabla_{F(X)} J_p(Y, F(X)) \Big|_{F=F^{(k-1)}} \right) = \left( z_i^{(k-1)} \right)_{i=x_{\min}, \dots, x_{\max}}$$

4:

$$f^{(k)}(X) \leftarrow X(XX^T + \lambda P)^{-1} X^T z^{(k-1)}$$

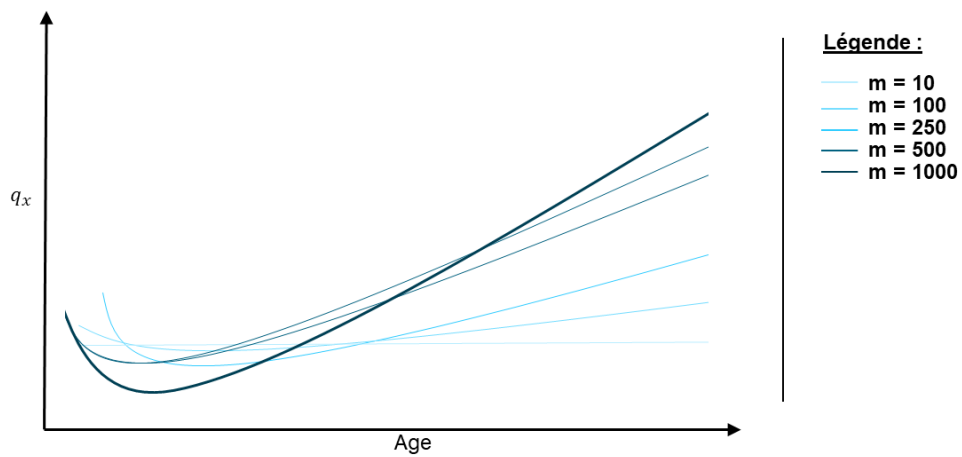
6:

$$\gamma^{(k)} = \underset{\gamma}{\operatorname{argmin}} J_p(Y, F^{(k-1)}(X) + \gamma f^{(k)}(X))$$

8:

$$F^{(k)} \leftarrow F^{(k-1)} + \nu \gamma^{(k)} f^{(k)}$$

10:

**end for****FIGURE 2.10** – Schéma de l'adaptation de la fonction lissée en fonction des itérations de  $m$

Cette méthode se compose en plusieurs étapes. La première consiste en l'initialisation de  $F_0$ , généralement  $F_0 = \bar{y}$ . Ensuite, on itère le processus d'optimisation jusqu'à  $m_{stop}$ , avec  $m$  qui correspondra au nombre d'itération. Tout d'abord on calcule  $z^{(k-1)}$ , cette valeur est en fait un gradient négatif qui nous donne des indications sur les directions (+/-) et l'ampleur dans lesquelles la fonction de coût peut être minimisée. Cette technique, où nous utilisons un gradient pour minimiser la perte sur notre modèle est très similaire à la technique de descente de gradient présentée Figure 2.12. Finalement la valeur  $z^{(k-1)}$  ne correspond qu'aux résidus du modèle. Avec, la fonction de coût d'erreur quadratique, on a :

$$\begin{aligned} z^{(k-1)} &= \left( -\nabla_{F(X)} J_p(Y, F(X)) \right)_{F=F^{(k-1)}} \\ &= Y - F^{(k-1)}(X) \end{aligned} \tag{2.66}$$

En fin de compte, ce qui donne la direction de descente et la « magnitude » de la minimisation de la fonction de coût n'est rien d'autre que les résidus du modèle.

Ensuite, nous évaluons nos résidus en utilisant nos apprenants faibles, qui dans notre cas sont des splines cubiques. On utilise ces apprenants faibles, car ils présentent de nombreuses qualités de régularité grâce à des propriétés issues de la récurrence de De Boor. Avec,  **$P$  : le terme de pénalisation et  $X$  : la matrice de design constitué des splines évaluées aux points d'interpolations.**

On ajoute ensuite à la prédiction les modifications réalisées tout en contrôlant l'apprentissage du modèle grâce à  $\nu$ , un taux d'apprentissage (*learning rate*) compris entre 0 et 1 qui contrôle le degré de contribution de la régression supplémentaire  $\gamma f^{(k)}$  à la prédiction combinée  $F^{(m)}$ . Un taux d'apprentissage plus petit réduit l'effet de la régression supplémentaire sur notre taux lissé final mais il réduit aussi fondamentalement le risque de sur-adaptation du modèle aux données d'apprentissage. On observe d'ailleurs que notre fonction lissée finale sera définie de la manière suivante :

$$F^{(m)}(X) = \nu \sum_{k=0}^{m-1} \gamma^k f^k(X) \tag{2.67}$$

On retrouve l'expression de l'équation 2.62 si notre taux d'apprentissage vaut 1 et en posant que  $g^{(k)}(X) = \gamma^k f^k(X)$ , c'est à dire que :

$$F^{(m)}(X) = G^{(m)}(X) = \sum_{k=0}^{m-1} g^{(k)}(X) \tag{2.68}$$

On peut aussi spécifier le degré de liberté  $df$  du modèle ce qui permet d'éviter un sur-apprentissage. En apprentissage automatique, les degrés de liberté peuvent faire référence au nombre de paramètres du modèle, comme le nombre de coefficients dans un modèle de régression linéaire ou le nombre de poids dans un réseau neuronal d'apprentissage profond. Le problème est que s'il y a plus de degrés de liberté (paramètres du modèle) dans l'apprentissage automatique, on s'attend à ce que le modèle sûr ajuste l'ensemble des données d'apprentissage. On se référera à Hofner et al. [2011] qui explique que le fait de rendre tous les apprenants de base comparables en termes de degrés de liberté réduit fortement le biais de sélection observé dans les spécifications du boosting naïf, et que plus le degré de liberté est faible moins le modèle sûr ajuste.

$$df = \text{Trace}(X(XX^T + \lambda P)X^T) \tag{2.69}$$

Ainsi la difficulté de du modèle GBM est l'étape de la calibration à nos données. En effet dans le cadre de portefeuilles avec peu de valeurs il est difficile pour l'algorithme de boosting de trouver les valeurs optimales rapidement. L'algorithme nécessitera alors un nombre élevé d'itérations  $m$ . Ce cas peut arriver lorsque nous étudions une granularité plus fine. Par exemple dans le cas en 2 dimensions présenté Section 3. Il faut alors avoir recours à des méthodes pour éviter de sûr ajuster (*overfit*).

Il existe diverses méthodes de **régularisation** ou de **contrainte** pour améliorer les performances de l'algorithme en matière de lutte contre le sûr ajustement, méthodes que nous utilisons dans le cadre de notre projet. D'autres, quant-à-elles, consistent à modifier l'échantillon de données, tirer parti du rétrécissement, de l'échantillonnage aléatoire ou encore de l'apprentissage pénalisé pour lutter contre le sûr dimensionnement. Nous utiliserons **l'échantillonnage pénalisé** dans le cadre de nos travaux.

Aussi, on étudiera dans ce mémoire un modèle proposé par Mayr et al. [2012] pour éviter cette problématique de sûr ajustement lorsqu'on utilise un modèle de boosting. Ce dernier consiste à transformer les apprenants faibles en prenant en considération des propriétés mathématiques des splines. Généralement on connaît la forme globale de notre courbe des taux lissés de mortalité : elle est généralement **croissante** et **convexe**.

Si l'une des conditions mentionnées n'est pas respectée (croissance ou convexité), les coefficients seront soumis à une contrainte. La formule de récurrence de De Boor (Section 2.5.3) permet de calculer la première et la deuxième dérivée de nos splines cubiques en conservant notre ensemble de nœuds  $\mathcal{U} = \{u_1, \dots, u_N\}$ , où  $n = N + k - 1$ ,

$$\begin{aligned}
 f'(x) &= \sum_{j=1}^n \beta_j B'_{j,k}(x) \\
 &= \sum_{j=1}^n \beta_j \left( \frac{k}{u_{j+k} - u_j} B_{j,k}(x) - \frac{k}{u_{j+k+1} - u_{j+1}} B_{j+1,k-1}(x) \right) \\
 &= \left( k \sum_{j=0}^{n-1} \frac{\beta_{j+1}}{u_{j+k+1} - u_{j+1}} B_{j+1,k-1}(x) \right) - \left( k \sum_{j=1}^n \frac{\beta_j}{u_{j+k+1} - u_{j+1}} B_{j+1,k-1}(x) \right)
 \end{aligned} \tag{2.70}$$

On rappelle qu'une spline d'ordre  $k$  correspond à un polynôme de degré  $k - 1$ . De plus, on a, grâce aux propriétés des splines de la récurrence de De Boor, que le premier terme de la somme de gauche vaut 0, ainsi que le dernier terme de la somme de droite, car nous avons des valeurs évaluées en  $0/0$  qui valent 0 par définition de la récurrence. On obtient alors la dérivée suivante, qui constitue une contrainte dite de **monocité**,

$$\begin{aligned}
 f'(x) &= k \sum_{j=1}^{n-1} \left( \frac{\beta_{j+1} - \beta_j}{u_{j+k+1} - u_{j+1}} B_{j+1,k-1}(x) \right) \\
 &= k \sum_{j=1}^{n-1} \left( \frac{\Delta^1 \beta_j}{u_{j+k+1} - u_{j+1}} B_{j+1,k-1}(x) \right)
 \end{aligned} \tag{2.71}$$

Pour finir on réarrange nos noeuds, car en dérivant notre fonction on réduit notre intervalle d'observations. Comme nous réalisons une dérivée première, nous transformons seulement notre ensemble de noeuds en  $\mathcal{U}' = \{u_2, \dots, u_{N-1}\}$  Le signe de la dérivée  $f(\cdot)$  dépend donc directement du signe de  $\Delta^1 \beta_j$  car les autres termes sont toujours supérieurs ou égaux à 0,  $k$  correspond à l'ordre du polynôme qui est toujours supérieur ou égal à 0 de même que les splines évalués  $B(\cdot)$ , mais aussi la distance  $h$ . On peut obtenir une relation de récurrence concernant nos splines cubiques en s'appuyant sur les travaux de Speleers [2019] on obtient,

$$\frac{\partial^d f(x)}{\partial x^d} = \sum_{j=1}^{n-d} \beta_{j,d} B_{i,k-d}(x) \quad (2.72)$$

Avec les coefficients  $\beta$  définis de la manière suivante sur l'ensemble de noeuds modifiés suite à la dérivée  $k$ -ème,  $\mathcal{U}^d = \{u_{d+1}, \dots, u_{N-d}\}$ ,

$$\beta_{j,d} = \begin{cases} \beta_j & \text{Si, } d = 0 \\ \alpha_{j,d} (\beta_{j+1,d-1} - \beta_{j,d-1}) = \alpha_{j,d} \Delta \beta_{j,d-1} & \text{Si, } d > 0 \end{cases} \quad (2.73)$$

Avec,

$$\alpha_{j,d} = \frac{k-d+1}{u_{j+k+1} - u_{j+d}} \quad (2.74)$$

Pour obtenir la contrainte de concavité, c'est à dire la dérivée seconde de notre fonction  $f(\cdot)$  on cherche à obtenir  $\beta_{j,2}$ . On obtient grâce à la relation 2.73 la valeur suivante,

$$\beta_{j,2} = \alpha_{j,2} \Delta \beta_{j,1} = \alpha_{j,2} \alpha_{j,1} \Delta^2 \beta_j \quad (2.75)$$

Ainsi on obtient la dérivée seconde de  $f(\cdot)$ ,

$$f''(x) = \sum_{j=1}^{n-2} \alpha_{j,2} \alpha_{j,1} \Delta^2 \beta_j B_{i,k-2}(x) \quad (2.76)$$

Ce résultat nous permet d'avoir une relation entre la convexité de notre fonction  $f(\cdot)$  et la pénalisation seconde des coefficients, soit  $\Delta^2 \beta_j$ . C'est à dire qu'en fonction du signe de la différence première et seconde notre fonction sera croissante, décroissante, convexe ou concave.

On cherchera donc, dans notre contexte spécifique d'une étude de **mortalité**, à poser deux contraintes associées à une fonction qui est censée être croissante et convexe. De cette manière, nous ajouterons à notre fonction de coût pénalisée une autre pénalité, qui pénalise les différences premières négatives (**contrainte de croissance**) et les différences secondes négatives (**contrainte de convexité**). Le principe consiste à ajouter le terme  $P_{mono}$  suivant :

$$P_{mono} = \sum_{j=1}^{n-z} v_j \Delta^z \beta_j \quad (2.77)$$

Avec,

$$v_j = \begin{cases} 0 & \text{Si, } \Delta^z \beta_j \geq 0 \\ 1 & \text{Si, } \Delta^z \beta_j < 0 \end{cases} \quad (2.78)$$

Ainsi, nous pouvons pénaliser davantage les coefficients qui ne respectent pas la structure générale des taux de mortalité en ajustant les valeurs de  $z$ . Lorsque  $z = 1$ , cela correspondra à la croissance, et lorsque  $z = 2$ , cela correspondra à la convexité. On obtient ainsi la formule suivante :

$\forall k \in \llbracket 1, m \rrbracket$ ,

$$f^{(k)}(X) \leftarrow X(XX^T + \lambda_1 P + \lambda_2 P_{mono})X^T z^{(k-1)} \quad (2.79)$$

L'intérêt de cette méthode est de contrebalancer le sûr ajustement du modèle GBM en lui imposant des contraintes. Le paramètre de lissage  $\lambda_2$  est choisi comme étant assez grand pour avoir un impact élevé sur la trajectoire des points lissés. De plus, cette méthode permet de contrôler la forme attendue de la courbe in fine et dans le cadre spécifique du lissage des taux de mortalité bruts, d'éliminer le bruit individuel associé aux taux bruts. Dans un cadre de réassurance, cette composante est essentielle pour notre modèle. Car en effet, on souhaite obtenir un tarif de réassurance pour notre portefeuille, il s'agit ainsi de capturer la tendance globale du portefeuille réassuré pour obtenir un tarif optimal à partir de l'expérience du portefeuille.

Enfin l'intérêt aussi du modèle est de pouvoir ajouter un avis plus « d'expert » dans la méthode de lissage à l'aide de contraintes sur la forme de la courbe à certains âges. Cet avis d'expert peut être en lien avec des tables standards de mortalités. Cependant, cette information ne peut pas être réellement prise en considération pour les âges avec peu de données ou il est difficile de distinguer des tendances « usuelles ». Nous présenterons cette adaptation du modèle GBM et la comparerons avec les méthodes « classiques » de lissage dans le cadre de la Partie 3.

### 2.10.1 Méthodes d'optimisation des paramètres et construction du caractère « adaptatif » de l'ensemble des techniques de lissage

Pour estimer les paramètres du modèle GAM et des modèles GLM locaux, plusieurs méthodes existent. On se concentrera dans le cadre de ce mémoire sur des méthodes de descente de gradient dites de second ordre. On utilise les méthodes du second ordre, car elles sont généralement plus précises que celles du premier ordre. Les méthodes d'estimation des paramètres sont essentielles et constituent un véritable enjeu, qu'il soit opérationnel vis à vis du temps de calcul dans le cadre de modèles avec plusieurs variables explicatives ou vis à vis de la qualité d'estimation et dans notre cas de lissage.

En effet, lorsque l'on souhaite lisser nos données avec une dimension, le choix de la méthode de lissage n'a pas de réelle utilité, car les temps de calcul sont similaires. Cependant, lorsque l'on souhaite augmenter le nombre de dimensions et de variables cela présente un intérêt en particulier pour estimer les paramètres du modèle GAM.

On posera dans cette partie pour alléger les notations :

- **La log-vraisemblance** :  $l(\beta) = l(\beta; y, \phi) = \ln(L(\beta; y, \phi))$
- **La fonction de coût** :  $J(\beta) = -l(\beta)$

#### Méthode de Newton

La méthode la plus connue et utilisée est la méthode de Newton. Elle nous servira de fondement pour introduire les autres méthodes et les comparer. On rappelle qu'elle consiste en une expansion de Taylor du second ordre de la fonction de coût. On aura en approximant  $J(\beta)$  en un point initial  $\beta_0$  :

$$J(\beta_0 + \delta\beta) \simeq J(\beta_0) + (\beta - \beta_0)^T \nabla_{\beta} J(\beta_0) + \frac{1}{2}(\beta - \beta_0)^T H(J)(\beta - \beta_0) \quad (2.80)$$

Avec,  $\delta\beta = \beta - \beta_0$

On peut ensuite obtenir à partir de l'équation 2.80, une relation de récurrence en évaluant la relation suivante au point critique  $\beta^*$ . On obtient ainsi :

$$\beta^* = \beta_0 - H(J)^{-1} \nabla_{\beta} J(\beta_0) \quad (2.81)$$

---

**Algorithm 2** – *Newton-Raphson avec la fonction de coût  $J_p(\beta)$*

---

**Require:** Valeurs initiales :  $\beta_0$

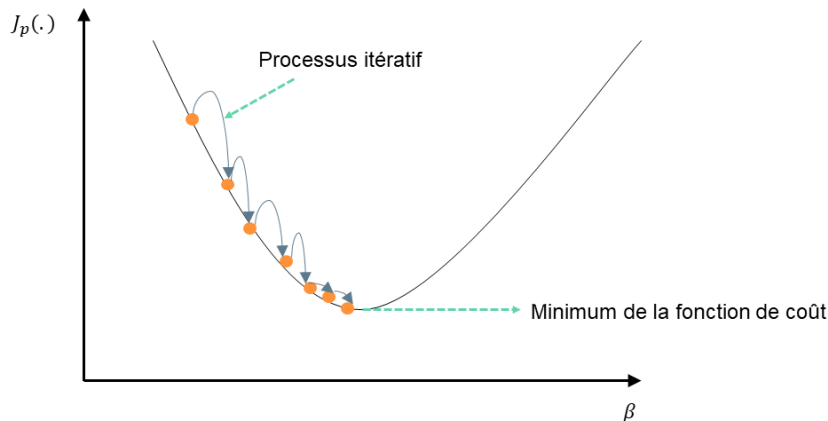
**Require:** Matrice de design :  $X$

**while**  $|\beta^{(k+1)} - \beta^{(k)}| \leq \epsilon$  **do**

2:  $\beta^{(k+1)} \leftarrow \beta^{(k)} - H(J_p)^{-1} \nabla_{\beta} J_p(\beta^{(k)})$

**end while**

---



**FIGURE 2.11** – Schéma de l'algorithme de Newton-Raphson avec la fonction de coût pénalisée

Généralement le pas  $\epsilon$  est choisi de manière arbitraire, mais doit nécessairement être très faible par exemple  $10e^4$ . Dans la suite de cette étude, on montrera que la plupart des méthodes d'optimisation sont mises en place avec l'hypothèse que l'on peut aisément calculer la valeur exacte de la matrice hessienne  $H$  et du gradient. Cependant, la grande majorité des méthodes approximent ces quantités. En effet, il est parfois difficile d'évaluer la matrice  $H$  et encore moins aisé de l'inverser : c'est même parfois impossible.

### Méthode du scoring de Fisher (GLM et GAM)

Ainsi, Ronald Fisher propose alors une méthode statistique pour éviter ce problème, celle-ci est la méthode dite du *scoring de Fisher* présentée par Jennrich et Sampson [1976].



Cette méthode consiste à approximer la matrice hessienne par la matrice d'information de Fisher noté  $\mathcal{I}(\beta)$ , de telle manière que :

$$-H(f) \simeq \mathbb{E} \left( -\frac{\partial^2 J_p(\beta)}{\partial \beta \partial \beta^T} \right) = \mathcal{I}(\beta) \quad (2.82)$$

Cette modification est intéressante, car il s'avère que l'évaluation et l'inversion empirique de l'information de Fisher est souvent moins exigeante en matière de calcul que l'évaluation et l'inversion de la matrice hessienne observée. De plus dans le cadre du GLM et du modèle GAM, on connaît la distribution sous-jacente de notre variable de comptage du nombre de sinistres  $Y$  qui fait partie de la famille exponentielle. Dans le cadre du mémoire, nous utiliserons cette méthode pour estimer les paramètres des modèles GLM locaux.

On obtient alors l'algorithme modifié avec le scoring de Fisher suivant :

---

**Algorithm 3** – *Scoring de Fisher avec la fonction de coût  $J_p(\beta)$*

---

**Require:** Valeurs initiales :  $\beta_0$   
**Require:** Matrice de design :  $X$   
**while**  $|\beta^{(k+1)} - \beta^{(k)}| \leq \epsilon$  **do**  
 2:  $\beta^{(k+1)} \leftarrow \beta^{(k)} + \mathcal{I}(\beta^{(k)})^{-1} \nabla_{\beta} J_p(\beta^{(k)})$   
**end while**

---

Ainsi, avec nos hypothèses de distribution du nombre de décès, on obtient en sommant tous les âges  $i$ ,

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket$$

$$\nabla_{\beta} J_p(\beta) = \sum_i (y_i x_i - x_i \mu_i) = X^T (y - \mu) \quad (2.83)$$

$$\mathcal{I}(\beta) = (X^T \Omega X + P)$$

Avec la matrice des poids ré-itérés notée :

$$\Omega = \text{diag}(\gamma_{x_{min}}, \dots, \gamma_{x_{max}}) \quad \text{et} \quad \gamma_i = \frac{1}{b''(\theta_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

À partir de ces expressions, on remplace leurs valeurs au temps  $t$  pour obtenir l'algorithme IRWLS (*Iteratively Re-weighted Least Squares*). Cet algorithme est encore utilisé pour déterminer les paramètres de la plupart des modèles GLM, car il est très peu coûteux en temps de calcul. Cet algorithme est apparu comme une solution à des contraintes de temps de calcul dans les années 70.

En effet au début des années 1970, il était clairement nécessaire de trouver une méthode d'optimisation permettant d'estimer des modèles non-linéaires à l'aide des méthodes MCO (Moindres Carrés Ordinaires) standards. Wedderburn et Nelder ont découvert que les méthodes utilisées pour estimer la régression linéaire pondérée pouvaient être ajustées pour modéliser de nombreuses situations de données qui étaient auparavant estimées par la méthode du maximum de vraisemblance, en particulier pour les modèles de maximum de vraisemblance basés sur la famille exponentielle des distributions.

$$\begin{aligned}
 \beta^{(k+1)} &\leftarrow \beta^{(k)} + \mathcal{I}(\beta^{(k)})^{-1} \nabla_{\beta} J_p(\beta^{(k)}) \\
 &\leftarrow \beta^{(k)} + (X^T \Omega^{(k)} X + P)^{-1} X^T (y - \mu^{(k)}) \\
 &\leftarrow (X^T \Omega^{(k)} X + P)^{-1} X^T \Omega^{(k)} \underbrace{\left[ X \beta^{(k)} + P \beta^{(k)} + (\Omega^{(k)})^{-1} (y - \mu^{(k)}) \right]}_{z^{(k)}} \\
 &\leftarrow (X^T \Omega^{(k)} X + P)^{-1} X^T \Omega^{(k)} z^{(k)}
 \end{aligned} \tag{2.84}$$

De plus, on remarque qu'une itération de Newton est équivalente à la résolution d'un problème de moindres carrés pondérés  $\sum_i \gamma_i (z_i - x^T \beta)^2$  d'où le nom de « *Iteratively Re-weighted Least Squares* ». L'algorithme se décompose de la manière suivante en supposant toujours que nos nombres de sinistres suivent une loi de poisson,

---

**Algorithm 4** – *Algorithme IRWLS avec la fonction de coût  $J_p(\beta)$*

---

**Require:** Valeurs initiales :  $\beta_0$   
**Require:** Matrice de design :  $X$   
**while**  $|\beta^{(k+1)} - \beta^{(k)}| \leq \epsilon$  **do**  
 2:  $\eta^{(k)} \leftarrow X^T \beta^{(k)} + \log(E)$   
 $\mu^{(k)} \leftarrow \exp(\eta^{(k)})$   
 4:  $\frac{\partial \eta^{(k)}}{\partial \mu^{(k)}} \leftarrow (\mu^{(k)})^{-1}$   
 $\Omega^{(k)} \leftarrow \mu^{(k)}$   
 6:  $z^{(k)} \leftarrow X^T \beta^{(k)} + (y - \mu^{(k)}) \times \frac{\partial \eta^{(k)}}{\partial \mu^{(k)}}$   
 $\beta^{(k+1)} \leftarrow (X^T \Omega^{(k)} X + P)^{-1} X^T \Omega^{(k)} z^{(k)}$   
 8: **end while**

---

Avec,  $E$  qui correspond à la variable *offset* et qui contient les expositions pour un âge et une durée donnée. Ainsi, au-delà du défi créé par certaines caractéristiques de la fonction objectif tel que le point de selle, la méthode de Newton classique a une charge de calcul trop élevée pour de nombreux paramètres. Le nombre d'éléments de la matrice hessienne étant le carré du nombre de paramètres par exemple avec  $k$  paramètres cette méthode devra inverser une matrice  $k \times k$ . De plus l'inversion de cette matrice présente une complexité élevée, de  $O(k^3)$ .

Dans le cadre du modèle GLM local des modifications devront être réalisées. En effet on ajoutera aux poids réitérés les poids associés à la fonction de poids en 1 et 2 dimensions (la fonction *kernel*) choisie et présentée dans le cadre de ce mémoire Section 2.3.

Avec la matrice de design  $X$  pour l'individu  $i$  suivante en 1 dimension (âge),

$$X = \begin{bmatrix} 1 & (x_{min} - x_i) & (x_{min} - x_i)^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & (x_{max} - x_i) & (x_{max} - x_i)^2 \end{bmatrix}^T$$

Avec la matrice de design  $X$  pour l'individu  $i$  suivante en 2 dimensions (âge et durée de la police d'assurance),

$$X = \begin{bmatrix} 1 & (x_{min} - x_i) & (t_{min} - t_i) & (x_{min} - x_i)^2 & (x_{min} - x_i)(t_{min} - t_i) & (t_{min} - t_i)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (x_{max} - x_i) & (t_{max} - t_i) & (x_{max} - x_i)^2 & (x_{max} - x_i)(t_{max} - t_i) & (t_{max} - t_i)^2 \end{bmatrix}^T$$

À partir des matrices de design et de la vraisemblance, on obtient notre valeur du scoring de Fisher pour l'individu avec les caractéristiques  $i$ . On conservera le paramètre  $\phi$  car nous présentons le modèle général avec la méthode de quasi-poisson.

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$I_{j,k} = \mathbb{E} \left( -\frac{\partial l(\beta)^2}{\partial \beta_j \beta_k} \right) = \sum_i \frac{\omega_{i,i}}{\phi b''(\theta_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 x_{ij} x_{ik} = \sum_i \frac{\omega_{i,i} * \gamma_{i,i}}{\phi} x_{ij} x_{ik}$$

On notera  $W$  la matrice des poids issus de la vraisemblance locale. On obtient ainsi matriciellement le scoring de Fisher dans le cadre du modèle GLM local,

$$I_{j,k} = (X^t W \Omega X)_{j,k} \tag{2.85}$$

- $W$  : La matrice des poids issus de la vraisemblance locale
- $\Omega$  : La matrice des poids réitérés

On obtient ainsi nos paramètres  $\beta$  ré-itérés dans le cadre de l'algorithme du scoring de Fisher de la manière suivante (on notera qu'ici aucun terme de pénalité n'intervient dans le modèle GLM local) :

$$\forall k \in \llbracket 1, m \rrbracket,$$

$$\beta^{(k+1)} = (X^T W \Omega^{(k)} X)^{-1} X^T W \Omega^{(k)} Z^{(k)} \tag{2.86}$$

Nous présenterons par la suite une autre méthode d'approximation de cette matrice hessienne, inspirée de la méthode du scoring de Fisher, mais ajoutant des hypothèses sur la direction de notre descente de gradient. Cela permet alors d'augmenter la vitesse de notre algorithme d'optimisation en convergeant plus rapidement vers le résultat optimal.

### Méthode de quasi-newton et gradient conjugué (LOESS / B-P Spline / GBM / RBF Network)

À partir de la méthode de Quasi-Newton on peut revenir à une méthode dite de descente du gradient. La méthode du Quasi-Newton consiste à fixer la matrice hessienne comme un hyper paramètre, le taux d'apprentissage  $\alpha = H(J_p)$ . Ainsi, nous n'avons plus de matrice de poids d'apprentissage, mais seulement un scalaire comme valeur d'apprentissage. Il faudra d'ailleurs, dans cette situation, être vigilant quant à la fonction que nous souhaitons optimiser, si cette fonction est convexe ou concave. En effet, dans ce cas-là, il faudra modifier le signe du paramètre  $\alpha$  pour que la descente soit bien ascendante ou descendante.

Dans le cadre de ce mémoire et pour optimiser les paramètres, en particulier en 2 dimensions, nous nous tournerons plutôt vers les méthodes dites de gradients conjugués. Ce sont des méthodes qui permettent d'éviter efficacement le calcul de l'hessienne inverse en descendant itérativement des directions conjuguées. L'inspiration de cette approche découle d'une étude sur la faiblesse des

méthodes par descente de gradient simple avec un taux d'apprentissage fixe. Cela permet de résoudre certains problèmes liés à l'utilisation d'un taux d'apprentissage fixe.

En effet, la différence au début de l'apprentissage est légère, mais cette méthode permet d'éviter de réaliser des allers-retours vers l'optimum lorsqu'on est proche de la solution ; et ainsi de converger plus rapidement vers l'optimum global. Le principe de cette méthode est que chaque direction de recherche des paramètres optimaux est orthogonale à la précédente. On présentera la recherche des paramètres optimaux dans un schéma explicatif Figure 2.12 dans un cas simple avec une fonction de coût et deux paramètres à estimer.

En notant **la direction de recherche** précédente,  $d^{(k)}$  au point minimum où l'on change de direction de recherche. A ce point on a que la dérivée de notre fonction de coût en les paramètres et dans la direction  $d^{(k)}$  vaut 0. C'est à dire que mathématiquement,  $\nabla_{\beta}(J_p(\beta^{(k)}) \cdot d^{(k)} = 0$ . On pose ainsi  $\nabla_{\beta}(J_p(\beta^{(k)})) = d^{(k+1)}$  et on obtient notre nouvelle direction de recherche dont on est assurés qu'elle est orthogonale à la précédente.

La principale difficulté de cette méthode est de trouver une expression convenable pour notre direction de recherche  $d$  des paramètres optimaux. Le lecteur intéressé pourra les différentes possibilités de directions de recherche. Dans le cadre de ce mémoire et pour l'estimation des paramètres du modèle LOESS nous avons considéré la direction de recherche conjuguée introduite par Fletcher et Reeves [1964].

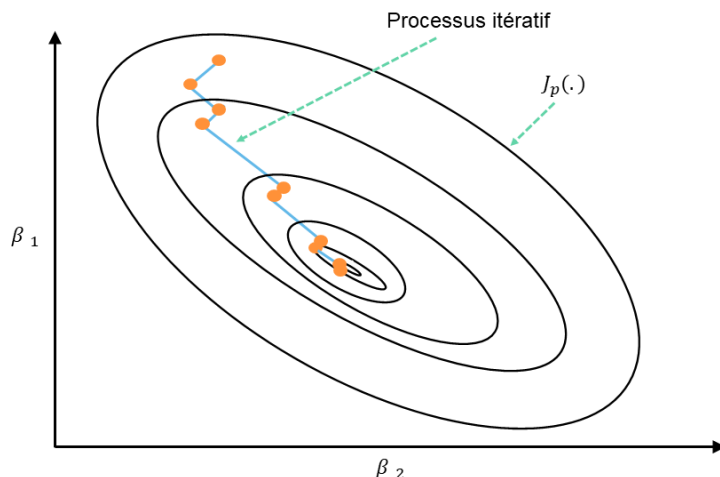


FIGURE 2.12 – Schéma explicatif de la méthode du gradient conjugué

Tout d'abord, on pose à l'itération  $k + 1$  la nouvelle direction de recherche,

$$d^{(k+1)} \leftarrow \nabla_{\beta} J(\beta^{(k)}) + \zeta^{(k)} d^{(k)} \quad (2.87)$$

Le paramètre  $\zeta$  contrôle le niveau d'importance que l'on donne à la dernière direction de recherche. Plus il est élevé plus celle-ci est importante. Pour satisfaire notre contrainte d'orthogonalité entre les directions de recherches dans l'algorithme d'optimisation, on impose selon Fletcher et Reeves [1964] l'expression suivante pour  $\zeta$ ,

$$\zeta^{(k)} = \frac{\nabla_{\beta} J(\beta^{(k)})^T \nabla_{\beta} J(\beta^{(k)})}{\nabla_{\beta} J(\beta^{(k-1)})^T \nabla_{\beta} J(\beta^{(k-1)})} \quad (2.88)$$

### Méthode d'estimation des paramètres $\lambda$

La section précédente a présenté plusieurs algorithmes pour estimer les paramètres  $\beta$  de nos modèles. Cependant, dans ces méthodes, aucun ne fait pas intervenir nos paramètres de pénalités. Or, l'objectif de ce mémoire est de proposer une méthode de lissage « adaptative » et « semi-automatique » pour les utilisateurs non-experts. Dans cette partie, nous présentons les stratégies d'estimation des paramètres de pénalisation développées pour le projet, ainsi que le critère choisi pour obtenir des estimations lisses tout en minimisant le biais.

