



**Mémoire présenté  
devant l'Institut de Science Financière et d'Assurances  
pour l'obtention  
du diplôme d'Actuaire de l'Université de Lyon**

**le 27 janvier 2010**

Par : Vivien GUETTE

Titre: LA PRISE EN COMPTE DES CATASTROPHES DANS LA MODELISATION DE LA  
MORTALITE

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

*Membre du jury de l'Institut des Actuaires*

M. Christian FETTIG

*Entreprise :*

WINTER & Associés

*Membres du jury I.S.F.A.*

M. Jean Claude AUGROS

M. Alexis BIENVENÛE

Mme Diana DOROBANTU

Mme Anne EYRAUD-LOISEL

M. Jean-Paul LAURENT

M. Nicolas LEBOISNE

M. Stéphane LOISEL

Mlle Esterina MASIELLO

Mme Véronique MAUME-DESCHAMPS

M. Frédéric PLANCHET

M. François QUITTARD-PINON

Mme Béatrice REY-FOURNIER

M. Didier RULLIERE

*Directeur de mémoire en entreprise :*

Frédéric PLANCHET

*Invité :*

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

Signature du responsable entreprise

*Secrétariat*

Mme Marie-Claude MOUCHON

Signature du candidat

*Bibliothèque :*

Mme Michèle SONNIER

« Comment se fait-il que l'espèce humaine, non seulement aujourd'hui mais durant toute son histoire, néglige délibérément les informations dont elle dispose, et qui lui permettraient de s'épargner certaines catastrophes ? »

*Le Moine et le Philosophe (1997), Jean-François Revel*

## RESUME

---

**Mots clés :** *risque catastrophe, taux de surmortalité / choc de mortalité, mortalité stochastique, modèles à sauts, processus CIR, modèle de Lee-Carter, risque de pandémie, lissage, splines pénalisés, modèles linéaires généralisés matriciels, extrapolation de surfaces, ajustement paramétrique, simulation surface de chocs.*

Face au nombre de catastrophes en constante augmentation depuis quelques années, la capacité du marché de la réassurance traditionnelle s'essouffle. Les assureurs doivent désormais accorder plus d'importance à la prise en compte des chocs et plus particulièrement aux chocs de mortalité en assurance de personnes. Le contexte normatif actuel va dans ce sens : le projet européen Solvabilité 2 impose désormais l'utilisation de sous-modules « catastrophes » pour le calcul du SCR.

Cependant, les incohérences observées dans ce projet démontrent bien que l'évaluation du risque catastrophe est assez délicate. La plupart des études existantes, notamment celles axées sur l'étude des effets d'une pandémie sur la surmortalité, conduisent à des résultats hétérogènes. En effet, les taux de surmortalité calculés pour une pandémie de période de retour de 200 ans varient suivant les études entre 0,15 % et 0,61 %. Cela n'est guère étonnant dans la mesure où les principales approches utilisées (approche épidémiologique ou approche « actuarielle » basée sur des modèles de mortalité stochastique) sont toutes fondées sur des hypothèses et des calibrages auxquels sont très sensibles les résultats.

Ce mémoire a pour objectif de proposer une synthèse de ces principales approches, d'identifier leurs difficultés de mise en œuvre et de proposer un modèle original et robuste de modélisation et de simulation de chocs de mortalité. Pour ce faire, nous dresserons dans une première partie un panorama de la situation actuelle concernant la prise en compte du risque catastrophe en assurance et mettrons en évidence les incohérences observées dans le QIS4. Dans une deuxième partie, nous aborderons les principales approches de l'étude de la surmortalité suite à une pandémie au travers de la récente approche épidémiologique de Swiss Re et d'un modèle de mortalité stochastique à sauts. Après avoir identifié les avantages et inconvénients de chaque méthode, nous proposerons dans une troisième partie un modèle original de simulation de surfaces de mortalité et de surmortalité. Ce modèle utilisera des techniques de lissage par splines pénalisés et recourra à l'utilisation des modèles linéaires généralisés matriciels. On déterminera alors un taux de surmortalité additif pour une catastrophe ayant une période de retour de 200 ans s'élevant à 0,25 % pour la France, soit un nombre de décès supplémentaire de près de 160 000.

## ABSTRACT

---

**Key words:** *catastrophe risk, excess mortality rate / mortality shock, stochastic mortality, jump diffusion processes, CIR process, Lee-Carter model, pandemic risk, smoothing, penalized splines, generalized linear array model, surface extrapolation, parametric adjustment, shock surface simulation.*

The number of catastrophe is constantly increasing since a few years so that the reinsurance market capacity is running out of steam. From now on, insurers have to pay more attention to consider shocks and more particularly mortality shocks in life insurance. In a context of increased normative pressure, the European Solvency 2 project imposes the use of capital charge for catastrophe risk in the SCR calculation.

However, the inconsistencies of the project manifest the fact that the catastrophe risk valuation is very delicate. Most existing studies, in particular these focused on the effects of a pandemic on the excess of mortality, lead to a lot of different results. Indeed, the excess of mortality rate calculated by these studies varies between 0.15 % and 0.61 %. This is not very surprising insofar as the main approaches used (epidemiological or actuarial with stochastic mortality models) are based on hypothesis and calibration at which the sensitivity of the results is very high.

The purpose of this report is to propose a synthesis of the main approaches, to identify the difficulties in carrying out them and to propose a hardy and original model of shocks mortality modeling and forecasting. To do so, in a first part we will present a panorama of the current situation concerning the recognition of catastrophe risk in insurance and we will underline the inconsistencies observed in the QIS4. In a second part, we will address the main approaches of studying the excess mortality following a pandemic event over the recent epidemiological approach of Swiss Re and a stochastic model with jump diffusion process. After having identified the advantages and the disadvantages of each method, we will present in a third part an original model of mortality and excess mortality surface simulation. This model will use smoothing techniques with penalized splines and the theory of the generalized linear array model. We will calculate an additional excess mortality rate for a 1-in-200 years severity catastrophe of 0.25 % for the France.

## REMERCIEMENTS

---

Je tiens tout d'abord à remercier Frédéric PLANCHET, directeur de ce mémoire, pour ses nombreuses relectures critiques et pour le temps qu'il a pu m'accorder tout au long du travail de réflexion et de rédaction de ce mémoire.

Je remercie également Pierre THÉRON, mon manager au sein du cabinet WINTER & Associés, pour ses conseils et pour l'année passée sous son encadrement.

Je remercie enfin l'ISFA pour la qualité de sa formation reçue ces trois dernières années.

# SOMMAIRE

---

RESUME.....	2
ABSTRACT.....	3
REMERCIEMENTS.....	4
SOMMAIRE.....	5
INTRODUCTION.....	7
<i>PARTIE I : LE RISQUE CATASTROPHE EN ASSURANCE DE PERSONNES.....</i>	<i>9</i>
CHAPITRE 1. DEFINITION DU RISQUE CATASTROPHE EN ASSURANCE DE PERSONNES.....	10
Section 1.1. Quelques remarques sur l'assurance de personnes.....	10
Section 1.2. La notion de catastrophe.....	11
Section 1.3. Les risques extrêmes en mortalité.....	12
CHAPITRE 2. PRISE EN COMPTE ET CALIBRAGE DU RISQUE CATASTROPHE.....	16
Section 2.1. Les états réglementaires C8 et C9.....	16
Section 2.2. Calibrage du risque catastrophe en assurance de personnes dans le QIS4.....	19
CHAPITRE 3. MISE EN EVIDENCE DES INCOHERENCES ENTRE LES DIFFERENTS MODULES DU QIS4 ET PISTES DE REFLEXION.....	26
Section 3.1. Mise en évidence des incohérences entre les différents modules du QIS4.....	26
Section 3.2. Quelques commentaires et pistes de réflexion.....	28
<i>PARTIE II : LES PRINCIPALES APPROCHES DE MODELISATION DE LA MORTALITE EXTREME.....</i>	<i>30</i>
CHAPITRE 1. L'APPROCHE EPIDEMIOLOGIQUE.....	31
Section 1.1. Présentation générale.....	31
Section 1.2. L'étude Swiss Re.....	33
CHAPITRE 2. L'APPROCHE « ACTUARIELLE ».....	39
Section 2.1. Présentation générale.....	39
Section 2.2. L'approche de Bauer et Kramer.....	41
Section 2.3. Amélioration du modèle de Bauer et Kramer : tendance de type « Lee-Carter »....	63
<i>PARTIE III : UNE APPROCHE ORIGINALE : UN MODELE DE LISSAGE A CHOCS.....</i>	<i>70</i>
CHAPITRE 1. LE LISSAGE PAR SPLINES.....	71
Section 1.1. Quelques généralités.....	71
Section 1.2. Définition d'un B-spline.....	72