Dans le cadre d'un besoin d'automatisation et d'adaptabilité des méthodes utilisées pour qu'elles soient proposées à différents utilisateurs dont des utilisateurs novices, il s'agit d'estimer les paramètres de lissage de la manière la plus cohérente possible, en réalisant toujours un compromis entre le biais et la variance. Pour cela, nous utilisons le **critère GCV** proposé par G. Wahba et Heath [1979] pour paramétrer et estimer les  $\lambda$  optimaux, qui permettront de minimiser à la fois le critère de fidélité, mais aussi de régularité/pénalité pour les différentes méthodes (voir comparaison Wahba [1985]) :

$$GCV(\lambda) = \frac{WRSS/n}{(1 - Tr(H(\lambda))/n)^2} \quad (2.89)$$

Avec  $H(\lambda)$  la matrice trouvée lors de la résolution des paramètres pour le lissage.  
Et le WRSS qui correspond à la somme des carrés des résidus pondérés :

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$WRSS = \sum_i w_i (y_i - f(x_i, \alpha))^2 \quad (2.90)$$

Nous choisirons le  $\lambda$  qui minimise la GCV (*Generalized Cross Validation*) de telle manière que :

$$\lambda^* = \operatorname{argmin}(GCV) \quad (2.91)$$

En appliquant une transformation logarithme à l'équation 2.89 on observe que l'on obtient :

$$\log(WMSE) + g(S) \quad (2.92)$$

C'est-à-dire que minimiser la GCV revient à minimiser les deux termes suivants :

- **Un critère de fidélité** :  $\log(WMSE)$  avec  $WMSE = \frac{1}{n} \sum_i w_i (y_i - f(x_i, \alpha))^2$
- **Un critère de régularité** :  $g(S) = -2\log(1 - Tr(H(\lambda)/n))$

Car minimiser  $g(\cdot)$  revient à maximiser la régularité de la fonction  $f(\cdot)$ . Un autre critère utilisé dans la littérature pour estimer de manière automatique les paramètres de pénalité est le **critère de Mallows**. On peut le retrouver en approximant la valeur  $\frac{1}{(1-Tr(H(\lambda))/n)} \simeq 1 + 2 Tr(H(\lambda))/n$  dans notre critère GCV :

$$\begin{aligned}
 GCV(\lambda) &= \frac{WRSS/n}{(1-Tr(H(\lambda))/n)^2} \\
 &\simeq \frac{WRSS}{n} + 2 \frac{Tr(H(\lambda))}{n^2} WRSS \\
 &\simeq \frac{WRSS}{n} + 2 \frac{\hat{\sigma}^2 Tr(H(\lambda))}{n} \\
 &= Cp(\lambda)
 \end{aligned} \tag{2.93}$$

Il convient d'être attentif à l'utilisation du critère GCV et des méthodes automatiques proposées. En effet, la minimisation du critère GCV peut être très variable : deux ensembles de données qui semblent visuellement similaires peuvent conduire à des résultats très différents. Il est donc recommandé d'utiliser le critère GCV pour chaque estimation de nos paramètres afin de comparer la cohérence globale de nos résultats avec d'autres critères, à la fois visuels et statistiques, que nous présenterons dans la Section 2.11.

Il existe généralement 2 stratégies de calcul pour (voir Fauvernier et al. [2019]) la sélection de  $\lambda$ , qui, si elles ne sont pas uniques, sont régulièrement utilisées :

**Les schémas à itération unique :** ils estiment le paramètre de pénalité  $\lambda$  à chaque étape d'itération d'estimation des paramètres  $\hat{\beta}$ . Dans notre cas on utilisera le critère GCV pour estimer cette valeur  $\lambda$  à chaque itération.

**Les schémas à itération imbriquée :** Ils définissent un critère de sélection  $\lambda$  en termes de déviance du modèle et l'optimisent directement. Par exemple, dans notre cas, on a décidé de choisir le critère GCV. Chaque évaluation du critère nécessite une estimation « interne » des paramètres  $\hat{\beta}$  avec une méthode d'estimation adaptée au modèle de lissage, on peut ainsi obtenir le terme  $\hat{\beta}_\lambda$ .

Généralement, le choix d'une méthode ou de l'autre dépend de notre jeu de données. La seconde méthode est assez longue - bien sûr cela est à pondérer par le choix de la méthode d'estimation des  $\beta$  que l'on choisit. Mais dans tous les cas pour des jeux de données importants, elle présente des temps de calculs longs. Cependant, on est presque certain de converger avec cette méthode, ce qui n'est pas le cas de la première méthode.

Finalement, la deuxième option est généralement préférable pour des raisons de fiabilité, mais la première peut-être rendue très efficace en termes de mémoire avec de très grands ensembles de données. Dans le cadre de ce mémoire, on utilisera le schéma imbriqué pour des raisons de convergence. L'algorithme avec notre critère GCV sera mis en place de la manière suivante :

On note dans l'Algorithme 5 :

- $H(GCV(\lambda^{(k)}))$  : La matrice hessienne du critère GCV par rapport à  $\lambda$
- $\nabla_\lambda GCV(\lambda^{(k)})$  : Le gradient du critère GCV par rapport à  $\lambda$

---

**Algorithm 5** – *Algorithme imbriqué pour estimation du paramètre de pénalité  $\lambda$  avec la fonction de coût  $J_p(\beta)$  pour  $\beta$  et  $GCV(\lambda)$  pour  $\lambda$*

**Require:** Valeurs initiales :  $\lambda_0$

**Require:** Matrice de design :  $X$

2: **while**  $|\lambda^{(k+1)} - \lambda^{(k)}| \leq \epsilon$  **do**

4: On choisit  $\beta_{\lambda_k}^{(0)}$

6: **while**  $|\beta^{(u+1)} - \beta^{(u)}| \leq \epsilon$  **do**

8: À partir de  $\beta_{\lambda_k}^0$  on applique un algorithme d'estimation de  $\beta$  (BFGS/Newton-Raphson/Scoring de Fisher/Quasi-Newton) en **interne** jusqu'à obtenir convergence.

10:  $\beta_{\lambda_k}^{(u+1)} \leftarrow \beta_{\lambda_k}^{(u)} - H(J_p(\beta_{\lambda_k}^{(u)}))^{-1} \nabla_{\beta} J_p(\beta_{\lambda_k}^{(u)})$

12: **end while**

14: On récupère  $\hat{\beta}_{\lambda_k}$  à convergence.

16: À partir de  $\hat{\beta}_{\lambda_k}$  on applique un algorithme d'estimation de  $\lambda$  (BFGS/Newton-Raphson/Scoring de Fisher/Quasi-Newton) en **externe** jusqu'à obtenir convergence.

18:  $\lambda^{(k+1)} \leftarrow \lambda^{(k)} - H(GCV(\lambda^{(k)}))^{-1} \nabla_{\lambda} GCV(\lambda^{(k)})$

20: **end while**

22: On récupère  $\hat{\lambda}$  à convergence et les  $\hat{\beta}_{\hat{\lambda}}$  correspondants.

---

La complexité de cet algorithme dépend ainsi de plusieurs critères : le nombre de paramètres de lissage  $\lambda$  à estimer, le nombre de paramètres  $\beta$  mais aussi de la méthode d'optimisation choisie. Généralement la méthode de Newton simple est trop longue. On se tournera plutôt vers une méthode de Quasi-Newton ou BFGS. Il faudra cependant être vigilant à la fois à la convergence de cet algorithme, mais aussi de la valeur de la matrice hessienne, qui sera perturbée dans le cas où elle ne serait pas définie positive.

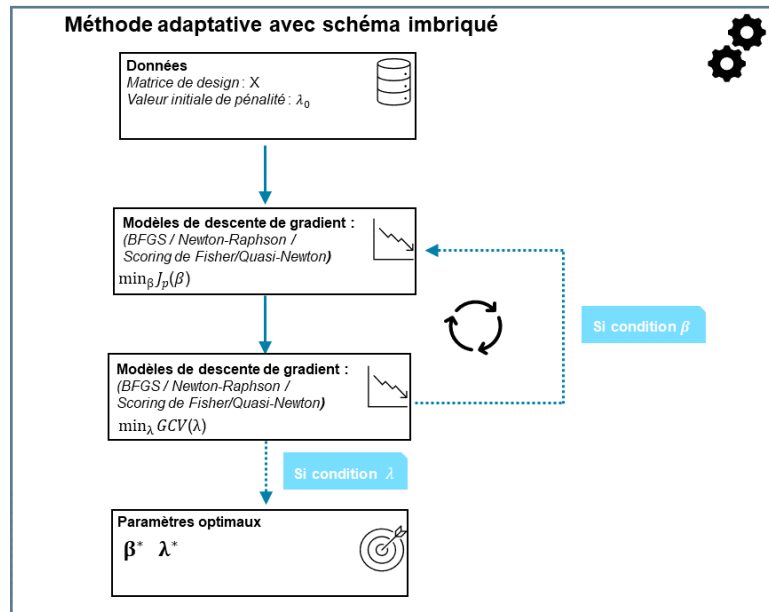


FIGURE 2.13 – Schéma explicatif de l'algorithme imbriqué



## 2.11 Critères d'adéquation et de choix de la méthode

### 2.11.1 La déviance D

Dans les modèles linéaires classiques nous considérons comme critère d'évaluation le somme résiduelle des carrés. Dans le cadre du GLM, nous considérons **la déviance**. Ce critère consiste à réaliser la distance/différence entre les vraisemblances de deux modèles. Un modèle est dit **saturé** s'il correspond à un modèle où il y a autant de paramètres que d'observations : soit un modèle où l'ajustement est parfait avec nos données. Un modèle peut également être dit estimé, c'est-à-dire en remplaçant nos données observées par nos estimations réalisées à l'aide des variables explicatives du modèle.

On définit le **modèle saturé** :

$$\ln(\mathcal{L}(y|y, \phi)) = \sum_i \left\{ \frac{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)}{\phi} + c(y_i, \phi) \right\} \quad (2.94)$$

Avec,  $\tilde{\theta}_i = g(y_i)$  et  $g(\cdot)$  une fonction qui dépend de la distribution choisie pour  $Y$ .

On définit le **modèle estimé** :

$$\ln(\mathcal{L}(\hat{\mu}|y, \phi)) = \sum_i \left\{ \frac{\hat{\mu}_i \hat{\theta}_i - b(\hat{\theta}_i)}{\phi} + c(\hat{\mu}_i, \phi) \right\} \quad (2.95)$$

Ce qui nous permet de définir **la déviance** :

$$\mathcal{D} = 2 \times (\ln(\mathcal{L}(y|y, \phi)) - \ln(\mathcal{L}(\hat{\mu}|y, \phi))) \quad (2.96)$$

Cette valeur est positive (le modèle saturé est censé ajuster parfaitement les données donc sa vraisemblance est toujours plus élevée que les autres modèles). De plus, plus cette valeur s'approche de 0, plus la vraisemblance du modèle choisi s'approche du modèle saturé, ce qui signifie qu'il est de bonne qualité.

Ainsi la déviance permet d'apprécier la qualité de l'ajustement de nos méthodes de lissages. On obtient alors, pour chacune des méthodes présentées avec le lien Poisson et les taux lissés les résultats suivants, en supposant que,  $D_{x,t} \sim \mathcal{P}(E_{x,t} \tilde{q}_x(t))$ . Dans le cas ou on souhaite estimer un modèle en 2 dimensions qui considère à la fois l'âge mais aussi la durée de la police d'assurance, on a :

- Si :  $D_{x,t} > 0$ , on a  $\text{Déviance}_{x,t} = 2 \left( D_{x,t} \ln \left( \frac{D_{x,t}}{E_{x,t} \tilde{q}_x(t)} \right) - (D_{x,t} - E_{x,t} \tilde{q}_x(t)) \right)$
- Si :  $D_{x,t} = 0$ , on a  $\text{Déviance}_{x,t} = 2E_{x,t} \times \tilde{q}_x(t)$

Si l'on considère un modèle avec 1 dimension on obtient la déviance suivante :

- Si :  $D_x > 0$ , on a  $\text{Déviance}_x = 2 \left( D_x \ln \left( \frac{D_x}{E_x \tilde{q}_x} \right) - (D_x - E_x \tilde{q}_x) \right)$
- Si :  $D_x = 0$ , on a  $\text{Déviance}_x = 2E_x \times \tilde{q}_x$

À partir des déviances individuelles, on déduit la déviance totale de notre méthode de lissage en sommant chacune des déviances, de telle manière que :  $\text{Déviance}_{\text{totale}} = \sum_{(x,t)} \text{Déviance}_{x,t}$

On peut alors, en repartant de cette déviance totale, réaliser un test d'ajustement : le test du rapport de vraisemblance. On cherche à estimer sous l'hypothèse  $H_0$  si nos nombres de décès lissés correspondent bien à une distribution de nombre de décès de Poisson. On définira alors la statistique de test suivante,

$$\forall(x, t) \in \llbracket x_{min}, x_{max} \rrbracket \times \llbracket t_{min}, t_{max} \rrbracket$$

$$\xi^{LR} = \sum_{(x,t)} \left( D_x \ln \left( \frac{D_x}{E_x \tilde{q}_x(t)} \right) - (D_x - E_x \tilde{q}_x(t)) \right) \quad (2.97)$$

### 2.11.2 Étude des résidus

Au niveau local, pour vérifier l'efficacité d'une méthode, on utilise l'interprétation graphique des résidus. Dans le cadre de notre travail de standardisation de méthodes, nous cherchons à automatiser l'analyse des résidus en proposant une option graphique alternative. Nous examinons les différents résidus en fonction de l'âge et de la durée des contrats étudiés :

$$\text{Résidus de la réponse : } r_{x,t}^{RESP} = \hat{q}_x(t) - \tilde{\mu}_{x,t}(t)$$

$$\text{Résidus de Pearson : } r_{x,t}^{PEARS} = (D_{x,t} - E_{x,t} \tilde{\mu}_{x,t}) / \sqrt{\mathbb{V}(E_{x,t} \tilde{\mu}_{x,t}(t))}$$

$$\text{Résidus de déviance : } r_{x,t}^{DEV} = \text{sign}(D_{x,t} - N_{x,t} \tilde{q}_x(t)) \times \sqrt{\text{Déviance}_{x,t}}$$

Nous sommes censés obtenir des résidus de Pearson sans tendance et hétérogène. En effet ils doivent aussi être concentrés entre l'intervalle -2 et 2, sinon cela signifie que nos taux lissés s'éloignent trop de nos taux bruts. De plus on est censés avoir des résidus de réponse avec des signes positifs et négatifs homogènes, de telle manière qu'il n'y a pas encore ici de tendance particulière. En effet, le lissage a pour but de supprimer le « bruit » lié à notre modélisation. Pour les résidus de déviance également, si nous avons une succession de signes identiques cela signifie que nos taux bruts sont sur-lissés.

### 2.11.3 La statistique du $R^2$

Le coefficient d'adéquation ou coefficient de détermination du  $R^2$  compris entre 0 et 1 mesure l'adéquation entre le modèle de lissage et les valeurs brutes. Il est défini mathématiquement comme la part de variance expliquée par la méthode de lissage sur la variance totale de nos taux bruts. Elle est surtout discriminante pour les modèles en 2 dimensions qui seront présentées Partie 3,

$$\forall(x, t) \in \llbracket x_{min}, x_{max} \rrbracket \times \llbracket t_{min}, t_{max} \rrbracket,$$

$$R^2 = 1 - \frac{\sum_{(x,t)} \left( \frac{D_{x,t}}{E_{x,t}} - \tilde{q}_x(t) \right)^2}{\sum_{(x,t)} \left( \frac{D_{x,t}}{E_{x,t}} - \frac{1}{N} \sum_{(x,t)} \frac{D_{x,t}}{E_{x,t}} \right)^2} \quad (2.98)$$

Ou en 1 dimension :

$$\forall x \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$R^2 = 1 - \frac{\sum_x \left( \frac{D_x}{E_x} - \tilde{q}_x \right)^2}{\sum_x \left( \frac{D_x}{E_x} - \frac{1}{N} \sum_x \frac{D_x}{E_x} \right)^2} \quad (2.99)$$

Avec, N le nombre d'observations associé à la dimension choisie.

### 2.11.4 La statistique de *MAPE*

Cette statistique (présentée par Felipe et al. [2002]) permet de mesurer l'exactitude de nos méthodes de lissage par rapport aux taux bruts. Elle correspond à la moyenne des écarts entre nos taux lissés et les taux bruts estimés, c'est une statistique qui s'exprime en pourcentage :

$$\forall x \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$MAPE = \frac{\sum_x \left| \frac{\frac{D_x}{E_x} - \tilde{q}_x}{\frac{D_x}{E_x}} \right|}{\sum_x D_x} \times 100 \quad (2.100)$$

Ou en 2 dimensions :

$$\forall (x, t) \in \llbracket x_{min}, x_{max} \rrbracket \times \llbracket t_{min}, t_{max} \rrbracket,$$

$$MAPE = \frac{\sum_{(x,t)} \left| \frac{\frac{D_{x,t}}{E_{x,t}} - \tilde{q}_x(t)}{\frac{D_{x,t}}{E_{x,t}}} \right|}{\sum_{(x,t)} D_{x,t}} \times 100 \quad (2.101)$$

### 2.11.5 Le test du *run* :

La déviance présente un problème lorsque l'on souhaite comparer des modèles paramétriques et non-paramétriques. En effet on ne peut pas calculer la déviance pour des modèles non paramétriques. Une possibilité pour comparer les méthodes à la fois paramétriques et non paramétriques est d'étudier différents tests qui permettent d'apprécier la qualité de lissage mais aussi la cohérence de la méthode. Enfin, l'intérêt opérationnel de ces tests est qu'ils peuvent être automatisés, ce qui présente un intérêt dans le cadre de notre étude visant à standardiser le choix des méthodes présentées.

Lors du test du *run* on évalue l'indépendance d'événements d'une séquence. Le test du *run* est un test statistique utilisé pour évaluer la randomisation dans une séquence de données binaires, c'est-à-dire des données qui ne peuvent prendre que deux valeurs possibles (par exemple, 0 et 1). Le test du *run* examine la longueur des séquences consécutives de 0 et de 1 dans les données. Si les données sont véritablement aléatoires, alors il devrait y avoir autant de séquences de 0 que de séquences de 1, et la longueur de ces séquences devrait être également répartie. Le test du *run* utilise une statistique de test appelée la statistique de Wald-Wolfowitz. Cette statistique mesure le nombre de fois où la séquence alterne entre 0 et 1 (c'est-à-dire le nombre de « *runs* » dans les données).

On compare ensuite cette statistique avec une distribution de probabilité connue pour déterminer si les données sont suffisamment aléatoires. Si la statistique de test est trop grande ou trop petite par rapport à cette distribution, on peut rejeter l'hypothèse nulle  $H_0$  selon laquelle les données sont aléatoires. Le test du *run* est souvent utilisé pour évaluer la qualité des générateurs de nombres pseudo-aléatoires, ou pour détecter des schémas dans des séries chronologiques. Dans notre d'étude, on représente une séquence de plusieurs changements de signe de la manière suivante :

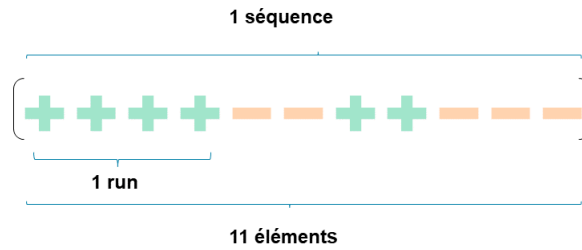


FIGURE 2.14 – Exemple d'une séquence

Dans le cadre de notre étude de résidus cela correspond au fait qu'il ne doit pas y avoir de relation ou de changement de signe qui suivent une logique particulière : ils doivent être aléatoires. On pose alors pour ce test, en notant  $n_+$ ,  $n_-$  et  $n = n_+ + n_-$  respectivement le nombre de signes positifs, de signes négatifs et total d'éléments de la séquence :

$$\mu = \frac{2n_+n_-}{n_+ + n_-} + 1 \quad \text{et} \quad \sigma^2 = \frac{2n_+n_-(2n_+n_- - (n_+ + n_-))}{(n_+ + n_-)^2(n_+ + n_- - 1)}$$

Cela nous permet d'en déduire la statistique pour le test du *run*, noté  $\xi^{RUN}$ .

$$\xi^{RUN} = \frac{\text{Nombre de runs} - \mu}{\sigma} \tag{2.102}$$

On a que si  $H_0$  est vraie alors la statistique suit une loi Normale centrée réduite de telle manière que :

$$\xi^{RUN} \sim N(0, 1) \tag{2.103}$$

Ainsi, l'hypothèse  $H_0$  ne sera rejetée que si :

$$|\xi^{RUN}| > N_{1-\alpha/2}(0, 1) \tag{2.104}$$

Avec,  $N_{1-\alpha/2}(0, 1)$  qui correspond au quantile  $(1 - \alpha/2)$  de la distribution Normale centrée réduite.

### 2.11.6 Le test des signes :

Le test dont il est question ici, tout comme le test du *run* (voir Section 2.11.5), se base sur l'observation de la fréquence de changements de signes entre les taux lissés et les taux bruts. Le

but de cette méthode est également de déterminer si notre modèle surestime ou sous-estime les taux bruts. Toutefois, contrairement au test du *run*, elle ne prend pas en compte l'indépendance entre les changements de signes. L'hypothèse  $H_0$  de ce test stipule que la médiane des signes doit être nulle, ce qui signifie qu'il doit y avoir autant de signes positifs que de signes négatifs. En effet, dans l'idéal, on s'attend à ce que le rapport entre *Actual/Expected* soit proche de 1, ce qui implique que la médiane des signes de la différence entre ces deux taux soit proche de 0. La statistique du test s'écrit de la manière suivante :

$$\xi^{SIG} = \frac{|n_+ - n_-| - 1}{\sqrt{n}} \quad (2.105)$$

De la même manière que précédemment on a que si  $H_0$  est vraie alors la statistique suit une loi Normale centrée réduite de telle manière que :

$$\xi^{SIG} \sim N(0, 1) \quad (2.106)$$

Par conséquent, l'hypothèse  $H_0$  ne sera rejetée que dans le cas où :

$$|\xi^{SIG}| > N_{1-\alpha/2}(0, 1) \quad (2.107)$$

### 2.11.7 Le ratio SMR

Le premier critère permettant d'évaluer l'efficacité d'une méthode de lissage consiste à vérifier si elle convient aux taux du portefeuille. Pour cela, en ce qui concerne le risque de mortalité, nous pouvons évaluer la qualité de l'estimation et de l'ajustement à nos données à l'aide de la métrique A/E ou *Actual/Expected*, qui représente le nombre de décès observé par rapport au nombre de décès modélisés ou lissés par nos méthodes. Ce ratio est appelé SMR (*Standardized Mortality Ratio*) :

$$\text{SMR} = \frac{\sum_x D_x}{\sum_x \mathbb{E}(D_x)}$$

Ou,

$$\text{SMR} = \frac{\sum_t \sum_x D_x}{\sum_t \sum_x \mathbb{E}(D_{x,t})}$$

Pour estimer la quantité  $\mathbb{E}(D_x)$ , nous supposons que nos taux de décès lissés suivent une loi de Poisson. En effet, en utilisant le lien entre la vraisemblance d'une loi de Poisson et celle d'une loi avec censure à droite et troncature à gauche, on retrouve une relation entre les deux vraisemblances. On obtient alors que la probabilité du nombre de décès est de la forme suivante :

$$\mathbb{P}(D_x = d_x) = \frac{(E_x \mu_x)^{d_x}}{d_x!} e^{-E_x \mu_x}$$

On obtient ainsi :

$$\mathbb{E}(D_x) = E_x \mu_x \quad \mathbb{V}(D_x) = E_x \mu_x$$

Cette approximation a le désavantage de prendre en considération l'**homoscédasticité** du nombre de décès à un âge  $x$ . Ceci permet d'obtenir le ratio SMR suivant en prenant seulement en considération l'âge sous certaines conditions d'approximations concernant la forme de la force de mortalité on obtient,

$$\text{SMR} = \frac{\sum_x D_x}{\sum_x E_x \tilde{\mu}_x} = \frac{\sum_x D_x}{\sum_x -E_x \ln(1 - \tilde{q}_x)} \simeq \frac{\sum_x D_x}{\sum_x E_x \tilde{q}_x} \quad (2.108)$$

En prenant en compte l'âge  $x$  et la durée  $t$  on obtient le SMR suivant :

$$\text{SMR} = \frac{\sum_x D_x}{\sum_x E_x \tilde{\mu}_x(t)} = \frac{\sum_{(x,t)} D_{x,t}}{\sum_{(x,t)} -E_{x,t} \ln(1 - \tilde{q}_x(t))} \simeq \frac{\sum_{(x,t)} D_{x,t}}{\sum_{(x,t)} E_{x,t} \tilde{q}_x(t)} \quad (2.109)$$

On interprète le SMR de la manière suivante :

- Si  $\text{SMR} > 1$  : Les décès ajustés sont **sous-estimés**
- Si  $\text{SMR} < 1$  : Les décès ajustés sont **sur-estimés**

Liddell [1984] propose alors de calculer une statistique pour savoir si le ratio SMR est significativement différent de 1. Cette méthode permet alors d'obtenir un intervalle de confiance à notre ratio SMR.

On procède ensuite à l'évaluation pour déterminer si notre ratio SMR diffère significativement de 1 ou non. L'hypothèse  $H_0$  émise stipule que le ratio SMR est significativement proche de 1. La statistique de test suivante est alors calculée de la manière suivante :

- Si  $\text{SMR} > 1$ ,  $\xi^{SMR} = 3 \times D^{\frac{1}{2}} (1 - (9D)^{-1} - (D/E)^{\frac{1}{3}})$
- Si  $\text{SMR} < 1$ ,  $\xi^{SMR} = 3 \times D^{*\frac{1}{2}} (1 - (9D^*)^{-1} - (D^*/E)^{\frac{1}{3}})$

Avec en fonction de si on étudie nos relations en 1 ou 2 dimensions,  $D = \sum_{(x,t)} D_{x,t}$  ou  $D = \sum_x D_x$ ,  $D^* = \sum_{(x,t)} D_{x,t} + 1$  ou  $D^* = \sum_x D_x + 1$  et  $E = \sum_{(x,t)} -E_{x,t} \times \ln(1 - \tilde{q}_x(t))$  ou  $E = \sum_x -E_x \times \ln(1 - \tilde{q}_x)$ .

Si le ratio SMR est significativement proche de 1 on a que la statistique de test suit une loi Normale centrée réduite,

$$\xi^{SMR} \sim \mathcal{N}(0, 1) \quad (2.110)$$

On testera donc que cette hypothèse n'est pas vérifiée à un niveau  $(1 - \alpha)$  si notre statistique dépasse  $\mathcal{N}_{(1-\alpha)}(0, 1)$  qui correspondra au quantile  $(1 - \alpha)$  d'une loi Normale centrée réduite.

$$\xi^{SMR} > \mathcal{N}_{(1-\alpha)}(0, 1) \quad (2.111)$$

On peut aussi obtenir la p-value associée,

$$\text{p-value} = 1 - F_{\mathcal{N}(0,1)}(\xi^{SMR})$$

L'intérêt de ce test est aussi d'obtenir un intervalle de confiance à partir du ratio comme présenté par Liddell [1984]. On obtient en effet les bornes supérieure et inférieure du ratio SMR définies de la manière suivante avec un intervalle de confiance à 0,5%,

- $SMR_{inf} = \frac{D}{E} \left(1 - \frac{1}{9D} - 1,96 \times (9D)^{-\frac{1}{2}}\right)^3$
- $SMR_{sup} = \frac{D^*}{E} \left(1 - \frac{1}{9D^*} + 1,96 \times (9D^*)^{-\frac{1}{2}}\right)^3$

### 2.11.8 Le critère d'adéquation du $\chi^2$

Le critère d'adéquation du  $\chi^2$  (présenté par Forfar et al. [1988]) est intéressant pour comparer nos estimations de mortalité entre nos taux lissés et la mortalité observée pour un âge x.

$$\chi_{test}^2 = \sum_x \frac{(D_x - E_x \tilde{q}_x)^2}{E_x (1 - \tilde{q}_x) \tilde{q}_x} \quad (2.112)$$

Nous obtenons le même critère en prenant en considération la duration en année t du contrat pour notre taux :

$$\chi_{test}^2 = \sum_{(x,t)} \frac{(D_{x,t} - E_{x,t} \tilde{q}_x(t))^2}{E_{x,t} (1 - \tilde{q}_x(t)) \tilde{q}_x(t)} \quad (2.113)$$

Selon le modèle, notre statistique de test est censée suivre une loi du  $\chi_{df}^2$ . C'est à dire une loi du  $\chi^2$  avec df degré de liberté. Cette valeur sera plus facile à calculer pour les modèles paramétriques (qui correspondra à  $n - r - 1$  avec  $n$  le nombre d'observations,  $r$  le nombre de paramètres) que ceux qui sont non-paramétriques. Plus la valeur du  $\chi^2$  est faible plus notre modèle est en adéquation avec les taux bruts. On préférera retenir un modèle avec un  $\chi^2$  qui est faible.

### Le critère de qualité de lissage

Pour observer la qualité du lissage c'est à dire à quel point notre modèle de lissage s'ajuste avec nos données, nous étudions la valeur de pénalisation pour chaque modèle avec un niveau de différenciation donné :

$$P = \sum_{x=x_{min}}^{x_{max}-z} (\Delta^z \tilde{q}_x)^2 \quad (2.114)$$

**Plus cette valeur est proche de 0 plus le critère de lissage est respecté.** On compare alors nos modèles en construisant cette valeur, en prenant en considération que pour certains modèles

nous minimisons d'ores et déjà cette valeur dans l'estimation de nos paramètres. Cette valeur correspond à la version discrète de la dérivée de nos taux lissés.

## 2.12 Analyse de sensibilité du portefeuille étudié

### Le niveau de mortalité du portefeuille

Pour comparer le niveau de mortalité de notre portefeuille, nous utilisons les tables de référence dans les marchés pour un type de produit et de risque donné. Cela signifie que nous étudions le niveau de mortalité de notre portefeuille en comparaison avec **une table de référence** du marché étudié. Cela nous permettra d'observer si le portefeuille étudié a une trajectoire de mortalité semblable ou différente des références.

Il convient toutefois d'être prudent lors de cette évaluation, car le portefeuille peut être complètement différent des références de mortalité. Cela peut s'expliquer par des différences de catégories socio-professionnelles (et donc des modes de vie différents). Afin de mesurer cet écart, nous utilisons un indicateur appelé  $LoM_x$ . Pour ce faire, nous divisons notre portefeuille en sous-groupes présentant des critères communs tels que la somme assurée, le type de produit, l'année de souscription, etc. Nous construisons ensuite :

Avec **sub** qui représente un sous jeu de données étudié dans le portefeuille :

$$LoM(sub) = \frac{D_x(sub)}{E_x(sub)q_{x,ref}} \quad (2.115)$$

On interprète ce ratio de la manière suivante :

- Si  $LoM > 1$  : Cela signifie que nous avons plus de décès observés dans notre sous jeu de données (sur-mortalité) que ce que les tables de référence pouvaient prédire.
- Si  $LoM = 1$  : Nous avons le même nombre de décès que ce que les tables de référence pouvaient prédire.
- Si  $LoM < 1$  : Nous avons moins de décès observés (sous-mortalité) dans notre sous jeu de données que ce que les tables de référence pouvaient prédire.

## 2.13 Obtention des intervalles de confiance des méthodes

Dans le cadre d'une méthode de lissage avec notre nombre de décès qui suivent une loi de Poisson, nous pouvons obtenir des intervalles de confiance entre les nombres de décès observés et lissés. Nous présenterons dans ce mémoire trois méthodes d'obtention d'intervalles de confiance. Une méthode par approximation de la loi de Poisson, une autre méthode à l'aide de la méthode Delta pour obtenir des intervalles de confiance pour les modèles généralisés lorsque le nombre d'individus étudiés est élevé, puis une méthode pour les modèles non-paramétriques.