Section 1.3. Pénalités .....	74
CHAPITRE 2. RAPPELS SUR LES MODELES LINEAIRES ET EXTENSION .....	76
Section 2.1. Le modèle linéaire généralisé ( <i>GLM</i> ) .....	76
Section 2.2. Le modèle linéaire généralisé matriciel ( <i>GLAM</i> ).....	78
CHAPITRE 3. APPLICATION AU LISSAGE D'UNE COURBE ET D'UNE SURFACE DE MORTALITE .....	83
Section 3.1. Présentation des données.....	83
Section 3.2. L'utilisation de P-splines unidimensionnels .....	83
Section 3.3. L'utilisation de P-splines bidimensionnels .....	86
CHAPITRE 4. LISSAGE D'UNE SURFACE DE MORTALITE AVEC COMPOSANTE DE CHOCS .....	92
Section 4.1. Mise en œuvre .....	92
Section 4.2. Extrapolation de la surface de mortalité et détermination d'un quantile extrême ...	97
Section 4.3. Avantages et inconvénients .....	109
CONCLUSION .....	110
BIBLIOGRAPHIE .....	112
ANNEXES.....	116
ANNEXE 1. CALCUL DU TAUX DE SURVIE DU MODELE DE BAUER.....	116
ANNEXE 2. RESOLUTION D'UNE EQUATION DIFFERENTIELLE ORDINAIRE DE RICCATI.....	118
ANNEXE 3. MORTALITE AUX ETATS-UNIS ET EN EUROPE ENTRE 1918 ET 1919.....	120
ANNEXE 4. LOIS UTILISEES .....	120
ANNEXE 5. DETAILS DES AJUSTEMENTS PARAMETRIQUES.....	121

## INTRODUCTION

---

Le risque de mortalité est le risque principal auquel doivent faire face les assureurs dans le cadre de l'assurance de personnes pour des contrats qui offrent un capital en cas de décès. Si la mortalité réellement observée s'écarte de manière significative de la mortalité anticipée lors de la tarification d'un contrat, l'assureur peut théoriquement avoir à piocher dans ses fonds propres pour verser les capitaux promis avec comme risque ultime la ruine. Traditionnellement, l'assureur se souciait peu de ce type de phénomène et avait avant tout pour priorité de souscrire des risques à son bilan, de gérer son équilibre actif / passif et de pratiquer jusqu'à récemment des tarifs très prudents. L'assureur espérait faire face à ces éventuels pics de mortalité en mettant en place des programmes de réassurance qui avaient pour but d'écarter les pointes de sinistres.

Cependant, les risques qui peuvent affecter de manière significative la mortalité sont très nombreux. On peut citer entre autres les catastrophes naturelles, les pandémies ou encore les accidents technologiques. L'Histoire est pleine de ces événements qui ont augmenté de manière exceptionnelle la mortalité d'un pays. Certes aujourd'hui, les populations sont moins vulnérables et la sécurité beaucoup plus importante mais le nombre croissant d'événements majeurs survenus ces dernières années laisse à penser que les risques sont toujours présents (voir par exemple le rapport de la Fédération Internationale des Sociétés de la Croix Rouge (2007)). Ils sont même en pleine mutation du fait, entre autres, du réchauffement climatique, de l'industrialisation et de la technicisation des sociétés ainsi que de la concentration urbaine des populations. Rien qu'en 2009, on peut déjà dénombrer une quinzaine de catastrophes dont la tempête Klaus en Europe en janvier, la rupture d'une digue dans la banlieue de Jakarta en mars, le séisme dans les Abruzzes en avril sans compter la menace de la grippe A(H1N1). Dans ce contexte, la part des risques prise en charge par les réassureurs devient de plus en plus limitée. Bien que la gestion de ces risques commence aujourd'hui à être transmise aux marchés financiers par le biais de la titrisation, les assureurs ont tout intérêt à s'intéresser de plus près à une meilleure prise en compte des chocs (et particulièrement des chocs de mortalité) dans leurs modèles.

Le législateur en a lui aussi conscience si bien que la prise en compte du risque catastrophe apparaît de manière explicite dans le projet européen Solvabilité 2. Les assureurs qui appliqueront la formule standard devront prendre en compte dans leurs calculs un sous-module « catastrophe ». Cependant, les propositions faites par le QIS4 sur ce point sont insuffisantes à l'heure actuelle. Le calibrage du risque catastrophe en santé est largement excessif. Le CEIOPS propose en effet une majoration de 10 % des primes alors que les mutuelles santé tablent plutôt sur un taux de majoration compris entre 2,7 % et 3,9 % (voir l'étude de la Mutualité Française (2008)). Les calibrages sont aussi incohérents entre le risque catastrophe en santé et le risque catastrophe en vie avec pourtant le même sous-jacent qu'est la pandémie. De plus, les préconisations actuelles dans le cadre du modèle interne sont très floues et ne fournissent aucune orientation précise sur la manière de modéliser ces risques catastrophiques.

Les études sur l'impact d'une catastrophe en termes de mortalité et les modèles développés par les assureurs et les réassureurs pour prendre en compte ces catastrophes sont déjà nombreux (on peut citer les études de Swiss Re (2007), Gen Re (2006), INVS (2005)). Cependant, leur mise en œuvre reste le plus souvent confidentielle et les résultats qui en découlent sont souvent relativement différents d'une étude à l'autre.

Dans ce mémoire, on se propose de s'intéresser de plus près aux approches qui existent actuellement pour modéliser la surmortalité. Une étude synthétique de ces approches permettra de déceler les avantages et inconvénients de chaque méthode et aussi de comprendre d'où vient l'hétérogénéité des résultats. L'objectif du mémoire est également de proposer une approche originale de prise en compte des chocs dans la modélisation de la mortalité. L'approche recherchée doit permettre de produire des résultats robustes et d'éviter les inconvénients des méthodes présentées par ailleurs.

Ce mémoire s'articule autour de trois parties principales.

La première partie est consacrée à la manière dont est appréhendé le risque catastrophe dans nos sociétés modernes, que ce soit d'un point de vue juridique dans le Code des Assurances et les états règlementaires ou d'un point de vue normatif avec l'examen du projet de Solvabilité 2. Nous mettrons en évidence les incohérences observées dans le QIS4 entre les différents modules « catastrophe » en vie et en santé.

La deuxième partie s'intéresse de plus près aux principales approches qui existent pour modéliser la mortalité extrême à savoir l'approche épidémiologique et l'approche « actuarielle ». La récente étude de Swiss Re est une très bonne illustration d'un modèle épidémiologique et est aujourd'hui une référence (c'est en partie sur ces résultats que le CEIOPS a calibré le risque catastrophe dans Solvabilité 2). L'approche actuarielle se base quant à elle sur des modèles de mortalité stochastique. Nous étudierons de plus près le modèle proposé par Bauer et Kramer auquel nous apporterons une amélioration. Pour ce type de modèle, on remarquera que la sensibilité des résultats à l'épidémie de grippe espagnole de 1918 est très élevée.

La dernière partie est la plus technique ce travail. Nous présenterons tout d'abord des techniques de lissage par splines puis dans un second chapitre une extension des modèles linéaires généralisés que sont les modèles linéaires généralisés matriciels et nous mettrons en évidence leur utilité dans la construction de modèles avec un grand nombre de données et de paramètres. L'utilisation des techniques de lissage par splines et des modèles linéaires généralisés matriciels permettront d'aborder une approche originale de modélisation de la mortalité. Cette approche a pour caractéristique d'isoler dans deux composantes une surface de mortalité lisse jouant le rôle de la tendance ainsi qu'une surface de chocs. L'extrapolation de ces deux surfaces et le calcul de quantile à 99,5 % seront aussi présentés. Pour la France, nous serons en mesure de déterminer un taux de surmortalité additif de 0,25 % pour une catastrophe ayant une période de retour de 200 ans.

## ***PARTIE I :***

### ***LE RISQUE CATASTROPHE EN ASSURANCE DE PERSONNES***

Dans cette partie, nous allons présenter de quelle manière est appréhendé le risque catastrophe en assurance (par le biais, entre autres, des états réglementaires C8 et C9) et plus particulièrement en assurance de personnes dans le QIS4. Nous mettrons également en évidence quelques incohérences que l'on retrouve dans le QIS4 qui traduisent la difficulté de quantifier le risque de surmortalité.

## **CHAPITRE 1. DEFINITION DU RISQUE CATASTROPHE EN ASSURANCE DE PERSONNES**

Dans ce chapitre, on se propose de présenter succinctement quelques généralités sur l'assurance de personnes ainsi qu'un éclaircissement sur la notion de catastrophe en assurance.

### **Section 1.1. Quelques remarques sur l'assurance de personnes**

Communément, on appelle assurance de personnes tout type d'assurance lié à un risque qui pourrait porter atteinte à la vie, la santé ou l'intégrité physique d'un assuré. L'assurance de personnes englobe donc l'assurance vie, l'assurance maladie, l'assurance invalidité, l'assurance accident et l'assurance dépendance.

Le marché total de l'assurance française en 2008 représentait 183,3 milliards d'euros en termes de cotisations (affaires directes) répartis en 75,6 % pour l'assurance de personnes et en 24,4 % pour l'assurance dommages. L'assurance vie représentait à elle seule 66,8 % du marché (chiffres issus du site internet de la Fédération Française des Sociétés d'Assurances). L'assurance vie représente aujourd'hui près de 35 % du patrimoine financier des ménages si bien qu'elle peut être considérée comme un outil de financement de l'économie et de préparation à la retraite. Cependant, du fait de l'aversion au risque croissante des ménages et de la baisse du pouvoir d'achat, la tendance est plutôt à la baisse depuis 2007. En cause aussi le contexte financier particulièrement perturbé pour l'assurance vie à vocation d'épargne ou de retraite. Cependant, du côté de la prévoyance (décès, maladie, incapacité, invalidité et dépendance) les cotisations ont continué à croître. L'assurance de personnes conserve donc encore de très loin la part la plus importante sur le marché de l'assurance.

Dans la suite de ce mémoire, on s'intéressera principalement aux contrats pour lesquels il existe une garantie en cas de décès. D'un point de vue un peu plus technique, si l'on s'intéresse donc de plus près à une assurance de type « vie entière », « temporaire décès » ou encore à une assurance emprunteur, l'évènement déclencheur du versement de la prestation par l'assureur est le décès de l'assuré. Afin d'équilibrer ses revenus (prime) et ses dépenses (prestations), l'assureur mutualise les primes qu'il reçoit au sein d'une réserve qu'il place sur les marchés financiers. Lorsqu'un décès survient, l'assureur finance la prestation qu'il doit verser aux ayant droits en puisant dans cette réserve et en vendant les actifs correspondants. Cet équilibre peut-être rompu par le biais de deux risques principaux à savoir le risque financier et le risque viager. En effet, si le rendement financier des placements est inférieur au rendement anticipé ou si la mortalité réelle augmente par rapport à la mortalité anticipée à la souscription du contrat, l'assureur devra piocher plus que prévu dans ses placements. Dans les cas les plus extrêmes, si les placements ne suffisent plus pour financer le règlement de tous les sinistres, l'assureur doit prélever les sommes dues sur ses fonds propres avec comme risque ultime la ruine.

Précisons que dans la suite, nous ne nous intéresserons qu'au risque de passif, c'est-à-dire au risque de déviation de la mortalité effective par rapport à la mortalité théorique déterminée en amont de la réalisation du risque. Théoriquement, une déviation très importante de la mortalité peut donc entraîner la ruine d'un assureur. Les événements qui peuvent entraîner une telle déviation sont nombreux et sont souvent qualifiés de « catastrophes » dans le langage courant lorsqu'ils surviennent. Nous allons maintenant apporter quelques éclaircissements sur la notion de catastrophe dans la section suivante.

## Section 1.2. La notion de catastrophe

Dans nos sociétés modernes, le risque est devenu omniprésent et multiforme. Il peut être naturel ou technologique, lié à l'aménagement de l'espace, aux enjeux sociaux et économiques, financiers ou politiques. Certes, les populations sont moins vulnérables qu'autrefois, même si cette affirmation mérite d'être nuancée. En effet, dans son livre *Economie des extrêmes*<sup>1</sup>, Daniel Zajdenweber souligne avec justesse que la crainte de la population et des assureurs quant à une aggravation des conséquences des catastrophes ne doit pas être simplement justifiée par le fait que leur nombre augmente mais aussi par le fait que le nombre d'assuré augmente également. On pourrait rajouter que l'urbanisation et que la concentration des populations dans les grandes villes augmentent l'exposition au risque. Déjà dans sa *Lettre sur la Providence* (1756), Rousseau répond à Voltaire après le tremblement de terre de Lisbonne de 1755 en écrivant ceci : « *Convenez que la nature n'avait point rassemblé là 20 000 maisons de six à sept étages, et que si les habitants de cette grande ville eussent été dispersés plus également et plus légèrement logés, le dégât eût été beaucoup moindre et peut-être nul* ».

Les citoyens modernes ont des exigences de sécurité et de refus des risques de plus en plus importantes. Cependant, ces exigences n'empêchent pas que parfois, un risque latent se cristallise à l'occasion d'une crise que l'on appellera catastrophe. La définition d'une catastrophe est assez délicate. Les catastrophes n'ont pas encore fait leur apparition dans les dictionnaires de vocabulaire juridique et l'approche médiatique ne fait que troubler la réflexion. En effet, les médias utilisent trop souvent ce terme pour désigner des événements qui ne mériteraient pas forcément ce qualificatif d'un point de vue juridique ou assurantiel. Pour le langage courant, la définition donnée par le Conseil National de l'aide aux Victimes est plutôt bonne : une catastrophe est un « *événement soudain qui provoque directement des atteintes graves à la personne de plusieurs victimes ou à des biens et dont l'origine peut être un phénomène naturel, une intervention humaine ou la combinaison des deux* ». Le crash d'un avion est un exemple de catastrophe typique qui va dans le sens de la définition précédente : il se produit soudainement à un lieu et à un instant donné.

Si l'on regarde de plus près le droit des assurances, on se rendra compte que diverses catégories existent pour cette dénomination trop large et médiatique. Ainsi, un événement qualifié ou non de catastrophe dans le langage courant et ayant des effets directs ou

---

<sup>1</sup> Voir Zajdenweber (2000)

indirects sur les personnes peut constituer pour le droit des assurances indifféremment une catastrophe naturelle de la loi de 1982, un attentat, un dommage causé par les effets du vent, une calamité agricole ou encore un sinistre d'ampleur exceptionnelle engendrant une indemnisation fondée sur le dommage causé (art. 1382 du Code Civil). C'est sans cynisme que Bertrand Pauvert, membre du CERDACC, (Centre Européen de Recherches sur le Droit des Accidents Collectifs et des Catastrophes) fait la remarque suivante (voir Pauvert (2003)) : au titre du droit des assurances, une catastrophe ne mérite ce qualificatif qu'à partir du moment où elle génère un coût beaucoup trop élevé pour la société d'assurance. Si l'on reprend l'exemple du crash d'avion précédent, pour un assureur n'ayant que deux ou trois passagers victimes possédant un contrat avec une garantie en cas de décès, cet évènement ne sera pas considéré comme étant une catastrophe au sens assurantiel.