### Intervalle de confiance ponctuel par approximation

On peut obtenir en approximant **la loi de Poisson** que suivent nos nombres de décès par une **loi Normale**, on obtient que :

$$D_x \sim \mathcal{N}(E_x \tilde{q}_x, E_x \tilde{q}_x (1 - \tilde{q}_x)) \quad (2.116)$$



On obtient alors les intervalles de confiance suivants :

$$IC_{(1-\alpha)}(D_x) = E_x \tilde{q}_x \pm z_{1-\frac{\alpha}{2}} \sqrt{E_x \tilde{q}(1-\tilde{q}_x)} \quad (2.117)$$

Avec,  $z_{(1-\frac{\alpha}{2})}$  est le quantile  $(1 - \frac{\alpha}{2})$  de la loi normale centrée réduite. Pour obtenir un résultat cohérent, on est supposés obtenir des nombres de décès bruts qui sont compris entre la borne inférieure et supérieure des nombres de décès lissés.

### Intervalle de confiance par la méthode delta

Dans les méthodes GLM local et GAM on suppose une distribution de Poisson sous-jacente puis on réitère le processus à partir de l'algorithme du scoring de Fisher. À partir de cette distribution, on peut ainsi obtenir un intervalle de confiance. Pour cela, on calcule  $Var(\eta)$ . On a, dans le cas général avec la matrice de coefficients  $H$  spécifique à chaque modèle,

$$\begin{aligned} Var(\eta) &= Var(X\hat{\beta}) \\ &= Var(H\hat{z}) \\ &= HVar(\hat{z})H^t \end{aligned}$$

On a dans les matrices  $H$  suivantes dans les modèles :

- GLM local :  $H = X(X^t W \Omega X)^{-1} X^t W \Omega$
- GAM :  $H = X(X^t \Omega X + P)^{-1} X^t \Omega$

Pour déterminer  $Var(\eta)$  on utilise l'itération de l'algorithme du scoring de *Fisher*, qui permet d'avoir la relation suivante pour un individu d'âge  $i$  :

$$\forall x_i \in \llbracket x_{min}, x_{max} \rrbracket,$$

$$Var(\eta_i) = Var\left(z_i + \frac{(y_i - \mu_i)}{\mu_i}\right) = \frac{Var(y_i)}{\mu_i^2}$$

Car la seule variable aléatoire dans l'équation 2.13 correspond à  $y_i$  qui caractérise le nombre de décès à l'âge  $i$  dans le cadre de notre étude de mortalité et suit une distribution de Poisson. Or, on a la valeur suivante pour **les poids réitérés** dans le cadre de la distribution de poisson qui constituent la matrice  $\Omega$  des modèles.

$$\gamma_{ii} = \frac{1}{b''(\theta_i)} \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \frac{\mu_i^2}{Var(\mu_i)}$$

Ainsi on obtient finalement que :

$$Var(\eta) = \Omega^{-1} H H^t \quad (2.118)$$

On a donc l'intervalle de confiance suivant pour  $\eta_i$  :

$$IC_{1-\alpha}(\hat{\eta}_i) = \hat{\eta}_i \pm z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\eta}_i)}$$

On obtient l'intervalle de confiance pour nos taux lissés grâce à la méthode delta et en inversant la fonction de lien  $g(\cdot)$  dans le cadre des modèles GLM local et GAM. En effet on a que pour un âge  $i$  :

$$\forall x_i \in [x_{min}, x_{max}], \quad \tilde{q}_i = g^{-1}(\hat{\eta}_i) \quad (2.119)$$

On obtient ainsi l'intervalle de confiance final pour nos estimations dans le cadre de ces modèles :

$$IC_{1-\alpha}(\tilde{q}_i) = \exp\left(\hat{\eta}_i \pm z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\eta}_i)}\right) \quad (2.120)$$

---

## Chapitre 3

# Application des méthodes à un portefeuille

Cette partie cherchera à appliquer de manière plus opérationnelle les méthodes présentées précédemment au travers d'un portefeuille d'une compagnie d'assurance. Nous réaliserons une étude complète d'expérience à partir des données individuelles des assurés pour confronter les différentes méthodes à la fois d'estimation de taux bruts, mais aussi de lissage. Nous réaliserons notre étude sur un portefeuille d'origine portugaise. Ce portefeuille est un portefeuille recueilli entre 2007 et 2010.

### 3.0.1 Présentation de la mortalité au Portugal

Selon l'institut national de statistique portugaise (*Instituto Nacional de Estatística*), en 2019 l'espérance de vie à la naissance, tant pour les hommes que pour les femmes, était estimée à 80,80 ans. À la naissance, les hommes et les femmes pouvaient respectivement espérer vivre jusqu'à 77,78 ans et 83,43 ans. Par rapport aux valeurs estimées pour 2015-2017, cela représente une augmentation de 0,04 an et 0,02 an, respectivement, et un ralentissement des gains de longévité (1,56 et 0,96 mois, respectivement, pour les hommes et les femmes en 2015-2017).

Au cours de la dernière décennie, l'espérance de vie à la naissance au Portugal a augmenté de 2,06 ans, soit une augmentation de 2,29 ans pour les hommes et de 1,62 an pour les femmes, par rapport aux valeurs de 2006-2008 (75,49 et 81,81 ans, respectivement). Cette augmentation de l'espérance de vie doit être considérée si on souhaite comparer nos résultats à un portefeuille actuel. Nous prendrons aussi cette variation en compte lors de l'estimation de notre traité de réassurance dans le cadre de l'application Section 3.4.2.

### 3.0.2 Présentation des produits

Le portefeuille étudié est constitué de différentes garanties de protection liées à des crédits. Nous allons donc étudier un portefeuille d'assurance emprunteur. Ce sont des assurances - dans le cadre du portefeuille étudié de crédits - liés à des prêts à la consommation, particuliers, hypothèques. Un prêt hypothécaire vous permet d'acheter un logement. Il est proposé par des banques, des sociétés de crédit immobilier ou d'autres prêteurs et est souvent garanti par votre propriété.

Un prêt hypothécaire est généralement assorti d'un taux d'intérêt plus faible et d'une période de remboursement plus longue que le crédit à la consommation. Toutefois, si vous ne respectez pas vos obligations de remboursement et que votre prêt hypothécaire a été garanti par votre propriété, les prêteurs peuvent saisir et revendre votre maison pour rembourser le prêt.

Les banques sont libres d'accepter ou de refuser votre demande de prêt hypothécaire. Avant de vous proposer un prêt hypothécaire, le prêteur doit évaluer votre solvabilité, c'est-à-dire déterminer si vous pouvez réellement vous le permettre. Ainsi, on peut s'attendre à un portefeuille d'assurance

---

---

qui sera peut-être en meilleure santé que la moyenne nationale grâce à cette sélection.

L'assurance-crédit hypothécaire entre en jeu si vous êtes confronté à des circonstances qui vous empêchent de rembourser votre dette, par exemple en cas de décès, de maladie ou de perte d'emploi. Ceci prend toute son importance lorsque l'on sait que pour la plupart des Portugais, l'assurance-vie est étroitement liée aux prêts immobiliers, mais aussi à d'autres formes de prêts ; une grande majorité des banques au Portugal exigeant une police d'assurance-vie comme garantie pour un prêt immobilier. En cas de décès ou d'invalidité de l'un des titulaires, les dettes sont remboursées sans hypothèque. Toutefois, l'utilisation de l'assurance-vie au Portugal peut également être liée à un besoin de protection et de prévention, à savoir assurer la sécurité des actifs financiers de la famille. Dans notre cas d'étude, nous nous concentrerons sur un portefeuille constitué de prêts hypothécaires, de consommations et de particuliers.

Au niveau légal, il y a aussi eu de nombreuses évolutions pour les institutions assurantielles au Portugal, le 22 octobre 2021, le gouvernement portugais a approuvé une loi (loi n° 75/2021 du 18 novembre), établissant le « droit à l'oubli ». La nouvelle loi a modifié la loi n° 46/2006 du 28 août, emportant la modification du régime juridique du contrat d'assurance qui a été approuvé par le décret-loi n° 72/2008 du 16 avril, afin d'interdire et de punir la discrimination fondée sur le handicap et/ou l'existence de risques de santé aggravés. Concrètement, cela signifie que le « droit à l'oubli », impose en partie aux assureurs de ne plus refuser de délivrer des polices d'assurance à des personnes ayant souffert de maladies susceptibles de modifier d'une manière ou d'une autre l'espérance de vie ou le taux d'incidences d'un individu, par exemple le VIH, le diabète, un cancer.

Plus exactement, aucune information de santé relative à un état pathologique ayant entraîné ou aggravé un risque de santé ou un handicap ne peut être collectée ou traitée par les établissements de crédit/assureurs dans un contexte pré contractuel après une période ininterrompue de :

**Dix ans** depuis la fin du protocole thérapeutique, en cas d'aggravation du risque de santé ou de dépassement du handicap.

**Cinq ans** depuis la fin du protocole thérapeutique, en cas de dépassement de la pathologie avant l'âge de 21 ans.

**Deux années** de protocole thérapeutique continu et efficace écoulées, en cas de risque aggravé pour la santé ou de handicap atténué.

Cette information peut être pertinente car elle peut avoir un impact indirect sur la mortalité globale de notre segment ou portefeuille. Toutefois, étant donné que notre portefeuille couvre la période de 2007 à 2010, ces antécédents ne sont pas pris en compte.

### 3.0.3 Présentation des données

#### Présentation des variables

Lors d'une étude d'analyse d'expérience, nous pouvons étudier de nombreuses variables. Cependant, dans cette étude, nous nous concentrerons seulement sur certaines d'entre elles, essentielles. En entrée de notre modèle, les données sont donc fournies de la même façon que la Figure 3.1.

On estime ensuite en sortie de la première phase de l'analyse d'expérience notre exposition en fonction de la méthode souhaitée, la méthode par exposition initiale ou centrale. Pour calculer cette valeur nous avons besoin d'une indicatrice concernant le statut de fin de l'intervalle d'étude, si c'est un rachat, une censure, expiration mais aussi le statut au début de l'intervalle d'étude. Cela nous permet alors de construire notre estimateur  $\hat{q}_x$  avec les méthodes présentées Partie 1. On récupère alors des données avec plusieurs variables qui nous indiquent le montant, la durée de la police, l'âge atteint de l'assuré... La granularité dépend du niveau de précision souhaité. Voici une liste non exhaustive des variables estimées et conservées dans le cadre de ce mémoire :

Police	Sexe	Date de naissance	Statut	Date de début de condition	Somme assurée (€)	Date de fin de condition	Date d'incident	Type de produit
1	F	15/04/1977	Décès	01/02/2010	100'000	10/03/2015	10/03/2015	Crédit conso
2	M	15/02/1961	Actif	10/03/2010	150'000			Crédit hypothécaire
3	F	15/12/1982	Actif	02/04/2010	50'000			Credit indiv
...	...	...	...	...	...	...	...	...

**FIGURE 3.1** – *Format des données en input de l'exemple d'analyse d'expérience*

**Actual Count** : Les valeurs actuelles de nombres de décès

**Exposure Count** : L'exposition initiale ou centrale de notre portefeuille à différentes granularités

**Expected Count** : Les valeurs projetés de nombres de décès

**Face Amount Upper Band** : Le montant assuré de la police assurée

**Policy Duration Year** : La durée de la police d'assurance en années

**Age Attained** : L'âge atteint de l'assuré

**Product Name** : Le type de produit de la police d'assurance

**Decrement** : Le type de risque estimé associé au taux étudié, dans notre cas  $\hat{q}_x$  pour le risque de mortalité

## Répartition des âges

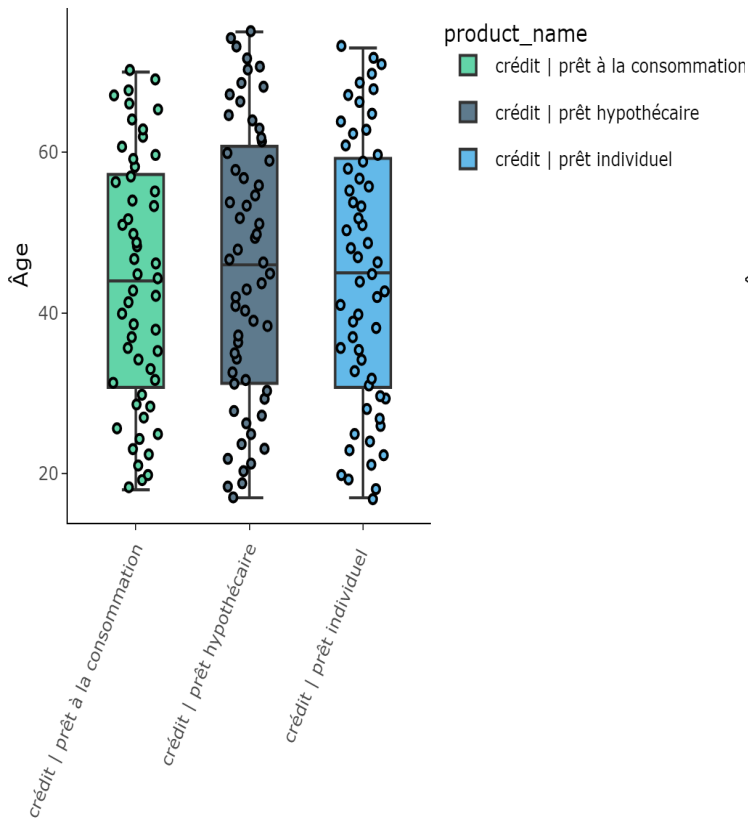
La répartition des âges est une caractéristique importante et à étudier dans le cadre d'une analyse d'expérience. En ce qui concerne les âges élevés, leur lissage n'est pas évident, car de manière générale, ils ont un faible nombre d'incidents, ils sont en dehors des limites du critère de Cochran. La définition de l'âge dans notre étude est aussi une caractéristique importante et prise en compte dans le cadre de l'application APEX. En effet l'application des taux de risque aux *model points* ou aux polices projetées doit être cohérente avec la dimension âge et la définition de l'âge de la table de taux sous-jacente. Cela devient particulièrement important si la dimension de l'âge ou la définition de l'âge diffère entre la table des primes et la table des taux de risque. Par exemple : les taux de prime peuvent être définis comme une prime annuelle basée sur l'âge au dernier anniversaire de la police alors que les taux de risque sont définis comme l'âge exact. Cela modifiera alors nos valeurs de *Actual/Expected* que l'on sur ou sous estimera, ce qui pourrait avoir un impact notable sur le prix de notre traité de réassurance.

L'âge correspond dans cette étude à la manière la plus courante dans les sociétés d'assurance et de réassurance de définir l'âge d'une personne, l'âge au dernier anniversaire de la police. Cependant, cela signifie que chaque  $q_x$  est censé s'appliquer à un ensemble de vies qui ont toutes la valeur «  $x$  » à leur dernier anniversaire et se situent donc en réalité entre  $x$  et  $x + 1$  exactement au début de l'année et entre  $x + 1$  et  $x + 2$  à la fin si elles survivent. Par exemple, pour une distribution uniforme des âges,  $q_{50}$  s'applique en moyenne à une personne qui a 50,5 ans au début de l'année considérée et 51,5 ans à la fin si elle survit. On notera la différence moyenne d'une demi-année par rapport à l'âge exact.

Concernant la distribution des âges par type de produits présentées à partir des Figures 3.3 et 3.2, on a des répartitions assez homogènes et centrées en les âges 43-45 ans environ. On a des âges légèrement plus élevés pour les produits de prêt hypothécaire mais dans l'ensemble ils sont répartis de la même manière. Ce qui est notable c'est qu'on a moins d'individus qui ont souscrit au produit lié aux prêts à la consommation.

Le produit qui a la population la plus âgée correspond au produit lié aux prêts hypothécaires, c'est plus particulièrement le cas pour les hommes.

Boxplot des âges par type de produits | femme



Boxplot des âges par type de produits | homme

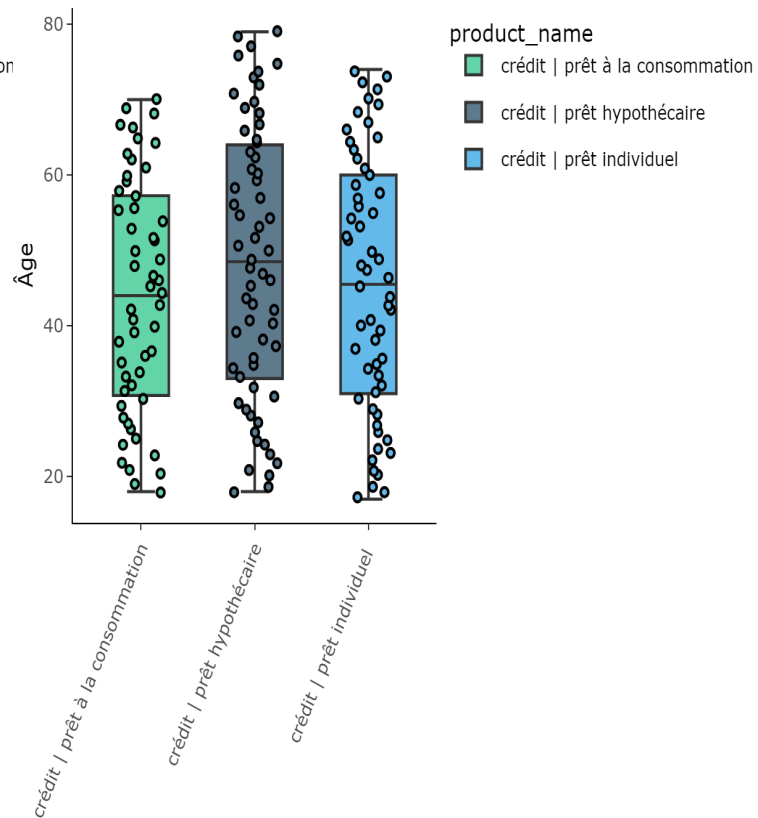


FIGURE 3.2 – Boxplot des âges pour les femmes

FIGURE 3.3 – Boxplot des âges pour les hommes

Concernant notre portefeuille étudié d'un point de vue plus général l'âge moyen est de 43,98 ans et la médiane est à 44 ans (voir Table 3.11).

La répartition des âges est plutôt hétérogène, les âges sont bien étendus on a un 1er quartile (Q1) à 34 ans et un 3e quartile (Q3) à 53 ans. Enfin, on a portefeuille avec des âges qui sont compris de 17 ans à 79 ans. On pourra finir par noter que les hommes ont des âges plus élevés que les femmes dans ce portefeuille.

## 3.1 Estimation des taux bruts

### 3.1.1 Estimation des taux bruts par exposition initiale

Dans le cadre de cette partie nous estimons les taux bruts de mortalité (par **exposition initiale et centrale**) associés à notre portefeuille. On étudiera les résultats par sexe mais aussi par type de produits.

#### Répartition homme-femme et type de produits

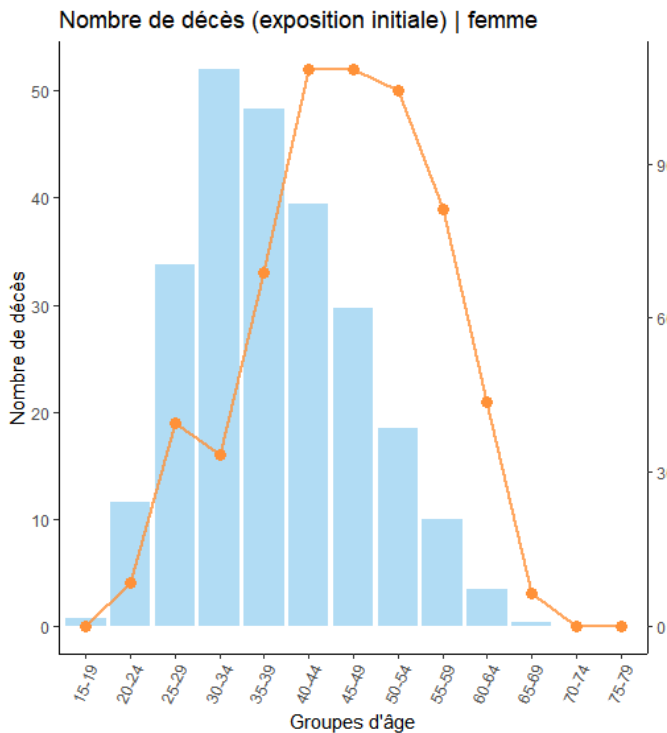
Lorsqu'on étudie la mortalité il est important d'étudier la répartition homme-femme de notre portefeuille, les niveaux de mortalité pouvant varier de manières significative. Cependant on notera que le reste de l'étude se fera en analysant notre table de mortalité construite de manière unisexe. Notre portefeuille est constitué de 45,1% de femmes et 54,9% d'hommes pour un total de 72 556 assurés. Ce portefeuille ne correspond pas au portefeuille total mais à un certain pourcentage d'un

portefeuille plus global. Concernant la répartition des décès, elle diffère entre les hommes et les femmes, mais aussi par type de produits souscrit par les assurés elle est répartie de la manière suivante Table 3.1 :

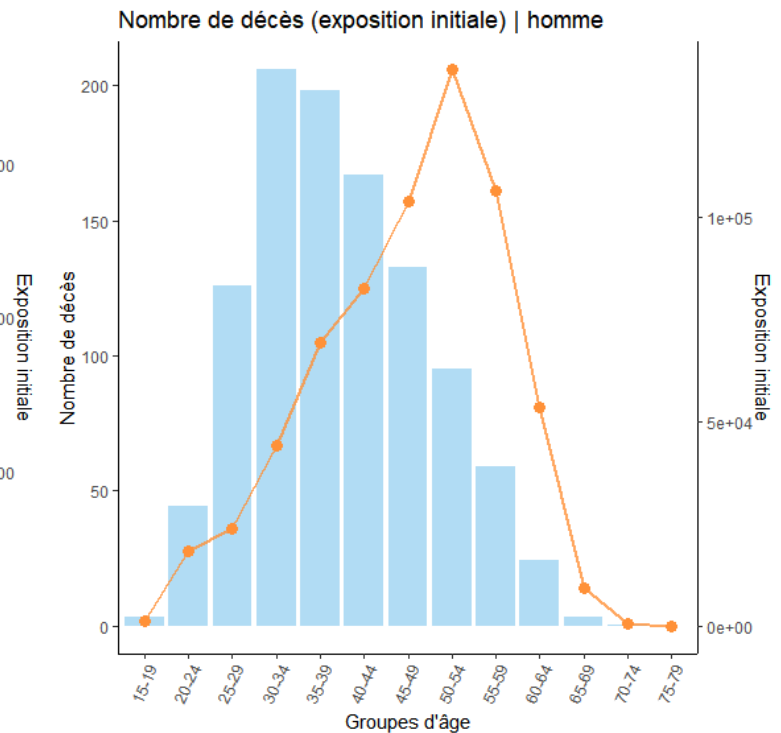
Sexe	$D_{total}$	Type de produit	$D_{total}$
Femme	289 (22,72%)	Prêt à la consommation	276 (21,69%)
Homme	983 (77,28%)	Prêt individuel	450 (35,37%)
Total	1272	Prêt hypothécaire	546 (42,92%)
		Total	1272

**TABLE 3.1** – Nombre de décès total  $D_{total}$  par hommes et femmes, mais aussi par type de produits

Les nombres de décès sont identiques pour les méthodes initiale et centrale (voir Section 1.16), ce qui est attendu car la différence réside dans la notion d'exposition plutôt que dans le nombre de décès. La différence entre les deux méthodes se manifeste lors de l'analyse des taux de mortalité, car la méthode centrale utilise une métrique plus réaliste, celle de la constance de la force de mortalité sur un intervalle donné. La Figure 3.1 montre que la mortalité des femmes est généralement plus faible que celle des hommes, accentuée par leur exposition déjà plus faible. Cette tendance est confirmée par une étude menée par Simoes et Correia [2019] sur une population portugaise en 2018, qui a estimé une espérance de vie de 83,4 ans pour les femmes contre 77,8 ans pour les hommes.



**FIGURE 3.4** – Nombre de décès  $D_x$  bruts estimés par exposition initiale pour les femmes



**FIGURE 3.5** – Nombre de décès  $D_x$  bruts estimés par exposition initiale pour les hommes

Les Figures 3.4 et 3.5 présentent le nombre de décès pour les hommes et les femmes calculé à partir de la méthode par **exposition initiale** présentée Section 1.2.2 (équation 1.16). La tendance générale des nombres de décès est plutôt logique : on a une augmentation au fil des âges puis un changement de tendance aux tranches d'âges de 50 – 54 et 55 – 59 ans. Ceci est en partie dû à une diminution de l'exposition à ces âges : pour les âges « élevés » l'exposition de notre portefeuille est plus faible. Il est également pertinent de noter que le nombre de décès est plus élevé pour les hommes que pour les femmes, jusqu'à 200 décès pour certaines tranches d'âges pour les hommes contre 50 pour les femmes. Ceci est particulièrement dû au fait qu'on se trouve face à une exposition plus importante pour les hommes.

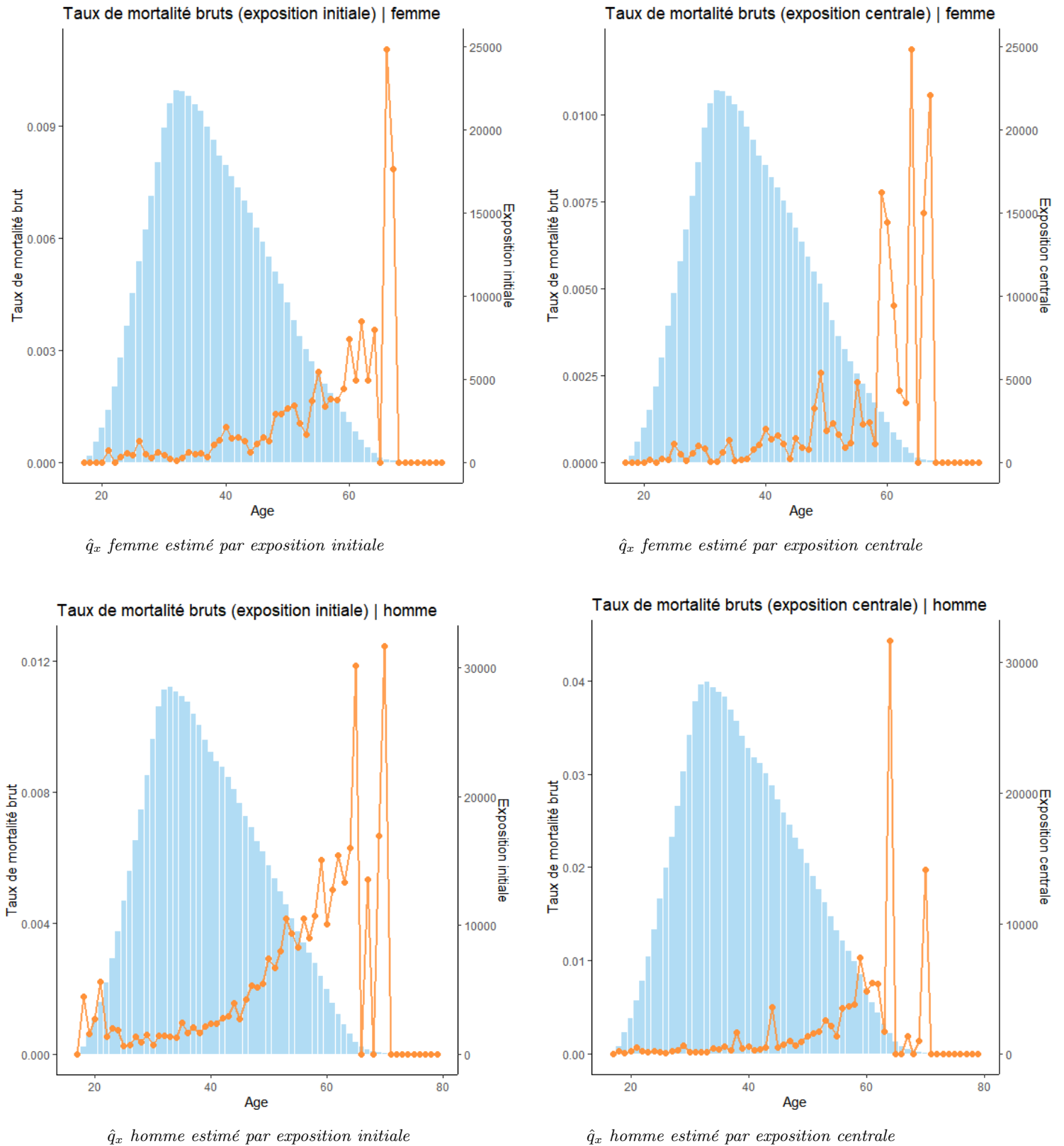


FIGURE 3.6 – Comparaison des taux bruts  $\hat{q}_x$  par type d'expositions (initiale et centrale)

La Figure 3.6 permet d'observer la différence entre les deux méthodes d'estimations des taux bruts présentées Partie 1 (méthode par exposition initiale et centrale). On peut observer une tendance générale croissante en 1 dimension quant à nos taux bruts estimés par la méthode par exposition initiale. On a une variabilité plus importante et des valeurs plus importantes aux âges faibles et élevés pour les femmes. De plus, la figure permet d'attester de la nécessité de réaliser un lissage



car nos taux bruts sont très variables d'un âge à l'autre. Le critère de Cochran (Section 2.2) est nécessaire pour ce type de portefeuille en raison du faible nombre de sinistres à certains âges (ici plus de 66 – 67 ans). On l'appliquera alors pour la suite de l'étude pour obtenir des résultats consistants.

### 3.1.2 Analyse du niveau de mortalité général de notre portefeuille étudié (LoM)

Le niveau de mortalité (*Level of Mortality*) présenté Section 3.8 permet de comparer la mortalité de notre portefeuille à celle au niveau national. C'est un indicateur important que l'on doit considérer dans le cadre de la tarification d'un traité de réassurance. On s'attend généralement à une mortalité 20% plus faible pour notre portefeuille que pour étude au niveau national. Ceci est dû au fait qu'il y a une sélection médicale mise en place avant la souscription de l'assurance emprunteur ; mais aussi au fait que la population constituant notre portefeuille correspond à une population avec un niveau de vie plus élevé. Concernant la table de référence nationale nous avons utilisée dans le cadre de ce mémoire elle est fournie par *L'Instituto Nacional de Estatística (INE)*, organisme officiel portugais chargé de la collecte, de la production et de la publication des statistiques. C'est un organisme comparable à l'Insee en France.