La prise en compte des dommages survenus lors de catastrophes ne traite pas de manière égalitaire les deux branches principales de l'assurance que sont l'assurance de personnes et l'assurance dommages. Il semble en effet que cette prise en compte soit très satisfaisante pour les assurances relatives à la couverture des biens et de la responsabilité, cependant elle demeure insuffisante pour les assurances de personnes. En effet, dans le cadre d'une catastrophe industrielle ou technologique, l'indemnisation des personnes victimes des conséquences d'une catastrophe de ce type réside dans la mise en jeu de l'assurance responsabilité civile des responsables de la catastrophe si un responsable peut être déterminé. Cependant, si un grand nombre d'individus est touché, l'entreprise juridiquement responsable peut très rapidement être dans l'impossibilité d'indemniser les victimes. En ce qui concerne les catastrophes naturelles, seuls les biens se trouvent généralement convenablement pris en charge alors qu'une personne victime d'une telle catastrophe ne dispose d'aucun moyen direct de bénéficier d'une indemnisation, si ce n'est en recherchant à mettre en cause un responsable administratif ou politique. Le seul système qui donne actuellement satisfaction concerne le mécanisme d'indemnisation des victimes d'accidents par recours à un fonds dans le cadre d'un attentat. On notera aussi que la Fédération Française des Sociétés d'Assurances a élaboré un produit « Garanti des Accidents de la Vie » qui permet d'indemniser la personne assurée qui a subi une catastrophe.

On retiendra donc que du point de vue assurantiel, ne constituent des catastrophes que les évènements qui entraînent un coût anormalement élevé pour l'assureur. Dans le cas de garanties en cas de décès, une catastrophe sera par conséquent un évènement qui entraîne un nombre anormalement élevé de décès sur une période très courte.

### **Section 1.3. Les risques extrêmes en mortalité**

Dans cette section, on se propose d'identifier tous les risques qui peuvent faire varier de manière significative ou non le nombre de décès observés dans le portefeuille d'un assureur.

Le risque qui pèse sur la mortalité peut ou ne peut pas être *a priori* contrôlé par l'assureur. En ce qui concerne les risques contrôlables, on distingue :

- Le risque d'auto sélection : un assuré a le comportement classique d'un individu qui cherche à maximiser l'utilité de son contrat. Ainsi, des garanties en cas de décès peuvent attirer des personnes en mauvaise santé et par conséquent augmenter la sinistralité d'un portefeuille d'assurés. Ce risque est cependant contrôlé par l'assureur qui impose généralement une sélection médicale à la souscription et le choix d'une table de mortalité prudente pour calculer les primes ;
- Le risque de fraude : ce risque est relativement bien maîtrisé puisque les preuves et les conditions du décès doivent satisfaire les clauses du contrat ;
- Les comportements à risque : la pratique d'activités dangereuses (sports aériens, sports mécaniques) ou les professions dangereuses (pompiers, militaires, policiers) sont généralement exclues du contrat ou peuvent faire l'objet d'une prime majorée.

Pour ce qui est des risques incontrôlables, on peut répertorier :

- Le risque structurel (risque de masse) : la mortalité augmente brutalement au sein de la population assurée du fait de facteurs qui touchent indistinctement toutes les catégories de cette population (pandémies par exemple) ;
- Le risque lié à un évènement (risque de concentration) : la mortalité augmente ponctuellement sur un groupe de tête *a priori* indépendant au sein de la population assurée, du fait de la survenance d'évènements naturels (tsunami, tremblements de terre, vague de chaleur), accidentels (crash d'avion, explosion d'usine) ou encore d'attaques terroristes.

Ces deux derniers risques incontrôlables sont particuliers dans la mesure où ils peuvent entraîner des chocs de mortalité. Un choc de mortalité est caractérisé par une très courte période brutale présentant une mortalité anormalement élevée. On distingue :

- Les pandémies : historiquement, plusieurs types de pathologies ont très fortement touché l'humanité à savoir la peste, le choléra, la variole ou encore la tuberculose. Aujourd'hui, le virus de la grippe reste un des rares virus à toujours rester potentiellement endémique. Les spécialistes ont recensé « officiellement » 31 épidémies de grippe depuis 1580 (bien que les archives témoignent de décès anormalement élevés en 412 avant J.C, au Moyen-âge au retour des croisades et pendant la Guerre de Cent Ans). Selon l'agence de notation Moody's (2007), pour l'industrie d'assurance aux Etats-Unis : « *In a severe pandemic influenza, the projected results for the industry could prove devastating, resulting in the elimination of almost a quarter of its statutory surplus on a gross basis as well as downgrades (multi-notch) of a significant number of life insurers / reinsurers* ». La grippe espagnole de 1918 est aujourd'hui une référence. Le plan d'épidémie<sup>2</sup> du US department of health and human services de novembre 2005 estime que :

---

<sup>2</sup> Voir HHS Pandemic Influenza Plan (2005)

« *uncertainty about the magnitude of the next pandemic requires plans to be made to cater for a severe pandemic such as 1918* » ;

- Les catastrophes naturelles : les phénomènes atmosphériques (ouragans, tornades et typhons) sont aujourd'hui de plus en plus nombreux car en lien avec le réchauffement climatique. De plus, l'urbanisation poussée de certaines zones sujettes à tremblements de terre (Californie et Japon) est un facteur qui expose les assureurs.

Voici une liste des évènements de surmortalité liés à des phénomènes naturels :

Evènement	Pays	Année	Nombre de décès
Inondations (Yang Tsé Kiang)	Chine	1933	3 700 000
Sécheresse	Inde	1942	1 500 000
Inondations (Honan)	Chine	1939	500 000
Tremblement de terre (plaine du Kanto)	Japon	1923	142 000
Tremblement de terre	Chine	1933	9 300
Ouragan (Galveston)	Etats-Unis	1900	6 000-12 000
Tempête	Japon	1959	5 098

Tableau 1 - Liste des principaux phénomènes naturels antérieurs à 1980

Evènement	Pays	Année	Nombre de décès
Tsunami (Sumatra)	Asie	2004	225 000
Vague de chaleur	France	2003	15 000
Tremblement de terre (Kobe)	Japon	1995	5 400
Ouragan (Katrina)	Etats-Unis	2005	1 836
Vague de chaleur	Etats-Unis	1995	670
Tornade	Etats-Unis	1984	600
Blizzard	Etats-Unis	1983	500
Inondations	Japon	1982	345

Tableau 2 - Liste des principaux phénomènes naturels postérieurs à 1980

(source : site internet *Death Toll from Disasters, War & Accidents*)

- Le crash de météorites : il y a 65 millions d'année, l'extinction des dinosaures aurait été provoquée par le crash d'une météorite. Plus récemment, différents crashes ont été répertoriés (la puissance est donnée en kilotonnes de TNT. A titre d'exemple, la puissance de l'explosion de la bombe d'Hiroshima a été de 15 kT).

Evènement	Pays	Année	Puissance (en kT)
Crash d'une météorite de 60 mètres de diamètre	Sibérie orientale	1908	15 000
Crash d'une météorite	Sibérie orientale (Shikhote Alin)	1947	20
Crash d'une météorite	Sibérie orientale (région d'Irkoutsk)	2002	205

Tableau 3 - Liste des crashes de météorites

(source : Bourgeois B. et Pougnet G. (2008))

- Les accidents industriels : malgré un fort degré de prévention et de protection, l'accident industriel est toujours une éventualité. Tchernobyl est bien sûr la catastrophe nucléaire la plus importante de l'histoire humaine. Selon des chiffres officiels certainement sous estimés, l'accident a provoqué 1 000 décès et a augmenté de 2 % le taux de cancer dans la région ;
- Les guerres : elles ont été relativement fréquentes dans l'Histoire et aujourd'hui encore, de très fortes tensions subsistent dans certaines régions du globe (Moyen-Orient, Inde, Pakistan, Corée du Nord) avec des pays qui possèdent l'arme nucléaire ;
- Les attaques terroristes : elles sont aujourd'hui le moyen d'action privilégié de certains groupes radicaux. Depuis l'attaque du World Trade Center du 11 septembre 2001 (qui a occasionné 3 000 décès), la menace apparaît désormais mondiale ;
- Les risques émergents : d'autres phénomènes émergents peuvent à l'avenir augmenter de manière significative le niveau de mortalité. On peut citer entre autres les phénomènes d'obésité et de diabète, les revers technologiques possibles (champs électromagnétiques, OGM, changement climatique)...