Cependant il faut noter que généralement les tables publiées par l'INE pour la population générale ne sont pas utilisées pour la population assurée. Les tables de mortalité de la population portugaise sont également publiées par *l'Instituto Nacional de Estatística/Statistics Portugal (INE)* et constituent la source d'information de base et l'unique source officielle concernant les taux de mortalité dans le pays. D'autres tables sont choisies pour inclure les facteurs d'ajustement de la population et la structure des sexes, dans le but d'une tarification égale pour les deux sexes, comme l'exige la loi. De plus il n'existe pas de table de mortalité projetée officielle au Portugal. Bien que l'INE mette régulièrement à jour les tables de mortalité pour la population globale - principalement parce que le montant des pensions versées par le système de sécurité sociale de l'État dépend de ces facteurs - les prestataires de pensions privées utilisent également leur propre expérience en matière de mortalité pour fixer l'hypothèse la plus adéquate sur la mortalité et choisir la table de mortalité appropriée. En effet, les tables publiées au niveau national ne sont souvent pas représentatives du profil de risque du portefeuille assuré. On modifie les taux de référence pour qu'ils aient la même proportion d'homme et de femme que ceux de notre portefeuille c'est à dire que,

$$q_x^{ref} = 0,459 \times q_{x,femme}^{ref} + 0,541 \times q_{x,homme}^{ref} \tag{3.1}$$

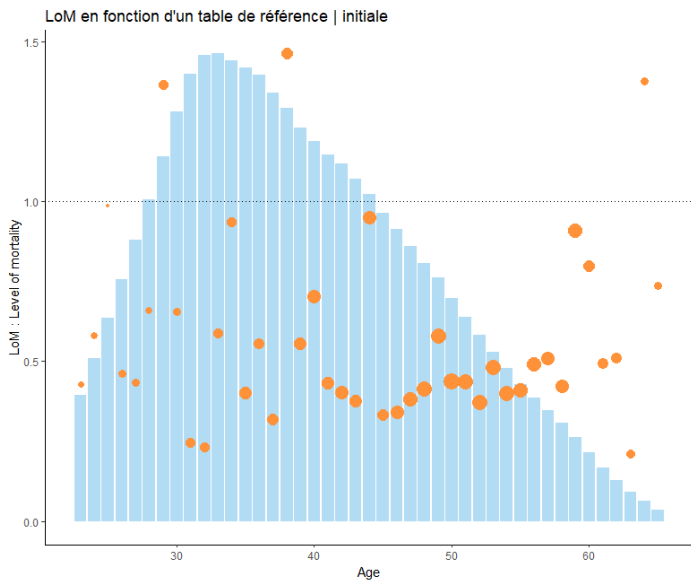


FIGURE 3.7 – LoM par rapport à la table de référence avec les  $\hat{q}_x$  bruts estimés par exposition initiale

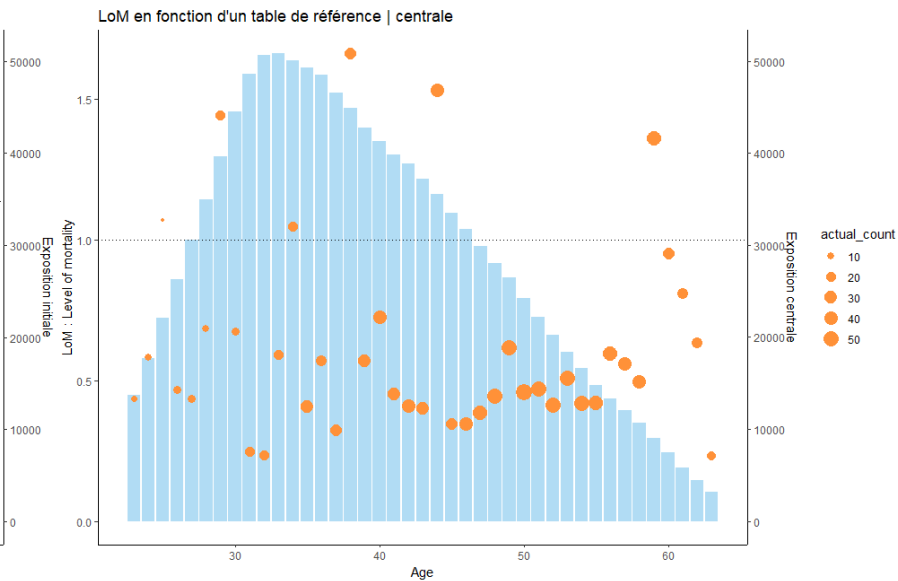


FIGURE 3.8 – LoM par rapport à la table de référence avec les  $\hat{q}_x$  bruts estimés par exposition centrale

## 3.2 Modélisation des taux lissés

Dans cette partie nous nous attacherons à modéliser et étudier les résultats des méthodes de lissage proposées le long du mémoire. Toutes ces méthodes sont implémentées de manière adaptatives pour répondre à la problématique du mémoire qui était de proposer des méthodes semi-adaptatives dans le cadre du lissage. On commentera aussi la pertinence des résultats obtenus en 1 dimension puis en 2 dimensions à partir de la liste de critères mentionnés Section 2.11.

### 3.2.1 Analyse de la dispersion dans le nombre de décès

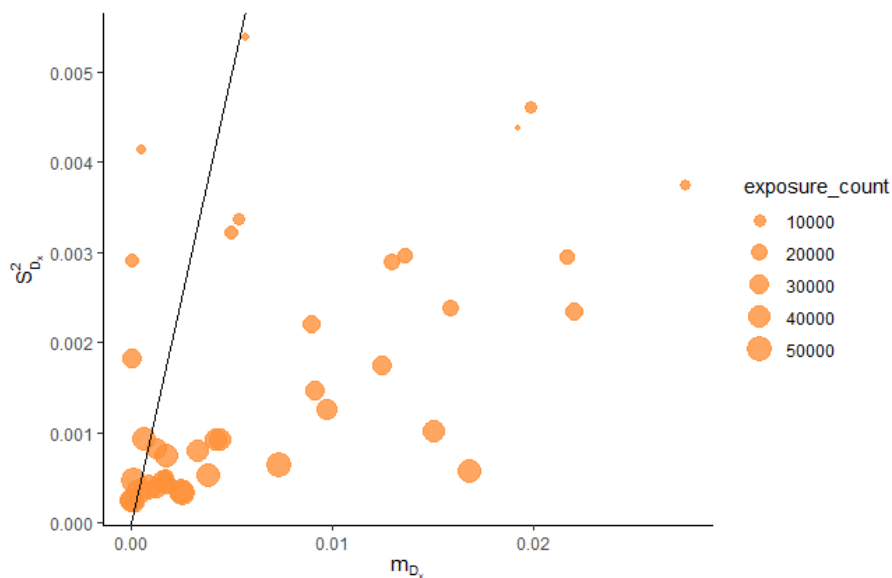


FIGURE 3.9 – Analyse de la dispersion des décès  $D_x$  dans le portefeuille étudié

La Figure 3.9 représente le test présenté dans la Section 2.7.3 pour évaluer la dispersion des décès. Nous traçons la droite  $(S_{D_x}^2, m_{D_x})$  et nous observons si les points de données (où chaque point correspond à un âge) sont situés en dessous ou au-dessus de la droite de régression. Il est intéressant de noter que les points sont situés en dessous de cette droite, ce qui indique la présence d'une **sur-dispersion** dans nos nombres de décès. Ainsi, pour mieux modéliser le nombre de décès, nous nous tournerons vers une loi de quasi-Poisson, en se référant à la Figure 2.5, pour sa simplicité d'implémentation par rapport à une distribution binomiale négative.

### 3.2.2 Analyse graphique des taux lissés avec les méthodes de lissage en 1 dimension

L'ensemble des méthodes présentées ci-dessous ont été soumises au préalable au critère de Cochran (Section 2.2) pour conserver seulement des données avec une crédibilité satisfaisante (c'est à dire plus de 5 décès par âges) pour appliquer les méthodes de lissage présentées précédemment. On retiendra alors, pour appliquer nos méthodes, les âges qui constituent l'intervalle [23 – 65] ans, les données retenues avec leur exposition et leurs taux bruts estimés à partir de la méthode par exposition initiale. Il faudra donc noter qu'on appliquera les méthodes de lissage associées avec **une exposition initiale**. Ce choix présente un certain biais quant au choix de la distribution de Poisson pour compter nos nombres de décès. Pour les âges de [17 – 22] ans et [66 – 79] ans on utilisera des méthodes d'interpolation associées aux âges élevés.

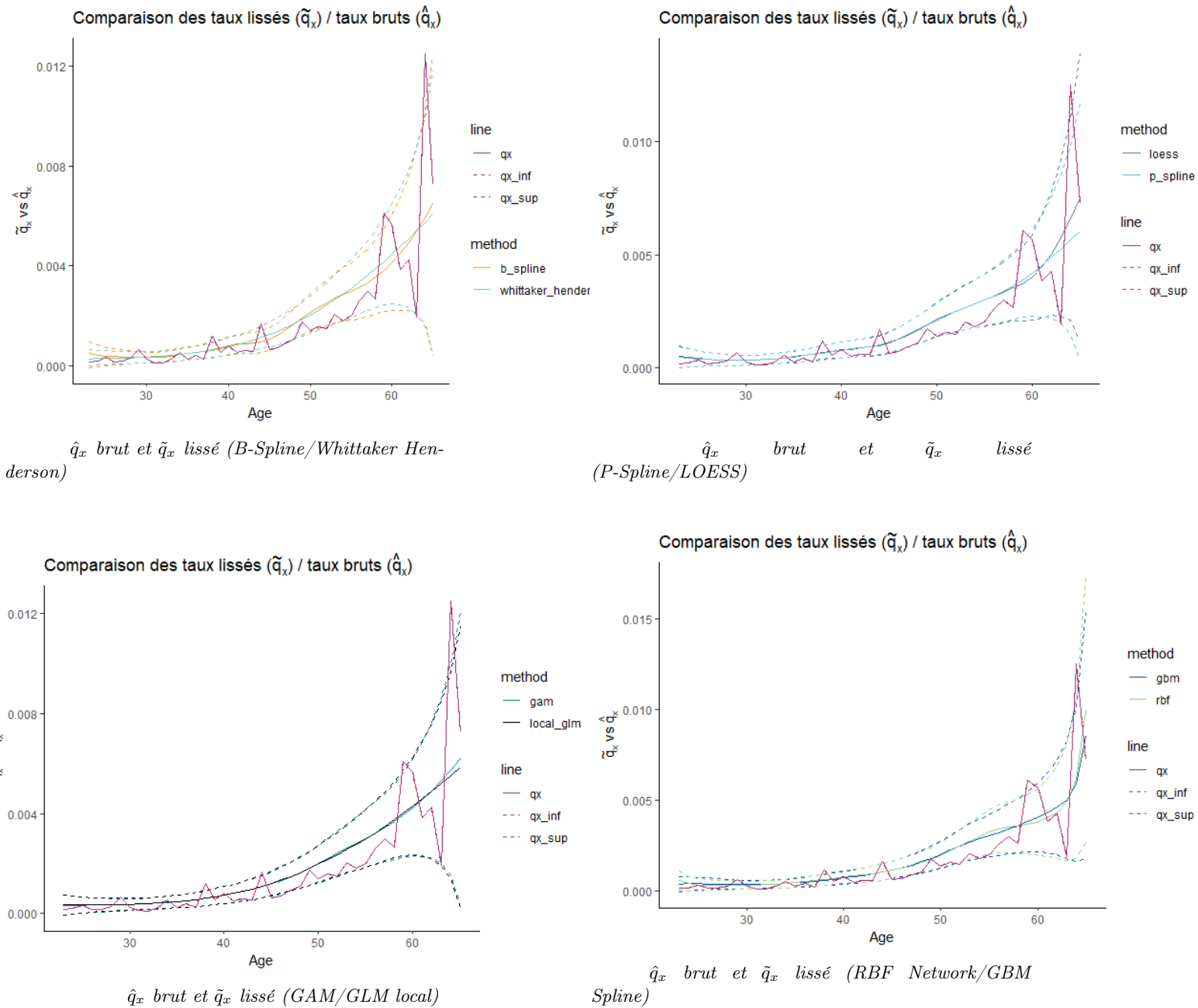
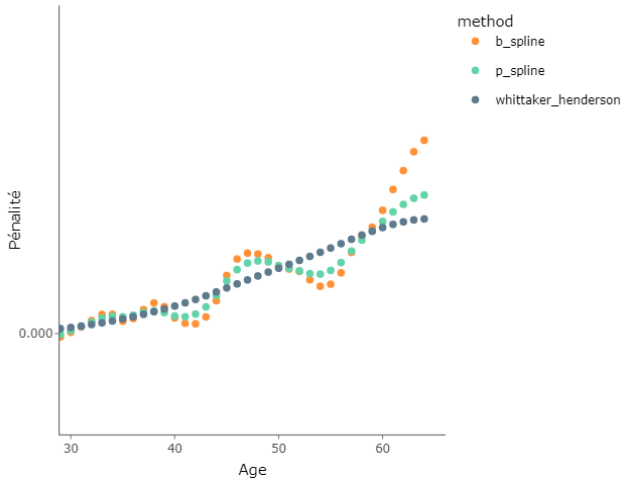


FIGURE 3.10 – Comparaison des taux lissés et des taux bruts avec des intervalles de confiance

La Figure 3.10 correspond à la comparaison des différentes méthodes de lissage en 1 dimension présentées Partie 2 avec les intervalles de confiance associés. Chaque méthode dans le cas à 1 dimension a des résultats qui sont plus ou moins cohérents. En effet la courbe est lisse et on a réduit la variabilité des estimations. Certains modèles sont moins adaptés que d'autres pour les âges élevés comme par exemple le modèle RBF Network. De plus toutes les courbes paraissent croissantes ce qui va dans la logique de l'augmentation de la mortalité au fil des âges. Il faudra étudier les critères mentionnés Section 2.11 pour tirer des conclusions quant à la meilleure méthode adaptative et semi-automatique.

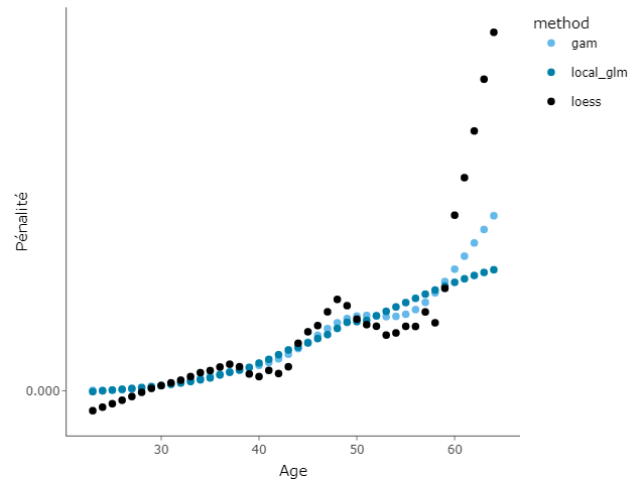
### 3.2.3 Analyse du critère de pénalité des méthodes de lissage en 1 dimension

Graphique représentant la pénalité en fonction de l'âge



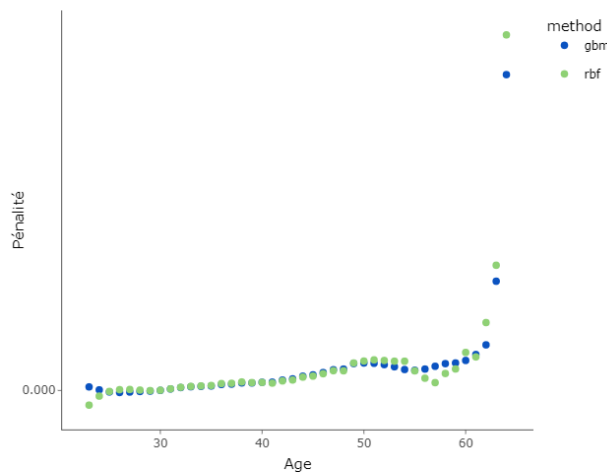
**FIGURE 3.11** — Critère de pénalité pour les modèles B-Spline/P-Spline/Whittaker-Henderson

Graphique représentant la pénalité en fonction de l'âge



**FIGURE 3.12** — Critère de pénalité pour les modèles GLM local/GAM/LOESS

Graphique représentant la pénalité en fonction de l'âge



**FIGURE 3.13** — Critère de pénalité pour les modèles RBF Network/GBM Spline

Les Figures 3.11, 3.12 et 3.13 représentent les critères de pénalités des différents modèles en 1 dimension présentés lors de la Section 2.11.8. Le critère de pénalité permet d'apprécier la régularité de nos taux lissés. Il correspond donc à la version discrète de la dérivée première. On s'attendrait à avoir un graphique de pénalité croissant car notre fonction est censée être croissante au fil des âges et augmenter légèrement aux âges élevés. On pourrait cependant avoir des pénalités décroissantes aux bords, on pensera par exemple pour les âges faibles à la diminution de la mortalité à la suite du phénomène de mortalité infantile. À travers nos résultats, on peut observer que les modèles B-Spline, LOESS et RBF Network présentent des taux qui ont des écarts de pentes importants. Ils ont des pentes qui varient en fonction des âges, ce qui est caractéristique d'un mauvais lissage des données. On notera de plus que les modèles B-Spline et LOESS ont des pentes négatives aux bas âges ce qui signifie que la mortalité est décroissante à ces âges. Le modèle GBM Spline a une pente qui n'augmente pas au fil des âges. Ceci est en partie dû à la contrainte de pénalisation *mono* qu'on lui a imposée dans le modèle de boosting. Les modèles avec les résultats les plus appréciables sont les modèles GAM et GLM local.

### 3.2.4 Analyse des résidus de Pearson pour les méthodes de lissage en 1 dimension

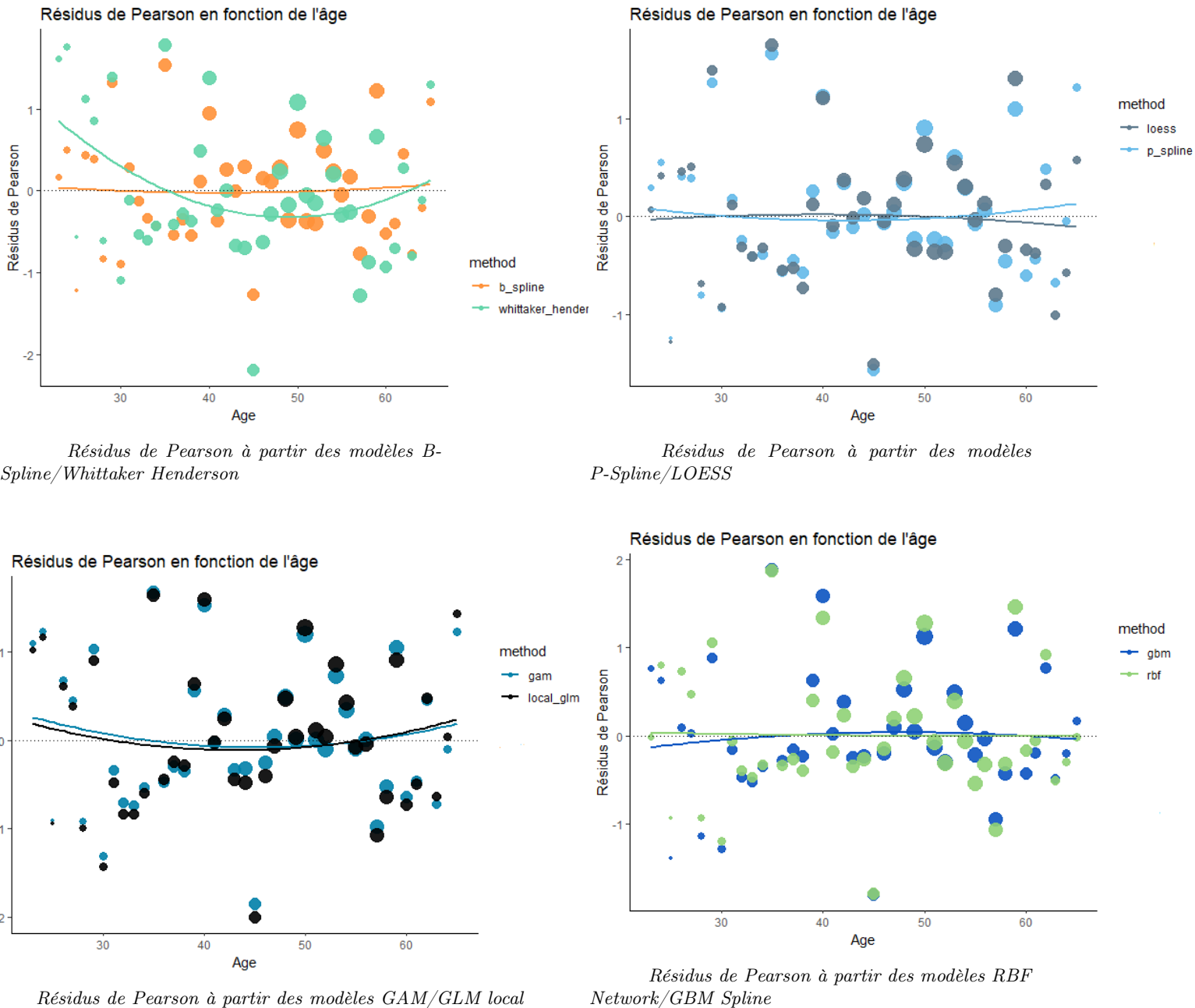
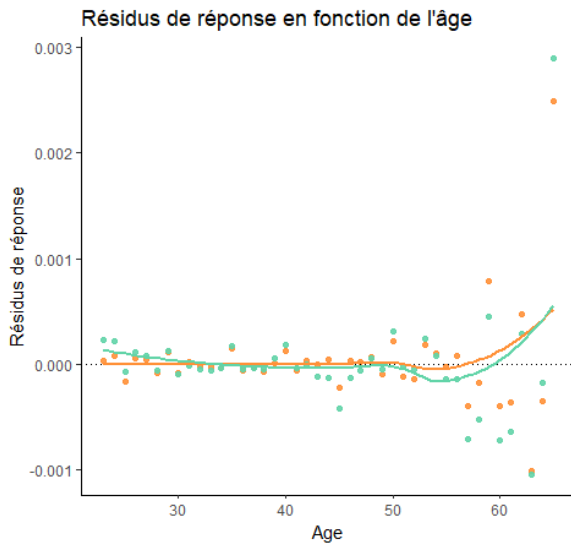


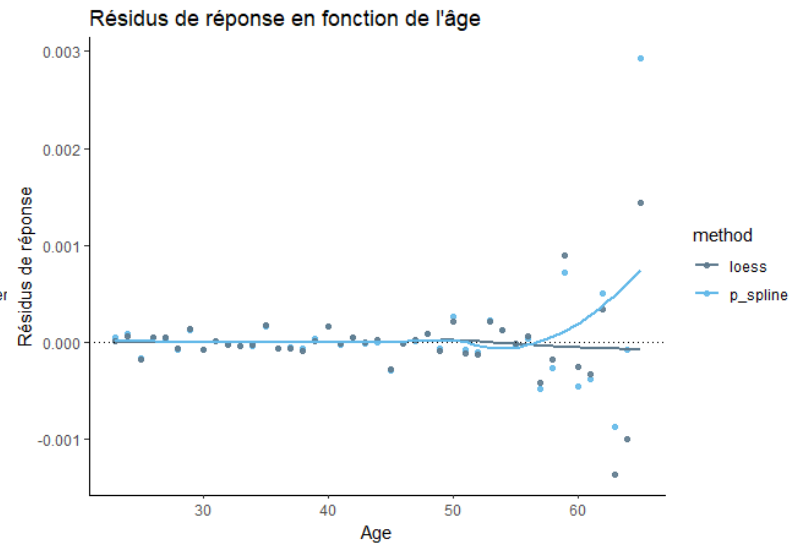
FIGURE 3.14 – Comparaison des résidus Pearson pour les différentes méthodes de lissage

La Figure 3.14 présente les résidus de Pearson, lesquels permettent de vérifier que nous n'avons pas de motifs particuliers, ce qui indique que les différents lissages sont globalement bien réalisés par les différentes méthodes. En effet, nous devrions obtenir des résidus distribués aléatoirement et compris principalement entre 1 et -1, étant donné que nous avons un portefeuille avec peu de décès. Nous constatons cependant que les modèles RBF Network et GBM Spline présentent des résidus élevés pouvant aller jusqu'à 2. Malgré cela, la plupart des méthodes présentent des résidus cohérents. En somme, nos modèles semi-automatiques sont capables de capturer relativement bien la variabilité des taux bruts.

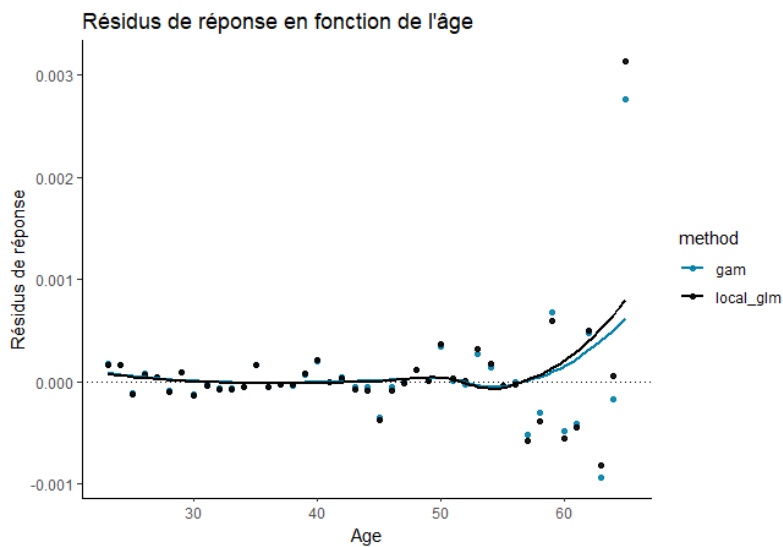
### 3.2.5 Analyse des résidus de la réponse pour les méthodes de lissage en 1 dimension



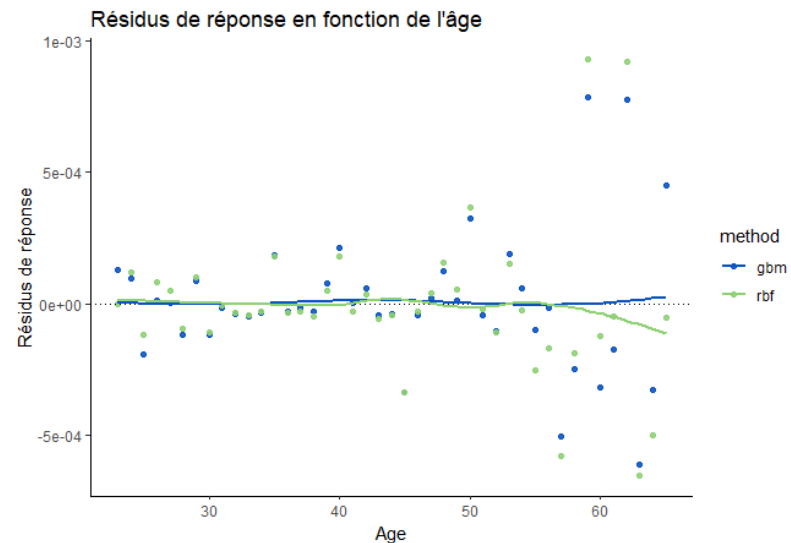
Résidus de la réponse à partir des modèles B-Spline/Whittaker Henderson



Résidus de la réponse à partir des modèles P-Spline/LOESS



Résidus de la réponse à partir des modèles GAM/GLM local

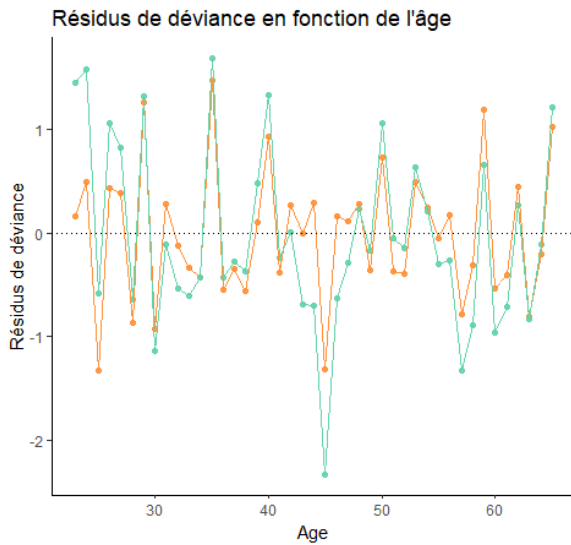


Résidus de la réponse à partir des modèles RBF Network/GBM Spline

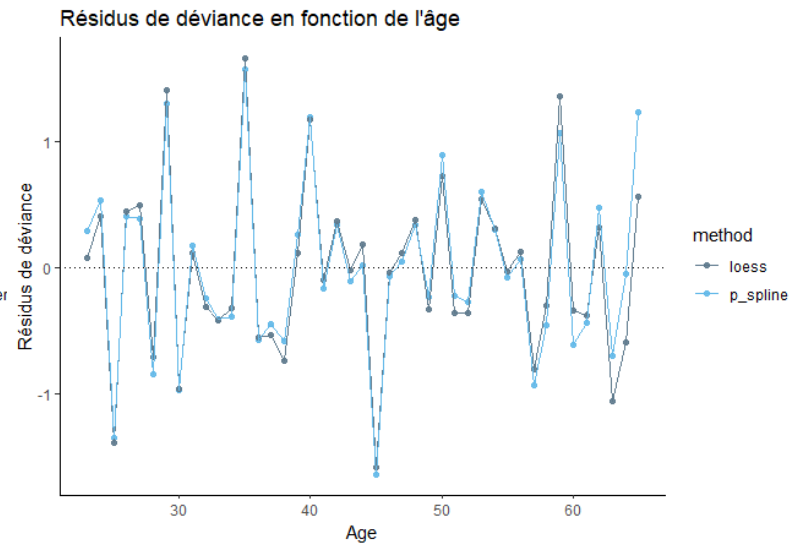
**FIGURE 3.15** – Comparaison des résidus de la réponse pour les différentes méthodes de lissage

La Figure 3.15 révèle que les résidus de la réponse présentent une tendance à la hausse pour les âges compris entre 55 et 63 ans avec les méthodes de lissage B-Spline, Whittaker-Henderson, P-Spline, GAM et GLM local. Cette observation suggère que ces modèles ont réalisé un ajustement inapproprié pour ces valeurs d'âge. En outre, le modèle RBF Network présente une mauvaise estimation pour les âges élevés, avec une tendance à sur-apprendre grâce aux biais, dépassant même les taux bruts initiaux, comme en témoignent les résidus de réponse négatifs. En revanche, les modèles GBM Spline et LOESS ont des résultats satisfaisants, sans sur ou sous ajustement des taux bruts.

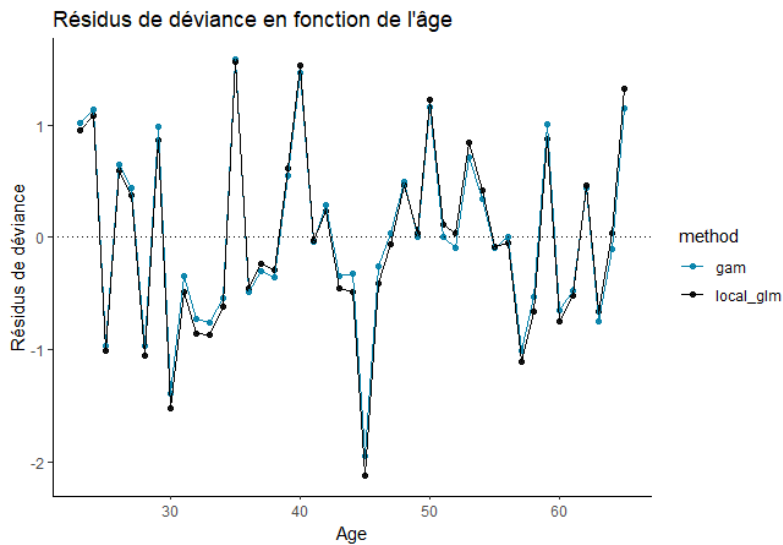
### 3.2.6 Analyse des résidus de la déviance pour les méthodes de lissage en 1 dimension



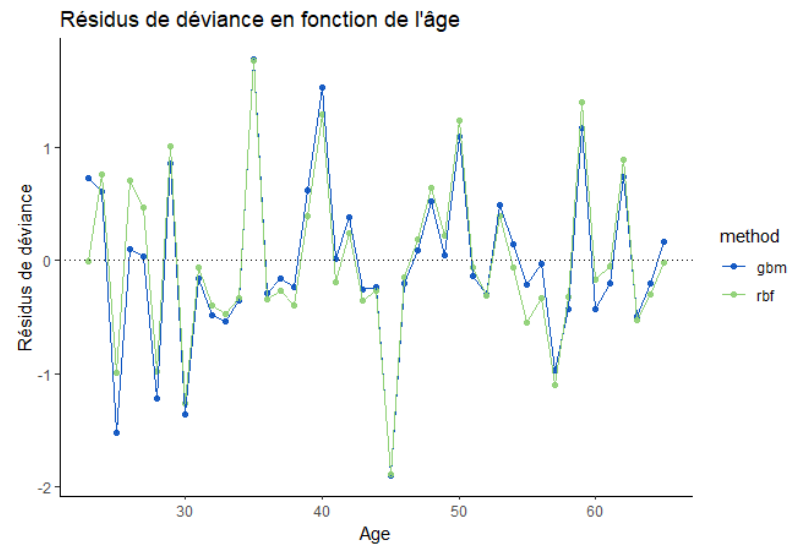
Résidus de la déviances à partir des modèles B-Spline/Whittaker Henderson



Résidus de la déviances à partir des modèles P-Spline/LOESS



Résidus de la déviances à partir des modèles GAM/GLM local

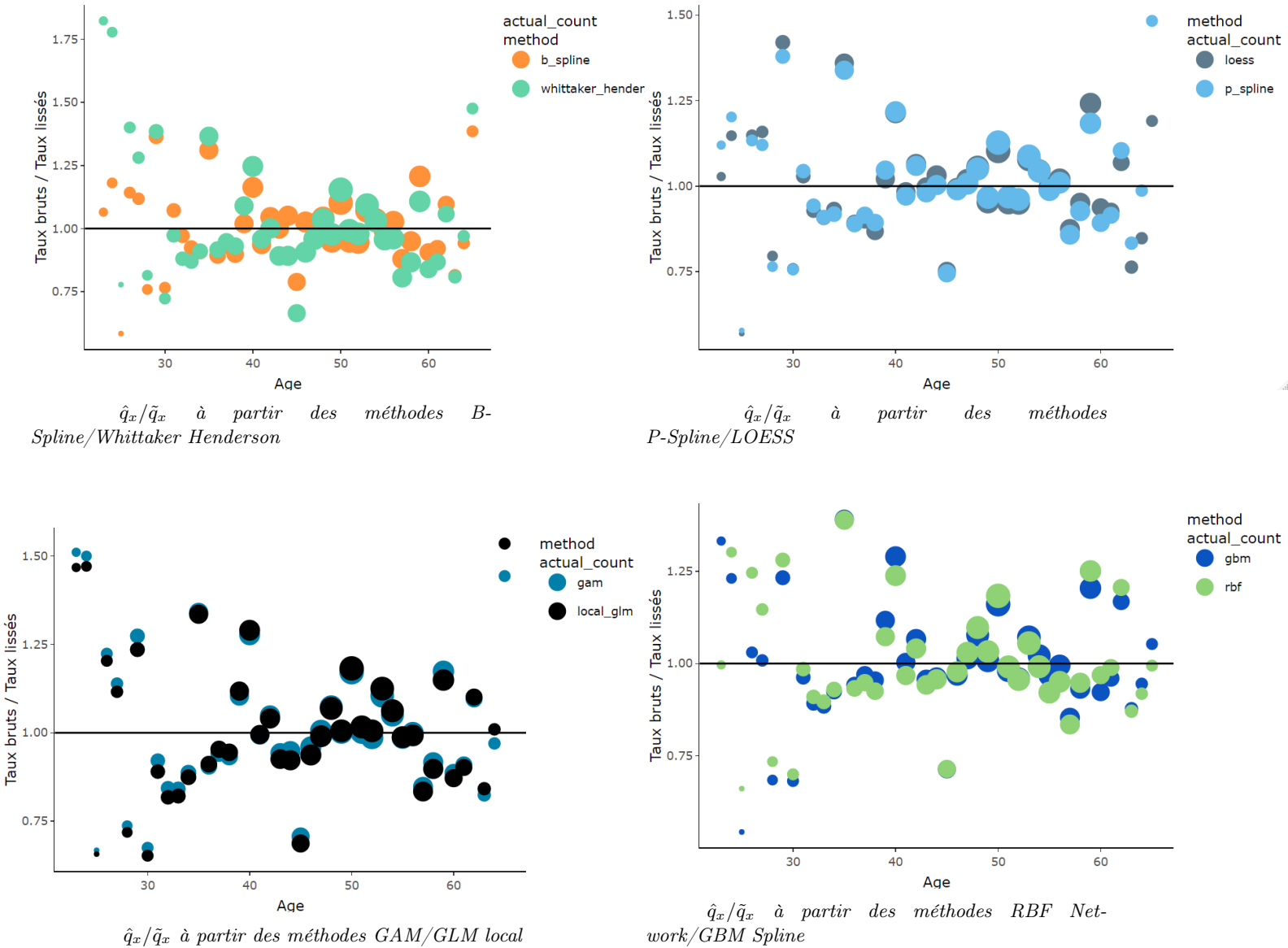


Résidus de la déviances à partir des modèles RBF Network/GBM Spline

**FIGURE 3.16** – Analyse des résidus de la déviance pour les différentes méthodes

Les résidus de la déviance, présentés dans la Figure 3.16, indiquent de bons résultats. On observe une alternance de signe régulière et aucun motif particulier ne se dégage pour une tranche d'âge donnée. La déviance est calculée en supposant que les nombres de décès suivent une distribution de Poisson, il s'agit donc d'être vigilant, car toutes les méthodes ne présupposent pas une distribution de Poisson. Les tests de *run* et de signe pourront être utilisés pour étudier plus précisément ces alternances de signe et leur fréquence. Le lecteur intéressé pourra aussi observer les valeurs de déviance fournies en annexe avec la Figure 3.42.