Nous avons redéfini dans ce chapitre le contexte dans lequel s'inscrit l'assurance de personnes et nous avons apporté quelques précisions concernant la notion de catastrophe en assurance et l'ensemble des risques qui peuvent affecter la mortalité. Dans le chapitre suivant, nous allons décrire de quelle manière sont appréhendés ces risques dans le Code des Assurances et dans le cadre du projet Solvabilité 2.

## CHAPITRE 2. PRISE EN COMPTE ET CALIBRAGE DU RISQUE CATASTROPHE

Dans ce chapitre, nous présentons tout d'abord les états réglementaires C8 et C9 qui concernent la couverture de réassurance des assureurs. L'état C9 est plus particulièrement tourné vers la simulation de charges de sinistres lors de la survenance de catastrophes. Nous présenterons ensuite les recommandations actuelles du projet Solvabilité 2 en termes de prise en compte du risque catastrophe en assurance de personnes. Nous regarderons de plus près de quelle façon sont calibrés les coefficients de majoration utilisés.

### Section 2.1. Les états réglementaires C8 et C9

#### 2.1.1. Contexte et présentation

Une compagnie d'assurance est soumise à diverses obligations réglementaires en matière d'information financière. Elle doit remettre annuellement à l'Autorité de Contrôle des Assurances et des Mutuelles certains documents. Dans les quatre mois qui suivent la date d'arrêté des comptes, elle doit remettre :

- l'état des renseignements généraux ;
- le bilan, le compte de résultat et les annexes ;
- les états d'analyse des comptes ;
- un rapport de gestion du chiffre d'affaires ;
- le rapport du commissaire aux comptes ;
- de nouveau le bilan, le compte de résultat et les annexes.

Au plus tard le 30 juin de chaque année, elle doit fournir un rapport de solvabilité, un rapport sur le contrôle interne et un rapport sur la politique de réassurance.

Parmi les états d'analyse des comptes, on trouve les états réglementaires C8 et C9 dont la forme est standardisée. Ces états ainsi que le rapport sur la politique de réassurance ont pour but d'amener les entreprises à s'interroger chaque année sur la pertinence de leur plan de réassurance en constatant dans plusieurs scénarios extrêmes le degré de protection que leur confèrent les couvertures allouées. Ils doivent également conduire à s'interroger sur le risque de contrepartie auquel elles s'exposent en se réassurant. L'état C8 permet d'apporter une vue d'ensemble du programme de réassurance. Pour chaque catégorie de risque figurant dans une liste précisée à l'annexe de l'article A. 344-10 du Code des Assurances, l'entreprise doit renseigner plusieurs tableaux décrivant son programme de réassurance pour la catégorie considérée. L'état C9 quant à lui traite de la dispersion des provisions techniques cédées et rétrocédées et du niveau de protection que confèrent à l'entreprise ses couvertures en réassurance.

## 2.1.2. Principes de renseignement des états C8 et C9

### 2.1.2.1. L'état C8

Chaque état C8 est constitué d'un tableau décrivant les couvertures proportionnelles (tableau A), d'un tableau portant sur les couvertures non proportionnelles par risque (tableau B), d'un tableau traitant des couvertures non proportionnelles par évènement (tableau C), d'un tableau de synthèse des couvertures par risque et des couvertures par évènement (tableau D) et d'un tableau décrivant les couvertures en excédent de perte (tableau E). Si l'entreprise souscrit des affaires directes dans un type de risque figurant dans la liste précisée à l'annexe de l'article A. 344-10 du Code des Assurance sans pour autant se réassureur, elle doit tout de même adresser à l'ACAM l'état C8 en ne remplissant que le tableau D.

### 2.1.2.2. L'état C9

L'état C9 est composé d'un tableau portant sur la dispersion des cessionnaires et rétrocessionnaires de l'entreprise (tableau A) et d'un tableau permettant de comparer la charge brute pour l'entreprise à la charge nette de réassurance en cas de survenance d'évènements défavorables (tableau B). Intéressons-nous plus particulièrement au tableau B. Il se présente de la manière suivante :

B – Simulation d'évènements		Charge de sinistre	
		Brute	Nette
Pire évènement survenu (1)	21		
Tempêtes Lothar et Martin (2)	22		
Evènement centenaire (3) Tempête-Ouragan-Cyclone (4)	23		
Evènement centenaire (3) inondations (4)	24		
Evènement centenaire (3) cataclysme (4)	25		
Evènement majeur responsabilité civile (5)	26		
Evènement majeur accident technologique (6)	27		
Evènement majeur épidémie (7)	28		
Evènement majeur garantie plancher (8)	29		

Tableau 4 - Tableau B de l'état C9

(1) Evènement qu'a connu l'entreprise dans le passé et qui conduirait, s'il survenait dans l'exercice en cours, à la charge de sinistres brute de réassurance la plus importante,

compte tenu de l'actuel portefeuille de risques de l'entreprise. La charge nette doit tenir compte des couvertures actuelles en réassurance ;

(2) Les charges de sinistres simulées relatives à ces deux événements (tempêtes du 26 et du 27 décembre 1999) doivent tenir compte de l'actuel portefeuille de risques de l'entreprise ainsi que de ses actuelles couvertures en réassurance ;

(3) Événement dont la période de retour, au regard du portefeuille de risques de l'entreprise, est égale à 100 années et dont la charge de sinistre brute de réassurance est la plus élevée ;

(4) Sont à exclure les risques de la société qui sont cédés de manière illimitée à la Caisse centrale de réassurance avec la garantie de l'Etat ;

(5) Il s'agit d'un scénario défavorable concernant le risque de responsabilité civile et utilisé par l'entreprise pour établir et analyser son programme de réassurance (sinistre de type « AZF » par exemple) ;

(6) Il s'agit d'un scénario défavorable de type accidents technologiques utilisé par l'entreprise pour établir et analyser son programme de réassurance ;

(7) Il s'agit d'un scénario défavorable de type épidémie utilisé par l'entreprise pour établir et analyser son programme de réassurance ;

(8) Il s'agit d'un scénario défini par la combinaison d'hypothèses financières standardisées. Ces hypothèses consistent, par rapport à leur moyenne respective constatée sur les trois dernières années, en une baisse de l'indice boursier de référence de 30 %, en une baisse de deux points des taux d'intérêt de l'obligation de référence et en une baisse de 20 % du prix des transactions immobilières. La charge de sinistres à renseigner correspond ici à la valeur actuelle probable, calculée au 1<sup>er</sup> janvier de l'exercice en cours, des prestations (nettes des prélèvements effectués au titre de ces garanties) associées aux garanties plancher jusqu'à leur extinction sous les hypothèses financières précédentes. Le taux d'actualisation à retenir est égal au minimum entre 3,5 % et 60 % du TME. Entre la table de mortalité TH 00-02 et la table TF 00-02 doit être retenue celle donnant la valeur actuelle probable des prestations la plus élevée. Pour le calcul de la charge nette sont prises en compte les primes cédées et les prestations cédées, en valeur actuelle probable, et sont appliquées les conditions des traités en vigueur, notamment celles concernant la durée de la garantie, sans tenir compte des renouvellements éventuels de ces traités.

Les scénarios et les méthodes utilisés par les entreprises pour réaliser ces simulations doivent être décrits dans le rapport de réassurance. Cependant, aucune préconisation exacte ni aucune pratique standard n'existent. Cela entraîne d'ailleurs beaucoup de difficultés pour certaines compagnies. En effet, dans son rapport de 2008 « Renouvellement des programmes de réassurance Catastrophe Décès Invalidité en France », le courtier en réassurance Benfield remarque que parmi les cédantes qui ont bien voulu répondre à son questionnaire, une majorité d'entre elles ont des difficultés à estimer leur rétention en cas

d'évènement majeur de type épidémie, du fait entre autres de la difficulté de simuler l'effet d'un tel évènement sur la mortalité de leur portefeuille. On notera d'ailleurs au passage que selon cette étude, 15 % des acteurs du marché de l'assurance (mutuelles, institutions de prévoyance et compagnies d'assurance) ne possèdent pas de couverture catastrophe.