### 3.2.7 Analyse du ratio taux bruts et taux lissés (pondéré par l'exposition) par âge des méthodes de lissage



**FIGURE 3.17** – Comparaison des taux bruts sur les taux lissés pondérés par le nombre de décès associés

La Figure 3.17 permet d'observer si notre lissage sur-estime ou sous-estime notre mortalité brute et pour quels âges. Ce critère doit être mis en lien avec l'ensemble des méthodes de lissage a tendance à sous-estimer la mortalité pour les âges les plus jeunes entre 20 à 23-24 ans environ. Pour les âges entre 30 et 60 les méthodes B-Spline, Whittaker-Henderson, LOESS et P-Spline sont proches de 1, c'est-à-dire que nos taux lissés sont proches de nos taux bruts. Tandis que les méthodes GAM, GLM local, GBM Spline et RBF Network sont plus éloignées de 1 sur cette même tranche d'âge.



### 3.2.8 Analyse des ratios SMR et des valeurs du $\chi^2$ des méthodes de lissage

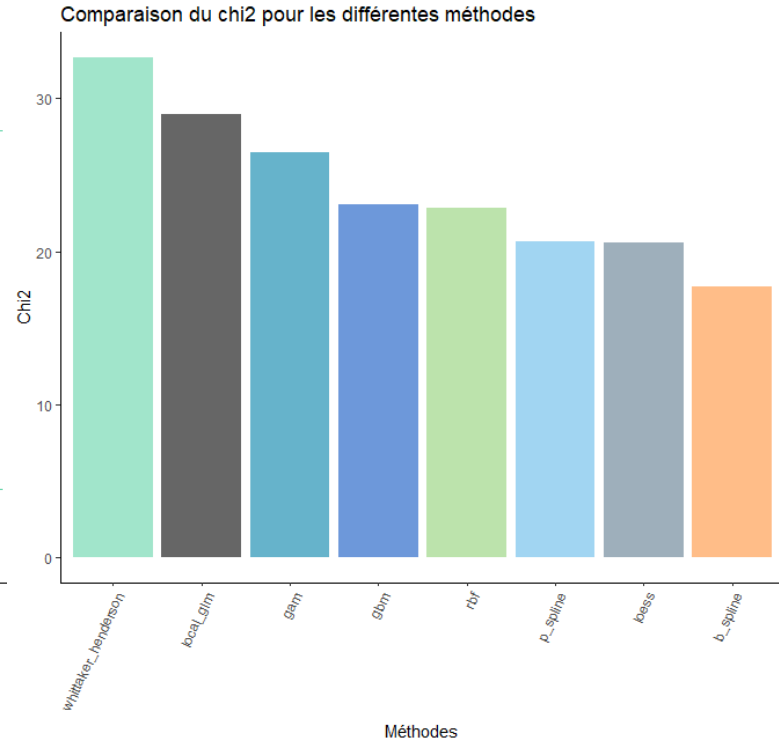
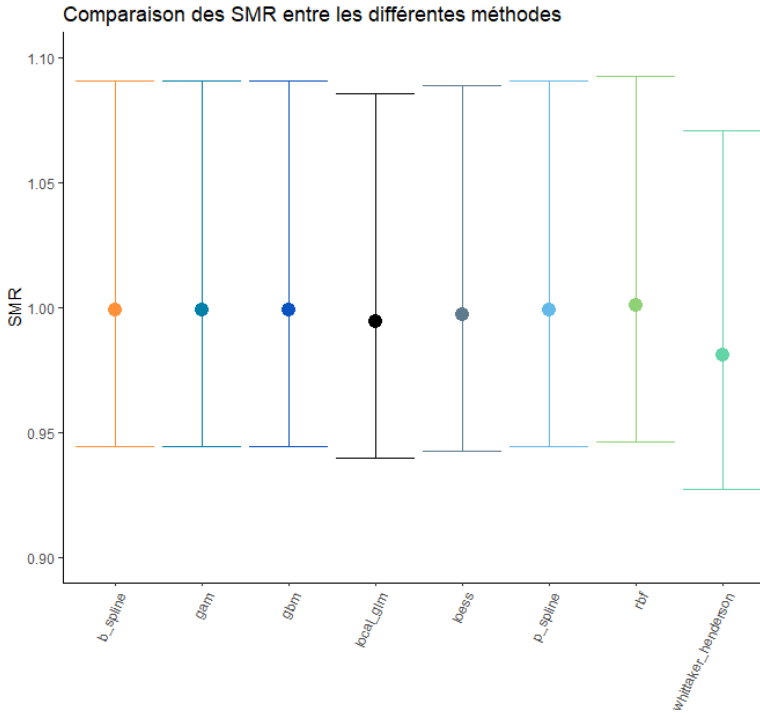


FIGURE 3.18 – Comparaison des ratios SMR pour les différentes méthodes

FIGURE 3.19 – Comparaison des valeurs du  $\chi^2$  pour les différentes méthodes

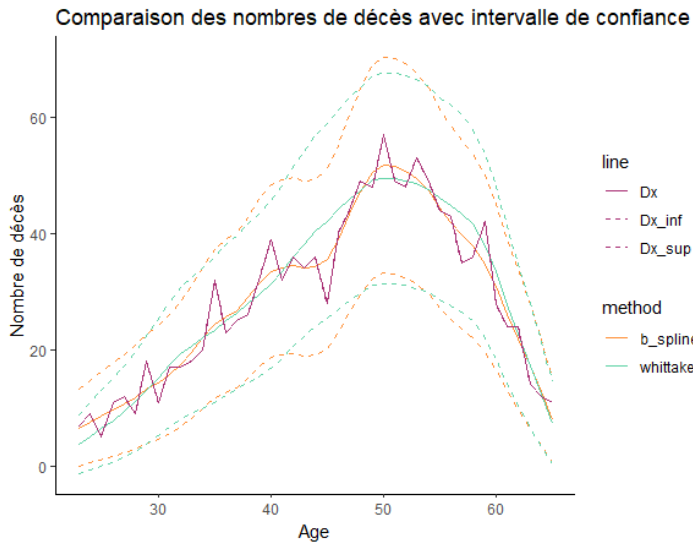
La Figure 3.18 représente les SMR (*Standardized Mortality Ratio*) des différentes méthodes. Les ratios SMR plutôt satisfaisants pour la plupart des méthodes semi-adaptatives. En effet, ils sont proches de 1 et l'intervalle de confiance du SMR est généralement compris entre  $[0,96 - 1,08]$ . Il faut noter que la méthode de Whittaker-Henderson a un SMR qui est plus éloigné de 1 (de 0.98) que les autres méthodes. En effet, il faut être vigilant car le lissage peut introduire un biais systématique. Par exemple, si le SMR diffère de plus de 2% cela peut avoir un impact non-négligeable et matériel en termes de business. C'est donc un critère essentiel et quantitatif pour les méthodes de lissage adaptatives. La Figure 3.18 permet ainsi d'attester de la qualité globale des méthodes B-Spline, GBM Spline, GAM, RBF Network et P-Spline qui ont des ratios quasi égale à 1. Cependant ce critère seul ne permet pas d'observer si une méthode est optimale.

Concernant la statistique du  $\chi^2$ , elle permet aussi d'apprécier de la qualité de fit de nos méthodes de lissage. La meilleure méthode en termes de qualité de fit est la méthode B-Spline. Cela est logique au vu de la valeur de son coefficient de pénalité : c'est une méthode qui a tendance à s'ajuster avec nos données. Enfin les méthodes GBM Spline et RBF Network présentent des résultats plutôt bons. Les résultats des modèles GAM et du modèle GBM Spline sont proches. Cependant, on notera qu'il est difficile de distinguer la meilleure méthode adaptative au travers de ces critères d'ajustements.

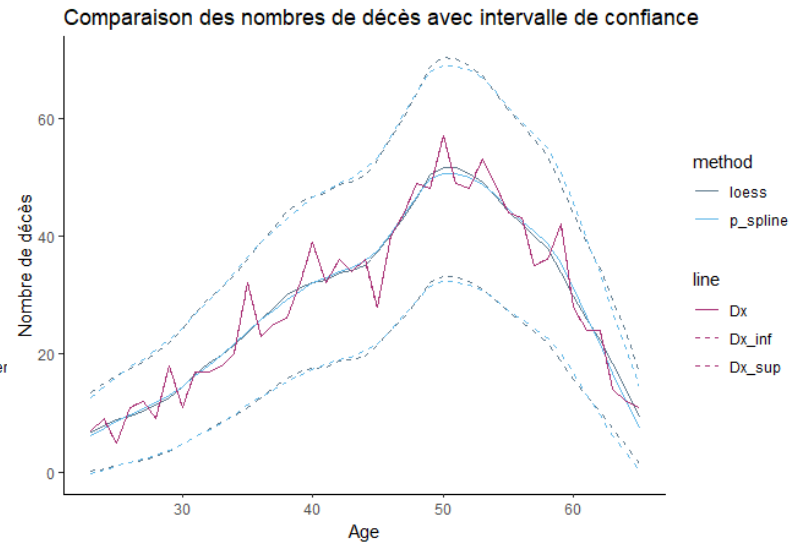
	GAM	B Spline	P Spline	LOESS	WH	GLM local	GBM	RBF
SMR	0.999	0.999	0.999	0.997	0.980	0.994	0.999	0.999
$SMR_{inf}$	0.944	0.944	0.944	0.943	0.927	0.939	0.944	0.945
$SMR_{sup}$	1.090	1.091	1.091	1.088	1.071	1.085	1.090	1.091
Test	H0 True	H0 True	H0 True	H0 True	H0 True	H0 True	H0 True	H0 True
Statistique	1.19	1.20	1.20	1.13	0.53	1.03	1.20	1.19
p_valeur	0.11	0.11	0.11	0.13	0.29	0.15	0.11	0.12

TABLE 3.2 – Résultat du critère SMR et du test associé par méthodes

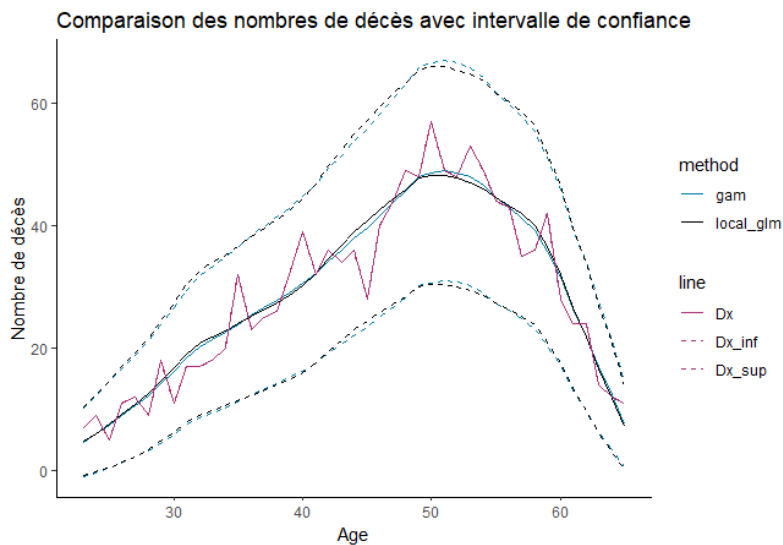
## 3.2.9 Nombre de décès des méthodes de lissage en 1 dimension



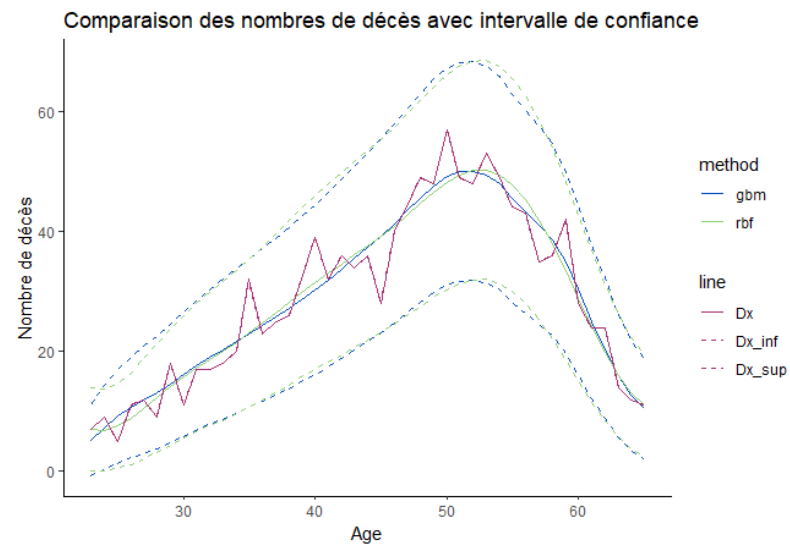
Nombres de décès  $D_x$  à partir des méthodes B-Spline/Whittaker Henderson



Nombres de décès  $D_x$  à partir des méthodes P-Spline/LOESS



Nombres de décès  $D_x$  à partir des méthodes GAM/GLM local



Nombres de décès  $D_x$  à partir des méthodes RBF Network/GBM Spline

**FIGURE 3.20** – Comparaison des taux lissés et des taux bruts avec des intervalles de confiance

La Figure 3.20 représente les nombres de décès par âge  $D_x$  pour les différentes méthodes avec leurs intervalles de confiance respectifs présentés lors de la Section 2.13. C'est-à-dire qu'on évaluera si nos nombres de décès bruts  $D_x$  sont bien compris entre les bornes sup et inf des nombres de décès lissés. En effet, un critère exigible est que notre modèle ne s'éloigne pas trop du nombre de décès des valeurs brutes estimées. On réalisera la même opération avec les taux lissés  $\hat{q}_x$  pour tester la concordance de nos taux lissés. On observe sur la Figure 3.20 que nos valeurs lissées entourent bien les nombres bruts de décès. On observe aussi que les méthodes LOESS, B-Spline, P-Spline enveloppent de manière plus proche les nombres de décès bruts que les autres méthodes. Cela confirme les résultats précédents (ratio SMR et la valeur du  $\chi^2$ ) sur le sûr ajustement potentiel qu'elles peuvent réaliser. Pour la méthode LOESS cela peut être dû à un voisinage  $h$  trop faible avec la minimisation du critère GCV et pour les méthodes par splines cela peut être dû au nombre de noeuds.

### 3.2.10 Application des tests et synthèse des résultats pour les méthodes de lissage en 1 dimension

	GAM	B-Spline	P-Spline	LOESS	WH	GLM local	GBM	RBF
Nber of runs	23	29	29	29	23	19	19	20
Signs (-)	23	22	23	23	27	23	24	23
Signs (+)	20	21	20	20	16	20	19	20
Statistic	0.187	2.011	2.049	2.049	0.630	-1.053	-1.004	-0.743
p.value	0.851	0.044	0.040	0.040	0.528	0.292	0.315	0.457
Test	H0 True	H0 False	H0 False	H0 False	H0 True	H0 True	H0 True	H0 True

**TABLE 3.3** – Résultats du test du Run en 1 dimension avec un niveau de confiance à 95 %

	GAM	B-Spline	P-Spline	LOESS	WH	GLM local	GBM	RBF
Signs (-)	23	22	23	23	27	23	24	23
Signs (+)	20	21	20	20	16	20	19	20
Statistic	0.304	0	0.304	0.304	1.524	0.304	0.609	0.3049
p.value	0.760	1	0.760	0.760	0.127	0.760	0.541	0.760
Test	H0 True	H0 True	H0 True	H0 True	H0 True	H0 True	H0 True	H0 True

**TABLE 3.4** – Résultats du test du Signe en 1 dimension avec un niveau de confiance à 95 %

Le test du run Table 3.3 apparaît plus discriminant que le test du signe Table 3.4 On peut noter que toutes les méthodes ont une probabilité égale d’obtenir des valeurs positives et négatives. Nous n’avons donc pas de sûr ou sous ajustement de nos modèles quant à nos taux bruts. De plus, pour certaines méthodes telles que les modèles GAM, GLM local, GBM Spline et RBF Network, on a des variations de signes qui sont considérées comme aléatoire. En effet, en ce qui concerne le test du *run* Table 3.3 on note que ces méthodes vérifient  $H_0$  (à 95%) et ont donc bien des séquences de valeurs qui se rapprochent de l’aléatoire.

La synthèse des résultats en une dimension est présentée dans la Table 3.5.

**TABLE 3.5** – Synthèse des résultats obtenus en 1 dimension pour les méthodes de lissage adaptatives (le test du SMR est réalisé à un niveau de confiance de 95 %)

Critères		GAM	B-Spline	P-Spline	LOESS	WH	GLM local	GBM	RBF
Test du SMR	SMR (%)	99,9%	99,9 %	99,9%	99,7%	98,0%	99,4%	99,9%	99,9%
	$\xi^{SMR}$	1.19	1.20	1.20	1.13	0.53	1.03	1.20	1.19
	p-valeur	0.11	0.11	0.11	0.13	0.29	0.15	0.11	0.12
Déviante		26.13	17.56	20.48	20.47	31.52	28.87	23.34	23.61
MAPE (%)		0.49 %	0.41 %	0.43 %	0.42 %	0.53 %	0.52 %	0.44 %	0.44 %
Résidus standardisés	> 2	0	0	0	0	1	0	0	0
	> 1	10	6	7	7	11	0	8	8
$R^2$		92.96%	93.97%	92.55%	95.59%	91.89%	91.49%	98.10%	92.64%

Il convient de noter que le modèle de Whittaker-Henderson présente le SMR le plus éloigné de 1. Par ailleurs, la méthode GBM Spline obtient le meilleur pouvoir explicatif avec un coefficient de détermination ( $R^2$ ) de 98,1%. Il est intéressant de souligner que toutes les méthodes ont donné des résultats plutôt satisfaisants avec des SMR relativement proches de 1 et ils vérifient  $H_0$  à un niveau de confiance de 95%. Cependant, il faut tenir compte du fait que les données unidimensionnelles présentaient peu de valeurs aberrantes grâce à l’application préalable du critère de Cochran.

### 3.2.11 Résultat des méthodes de lissage adaptatives en 2 dimensions

Comme précédemment évoqué, dans le cas unidimensionnel, les taux restent stables et les méthodes de lissage adaptatives fonctionnent bien. Toutefois, pour certains types de produits d'assurance, les effets aléatoires d'autres variables telles que la durée de la police ou la somme assurée peuvent être nécessaires. Les effets de ces variables sont à double tranchant : si cela nous permet d'être plus précis et d'avoir des taux de mortalité plus représentatifs, nos données ont une expérience plus faible et donc une estimation plus fragile. En effet, un article de Planchet et Leroy [2009] met en avant ce risque d'une segmentation trop fine pour le calcul des taux de mortalité. Le lissage prend alors tout son sens dans cette situation, car il permet d'effacer les fluctuations de mortalité. Par exemple, toujours avec notre portefeuille étudié en considérant l'âge  $x$ , mais aussi la durée  $t$ , on obtient la surface de mortalité suivante (Figure 3.21 et Figure 3.22),

Le graphique en deux dimensions, obtenu à partir de la méthode par exposition initiale, montre une variabilité importante des nombres de décès. De plus, les âges n'ont pas les mêmes durées de polices, la durée de police maximale étant de 17 ans. On observe toujours une augmentation principalement due à l'âge et une légère augmentation en fonction de la durée. Cependant, celle-ci n'est pas totalement visible, à cause, d'une exposition très faible pour les âges et durées élevés.

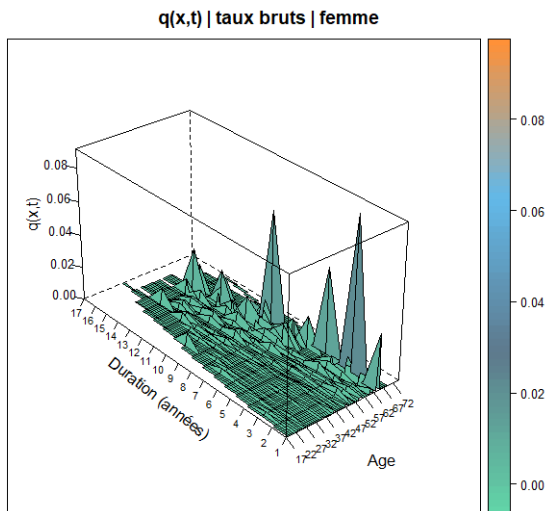


FIGURE 3.21 — (Vue 1)  $\hat{q}_x(t)$  par exposition initiale femme

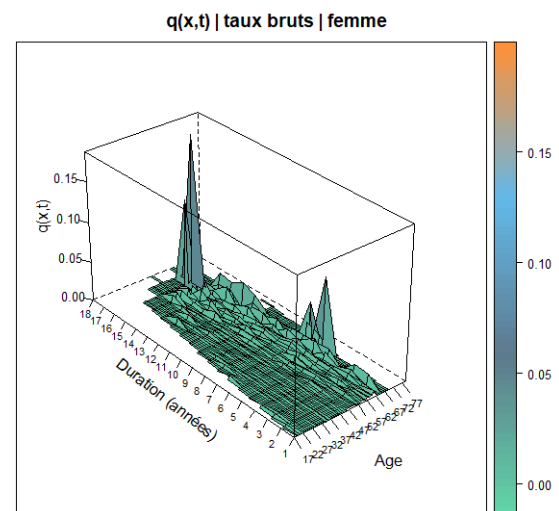


FIGURE 3.22 — (Vue 1)  $\hat{q}_x(t)$  par exposition initiale homme

Nous remarquons que les surfaces de mortalité en deux dimensions sont extrêmement instables en raison du faible nombre de décès. Afin de produire des lissages stables et fiables, nous avons décidé d'agrémenter nos âges par tranche de cinq ans. Par exemple, nous avons remplacé les âges compris entre 20 et 24 ans par un seul âge, à savoir 22 ans. Dans nos surfaces de mortalité, nous avons noté que les hommes ont un effet de durée ou duration plus important que les femmes, en particulier pour les durations élevées. Nous avons également constaté que la mortalité augmente avec l'âge pour les deux sexes, l'effet de l'âge étant plus prononcé que celui de la durée, du moins pour les taux bruts. Enfin, pour les femmes et les hommes, nous avons observé des pics de mortalité avec une tendance générale à l'augmentation en fonction de l'âge, comme le montre la Vue 2.

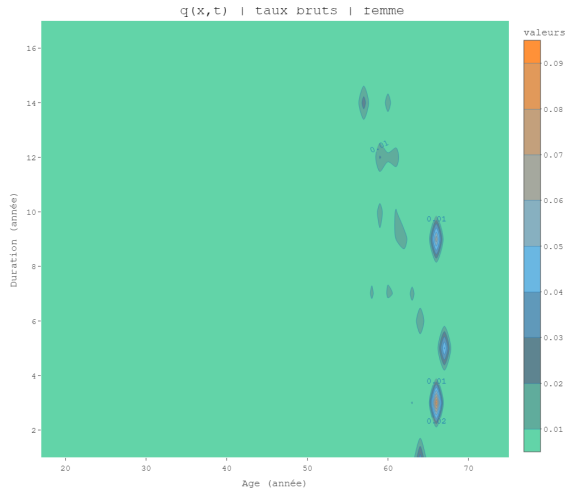


FIGURE 3.23 – (Vue 2)  $\hat{q}_x(t)$  par exposition initiale femme

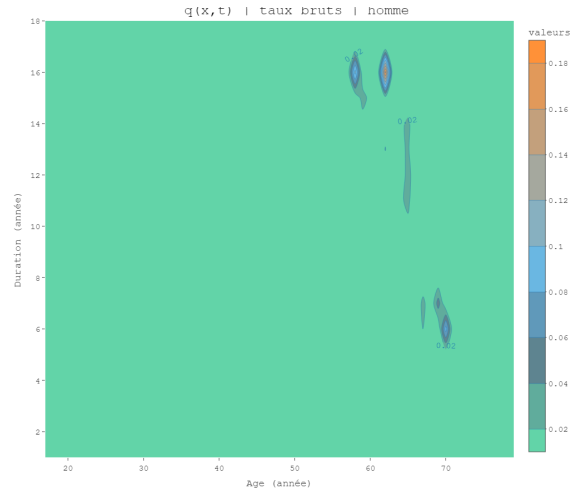


FIGURE 3.24 – (Vue 2)  $\hat{q}_x(t)$  par exposition initiale homme

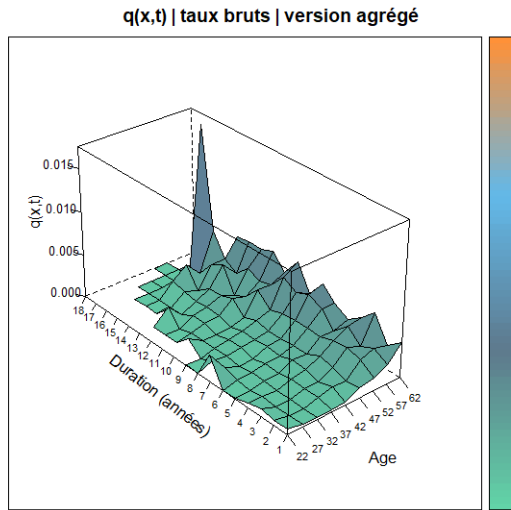


FIGURE 3.25 – (Vue 1)  $\hat{q}_x(t)$  agrégé par exposition initiale

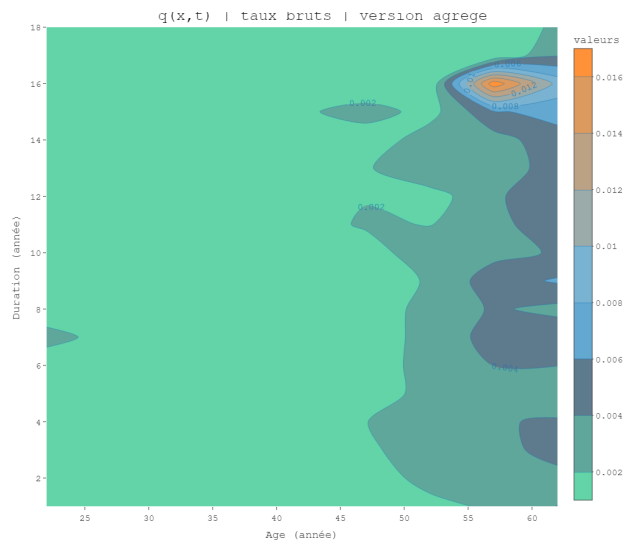


FIGURE 3.26 – (Vue 2)  $\hat{q}_x(t)$  agrégé par exposition initiale

### 3.2.12 Analyse graphique des taux lissés avec les méthodes de lissage en 2 dimension

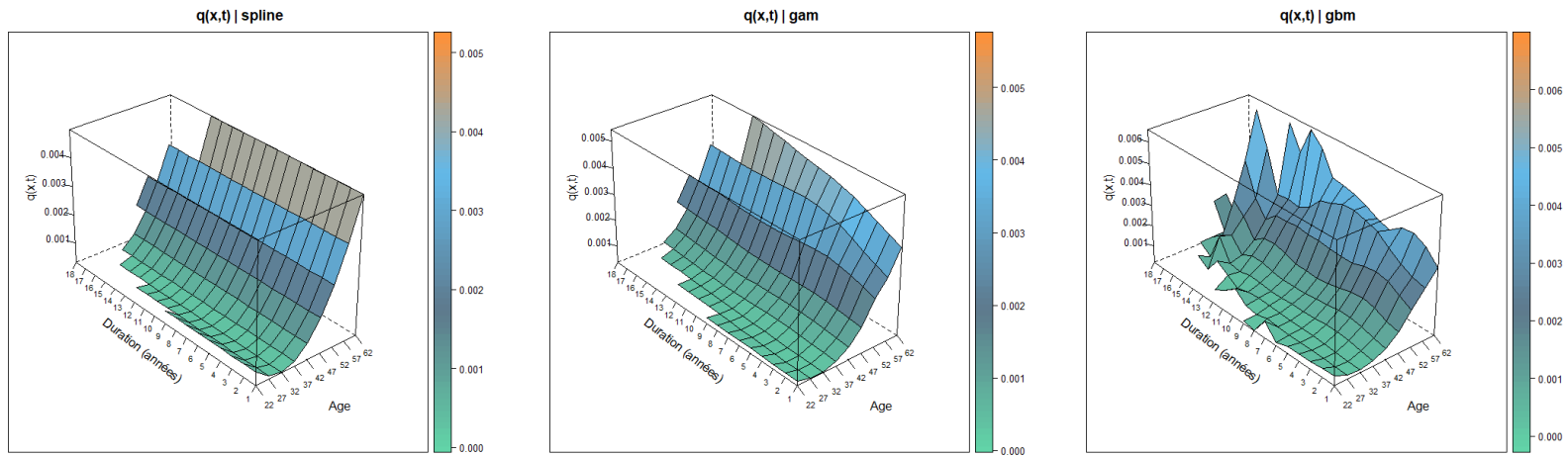


FIGURE 3.27 – (Vue 1)  $\tilde{q}_x(t)$  par méthode adaptative P-Spline/GAM/GBM

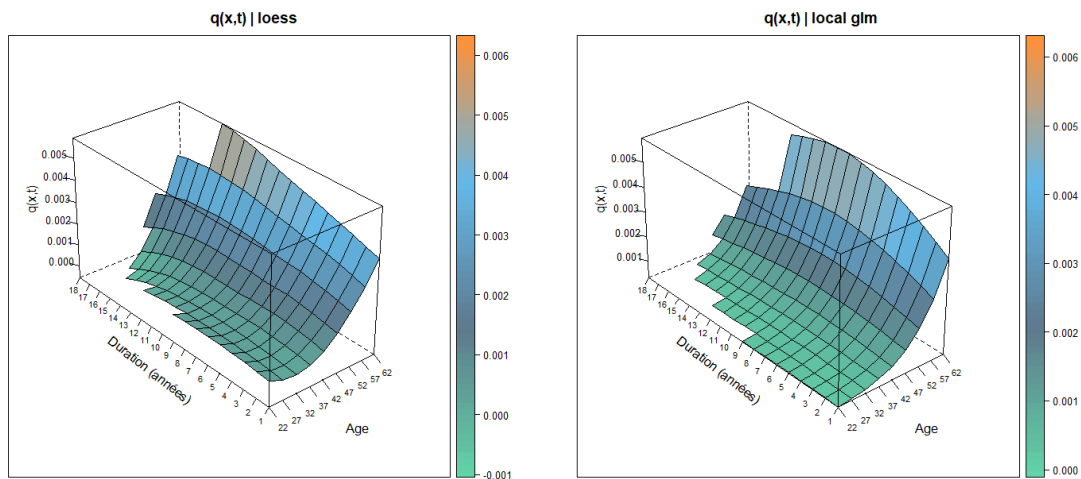


FIGURE 3.28 – (Vue 1)  $\tilde{q}_x(t)$  par méthode adaptative LOESS/GLM local

Les Figures 3.27 et 3.28 permettent de constater que les résultats sont plutôt cohérents et que les irrégularités sont gommées. On a des valeurs qui sont plus regroupées avec moins de volatilité. On peut observer que certaines méthodes accordent plus ou moins d'importance à l'effet de durée de la police. La méthode P-Spline, par exemple n'accorde pas d'importance à la durée ; tandis que les autres méthodes ont des effets de durée.