Nous avons vu dans cette section que dans la réglementation actuelle, la prise en compte d'évènements catastrophiques est très bien encadrée sur la forme mais très peu sur le fond ce qui n'est pas sans poser de difficultés à certaines compagnies d'assurance qui ont du mal à déterminer leur exposition dans le cas d'évènements majeurs tels que les épidémies. Comme nous allons le voir, la logique du projet de Solvabilité 2 est toute autre. Les risques à prendre en compte sont explicitement mentionnés et la manière de les appréhender précisément encadrée.

## **Section 2.2. Calibrage du risque catastrophe en assurance de personnes dans le QIS4**

### *2.2.1. Quelques rappels sur le projet Solvabilité 2*

Rappelons brièvement dans quel contexte s'inscrit la directive européenne Solvabilité 2.

A la suite du projet Solvabilité 1 qui était consacré à la marge de solvabilité, le projet Solvabilité 2 a pour but de mettre à jour le système de solvabilité européen en intégrant tous les risques qui pèsent de manière effective sur les sociétés d'assurance.

Le projet Solvabilité 2 repose sur trois piliers :

- le premier est l'exigence de fonds propres. Il a pour rôle de mettre en place des seuils de fonds propres et de provisions pour que les organismes assureurs puissent être solvables ;
- le deuxième est que les organismes assureurs puissent être mieux contrôlés par l'autorité de contrôle qui pourra infliger des sanctions en cas de non respect des normes ;
- le dernier est l'obligation pour les organismes assureurs de publier certaines informations et de les détailler à l'autorité de contrôle.

Pour plus de détails concernant l'architecture du projet Solvabilité 2, le lecteur intéressé pourra consulter la section Assurances du site internet de la Commission Européenne ou encore la première partie de Sauvet (2006).

Un dialogue s'est instauré entre la Commission Européenne et le CEIOPS (Comité Européen des Contrôleurs d'Assurance et des Pensions Professionnelles : Committee of European Insurance and Occupational Pensions Supervisors). Plusieurs séries de tests ont été élaborées pour valider différents corps de règles d'évaluation des risques pris par les

organismes assureurs. En effet, la Commission Européenne a demandé au CEIOPS d'étudier les répercussions quantitatives du nouveau système par le biais d'études quantitatives d'impact (QIS, Quantitative Impact Studies). Les QIS permettent de recueillir des retours qualitatifs et quantitatifs auprès des acteurs du marché. Il s'agit d'un véritable aller et retour entre théorie et pratique :

- les QIS sont la traduction concrète des orientations incluses dans les réponses du CEIOPS aux demandes d'avis de la Commission ;
- l'exploitation des QIS permet au CEIOPS de faire évoluer et d'améliorer ses réponses initiales.

Les QIS successifs ont introduit entre autres des recommandations pour le calcul du SCR (*Solvency Capital Requirement*) qui représente le capital minimum exigé pour garantir la solvabilité de l'assureur à un niveau satisfaisant. Cette quantité est calculée en agrégeant différents sous-modules de risque : le risque opérationnel, le risque de marché, le risque de contrepartie, le risque de souscription vie, le risque de souscription santé et le risque de souscription non-vie. La dernière étude d'impact (QIS4) a été publiée en mars 2008 par le CEIOPS. En novembre 2008, le CEIOPS a publié un compte-rendu des réponses apportées à cette étude par les organismes assureurs. Voici un schéma récapitulatif de l'agencement des modules dans le QIS4. Les modules entourés de rouge sont ceux qui nous intéressent et qui seront étudiés plus en détail :

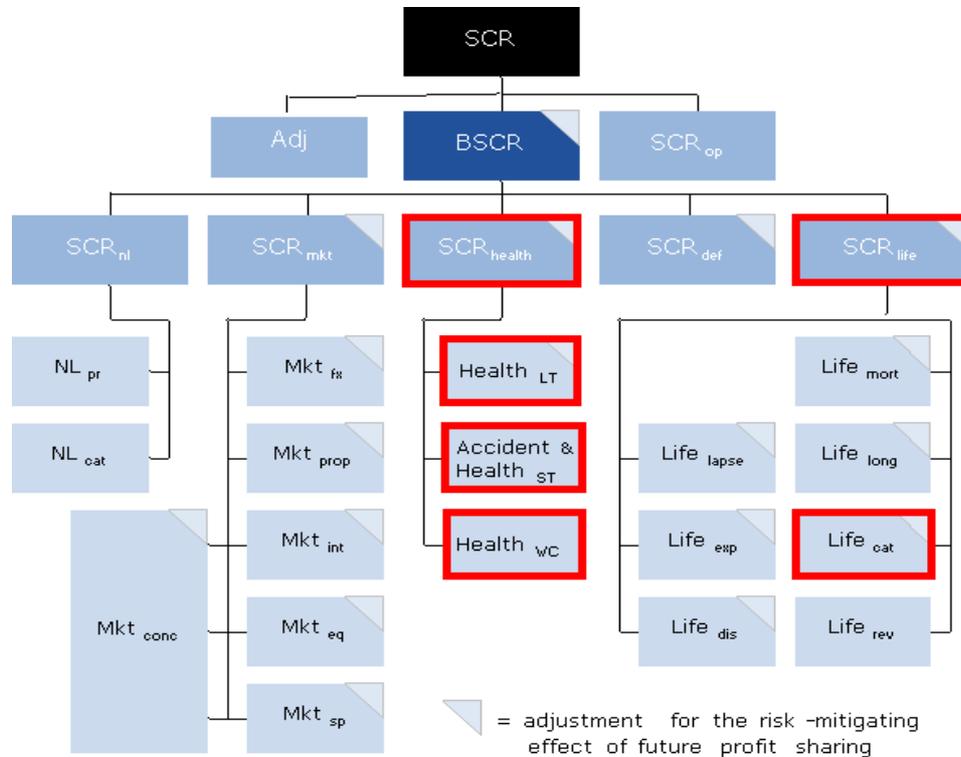


Schéma 1 - Agencement des modules dans le QIS4

Dans la suite nous allons nous intéresser au risque catastrophe ayant un impact sur l'activité d'assurance de personnes. Nous nous intéresserons donc seulement aux modules risque de souscription vie et risque de souscription santé. Bien que comme nous l'avons précisé au début du mémoire nous allons nous intéresser surtout au risque en cas de décès et aux situations de surmortalité, nous allons tout de même détailler dans cette partie les préconisations du QIS4 en santé pour mettre en évidence les incohérences que l'on peut relever.

### 2.2.2. Le module risque de souscription santé

Ce module est composé des sous-modules *Accident & Health Short Term* (accident et santé court terme), *Health Long Term* (santé long terme) et *Health Workers Compensation* (accidents du travail).

#### 2.2.2.1. Risque catastrophe en assurance santé au premier euro (Health Long Term)

Ce module concerne le risque de souscription associé à l'assurance santé pratiquée sur une base technique similaire à celle de l'assurance vie et assortie de restrictions supplémentaires prévues par la législation nationale comme il en existe en Allemagne et en Autriche. Il est de peu d'intérêt pour la France mais nous le présentons tout de même car nous y ferons référence par la suite lors de la mise en évidence des incohérences du projet.

Ce module est composé d'un sous-module risque d'épidémie / d'accumulation qui permet de prendre en compte les effets découlant d'épisodes épidémiques importants. C'est donc dans ce sous-module qu'est pris en compte le risque de catastrophe.

Ce sous-module requiert les données suivantes :

$claims_{ay}$	=	Frais de sinistre de l'exercice comptable sur le marché de l'assurance santé
$P_{ay}$	=	Primes brutes acquises sur l'exercice comptable
$MP_{ay}$	=	Total des primes brutes acquises sur l'exercice comptable sur le marché de l'assurance santé

Le chargement en capital au titre du risque d'épidémie / d'accumulation en santé est calculé comme suit :

$$Health_{ac} = \lambda_{ac} \cdot claims_{ay} \cdot \frac{P_{ay}}{MP_{ay}}$$

où  $\lambda_{ac}$  = facteur de risque  $Health_{ac}$ .

Le QIS4 fixe la valeur de ce coefficient  $\lambda_{ac}$  à 6,5 %. Pour fixer ce coefficient, le CEIOPS s'est basé sur les travaux menés en Allemagne par le Robert Koch Institute, une institution responsable du contrôle et de la prévention des maladies. Ces travaux peuvent être consultés sur le site Internet du RKI. Ils avaient pour but d'estimer le pourcentage de la population ayant effectué une consultation médicale ou ayant été pris en charge cliniquement suite à une épidémie de grippe de huit semaines avec un taux d'infection de

50 %. Il en résulte que 25 % de la population consulterait un médecin et 0,75 % de la population nécessiterait une prise en charge hospitalière. C'est sur la base de ces résultats que le CEIOPS a établi ce coefficient de 6,5 % (sans fournir d'explications supplémentaires dans le document dédié au calibrage dans le QIS3 : *QIS3 Calibration of the underwriting risk, market risk and MCR*), estimant que la vraisemblance d'un tel scénario correspondait à un quantile à 99,5 % (c'est-à-dire à un évènement ayant une probabilité de survenance de 0,5 % par an, soit de manière équivalente, une période de retour d'une fois tous les deux cents ans) ce qui est donc cohérent avec l'approche du calcul du SCR qui requiert que l'entreprise soit solvable à un horizon d'un an avec une probabilité de 99,5 %.

#### 2.2.2.2. Risque catastrophe en assurance complémentaire santé (Accident & Health Short Term)

L'activité d'assurance complémentaire santé se retrouve dans la *line of business (lob = branche) Short Term Health*. Cette *lob* a récemment migré du module *Non Life* vers le module *Health* entre le QIS3 et le QIS4.

Le QIS4 propose de calculer la risque catastrophe en complémentaire santé de la même manière que dans le module *Non Life*, c'est-à-dire selon trois principes :

- soit à l'aide d'une formule standard si aucun scénario catastrophe n'est formalisé par l'autorité de contrôle. La formule consiste à appliquer un coefficient forfaitaire sur le montant des primes pour calculer l'impact économique du risque catastrophe :

$$NL_{CAT} = c.P$$

Actuellement, le coefficient forfaitaire  $c$  est fixé à 10 % ;

- soit à l'aide de scénarios régionaux s'ils ont été fournis par l'autorité de contrôle. Ils remplacent alors la formule standard de la méthode précédente. Les scénarios régionaux comprennent les catastrophes naturelles et d'origine humaine. Voici quelques exemples de scénarios proposés par le CEIOPS : pandémie (grippe aviaire), épidémie de polio, accident industriel, action terroriste ayant des effets différés dans le temps (poison déversé dans le réseau d'eau potable)... ;
- soit à l'aide de scénarios catastrophes personnalisés selon les branches souscrites et la concentration géographique si les deux méthodes précédentes semblent peu pertinentes face à l'exposition aux catastrophes de la compagnie.

Il est difficile de justifier en citant des sources précises la valeur du coefficient  $c$  de 10 % appliqué dans la première méthode. Ce coefficient est d'ailleurs jugé beaucoup trop important par les mutuelles et les institutions de prévoyance (nous y reviendront par la suite). Il semblerait d'après le *QIS4 background document, Calibration of SCR, MCR and proxies* (voir CEIOPS (2008)) que ce coefficient ait été déterminé suite aux différentes

simulations de scénarios proposés par le CEIOPS dans le QIS3 pour le sous-module catastrophe du module risque de souscription non-vie.

### 2.2.2.3. Risque catastrophe en accident de travail (Health Worker Compensation)

Tout comme pour les modules précédents, un sous-module est spécifiquement dédié au risque catastrophe en accident de travail. L'activité d'assurance accident de travail a elle aussi migré du module *Non Life* vers le module *Health* entre le QIS3 et le QIS4. Le calcul se fait donc de la même manière que pour l'activité d'assurance complémentaire santé :

- soit à l'aide d'une formule standard si aucun scénario catastrophe n'est formalisé par l'autorité de contrôle :

$$Health_{WCCAT} = c.P$$

Actuellement, le coefficient forfaitaire  $c$  est fixé à 7 % ;

- soit à l'aide de scénarios régionaux s'ils ont été fournis par l'autorité de contrôle. Voici quelques exemples de scénarios : maladie professionnelle s'avérant très coûteuse et touchant un grand nombre de personnes, accident concentré important ou accident terroriste impliquant un large effectif pour une entreprise ou pour une zone donnée ;
- soit à l'aide de scénarios catastrophes personnalisés selon les branches souscrites et la concentration géographique si les deux méthodes précédentes semblent peu pertinentes face à l'exposition aux catastrophes de la compagnie.

Là encore, il est difficile de justifier en citant des sources précises la valeur du coefficient  $c$  de 7 % appliqué dans la première méthode. Il semblerait toujours que ce coefficient ait été déterminé suite aux différentes simulations de scénarios proposés par le CEIOPS dans le QIS3 pour le sous-module catastrophe du module risque de souscription non-vie.

### 2.2.3. Le module risque de souscription vie

L'assurance vie fait partie du module risque de souscription vie qui concerne le risque résultant de la souscription de contrats d'assurance vie. Ce risque est composé des risques biométriques (risque de mortalité, risque de longévité et risque d'invalidité/de morbidité), du risque de rachat, du risque de dépenses, du risque de révision et du risque catastrophe.

Les risques de catastrophe en vie ont trait à des événements extrêmes ou irréguliers qui ne sont pas suffisamment pris en compte par les chargements en capital des autres sous-modules du risque de souscription vie. On peut donc penser à des événements de type épidémie, incendie, tremblements de terre, attaques terroristes...Le QIS3 propose de ne considérer que le risque de pandémie dans le sous-module risque de catastrophe en

assurance vie, estimant que ce risque surpasse tous les autres risques potentiels en termes d'impact sur la santé ou la durée de vie humaine.

Le chargement en capital au titre du risque de catastrophe en vie est défini comme suit :

$$Life_{CAT} = \Delta NAV \mid \text{life CAT shock}$$

où  $\Delta NAV$  représente la variation d'actif net et où le choc  $Life_{CAT}$  est une combinaison des événements suivants survenant simultanément :

- une augmentation en valeur absolue de 1,5 pour mille du taux de souscripteurs décédant dans l'année suivante (par exemple de 1,0 pour mille à 2,5 pour mille) ;
- une augmentation en valeur absolue de 1,5 pour mille du taux de souscripteurs touchés par la maladie au cours de l'année suivante. Le cas échéant, les entreprises doivent supposer qu'un tiers de ces souscripteurs seront malades pendant six mois, un tiers pendant douze mois et un tiers pendant vingt quatre mois à compter du moment où ils tombent malades.

Le QIS4 propose une formule simplifiée. La formule suivante peut être appliquée pour le sous-module de catastrophe en vie : les données sont nécessaires pour chaque police pour laquelle le versement des prestations (sous forme de capital ou de rente) est subordonné à la mortalité ou à l'invalidité :

$$Life_{CAT} = \sum_i 0.0015 \cdot Capital\_at\_Risk_i$$

où l'indice  $i$  dénote chaque police pour laquelle le paiement de prestations (sous forme de capital ou de rente) est subordonné à la mortalité ou à l'invalidité et où  $Capital\_at\_Risk_i$  est déterminé comme suit :

$$Capital\_at\_Risk_i = SA_i + AB_i \cdot Annuity\_factor - TP_i$$

avec :

- $TP_i$  = Provision technique (nette de réassurance) pour chaque police  $i$  ;
- $SA_i$  = Pour chaque police  $i$  : somme assurée (nette de réassurance) au décès ou à l'invalidité lorsque les prestations sont dues sous forme de capital. Autrement, zéro ;
- $AB_i$  = Pour chaque police  $i$  : montant annualisé des prestations (net de réassurance) dues au décès ou à l'invalidité lorsque les prestations ne sont pas dues sous forme de capital. Sinon, zéro ;
- $Annuity\_factor$  = Facteur de rente moyen pour la durée de versement des prestations attendue en cas de sinistre.