La méthode GBM Spline présente des résultats satisfaisants : elle arrive à capter la forme originale des taux bruts tout en proposant un résultat plus lisse sans valeurs aberrantes. La condition de convexité qu'on lui a imposée prend ici tout son sens, car elle permet de contenir l'effet de sur-apprentissage ou de valeurs aberrantes qu'y pourrait avoir lieu avec cette méthode. Cependant pour certains portefeuilles d'assurance pour lesquels on cherchera à lisser des taux de rachats où il y a des effets spécifiques de durée de police d'assurance, due en partie à des raisons fiscales, il s'agira d'être vigilant quant à cette méthode pour ne pas imposer une tendance trop restrictive à notre modèle.

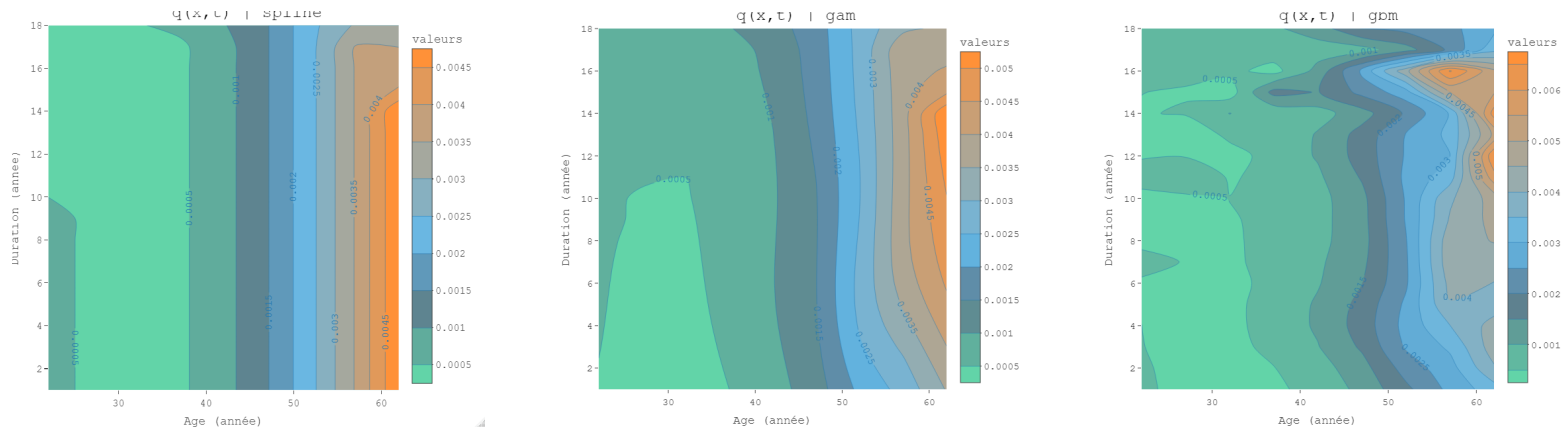


FIGURE 3.29 – (Vue 2)  $\tilde{q}_x(t)$  par méthode adaptative P-Spline/GAM/GBM

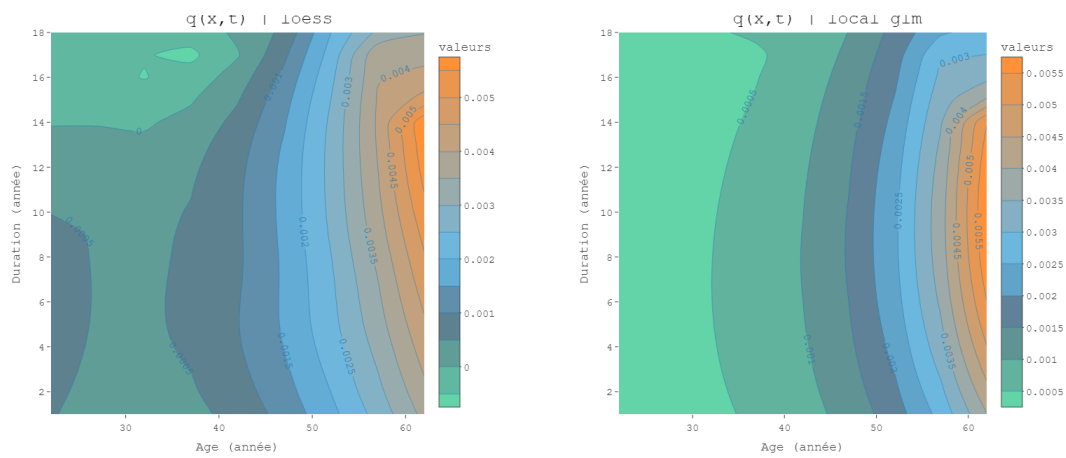
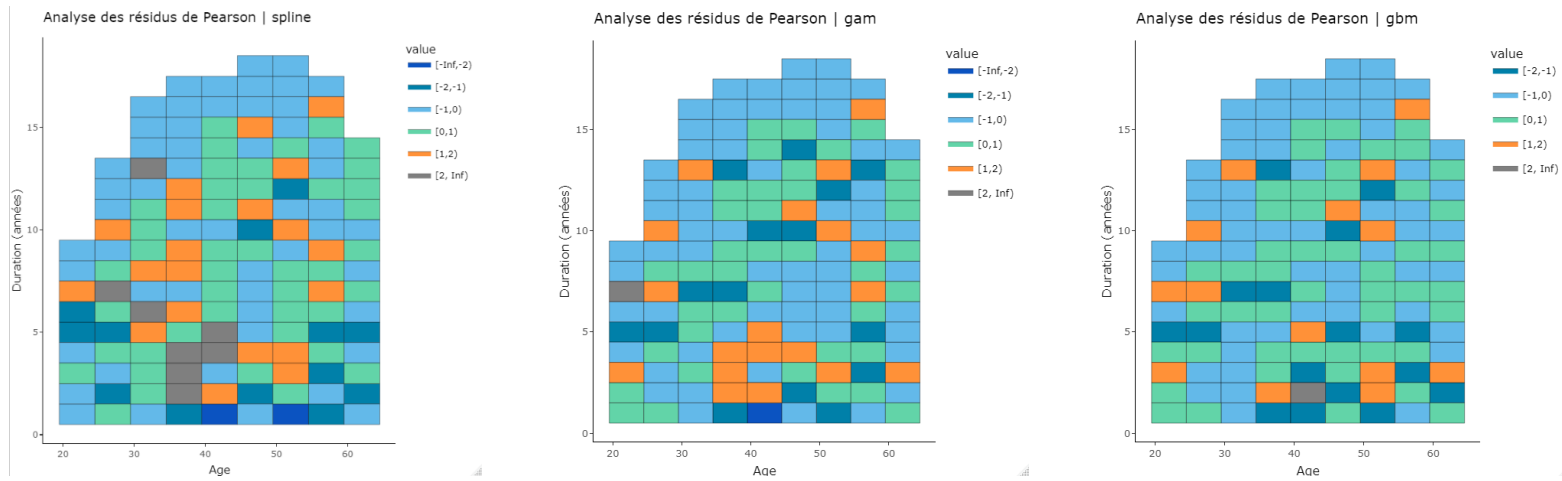


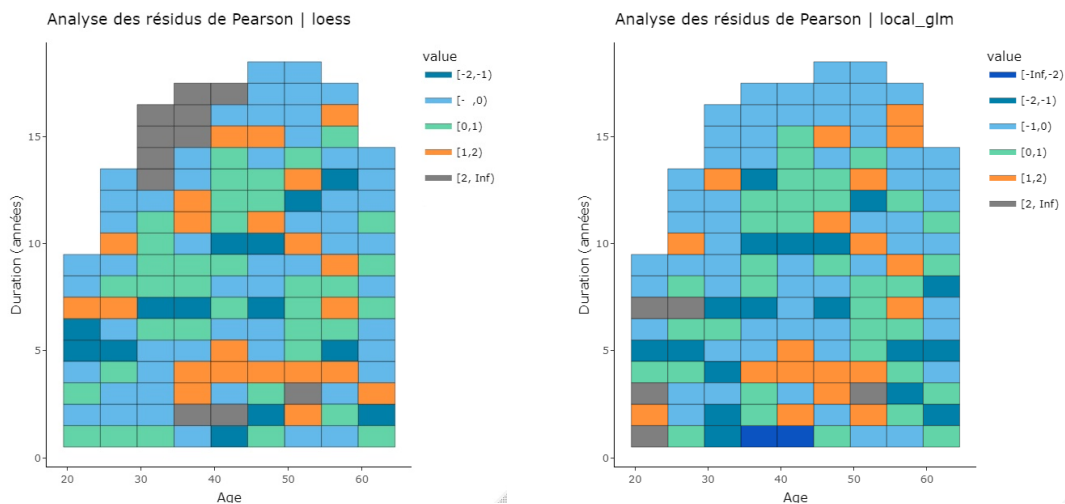
FIGURE 3.30 – (Vue 2)  $\tilde{q}_x(t)$  par méthode adaptative LOESS/GLM local

L'intérêt des Figures 3.29 et 3.30 est de mettre en évidence de manière plus précise les lignes de niveaux et les effets de durée et d'âge. On observe que les méthodes GAM et LOESS adaptatives lissent de manière correcte nos données avec plusieurs points sur le même niveau. Cependant ces méthodes sont moins proches de ce que l'on pouvait observer avant la phase de lissage Figure 3.26. On remarque que le modèle GBM Spline arrive, tout en lissant les données, à capter le pic aux âges élevés et aux durations élevées. Ceci est en partie dû au fait qu'il se construit par une combinaison de plusieurs modèles tout en apprenant de l'erreur des précédents.

### 3.2.13 Analyse des résidus de Pearson pour les méthodes de lissage en 2 dimensions



**FIGURE 3.31** – Comparaison des résidus de Pearson en 2 dimensions pour les méthode *adaptive P-spline*/GAM/GBM



**FIGURE 3.32** – Comparaison des résidus de Pearson en 2 dimensions pour les méthode *adaptive LOESS*/GLM local

Les Figures 3.31 et 3.32 concernent l'analyse des résidus de Pearson en 2 dimensions, qui est essentielle pour comparer les méthodes. On est censés obtenir des valeurs qui sont réparties de manière aléatoire entre les âges et les durations. De plus on n'est pas censés obtenir une suite de valeurs positives ou négatives qui se distingue lorsque l'on observe nos résultats à une durée donnée.

En effet, les résidus sont censés représenter la volatilité gommée par les méthodes de lissage. On observe que toutes les méthodes ont des grilles de résidus qui sont aléatoires. Il n'y a pas de forme particulière qui se distingue qui nous ferait dire que nos méthodes sûres où sous-estime la mortalité du portefeuille pour certaines plages d'âges et de durations.



### 3.2.14 Application des tests et synthèse des résultats pour les méthodes de lissage en 2 dimensions

	GAM	LOESS	P-Spline	GBM	GLM local
Nbr de runs	72	74	60	80	62
Signes (-)	80	69	68	76	80
Signes (+)	59	70	71	63	59
Statistique	0.54	0.59	-1.78	1.74	-1.2
p_valeur	0.59	0.55	0.07	0.08	0.23
Test	H0 True	H0 True	H0 False	H0 False	H0 True

**TABLE 3.6** – Résultat du test du Run en 2 dimensions avec un niveau de confiance à 95 %

	GAM	LOESS	P-Spline	GBM	GLM local
Signes (-)	80	69	68	76	80
Signes (+)	59	70	71	63	59
Statistiques	1.69	0	0.17	1.01	1.69
p_valeur	0.09	1	0.86	0.39	0.089
Test	H0 False	H0 True	H0 True	H0 True	H0 False

**TABLE 3.7** – Résultat du test du Signe en 2 dimensions avec un niveau de confiance à 95 %

Les Tables 3.6 et 3.7 permettent de comparer le caractère aléatoire des résultats des méthodes de lissage en 2 dimensions. C'est un critère qui est difficile à juger en 2 dimensions, car les données en entrée des modèles sont peu nombreuses, il y a donc très peu de sûr apprentissage des modèles. Cependant, le modèle LOESS a des valeurs qui sont considérées comme aléatoires par les deux tests. Les modèles GLM Local, LOESS et GAM ont des valeurs considérées comme aléatoires par le test du run. Tandis que les modèles LOESS, P-Spline, GBM eux sont considérés comme aléatoire par le test du signe.

La Table 3.8 met en évidence les résultats en deux dimensions.

**TABLE 3.8** – Synthèse des résultats obtenus en 2 dimensions pour les méthodes de lissage adaptatives (le test du SMR est réalisé à un niveau de confiance de 95 %)

Critères		GAM	LOESS	P-Spline	GBM	GLM local
<b>Test du SMR</b>	SMR (%)	99,9 %	103,2 %	104,5%	99,9%	99,9%
	$\xi^{SMR}$	3.973	-0.001	-0.0012	3.973	3.976
<b>Déviance</b>		106.95	114.38	151.10	86.46	142.62
<b>MAPE (%)</b>		2.58 %	2.74 %	2.89 %	2.29 %	2.95%
<b>Résidus standardisés</b>	> 2	2	0	10	1	7
	> 1	36	0	43	31	44
<b>R<sup>2</sup></b>		53,39%	55,94%	52,11 %	66,09 %	52,07%

On peut remarquer que les méthodes LOESS et P-Spline présentent des SMR très éloignés de 1. Comme prévu, les méthodes ont des MAPE plus élevées en raison de la granularité plus fine et donc d'une quantité moindre de données explicatives. Toutefois, la méthode GBM Spline affiche des résultats plutôt bons en termes de SMR, ainsi que le meilleur coefficient de détermination ( $R^2$ ). Cependant, il convient de rester vigilant car ces mesures quantitatives ne sont pas pertinentes sans une analyse visuelle approfondie.

### 3.3 Comparaison qualitative des méthodes de lissage

Les parties précédentes ont permis, au travers d'un exemple et par l'intermédiaire de critères statistiques de mettre en évidence les méthodes les plus performantes en 1 et 2 dimensions. Cependant, d'un point de vue plus opérationnel, il y a aussi de nombreux critères essentiels à considérer.

	WH	LOESS	P-Spline / B-Spline	Local GLM	GAM	GBM Spline	RBF Network
Calibration	++	++	+ -	+ -	+	--	+ -
Interprétabilité	1	2	5	3	6	7	4
Multiplicité des dimensions	--	--	+ -	++	++	-	--
Qualité de lissage	--	+	++	+	++	++	++
Auditabilité	++	++	--	+ -	+ -	--	--

FIGURE 3.33 – Comparaison qualitative des méthodes de lissage

Pour classifier les méthodes de lissage, 5 critères ont été retenus. La classification des critères plus « quantitatifs » s'est faite selon les règles suivantes :

Critère	Appréciation					
	+	+	+	+ -	-	- -
Calibration	Nbr param ≤ 2		2 < Nbr param ≤ 4	4 < Nbr param ≤ 5	5 < Nbr param ≤ 6	6 < Nbr param
Qualité de lissage	$\epsilon_{SMR} \leq 0,2\%$		$0,2\% < \epsilon_{SMR} \leq 0,5\%$	$0,5\% < \epsilon_{SMR} \leq 2\%$	$2\% < \epsilon_{SMR} \leq 5\%$	$5\% < \epsilon_{SMR}$

Avec,  $\epsilon_{SMR} = |1 - SMR|$  et Nbr param qui correspond aux nombres de paramètres des modèles.

Tout d'abord, le plus quantitatif des critères qualitatifs, correspond à **la qualité de lissage**. Le critère retenu est basé sur le SMR, avec ce critère, les résultats en 1 dimension des méthodes sont plutôt satisfaisants, à part la méthode de Whittaker-Henderson qui présente un SMR éloigné de 1.

**La calibration** est un critère important pour un modèle et les méthodes GBM Spline, B-Spline, P-Spline, GLM local et RBF Network requièrent à l'utilisateur de configurer plusieurs paramètres et parfois de manière arbitraire. C'est d'ailleurs pour cette raison que cette étude, c'est concentré sur des méthodes adaptatives avec le critère GCV.

En ce qui concerne le critère **d'auditabilité**, 3 niveaux de classification sont retenus dans cette étude. Le premier et le meilleur niveau : (+ +) correspond au fait que ces méthodes peuvent être implémentées sous Excel, mais aussi qu'il est facile à la fois de les comprendre et de tester

si les résultats sont corrects. Pour les méthodes de Whittaker-Henderson et LOESS, ce critère est vérifié, car elles permettent un contrôle des résultats qui est optimal et il est facile de les implémenter sous Excel. Le second niveau : (+ -) correspond au fait que pour implémenter ces méthodes, on doit où il est préférable de passer par un langage de programmation (R/Python/...) mais aussi que les résultats restent auditable, c'est-à-dire qu'on peut les comprendre et obtenir des tests de sensibilité. Les méthodes GLM local et GAM vérifient ces critères. Enfin le troisième et dernier niveau (- -) correspond aux méthodes plus difficilement auditables. Dans cette catégorie on a décidé de classer les méthodes GBM Spline, RBF Network et P-Spline/B-Spline. Ce choix a été motivé par le fait qu'il est difficile avec ces méthodes d'obtenir des valeurs de sensibilité à des variations de paramètres, mais aussi que pour les méthodes par B-Spline/P-Spline et RBF Network nous devons passer par certaines approximations pour obtenir des intervalles de confiance.

**La multiplicité des dimensions** ou plus grossièrement les modifications possibles aux modèles est un critère qualitatif important si on souhaite utiliser un modèle dans le temps. En effet, dans le cadre de ce mémoire, on a concentré l'étude à 1 ou 2 dimensions, mais il est parfois possible de rajouter une troisième dimension qui peut par exemple correspondre à la somme assurée des individus. On peut aussi imaginer que l'on souhaite gagner en qualité de lissage et rajouter une autre dimension qui nous paraît intéressante. Cependant certaines méthodes ne permettent pas de passer si facilement de 2 à 3 dimensions voir plus. Les méthodes les plus flexibles concernant ce critère sont les modèles GAM et GLM local. En effet pour le modèle GBM Spline, il est possible de passer à plus de dimensions, mais les temps d'exécution seront beaucoup plus longs. La méthode de Whittaker-Henderson et la méthode LOESS présentent aussi ce problème.

Enfin un critère non négligeable lorsqu'on choisi un modèle correspond à son **interprétabilité**. Dans ce domaine le modèle GBM Spline présente de nombreux points négatifs à cause de la méthode de boosting. Comme il ajoute plusieurs métas modèles pour obtenir un modèle final, il est difficile à interpréter. De plus on ne peut pas tester la significativité des coefficients in fine. Cependant, à l'aide de la valeur de Shapley par exemple il est possible de connaître « l'apport » d'une variable à la prédiction finale.

### 3.4 Illustration de l'intérêt des méthodes de lissage sur un exemple :

Cette partie illustrera l'importance des méthodes de lissage et viendra comparer les résultats des différentes méthodes adaptatives. On ajustera d'abord nos taux lissés avec des taux de référence issue des tables de mortalité « standards », qui correspondent dans notre étude à la table de mortalité de la population Portugaise fournie par l'INE présentée Table 3.10. On comparera ensuite les prestations actualisées pour connaître le prix de notre traité de réassurance. Cette application dans laquelle nous calculerons seulement la prime pure, qui est une simplification d'un processus de tarification, nous permettra d'ores et déjà de comparer d'un point de vue plus financier les résultats de nos méthodes semi-adaptatives de lissage in fine.

#### 3.4.1 Ajustement des niveaux de mortalité grâce à la théorie de la crédibilité

Une fois le rapport d'analyse de l'expérience achevé - dans notre cas de la construction de notre table de mortalité - les résultats doivent être pris en compte pour produire une base de tarification affinée pour notre traité de réassurance. Un élément majeur de ce processus consiste à déterminer le niveau de crédibilité à attribuer à l'expérience. La crédibilité vise à ajuster les taux considérés en fonction de la crédibilité et des profils de risque de nos portefeuilles étudiés. Dans le cas d'une mortalité très homogène par exemple, tous les portefeuilles ont un profil de risque proche, ils ont donc une mortalité similaire, proche de la table standard. Le plus judicieux est alors de seulement considérer la table standard qui a une expérience plus importante et des valeurs plus nombreuses.

En revanche, dans le cas d'une mortalité très hétérogène, on retrouve une variation importante des mortalités attendues, c'est-à-dire que la mortalité varie beaucoup d'un portefeuille à l'autre et de la table standard. Il serait alors dangereux de se fier à la table standard qui représenterait mal l'expérience et avec elle le profil de risque de notre portefeuille réassuré. La meilleure option est alors de se fier entièrement à l'expérience du portefeuille. Cela s'appelle la crédibilité totale. On a avec  $D_x^A$  : le **nombre de décès actuel** de notre portefeuille lissé et  $D_x^E$  : le **nombre attendu de décès, fournie par des tables de base ou de référence**,

$$\begin{aligned} D_x^{adjust} &= Z_x \times D_x^A + (1 - Z_x) \times D_x^E \\ \Leftrightarrow D_x^{adjust}/D_x^E &= Z_x \times \frac{D_x^A}{D_x^E} + (1 - Z_x) \end{aligned} \quad (3.2)$$

Dans notre exemple, l'expérience passée (expérience obtenue dans l'ensemble du mémoire) et les prévisions de mortalité actuelles (issues de tables standards de mortalité) sont données dans le cadre de notre étude. Le problème est de trouver le facteur de crédibilité,  $Z$ . Si l'expérience est totalement crédible, ce facteur sera égal à un. Si l'expérience n'est pas du tout crédible, le facteur sera de zéro. Dans la plupart des cas, le facteur se situe quelque part entre les deux. Le but des procédures de crédibilité est de mélanger l'information provenant de l'expérience du sujet avec l'information provenant d'un ou plusieurs ensembles d'expériences connexes lorsque l'expérience du sujet n'a pas une crédibilité totale afin d'améliorer l'estimation des valeurs attendues, ou de déterminer quand l'expérience du sujet devrait avoir une crédibilité totale et que le mélange est inutile.

Donc, le facteur de mélange  $Z$  équilibre la précision de l'estimation de l'expérience (liée à notre choix de méthode de lissage) et la pertinence de l'estimation de l'exposition (liée à notre choix de méthode d'exposition). On se contentera dans ce mémoire d'étudier les méthodes de crédibilité partielles. Pour cela, on réalisera l'hypothèse suivante, celle que  $D_x^A = \hat{\theta}_x \cdot D_x^E$ . La valeur  $\hat{\theta}_x$  désigne le niveau de mortalité estimé de notre portefeuille à l'âge  $x$ , c'est un multiple du taux de mortalité de la table de référence ou standard. On se contente d'utiliser cette méthode car on ne dispose pas de plusieurs portefeuilles différents auquel cas nous aurions pu nous tourner vers d'autres méthodes comme celle de Bühlmann (voir Bühlmann et Gisler [2005]). Nous aurions aussi pu utiliser

### 3.4. ILLUSTRATION DE L'INTÉRÊT DES MÉTHODES DE LISSAGE SUR UN EXEMPLE :

les différents produits dans le cadre de notre exemple pour créer plusieurs portefeuilles, cependant les données étaient insuffisantes.

Dans la méthode par fluctuation limitée on utilise la notion d'intervalle de confiance. On va chercher à estimer le nombre de données nécessaire pour avoir une crédibilité totale quant à notre estimation actuelle du nombre de décès (voir Gong et al. [2018] / Paunović et Gajovic [2020]). Cela revient au nombre de données nécessaire pour que nos nombres de décès soit compris avec certitude dans un intervalle de confiance (dans cet exemple à une distance sera à une distance  $c$  de la véritable moyenne) pour une certaine probabilité donnée. Dans toute cette partie on posera que,

$$D_x^{AE} = \frac{D_x^A}{D_x^E} = \frac{\sum_{i=1}^{n_x} D_{i,x}}{D_x^E}$$

Avec,

- $n_x$  : Le nombre d'individus à l'âge  $x$  dans notre portefeuille.
- $D_{i,x}$  : Une variable aléatoire de Bernoulli qui vaut 1 en cas de décès et 0 sinon.

$$\mathbb{P} \left( -c \leq Z_x \frac{(D_x^{AE} - \mathbb{E}(D_x^{AE}))}{\mathbb{E}(D_x^{AE})} \leq c \right) \geq p \quad (3.3)$$

A partir de cette égalité, grâce au **théorème central limite**, c'est à dire que avec un grand nombre d'individus à l'âge  $x$  on a,

$$\sqrt{n_x} \times \frac{D_x^{AE} - \mathbb{E}(D_x^{AE})}{\sqrt{\text{Var}(D_x^{AE})}} \sim \mathcal{N}(0, 1) \quad (3.4)$$

Dans le cadre de la méthode par fluctuation limitée, on désigne deux cas, le cas où nous vérifions ce qu'on appelle la crédibilité totale  $Z_x = 1$ , cela signifie que nous avons suffisamment de données. Et le cas où on se place dans un cadre de crédibilité partielle, c'est-à-dire que  $Z_x < 1$ .

La condition pour se placer dans un cadre dit de crédibilité totale (i.e  $Z_x = 1$ ) en utilisant le résultat du théorème central limite et l'équation 3.3 est que :

$$\sqrt{n_x} \cdot c \cdot \left( \frac{\mathbb{E}(D_x^{AE})}{\sqrt{\text{Var}(D_x^{AE})}} \right) \geq z_{(1+p)/2} \iff n_x \geq \left( \frac{z_{(1+p)/2}}{c} \right)^2 \cdot \left( \frac{\sqrt{\text{Var}(D_x^{AE})}}{\mathbb{E}(D_x^{AE})} \right)^2 \quad (3.5)$$

Si l'équation 3.5 n'est pas vérifié, dans le cadre de la crédibilité partielle la valeur de  $Z$  vaut :

$$Z_x = \frac{\mathbb{E}(D_x^{AE})}{\sqrt{\text{Var}(D_x^{AE})}} \cdot \left( \frac{c}{z_{(1+p)/2}} \right) \cdot \sqrt{n_x} \quad (3.6)$$

On prenant les deux cas (**totale et partielle**) on peut ainsi réécrire le coefficient  $Z_x$  de la manière suivante :

$$Z_x = \min \left\{ 1, \frac{\mathbb{E}(D_x^{AE})}{\sqrt{\text{Var}(D_x^{AE})}} \cdot \left( \frac{c}{z_{(1+p)/2}} \right) \cdot \sqrt{n_x} \right\} = \min \left\{ 1, \sqrt{\frac{n_x}{n^{cred}}} \right\} \quad (3.7)$$

Avec,  $z_{(1+p)/2}$  le quantile désigne le quantile d'ordre  $(1+p)/2$  de la loi normale centrée réduite.

Pour calculer cette valeur, on suppose que nos décès suivent une loi de Bernoulli,  $D_{i,x} \sim \text{Bern}(E_{i,x}, \tilde{q}_x)$ . Avec,  $E_{i,x}$  l'**exposition initiale** pour l'individu  $i$  à l'âge  $x$  dans notre exemple. On obtient ainsi, en supposant l'indépendance concernant la mortalité des individus à l'âge  $x$ ,

$$\begin{aligned}
 \mathbb{E}(D_x^{AE}) &= \frac{\sum_{i=1}^{n_x} \mathbb{E}(D_{i,x})}{D_x^E} \\
 &= \frac{\sum_{i=1}^{n_x} E_{i,x} \tilde{q}_x}{D_x^E} \\
 &= \frac{\sum_{i=1}^{n_x} E_{i,x} \theta_x q_x^s}{D_x^E} \\
 &= \theta_x
 \end{aligned} \tag{3.8}$$

De même on obtient la variance,

$$\begin{aligned}
 Var(D_x^{AE}) &= \frac{\sum_{i=1}^{n_x} Var(D_{i,x})}{(D_x^E)^2} \\
 &= \frac{\sum_{i=1}^{n_x} E_{i,x} \tilde{q}_x (1 - E_{i,x} \tilde{q}_x)}{(D_x^E)^2} \\
 &= \frac{\sum_{i=1}^{n_x} E_{i,x} \theta_x q_x^s (1 - E_{i,x} \theta_x q_x^s)}{(D_x^E)^2}
 \end{aligned} \tag{3.9}$$

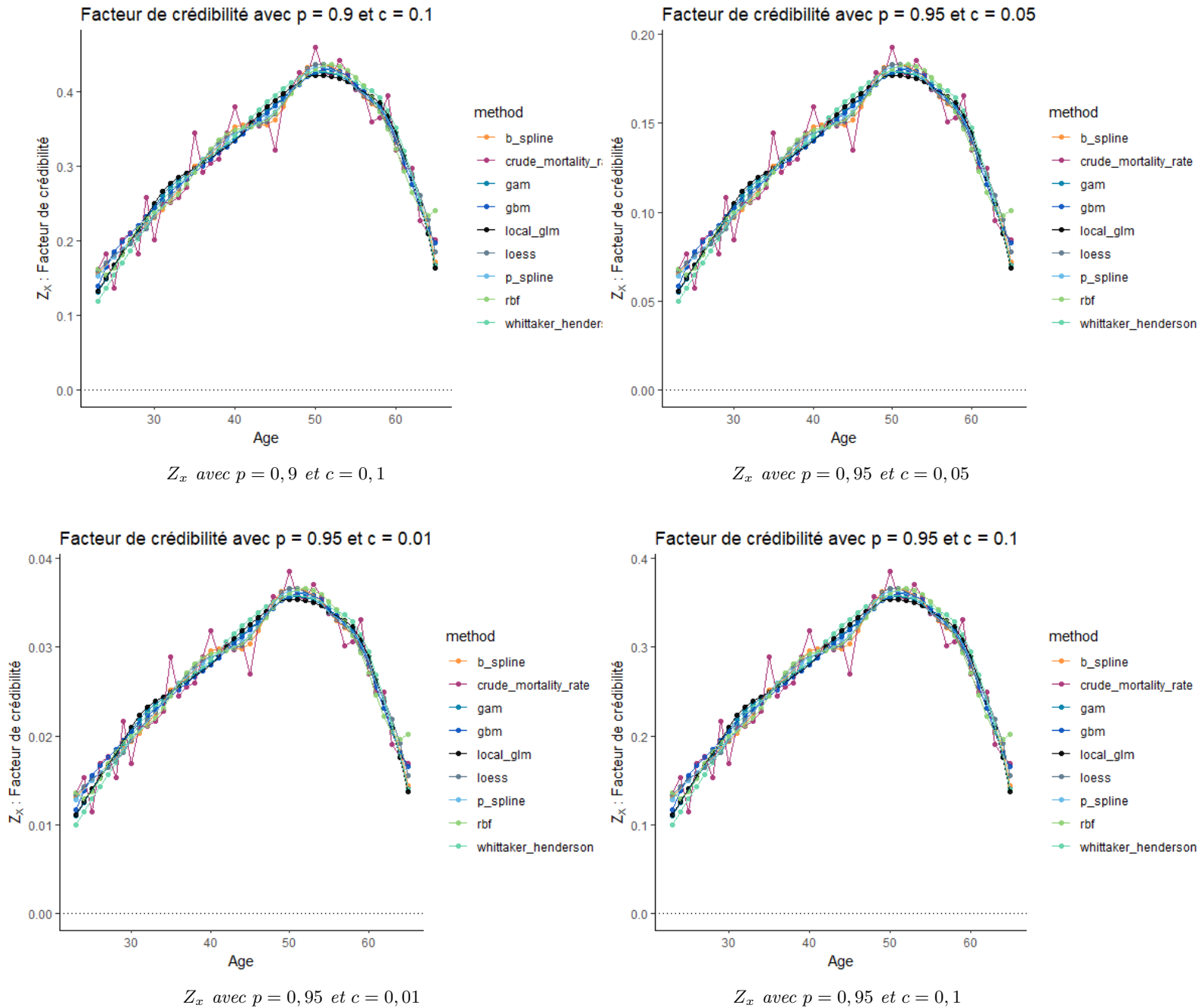
On réalise l'approximation selon laquelle, pour un individu  $i$  à l'âge  $x$ , la quantité  $(1 - E_{i,x} \theta_x q_x^s) \simeq 1$ . **Cette approximation peut ne pas être valable pour certains âges très élevés.** On la considérera ici car notre portefeuille n'est pas constitué d'âges élevés à l'aide du critère de Cochran. On obtient alors l'estimateur suivant pour la variance :

$$Var(D_x^{AE}) = \frac{\sum_{i=1}^{n_x} E_{i,x} \theta_x q_x^s}{(D_x^E)^2} = \frac{\theta_x}{D_x^E}$$

Cela nous permet de calculer notre coefficient de crédibilité  $Z$  toujours en utilisant l'approximation précédente et en remplaçant les termes manquants dans l'équation 3.7,

$$Z_x = \min \left\{ 1, \sqrt{\underbrace{\sum_{i=1}^{n_x} E_{i,x} \theta_x q_x^s}_{\text{Nombre de décès estimés à l'âge } x}} \cdot \left( \frac{z(1+p)/2}{c} \right) \right\} \tag{3.10}$$

### 3.4. ILLUSTRATION DE L'INTÉRÊT DES MÉTHODES DE LISSAGE SUR UN EXEMPLE :



**FIGURE 3.34** – Différentes valeurs du facteur de crédibilité  $Z_x$  à des âges  $x$  donnés en fonction de  $p$  et  $c$

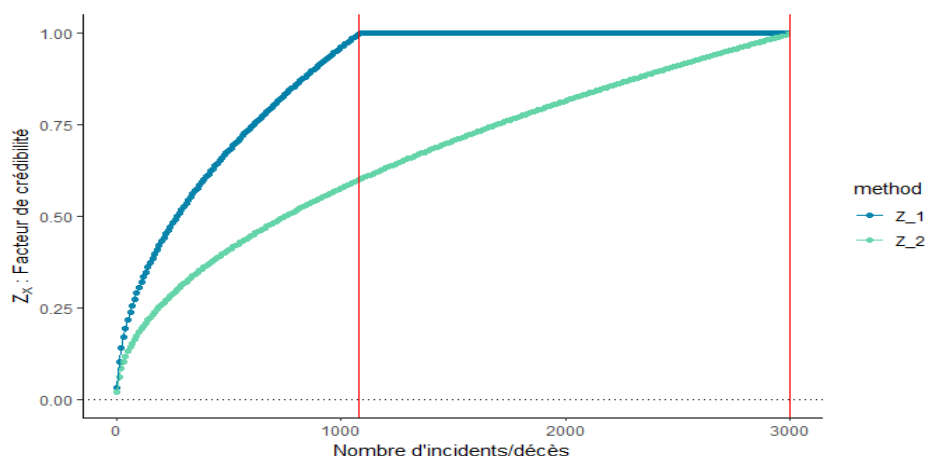
La Figure 3.34 permet d'observer la valeur du facteur de crédibilité  $Z_x$  en fonction du niveau de précision que l'on se donne. En effet plus  $c$  est faible, plus notre erreur est faible. De même plus la valeur  $p$  est élevée, plus on est exigeant sur le fait que l'erreur doit être comprise dans l'intervalle  $[-c, c]$ . On pourra noter que dans le cadre de notre étude nous manquons clairement de données pour appliquer la crédibilité totale. En effet, pour le cas le plus restrictif de nos graphiques ( $p = 0,95$  et  $c = 0,01$ ) on accorde comme facteur de crédibilité à l'expérience  $Z_x \simeq 0,04$  ce qui est très faible.

Le tableau suivant résume le nombre de décès minimal estimés à un âge  $x$  que nous devons obtenir pour qu'on considère une crédibilité totale  $Z_x = 1$  pour nos données :

### 3.4. ILLUSTRATION DE L'INTÉRÊT DES MÉTHODES DE LISSAGE SUR UN EXEMPLE :

Nombre minimum d'incidents/décès à l'âge $x$ pour obtenir la crédibilité totale			
$p \backslash c$	0,01	0,05	0,1
0,9	27055	1082	270
0,925	31700	1268	317
0,95	38414	1536	384
0,975	50238	2009	502

**TABLE 3.9** – Nombre minimal de d'incidents/décès pour appliquer la crédibilité totale



**FIGURE 3.35** – Nombre minimal d'incidents/décès pour les facteurs de crédibilités  $Z_{1,x}$  ( $p = 0,9$  et  $c = 0,05$ ) et  $Z_{2,x}$  ( $p = 0,9$  et  $c = 0,03$ )

La Figure 3.35 représente l'évolution pour des valeurs que l'on retrouve dans la littérature ( $p = 0,9$  et  $c = 0,05$  mais aussi  $p = 0,9$  et  $c = 0,03$ ). La Table 3.9 permet d'observer l'impact d'une variation de  $c$  sur le nombre minimal de données, si on cherche à être très précis dans notre modélisation on devra avoir une table d'expérience avec de nombreux incidents. On peut ainsi observer les exigences en termes de données, il nous faudrait un nombre de décès pour nos âges  $x$  bien plus élevé que ce que la table d'expérience que nous avons construite nous fournis, plus de 1000 pour  $p = 0,9$  et  $c = 0,05$  et plus de 3000 pour  $p = 0,9$  et  $c = 0,03$ .

Il n'apparaît alors pas judicieux dans le cadre de nos données d'utiliser cette méthode à une granularité aussi fine que l'âge  $x$  pour nos nombres de décès. En pratique, il est peu probable d'avoir exactement 1082 sinistres pour chaque tranche d'âge. Ainsi, nous allons adopter, dans le cadre spécifique de ce mémoire une approche plus globale en travaillant sur l'ensemble de notre portefeuille, plutôt que de traiter chaque tranche d'âge individuellement.



### 3.4.2 Application dans le cadre d'un contrat

Dans cette section, nous allons démontrer les écarts de primes pures qui peuvent être constatés en cas de non-lissage des taux, ainsi que les différences de primes en fonction des différentes méthodes de lissage examinées. Nous allons également nous pencher sur les résultats obtenus grâce à la méthode GBM Spline. Grâce à la modélisation des taux bruts et lissés, nous pourrions calculer la prime pure.

En effet, le processus vers le calcul de la prime pure peut être résumé par le **schéma - très simplifié** - suivant :

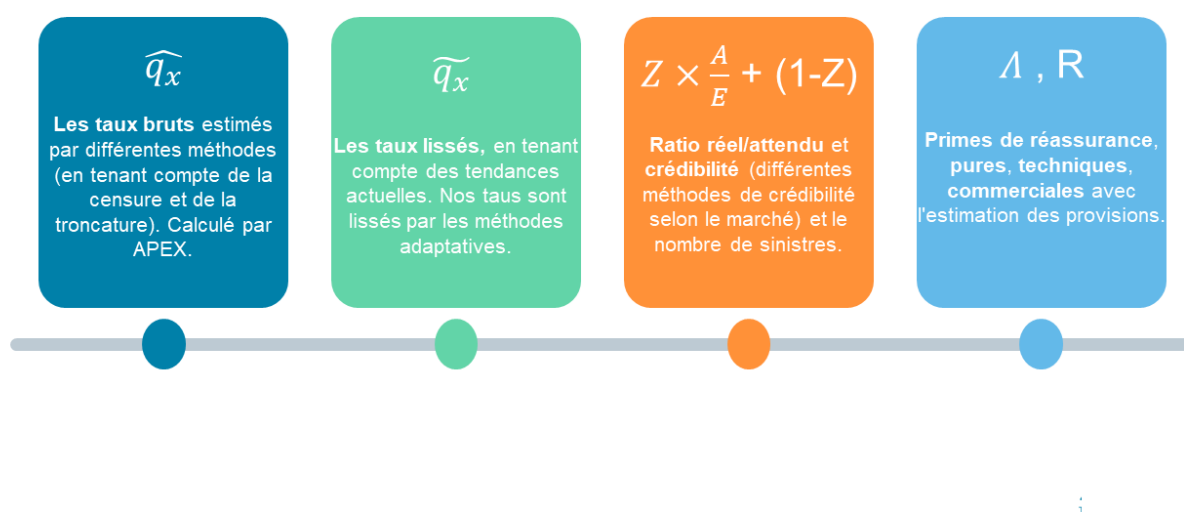


FIGURE 3.36 – Description des étapes jusqu'à la prime de réassurance dans le cadre du mémoire

#### Traité par Excédent de Sinistre :

Un traité excédant de sinistres est généralement découpé en plusieurs tranches, mais dans le cadre de ce mémoire nous nous intéresserons seulement à une tranche. Dans le cadre des traités par excédant de sinistres, on désigne par «**priorité**» la limite basse prise en compte par le traité de réassurance et par «**portée**» la limite haute prise en compte par le traité de réassurance.

On notera alors le traité de réassurance  $P$  xs  $F$  par risques de priorité  $F$  et de portée  $P$ . La notion de traité par événement signifie que le traité s'applique au montant cumulé de tous les sinistres causés par un même événement. Cette notion est essentielle et est mentionnée expressément dans le contrat. On a aussi les traités par risques, ce qui signifie que le traité s'applique à chaque sinistre survenu durant la période de couverture du traité. On aura alors une granularité plus fine car on prendra en compte les sinistres individuels. Il existe aussi des traités qui prennent en compte à la fois des couvertures par événement et par risques.

Le réassureur payera donc en conservant les notations précédentes de notre portefeuille :

$$R = ((\Lambda_x - F) \wedge P)^+ \quad (3.11)$$

### 3.4. ILLUSTRATION DE L'INTÉRÊT DES MÉTHODES DE LISSAGE SUR UN EXEMPLE :

Le calcul de la prime pure s'obtient ensuite en calculant l'espérance des prestations qui devront être fournies par le réassureur dans le cadre du traité. Cependant à cette prime pure on ajoute généralement des quantités qui permettent d'obtenir la prime technique puis la prime commerciale.

À partir de la prime pure, on ajoute les quantités suivantes :

**Le chargement de sécurité :** Il correspond à un pourcentage de la volatilité de la sinistralité.

Il est donc fonction de la variance et peut être calculé de plusieurs manières différentes en fonction du niveau de sécurité que l'on souhaite prendre. On obtient, à partir de ce chargement qu'on ajoute à la prime pure, le chargement sécurisé.

**Les frais de gestion :** Ce sont des frais qui rémunèrent les employés de la compagnie de réassurance c'est à dire des frais de gestion de police, de sinistres.

**Les frais de courtage :** Ils servent à rémunérer les courtiers en réassurance, généralement exprimés en pourcentage de la prime commerciale.

**Le coût du capital ou Cost of Capital (CoC) :** Il sert à rémunérer les actionnaires.

Cela nous permet d'obtenir la prime technique. À cette prime technique on ajoute ensuite un pourcentage pour obtenir la prime commerciale. Ce pourcentage dépend de la politique commerciale de la compagnie de réassurance, on a donc :

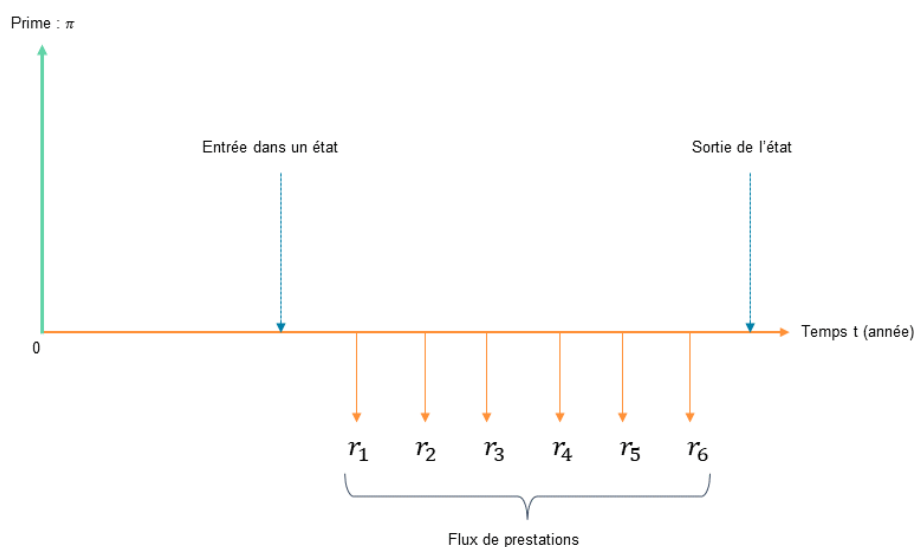


FIGURE 3.37 – Description des flux du contrat

À partir de nos taux lissés, nous pouvons déduire la somme des flux actualisés et estimer  $\Lambda_x$ . Dans le cadre de l'application numérique, on s'intéresse, par simplification à un produit au travers duquel un ensemble de personnes dans un état (survie, arrêt de travail, incapacité,..) perçoivent une prestation tant qu'ils sont dans cet état. On a le flux de prestation suivant lié à notre portefeuille pour tous les individus  $j$  dans un état donné, à l'instant  $t$  :

$$F_t = \sum_{j \in J} r_j \mathbb{1}_{]t; +\infty[}(T_j) \quad (3.12)$$

Avec,

- $J$  : Un ensemble qui regroupe les individus qui composent le portefeuille
- $T_j$  : La loi de maintien de l'individu  $j \in J$

On en déduit la charge totale du portefeuille en sommant sur chacun des temps le montant estimé à couvrir pour le réassureur. On note  $\Lambda$  ce montant total, de telle manière que :

$$\Lambda = \sum_{t=1}^{+\infty} \frac{F_t}{(1+i)^t} = \sum_{j \in J} r_j \sum_{t=1}^{+\infty} \frac{\mathbb{1}_{]t; +\infty[}(T_j)}{(1+i)^t} = \sum_{j \in J} r_j X_j \quad (3.13)$$

A partir de notre table d'expérience construite on pourra comparer les montants de sinistres en fonction des méthodes adaptatives de lissage choisies. On supposera un taux d'intérêt constant qui est de  $i = 0,05\%$  et une durée  $\mathbf{t} = \mathbf{5 \text{ ans}}$ . On obtient alors le montant total suivant pour un individu d'âge  $x$ ,

$$\pi_x = \mathbb{E}(\Lambda_x) = \sum_{t=1}^5 {}_tP_x q_{x+t} \frac{r_t}{(1+i)^t} \quad (3.14)$$

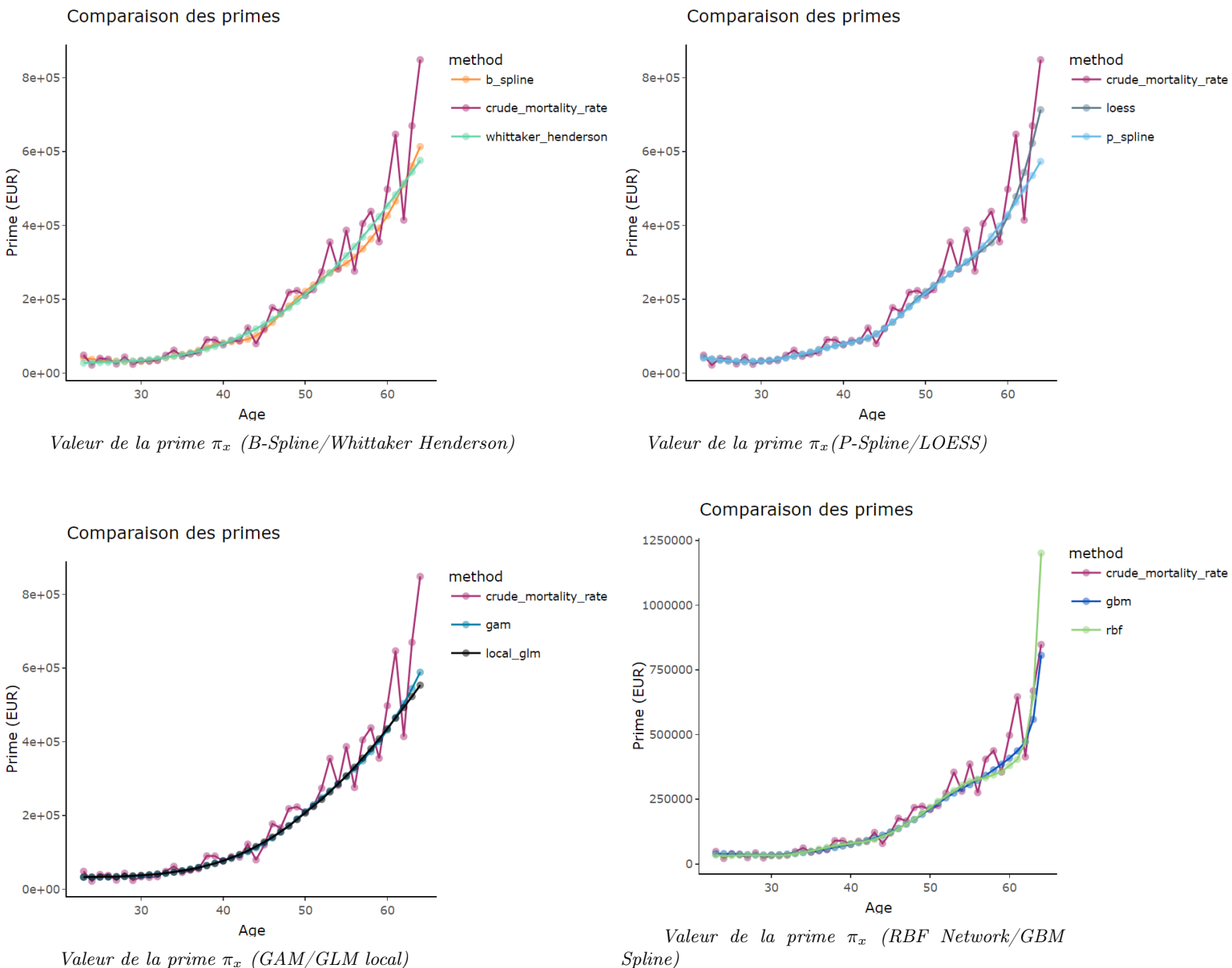
On calculera la valeur espérée des montants, c'est à dire les primes pour chaque âge. Le montant assuré  $r_t$  à l'année  $t$  est supposé déterminé et dans notre exemple on le considérera constant de telle manière que  $r_t = r = 1 \times 10^9 \text{ €}$ .

La Figure 3.38 permet de mettre en évidence les différences en termes de prime pure  $\pi_x$  des différentes méthodes de lissage, ce qui est intéressant à noter, ce sont les différentes tendances des méthodes. Avec des primes  $\pi_x$  qui augmentent significativement pour les âges élevés avec les méthodes adaptatives dans le cadre des modèles GBM Spline et RBF Network. On notera aussi que les méthodes GAM et GLM local ont des estimations à l'aide des méthodes adaptatives (associé au critère GCV) qui sont homogènes à la tendance générale du portefeuille.

La Figure présentée permet également de mettre en évidence la nécessité de lisser les taux bruts afin d'obtenir une cohérence réelle entre les primes des différentes tranches d'âge. En effet, si l'on se base uniquement sur les taux bruts, des incohérences peuvent apparaître entre les différentes tranches d'âge, comme par exemple le cas où un individu de 59 ans paie moins cher qu'un individu de 58 ans. La Figure 3.38 illustre clairement ce phénomène.

Nous pouvons également remarquer une légère augmentation de la prime pour les âges élevés avec les méthodes adaptatives de Whittaker-Henderson et B-Spline. Cette simple illustration met en évidence la différence importante entre les méthodes de lissage. Cependant, pour chaque méthode, les résultats obtenus avec les versions adaptatives reflètent la tendance générale et produisent des primes moins volatiles et plus cohérentes en fonction des âges.

### 3.4. ILLUSTRATION DE L'INTÉRÊT DES MÉTHODES DE LISSAGE SUR UN EXEMPLE :



**FIGURE 3.38** – Comparaison des primes  $\pi_x$  pour les différentes méthodes

En outre, en plus de risques de volatilités dans la valeur estimée de la prime  $\pi_x$ , la décomposition des risques présentée Figure 16 à la Section 0.1.3 a révélé plusieurs autres risques, notamment des risques de tendance liés aux risques d'estimation. Pour ce risque l'erreur est d'autant plus grande que la base de données est petite. Cette problématique est présentée par Delwarde [2005].

Cependant, nous n'avons pas suffisamment de données pour estimer une table prospective à l'aide de notre table d'expérience estimée par les différentes méthodes de lissage adaptatives. Ainsi, pour quantifier ce risque « d'estimation » pour chacune de nos méthodes, nous utiliserons un indicateur présenté par Planchet et Tomas [2011], qui permet d'évaluer la robustesse de nos méthodes de lissage.

Dans le cadre de la tarification d'un traité de réassurance, l'un des principaux risques consiste à estimer les paramètres d'un modèle de projection à partir de données issues d'une table d'expérience qui sont trop bruitées. Dans notre cas, cela correspond à des mauvaises estimations des taux de mortalité lissés.

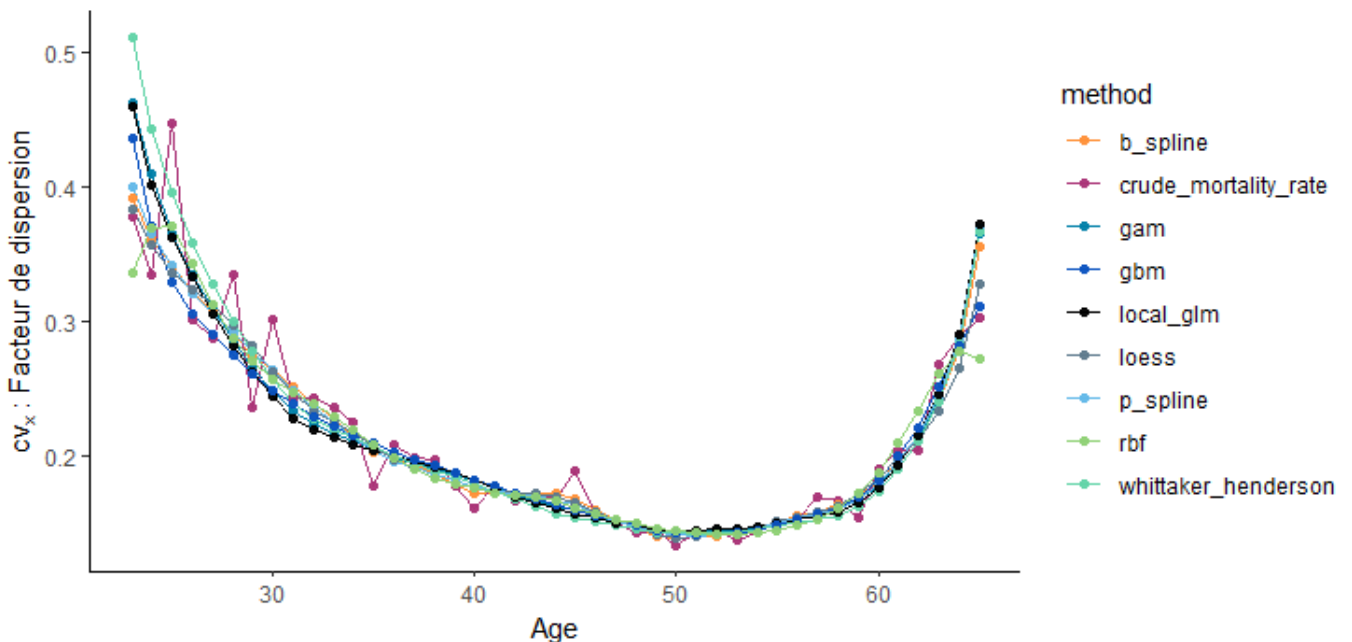
### 3.4. ILLUSTRATION DE L'INTÉRÊT DES MÉTHODES DE LISSAGE SUR UN EXEMPLE :

Pour tenter de capturer ce risque « d'estimation » on peut supposer que nos nombres de décès suivent une loi de Poisson,  $D_x \sim \mathcal{P}(E_x \tilde{q}_x)$ . Cette approximation permet de simuler par la méthode de Monte-Carlo pour chaque âges  $x$  nos nombres de décès  $D_x^k$  associés à une simulation  $k$ .

On peut ainsi calculer un indicateur du risque « d'estimation » ou « risque systémiques » liés à la dispersion des taux de mortalité  $q_x^k = D_x^k / E_x$  pour chaque simulations paramétrés à l'aide nos méthodes adaptatives :

$$cv_x = \frac{\sqrt{(1/(K-1)) \sum_k (\tilde{q}_x^k - (1/K) \sum_k \tilde{q}_x^k)^2}}{(1/K) \sum_k \tilde{q}_x^k} \quad (3.15)$$

L'indicateur  $cv_x$  peut-être interprété comme la **dispersion** liée aux simulations avec les taux de mortalité lissés estimés par les méthodes adaptatives.



**FIGURE 3.39** –  $cv_x$  pour tous les âges et en fonction des différentes méthodes de lissage (nombre de simulations  $K = 100000$ )

D'après la Figure 3.39, il est surprenant de constater qu'après simulation de Monte-Carlo, il n'y a pas de différence significative en termes de volatilité lorsqu'on ne considère que les taux bruts. Cependant, si l'on examine les taux bruts pour différents âges, on peut remarquer que la dispersion entre les taux est plus imprévisible, comme on aurait pu s'y attendre. Cette observation peut s'expliquer par le fait qu'on ne tient pas compte des effets temporels  $t$ , qui ont un impact significatif sur la variabilité. Lorsque nous avons deux dimensions, les valeurs sont beaucoup plus volatiles en raison de ces effets temporels.

Outre le risque d'estimation, il existe d'autres risques qui sont tout aussi importants et qui doivent être pris en compte lors de l'élaboration d'une table d'expérience. Le risque opérationnel, par exemple, est un risque important qui n'est pas traité dans le cadre de ce mémoire. Ce risque peut être lié en partie à la qualité des données fournies, et il est difficile à maîtriser et surtout à évaluer dans nos prédictions finales. Les lecteurs intéressés peuvent se référer à Horr [2009].

---

# Conclusion

En résumé, cette étude a permis de présenter la construction d'une table de mortalité ainsi que des méthodes adaptatives de lissage pour les assureurs et les réassureurs. Étant donné que cette construction peut être chronophage et complexe en raison de nombreux paramètres à prendre en compte, des méthodes basées sur des critères quantitatifs ont été proposées pour répondre aux problèmes tout en équilibrant le biais et la variance des taux lissés. Cela est particulièrement important lorsqu'on utilise ces tables d'expérience pour des projections à long terme, car il est essentiel que les estimations soient robustes et précises sur une période de temps étendue.

Cette étude a pu mettre en lumière que l'absence de lissage entraîne une volatilité excessive des taux de mortalité, ce qui a un impact direct sur le prix du traité de réassurance. Les méthodes de lissage adaptatives, pondérées par crédibilité, permettent d'obtenir des taux plus précis et en adéquation avec l'expérience du portefeuille. Ces méthodes peuvent également servir de base pour les paramètres des modèles, qui peuvent être optimisés manuellement par la suite par les utilisateurs. Le modèle GBM Spline offre une maîtrise optimale de la forme générale de la courbe en utilisant des contraintes pour contrôler l'ajustement dans les modèles de boosting.

Il est pertinent de mentionner les résultats encourageants obtenus en deux dimensions avec le modèle GBM Spline, qui parvient à capturer efficacement la tendance de mortalité avec un  $R^2$  plus élevé que les autres méthodes et à fournir un résultat cohérent et régulier. Toutefois, comme pour le modèle RBF Network, ces approches exigent des données hautement crédibles et peu de valeurs aberrantes. Par conséquent, leur utilisation devrait être limitée. En outre, leur temps de traitement est considérable et leur vulnérabilité aux données est trop importante. Ainsi, des techniques telles que les modèles GLM local et GAM peuvent être préférées, car ils donnent des résultats satisfaisants et leur interprétation est plus simple.

Il convient de souligner que les méthodes de modélisation telles que GBM Spline et RBF Network peuvent être confrontées à des difficultés d'interprétabilité, ce qui peut être problématique pour les praticiens cherchant à comprendre les variables « clés » ayant un impact significatif sur le processus de lissage. Une solution prometteuse pour améliorer la compréhension et l'audibilité de ces modèles serait d'utiliser des métriques telles que la valeur de Shapley, qui permettent de mieux comprendre les variables « discriminantes » et « significatives » dans la phase de lissage. Cela permettrait aux praticiens de mieux interpréter les résultats obtenus et d'ajuster les modèles de manière plus précise pour répondre aux exigences spécifiques du portefeuille étudié.

Il s'agit aussi de noter que cette étude ne traite pas d'une série de modèles relationnels (voir Barbi et Camarda [2011] ou Hannerz [2001]) qui permettent d'obtenir des taux de mortalité lissés à partir d'une table de référence. Ces modèles présentent de nombreux avantages et doivent être comparés aux méthodes proposées dans ce mémoire. En effet, cette étude se concentre sur le cas (restrictif) où nous ne disposons pas de tables de mortalité de référence « officielles ». Cependant, les modèles relationnels ont l'avantage d'un point de vue réglementaire de s'appuyer sur une table d'ores et déjà validée. Bien que les modèles de *boosting* puissent présenter des problèmes d'interprétation, nous pourrions envisager pour approfondir cette étude, un modèle qui combine les spécificités du modèle GBM Spline tout en minimisant une fonction de coût prenant en compte une table de référence.

---

---

# Table des figures

1	Schéma explicatif du cheminement du mémoire et d'une analyse d'expérience . . .	3
2	Schéma explicatif de l'algorithme imbriqué . . . . .	5
3	Dashboard du critère de pénalité et du SMR pour les différentes méthodes . . . . .	7
4	Dashboard de l'étude des résidus pour les différentes méthodes . . . . .	7
5	Dashboard permettant de comparer les taux lissés $\tilde{q}_x$ d'une partie des méthodes adaptatives présentés dans ce mémoire avec les taux bruts $\hat{q}_x$ . . . . .	7
6	Comparaison des primes $\pi_x$ pour les différentes méthodes . . . . .	8
7	Explanatory diagram of the study and of an experience analysis process . . . . .	9
8	Explanatory diagram of the nested algorithm . . . . .	11
9	Dashboard of penalty criteria and SMR for the different methods . . . . .	13
10	Dashboard of the residue study for the different methods . . . . .	13
11	Dashboard allowing to compare the smoothed rates $\tilde{q}_x$ of a part of the adaptive methods presented in this thesis with the raw rates $\hat{q}_x$ . . . . .	13
12	Comparison of the $\pi_x$ premiums for the different methods . . . . .	14
13	Analyse d'expérience au sein de <b>SCOR</b> . . . . .	19
14	Décomposition par type de risques sous <b>APEX</b> . . . . .	20
15	Motivation et points clés derrière cette étude et <b>APEX</b> . . . . .	21
16	Décomposition du risque . . . . .	21
1.1	Schéma explicatif de la notion de <i>censure</i> et de <i>troncature</i> . . . . .	24
1.2	Représentation de l'estimateur de <b>la fonction de survie</b> de Kaplan Meier avec les données de survie de la Figure 1.1 . . . . .	28
1.3	Schéma de l'exposition dans le cadre de l'exemple introductif . . . . .	29
1.4	Schéma explicatif de l'étude pour l'âge partielle 53 ans avec l'exposition centrale et initiale . . . . .	31
1.5	Schéma explicatif de l'hypothèse de constance de la force de mortalité . . . . .	33
2.1	Comparaison du log des taux bruts (à gauche) vs taux lissés (à droite) . . . . .	36
2.2	Comparaison des méthodes lorsque $\lambda = 0$ et $\lambda \neq 0$ . . . . .	38
2.3	Représentation des différentes fonctions de poids $W(\cdot)$ . . . . .	42
2.4	Représentation de la fonction $f(\cdot)$ . . . . .	46
2.5	Résumé du choix des lois en fonction de <b>l'hétérogénéité</b> et de la <b>distribution</b> du portefeuille pour le nombre de sinistres . . . . .	54
2.6	Schéma explicatif d'un réseau de neurones . . . . .	61
2.7	Schéma de la fonction lissée avec le modèle RBF . . . . .	63
2.8	Schéma du réseau de neurones RBF . . . . .	64
2.9	Schéma représentatif du principe de boosting . . . . .	66
2.10	Schéma de l'adaptation de la fonction lissée en fonction des itérations de $m$ . . . . .	68
2.11	Schéma de l'algorithme de Newton-Raphson avec la fonction de coût pénalisée . . . . .	73
2.12	Schéma explicatif de la méthode du gradient conjugué . . . . .	77
2.13	Schéma explicatif de l'algorithme imbriqué . . . . .	81
2.14	Exemple d'une séquence . . . . .	85
3.1	Format des données en input de l'exemple d'analyse d'expérience . . . . .	94
3.2	Boxplot des âges pour les femmes . . . . .	95
3.3	Boxplot des âges pour les hommes . . . . .	95
3.4	Nombre de décès $D_x$ bruts estimés par exposition initiale pour les femmes . . . . .	96
3.5	Nombre de décès $D_x$ bruts estimés par exposition initiale pour les hommes . . . . .	96

---

3.6	Comparaison des taux bruts $\hat{q}_x$ par type d'expositions (initiale et centrale) . . . . .	97
3.7	LoM par rapport à la table de référence avec les $\hat{q}_x$ bruts estimés par exposition initiale . . . . .	98
3.8	LoM par rapport à la table de référence avec les $\hat{q}_x$ bruts estimés par exposition centrale . . . . .	98
3.9	Analyse de la dispersion des décès $D_x$ dans le portefeuille étudié . . . . .	99
3.10	Comparaison des taux lissés et des taux bruts avec des intervalles de confiance . .	100
3.11	Critère de pénalité pour les modèles <i>B-Spline/P-Spline/Whittaker-Henderson</i> . . .	101
3.12	Critère de pénalité pour les modèles <i>GLM local/GAM/LOESS</i> . . . . .	101
3.13	Critère de pénalité pour les modèles <i>RBF Network/GBM Spline</i> . . . . .	101
3.14	Comparaison des résidus Pearson pour les différentes méthodes de lissage . . . . .	102
3.15	Comparaison des résidus de la réponse pour les différentes méthodes de lissage . .	103
3.16	Analyse des résidus de la déviance pour les différentes méthodes . . . . .	104
3.17	Comparaison des taux bruts sur les taux lissés pondérés pondérés par le nombre de décès associés . . . . .	105
3.18	Comparaison des ratios SMR pour les différentes méthodes . . . . .	106
3.19	Comparaison des valeurs du $\chi^2$ pour les différentes méthodes . . . . .	106
3.20	Comparaison des taux lissés et des taux bruts avec des intervalles de confiance . .	107
3.21	(Vue 1) $\hat{q}_x(t)$ par exposition initiale femme . . . . .	109
3.22	(Vue 1) $\hat{q}_x(t)$ par exposition initiale homme . . . . .	109
3.23	(Vue 2) $\hat{q}_x(t)$ par exposition initiale femme . . . . .	110
3.24	(Vue 2) $\hat{q}_x(t)$ par exposition initiale homme . . . . .	110
3.25	(Vue 1) $\hat{q}_x(t)$ agrégé par exposition initiale . . . . .	110
3.26	(Vue 2) $\hat{q}_x(t)$ agrégé par exposition initiale . . . . .	110
3.27	(Vue 1) $\tilde{q}_x(t)$ par méthode adaptative <i>P-Spline/GAM/GBM</i> . . . . .	111
3.28	(Vue 1) $\tilde{q}_x(t)$ par méthode adaptative <i>LOESS/GLM local</i> . . . . .	111
3.29	(Vue 2) $\tilde{q}_x(t)$ par méthode adaptative <i>P-Spline/GAM/GBM</i> . . . . .	112
3.30	(Vue 2) $\tilde{q}_x(t)$ par méthode adaptative <i>LOESS/GLM local</i> . . . . .	112
3.31	Comparaison des résidus de Pearson en 2 dimensions pour les méthode adaptative <i>P-Spline/GAM/GBM</i> . . . . .	113
3.32	Comparaison des résidus de Pearson en 2 dimensions pour les méthode adaptative <i>LOESS/GLM local</i> . . . . .	113
3.33	Comparaison qualitative des méthodes de lissage . . . . .	115
3.34	Différentes valeurs du facteur de crédibilité $Z_x$ à des âges $x$ donnés en fonction de $p$ et $c$ . . . . .	120
3.35	Nombre minimal d'incidents/décès pour les facteurs de crédibilités $Z_{1,x}$ ( $p = 0,9$ et $c = 0,05$ ) et $Z_{2,x}$ ( $p = 0,9$ et $c = 0,03$ ) . . . . .	121
3.36	Description des étapes jusqu'à la prime de réassurance dans le cadre du mémoire .	122
3.37	Description des flux du contrat . . . . .	123
3.38	Comparaison des primes $\pi_x$ pour les différentes méthodes . . . . .	125
3.39	$cv_x$ pour tous les âges et en fonction des différentes méthodes de lissage (nombre de simulations $K = 100000$ ) . . . . .	126
3.40	(Vue 3) $\tilde{q}_x(t)$ des méthodes adaptatives en 2 dimenions . . . . .	135
3.41	(Vue 3) $\tilde{q}_x(t)$ avec le modèle GBM Spline . . . . .	136
3.42	Comparaison des déviances pour les différentes méthodes de lissage . . . . .	139



---

# Liste des tableaux

1	Résumé des modèles utilisés et des algorithmes d'optimisations utilisés pour le principe « adaptatif » dans le cadre d'une construction de table de mortalité . . . . .	4
2	Résumé des résultats obtenus en 1 dimension pour les méthodes de lissage adaptatives	6
3	Résumé des résultats obtenus en 2 dimensions pour les méthodes de lissage adaptatives	6
4	Summary of the models used and the optimisation algorithms used for the « adaptive » principle in the context of a life table construction . . . . .	10
5	Summary of 1 dimension results for adaptive smoothing methods . . . . .	12
6	Summary of the results obtained in 2 dimensions for the adaptive smoothing methods	12
1.2	<i>Pseudo donnée</i> dans le cadre de notre exemple . . . . .	35
2.2	Exemple de fonction d'activations . . . . .	62
3.1	Nombre de décès total $D_{total}$ par hommes et femmes, mais aussi par type de produits	96
3.2	Résultat du critère SMR et du test associé par méthodes . . . . .	106
3.3	Résultats du test du <i>Run</i> en 1 dimension avec un niveau de confiance à 95 % . . .	108
3.4	Résultats du test du <i>Signe</i> en 1 dimension avec un niveau de confiance à 95 % . .	108
3.5	Synthèse des résultats obtenus en 1 dimension pour les méthodes de lissage adaptatives (le test du SMR est réalisé à un niveau de confiance de 95 %) . . . . .	108
3.6	Résultat du test du <i>Run</i> en 2 dimensions avec un niveau de confiance à 95 % . . .	114
3.7	Résultat du test du <i>Signe</i> en 2 dimensions avec un niveau de confiance à 95 % . . .	114
3.8	Synthèse des résultats obtenus en 2 dimensions pour les méthodes de lissage adaptatives (le test du SMR est réalisé à un niveau de confiance de 95 %) . . . . .	114
3.9	Nombre minimal de d'incidents/décès pour appliquer la crédibilité totale . . . . .	121
3.10	Données issues d'études nationales de mortalités de l'INE Portugaise filtrés avec le critère de <i>Cochran</i> et ré-équilibrés en fonction de la proportion homme-femme de notre portefeuille . . . . .	137
3.11	Statistiques descriptives des données du portefeuille . . . . .	138

---

---

# Bibliographie

- Barbi, E. et Camarda, C. G. [2011], ‘Period and cohort effects on elderly mortality : a new relational model for smoothing mortality surfaces’, *Statistica* .
- Baughman, D. et Liu, Y. [1995], ‘Classification : Fault diagnosis and feature categorization’, *Neural Networks in Bioprocessing and Chemical Engineering* .
- Bühlmann, H. et Gisler, A. [2005], ‘A course in credibility theory and its applications’, *Scandinavian Actuarial Journal* .
- Cleveland, W. S. [1979], ‘Robust locally weighted regression and smoothing scatterplots’, *Journal of the American Statistical Association* .
- de Boor, C. [1978], ‘A practical guide to splines’, *Berlin : Springer Verlag* .
- Delwarde, A. [2005], ‘Modèle log-bilinéaire pour l’élaboration de tables de mortalité prospectives’, *Mémoire d’actuariat - UCL* .
- Denuit, M., Maréchal, X., Pitrebois, S. et Walhin, J. [2007], *Actuarial Modelling of Claim Counts : Risk Classification, Credibility and Bonus-Malus Systems*.
- European, U. [17/12/2009], ‘Directive 2009/138/ec of the european parliament and of the council of 25 november 2009 on the taking-up and pursuit of the business of insurance and reinsurance (solvency ii)’.
- Fauvernier, M., Roche, L., Uhry, Z., Tron, L., Bossard, N. et Remontet, L. [2019], ‘Splines multidimensionnelles pénalisées dans les modèles de survie : applications en épidémiologie des cancers’.
- Felipe, A., Guillen, M. et Perez-Marin, A. M. [2002], ‘Recent mortality trends in the spanish population’, *British Actuarial Journal* .
- Fletcher, R. et Reeves, C. M. [1964], ‘Function minimization by conjugate gradients’, *The Computer Journal* .
- Forfar, D., McCutcheon, J. J. et Wilkie, A. D. [1988], ‘On graduation by mathematical formula’, *Journal of the Institute of Actuaries* **115**.
- G. Wahba, Golub, H. G. et Heath, M. [1979], ‘Generalized cross-validation as a method for choosing a good ridge parameter’, *Technometrics* .
- Gavin, J. B., Haberman, S. et Verrall, R. J. [1994], ‘On the choice of bandwidth for kernel graduation’, *Journal of the Institute of Actuaries* .
- Goepf, V., Bouaziz, O. et Nuel, G. [2018], ‘Spline regression with automatic knot selection’.
- Gong, Y. M., Li, Z., Milazzo, M., Moore, K. et Provencher, M. [2018], ‘Credibility methods for individual life insurance’, *Risks* .
- Grievink, J. et van Schijndel, J. [2002], ‘A new algorithm for developing dynamic radial basis function neural network models based on genetic algorithms’, *European Symposium on Computer Aided Process Engineering-12* .
- Hannerz, H. [2001], ‘An extension of relational methods in mortality estimation’.
- Hastie, T. et Tibshirani, R. [1986], ‘Generalized additive models’, *Statistical Science* .
-

- Henderson, W. [1924], ‘A new method of graduation’, *Transactions XXV of the Actuarial Society of America* .
- Hofner, B., Hothorn, T., Kneib, T. et Schmid, M. [2011], ‘A framework for unbiased model selection based on boosting’, *Journal of Computational and Graphical Statistics* .
- Horr, R. E. [2009], ‘Construction des tables d’expérience et étude de l’impact de l’incertitude des données’, *Mémoire d’actuariat - ISFA* .
- Huang, T.-M. [2021], ‘An estimation procedure with knot selection for multivariate regression splines’, *Journal of Physics : Conference Series* .
- Jennrich, R. I. et Sampson, P. F. [1976], ‘Newton-raphson and related algorithms for maximum likelihood variance component estimation’, *Technometrics* .
- Kamega, A. [2011], ‘Outils théoriques et opérationnels adaptés au contexte de l’assurance vie en Afrique subsaharienne francophone - analyse et mesure des risques liés à la mortalité’, *Thèse d’actuariat université Lyon I* .
- Kaplan, E. L. et Meier, P. [1958], ‘Nonparametric estimation from incomplete observations’, *Journal of the American Statistical Association* .
- la Commission d’Agrément [2006], ‘Lignes directrices mortalité’, *Recommandation dans les règles professionnelles de l’Institut par le Conseil d’Administration du 20 juin 2006* .
- Li, Z. et Cao, J. [2022], ‘General p-splines for non-uniform b-splines’.
- Liddell, F. D. [1984], ‘Simple exact analysis of the standardised mortality ratio.’, *Journal of Epidemiology & Community Health* .
- Local Likelihood Estimation* [1987], *Journal of the American Statistical Association* .
- Mayr, A., Fenske, N., Hofner, B., Kneib, T. et Schmid, M. [2012], ‘Generalized additive models for location, scale and shape for high dimensional data—a flexible approach based on boosting’, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* .
- McCullagh, P. et Nelder, J. [1989], ‘Generalized linear models’, *Chapman and Hall/CRC* .
- Paunović, M. et Gajovic, V. [2020], ‘Adjustment of mortality tables by limited fluctuation method’, *Tokovi osiguranja* .
- Planchet, F. et Leroy, G. [2009], ‘Quel niveau de segmentation pertinent ?’, *La Tribune de l’Assurance* **n°142**.
- Planchet, F. et Tomas, J. [2011], ‘Construction et validation des références de mortalité de place’.
- Simoës, O. et Correia, T. [2019], ‘Portugal mortality report 2019’, *IAA Mortality Working Group* .
- Speleers, H. [2019], ‘Algorithm 999 : Computation of multi-degree b-splines’, *ACM Trans. Math. Softw.* .
- Tomas, J. [2015], ‘A local likelihood approach to univariate graduation of mortality’, *Bulletin Français d’Actuariat* .
- Tomas, J. et Planchet, F. [2013], ‘Multidimensional smoothing by adaptive local kernel-weighted log-likelihood : Application to long-term care insurance’, *Insurance : Mathematics and Economics* .
- Wahba, G. [1975], ‘Smoothing noisy data with spline functions.’, *Numerische Mathematik* .
- Wahba, G. [1985], ‘A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem’, *The Annals of Statistics* .
- Wood, S. [2006], *Generalized Additive Models : An Introduction With R*.
- Wooldridge, J. [1997], ‘Quasi-likelihood methods for count data’, *Handbook of Applied Econometrics* .

---

# Abréviations

**GBM Spline** : *Gradient Boosting Model Spline*  
**GAM** : *Generalized Additive Model*  
**GLM** : *Generalized Linear Model*  
**RBF Network** : *Radial Basis Function Network*  
**LOESS** : *LOcally Estimated Scatterplot Smoothing*  
**SMR** : *Standardized Mortality Ratio*  
**GCV** : *Generalized Cross Validation*  
**WRSS** : *Weighted Residual Sum of Squares*  
**EA** : *Experience Analysis*  
**IRWLS** : *Iteratively Reweighted Least Squares*  
**BFGS** : *Broyden-Fletcher-Goldfarb-Shanno*  
**MAPE** : *Mean Absolute Percentage Error*  
**A/E** : *Actual / Expected*  
**LoM** : *Level of Mortality*  
**BE** : *Best Estimate*  
**INE** : *Instituto Nacional de Estadística*

---

---

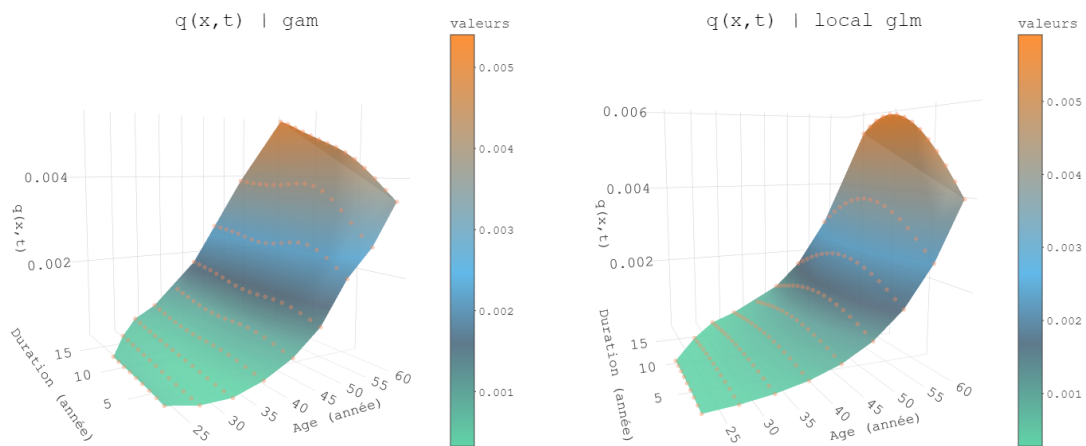
# Annexes

---

---

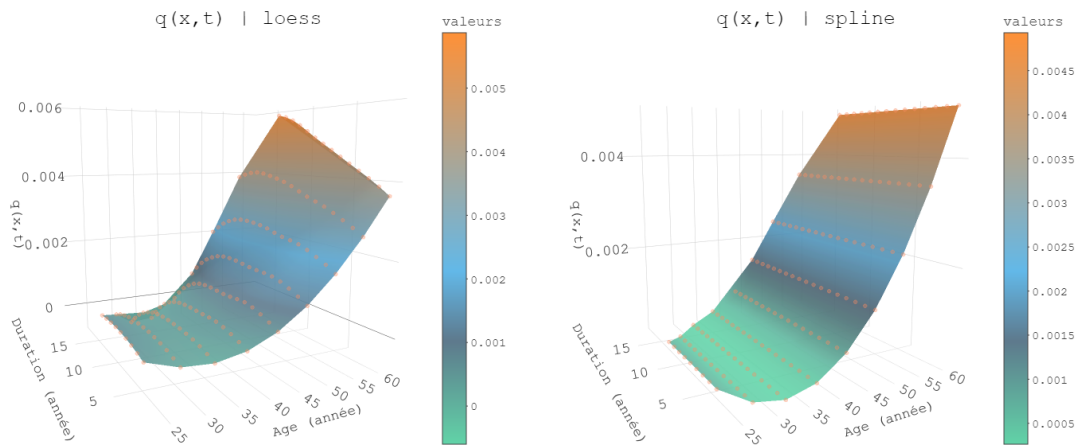
# Annexe A

## Résultat des méthodes de lissage adaptatives en 2 dimensions



(Vue 3)  $\tilde{q}_x(t)$  avec le modèle GAM

(Vue 3)  $\tilde{q}_x(t)$  avec le modèle GLM local



(Vue 3)  $\tilde{q}_x(t)$  avec le modèle LOESS

(Vue 3)  $\tilde{q}_x(t)$  avec le modèle P-Spline

**FIGURE 3.40** – (Vue 3)  $\tilde{q}_x(t)$  des méthodes adaptatives en 2 dimensions

---

---

# Annexe A

## Résultat des méthodes de lissage adaptatives en 2 dimensions

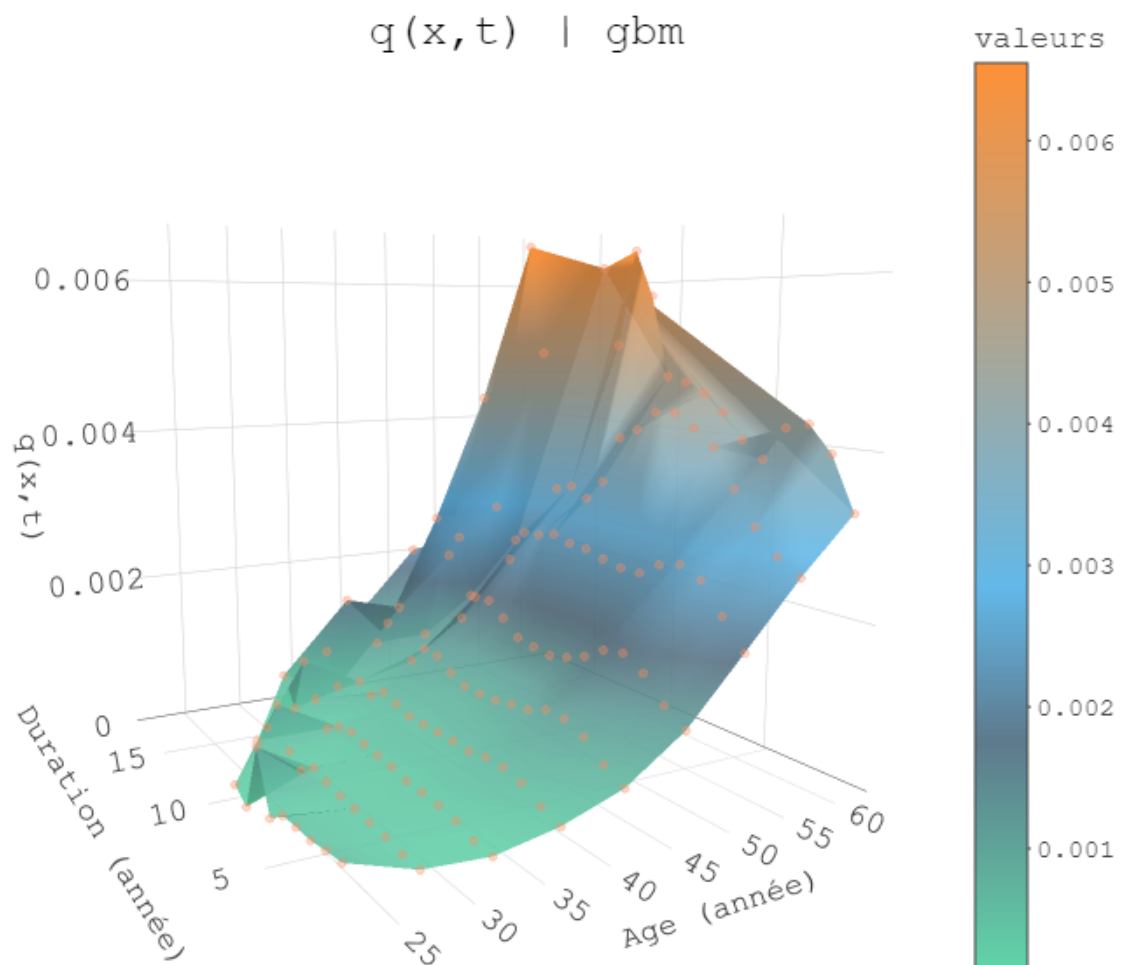


FIGURE 3.41 – (Vue 3)  $\tilde{q}_x(t)$  avec le modèle GBM Spline

---

---

# Annexe B

## Présentation des données issues de l'INE

Sexe	$E_x$ (Exposition)	x (Age)	$\hat{q}_x$ (Taux bruts)
MF	$9.9 \times 10^4$	23	0.00032
MF	$9.9 \times 10^4$	24	0.00031
MF	$9.9 \times 10^4$	25	0.0003
MF	$9.9 \times 10^4$	26	0.00029
MF	$9.9 \times 10^4$	27	0.00041
MF	$9.9 \times 10^4$	28	0.00042
MF	$9.9 \times 10^4$	29	0.00045
MF	$9.9 \times 10^4$	30	0.00037
MF	$9.9 \times 10^4$	31	0.00047
MF	$9.9 \times 10^4$	32	0.00044
MF	$9.9 \times 10^4$	33	0.00041
MF	$9.9 \times 10^4$	34	0.00053
MF	$9.9 \times 10^4$	35	0.00062
MF	$9.9 \times 10^4$	36	0.0007
MF	$9.9 \times 10^4$	37	0.00074
MF	$9.9 \times 10^4$	38	0.00078
MF	$9.9 \times 10^4$	39	0.00094
MF	$9.9 \times 10^4$	40	0.0011
MF	$9.8 \times 10^4$	41	0.0011
MF	$9.8 \times 10^4$	42	0.0014
MF	$9.8 \times 10^4$	43	0.0014
MF	$9.8 \times 10^4$	44	0.0017
MF	$9.8 \times 10^4$	45	0.0018
MF	$9.8 \times 10^4$	46	0.002
MF	$9.7 \times 10^4$	47	0.0023
MF	$9.7 \times 10^4$	48	0.0025
MF	$9.7 \times 10^4$	49	0.0028
MF	$9.7 \times 10^4$	50	0.003
MF	$9.6 \times 10^4$	51	0.0034
MF	$9.6 \times 10^4$	52	0.0037
MF	$9.6 \times 10^4$	53	0.004
MF	$9.5 \times 10^4$	54	0.0042
MF	$9.5 \times 10^4$	55	0.0046
MF	$9.5 \times 10^4$	56	0.0049
MF	$9.4 \times 10^4$	57	0.0055
MF	$9.4 \times 10^4$	58	0.0058
MF	$9.3 \times 10^4$	59	0.0062
MF	$9.3 \times 10^4$	60	0.0066
MF	$9.2 \times 10^4$	61	0.0072
MF	$9.1 \times 10^4$	62	0.0078
MF	$9.1 \times 10^4$	63	0.0086
MF	$9 \times 10^4$	64	0.0085
MF	$8.9 \times 10^4$	65	0.0092

**TABLE 3.10** – Données issues d'études nationales de mortalités de l'INE Portugaise filtrés avec le critère de Cochran et ré-équilibrés en fonction de la proportion homme-femme de notre portefeuille

---



# Annexe C

## Statistique descriptive des données étudiés

---

product_name	n	Valeurs manquantes	Valeurs distinctes
	73059	0	3
Valeurs	credit linked - consumer loan	credit linked - individual	
Fréquences	8402	26028	
Proportions	0.109	0.342	
Valeurs	credit linked - mortgage		
Fréquences	41780		
Proportions	0.548		

---

gender	n	Valeurs manquantes	Valeurs distinctes
	75685	525	2
Valeurs	female	male	
Fréquences	34716	40969	
Proportions	0.459	0.541	

---

age	n	Valeurs manquantes	Valeurs distinctes	Moyenne	.25	.50	.75
	76210	0	48	43.98	34	44	54

Les plus faibles : 17 18 19 | Les plus élevés : 77 78 79

---

actual_count	n	Valeurs manquantes	Valeurs distinctes	Moyenne
	76210	0	4	0.01716
Valeurs	0	1	2	3
Fréquences	74982	1186	40	2
Proportions	0.984	0.016	0.001	0.000

---

policy_duration_year	n	Valeurs manquantes	Valeurs distinctes	Moyenne	.25	.50	.75
	76210	0	18	6.318	3	5	10

Les plus faibles : 3 4 5 | Les plus élevés : 16 17 18

Valeurs	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Fréquences	8275	8424	8057	7221	6213	5525	5001	4343	3900	3793	3806	3812	3452	2558
Proportions	0.109	0.111	0.106	0.095	0.082	0.072	0.066	0.057	0.051	0.050	0.050	0.050	0.045	0.034

Value	15	16	17	18
Frequency	1366	388	69	7
Proportion	0.018	0.005	0.001	0.000

---

face_amount_upper_band	n	Valeurs manquantes	Valeurs distinctes	Moyenne
	76141	69	9	146355

Les plus faibles : 10000 25000 50000 | Les plus élevés : 1000000 2500000 5000000

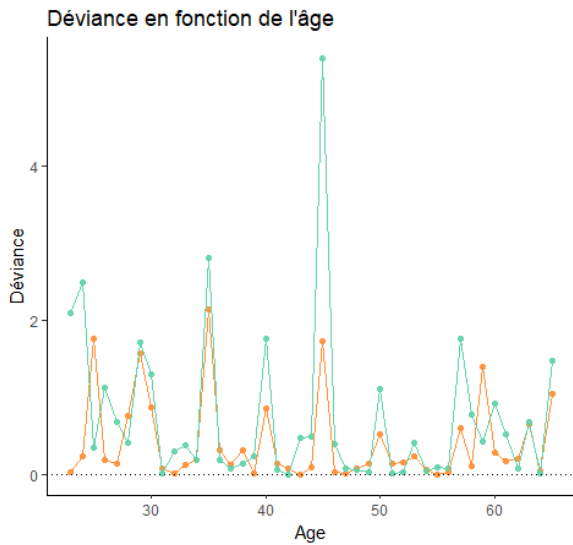
Valeurs	10000	25000	50000	100000	250000	500000	1000000	2500000	5000000
Fréquences	14100	15430	14847	13355	10460	5385	2199	317	48
Proportions	0.185	0.203	0.195	0.175	0.137	0.071	0.029	0.004	0.001

TABLE 3.11 – Statistiques descriptives des données du portefeuille

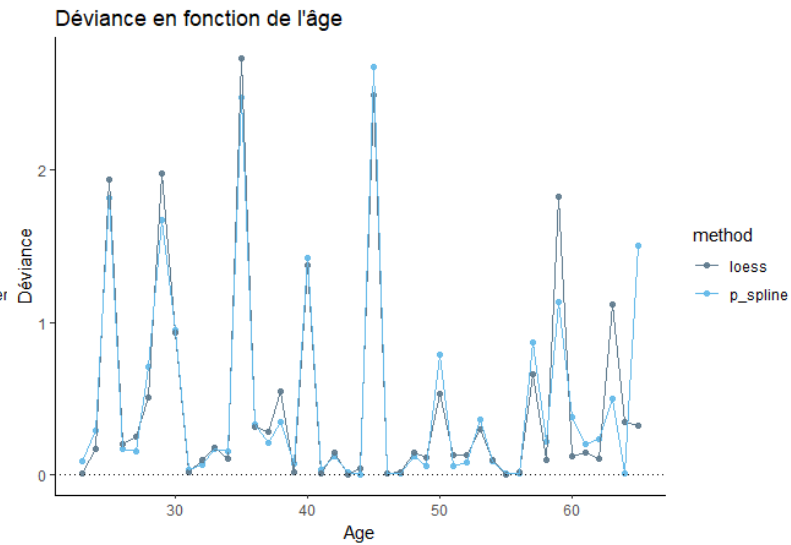
---

# Annexe D

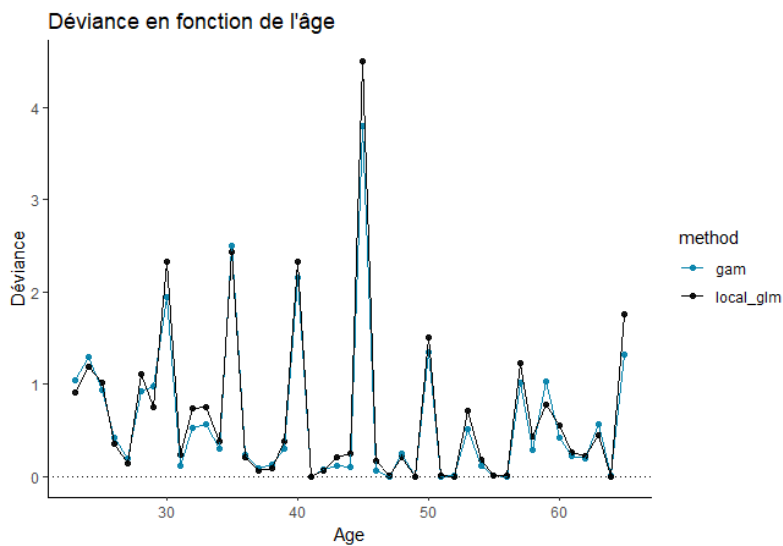
## Résultat des valeurs de déviance en 1 dimension



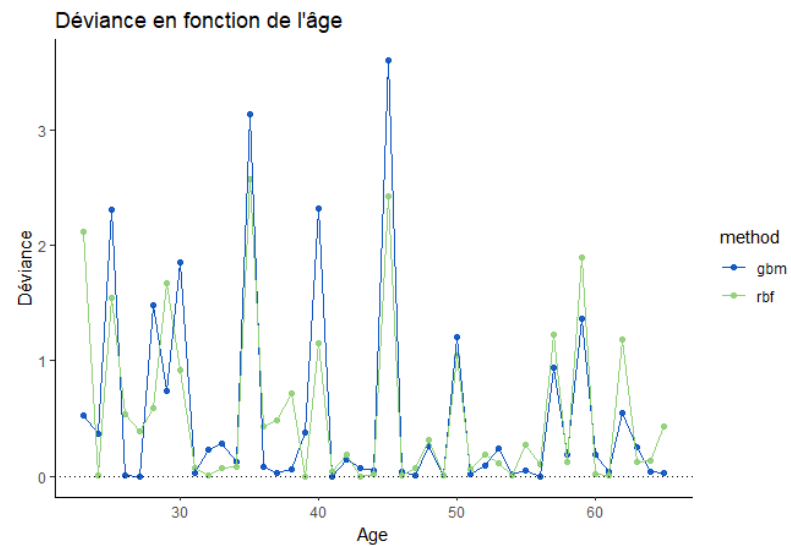
*Déviance<sub>x</sub> à partir des méthodes B-Spline/Whittaker Henderson*



*Déviance<sub>x</sub> à partir des méthodes P-Spline/LOESS*



*Déviance<sub>x</sub> à partir des méthodes GAM/GLM local*



*Déviance<sub>x</sub> à partir des méthodes RBF Network/GBM Spline*

**FIGURE 3.42** – *Comparaison des déviances pour les différentes méthodes de lissage*

---