Ce taux de mortalité en excès de la mortalité normale qui s'élève à 0,15 % dans le scénario adopté pour l'assurance vie dans le QIS4 reste inchangé par rapport à celui choisi dans le QIS3. Pour fixer ce coefficient de 0,15 %, le CEIOPS s'est basé sur la grippe espagnole de 1918. En Europe, le taux de surmortalité moyen suite à cette pandémie de 1918 est de 0,5 % selon le CEIOPS. En imaginant que le H5N1 pourrait à l'avenir se développer et muter en un virus beaucoup plus dangereux pour l'homme, le CEIOPS propose de retenir le même impact sur le taux de surmortalité suite à une épidémie de grippe aviaire que celui observé en 1918. Cependant, le CEIOPS a décidé de réduire le taux de 0,5 % à 0,15 % après avoir pris en compte les avancées médicales faites depuis 1918.

Le CEIOPS confirme son raisonnement en faisant référence à une étude de Swiss Re parue en juin 2007 : *Pandemic Influenza : a 21st century model for mortality shocks*. Basé sur un modèle épidémiologique pour une pandémie de période de retour de deux cents ans, l'étude conclut sur un taux de surmortalité compris en 0,1 % et 0,15 % pour la plupart des pays développés. Cette étude sera présentée plus en détails par la suite.

### **CHAPITRE 3. MISE EN EVIDENCE DES INCOHERENCES ENTRE LES DIFFERENTS MODULES DU QIS4 ET PISTES DE REFLEXION**

Maintenant que nous avons présenté les préconisations du QIS4 en termes de prise en compte du risque catastrophe en assurance de personnes dans le chapitre précédent, voyons ce qu'il en ressort de la cohérence des différentes majorations proposées par ce projet.

#### **Section 3.1. Mise en évidence des incohérences entre les différents modules du QIS4**

Cette partie s'appuie sur les résultats d'un pré-test du QIS4 de la Mutualité Française intitulé *Calibrage du risque catastrophe en assurance complémentaire santé* de février 2008. Dans cette note, la Mutualité Française se propose d'évaluer les conséquences d'une pandémie de grippe sur la sinistralité d'un assureur en complémentaire santé tout en restant cohérent avec la démarche retenue par le CEIOPS dans le module risque de souscription vie, c'est-à-dire en assimilant le risque de catastrophe au risque de pandémie de grippe.

Le modèle développé par la Mutualité Française reprend les résultats d'une étude menée par l'Institut National de Veille Sanitaire (INVS) sur l'estimation de l'impact d'une pandémie de grippe en France (Doyle et al. (2005)). Cette étude reprend elle-même une étude de Metzger et al. (1999) qui développe une modélisation stochastique du risque de pandémie de grippe afin d'obtenir une distribution des décès et des hospitalisations autour du nombre de cas « attendus ». Les paramètres de ce modèle sont calibrés à partir des données observées sur les pandémies des années 1918 (grippe espagnole), 1957 (grippe asiatique) et 1968 (grippe de Hong-Kong). Trois jeux d'hypothèses sur les taux d'infection, d'hospitalisation et de létalité sont distingués suivant la fréquence de l'évènement : l'hypothèse basse équivalente à un évènement catastrophique pris au quantile 1/30 ans (autrement dit une catastrophe de période de retour de 30 ans), l'hypothèse moyenne au quantile 1/60 ans et l'hypothèse haute au quantile 1/600 ans.

Après avoir estimé le nombre de cas attendus dans ces trois hypothèses, la Mutualité Française a estimé les coûts moyens d'une pandémie de grippe par personne pour les assureurs sur la base d'un autre modèle élaboré par MutRé, réassureur spécialisé en assurance complémentaire santé et en prévoyance.

Au final, la Mutualité Française conclut sur le fait que le coût d'une pandémie de grippe varie entre 2,7 % et 3,9 % des primes pour un assureur en complémentaire santé. De plus, le taux de surmortalité varie selon la sévérité de la pandémie entre 0,22 % et 0,50 % selon les hypothèses. Ce qui représente sur la population française un nombre de décès supplémentaires compris entre 141 000 et 321 000 (pour comparaison, près de 543 000 décès ont été enregistrés en France en 2008 selon l'INSEE).

Bien sûr, de nombreuses autres études ont été publiées concernant les taux de surmortalité attendus suite à une pandémie de grippe. Voici un tableau récapitulatif des résultats des différentes études :

Sources	Hypothèses du modèle	Taux de surmortalité additif
Swiss Re (2007)	USA, pondération spécifique de la grippe espagnole, prise en compte des politiques préventives et curatives de limitation de la dissémination	Entre 0,10 % et 0,15 %
Gen Re (Stracke et al. 2006)	Allemagne, prise en compte de la pandémie de grippe espagnole	Entre 0,41 % et 0,61 %
INVS (2005)	France, prise en compte des trois dernières pandémies, hors politique de santé publique	0,35 %
Mutualité Française (2008)	Nombre de cas attendus avec le modèle de l'INVS. Trois jeux d'hypothèses	Entre 0,22 % et 0,50 %

Tableau 5 - Taux de surmortalité de différents modèles

A titre de comparaison, voici des taux de surmortalité réellement observés pour des pandémies passées (pour une liste plus complète, on pourra se référer au Tableau 7 plus loin) :

Sources	Données historiques	Taux de surmortalité additif
Simonsen & al.	USA, pandémie de 1918	0,53 %
INVS	France, pandémie de 1968	0,06 %

Tableau 6 – Taux de surmortalité réellement observés

On remarque ici la grande hétérogénéité des résultats que l'on explique en partie par des jeux d'hypothèses différents pour chaque modèle. En effet, certains prennent en compte des avis d'expert, d'autres des politiques de santé publique. Certains sont calibrés sur trois pandémies passées alors que d'autres ne prennent en compte que la grippe espagnole de 1918.

A partir des résultats de son étude, la Mutualité Française a pu évaluer les conséquences économiques d'une pandémie de grippe sur l'activité d'assurance au premier euro en redressant les chiffres obtenus en assurance complémentaire santé (la Mutualité Française retient le coefficient le plus sévère de 3,9 %) pour tenir compte :

- du poids relatif des remboursements de soins optiques et dentaires dans les comptes des assureurs en complémentaire santé ;
- de la prise en charge des Affections de Longue Durée par le régime de Sécurité Sociale.

La méthode employée pour redresser les chiffres obtenus en assurance complémentaire santé n'est pas détaillée dans l'étude de la Mutualité Française. Il est simplement précisé que les données sont issues des prévisions 2007 de l'INSEE pour les *Comptes Nationaux de la Santé*. L'évaluation a alors été estimée à 10 milliards d'euros en France, soit 7 % des prestations versées.

La Mutualité Française affirme donc que cette estimation est cohérente avec le calibrage proposé par le CEIOPS pour le module *Long Term Health* (on rappelle que le coefficient forfaitaire appliqué est de 6,5 %). Cependant, le taux de 3,9 % sur lequel ce dernier calcul est basé n'est pas cohérent avec le coefficient de 10 % retenu par le CEIOPS en complémentaire santé. De plus, comme on l'a vu précédemment, le taux de mortalité en excès de la mortalité normale dans le cas le plus grave s'élèverait à 0,50 %, contre 0,15 % pour le scénario adopté en assurance vie. Ces différences font naître un doute sur la cohérence de la modélisation du risque catastrophe pour l'activité d'assurance de personnes dans le QIS4.

### **Section 3.2. Quelques commentaires et pistes de réflexion**

Après avoir observé de quelle manière est appréhendé le risque catastrophe en réassurance (dans les états réglementaires) et dans le QIS4 et après avoir mis en évidence les incohérences existantes entre les différents modules concernant ce risque, voici quelques commentaires et pistes de réflexion que l'on peut dégager.

Tout d'abord, comme nous l'avons déjà remarqué, le risque catastrophe en assurance vie n'est restreint qu'au risque de pandémie, le CEIOPS estimant que ce risque surpasse tous les autres en terme d'impact sur la mortalité et la santé. Ceci doit cependant être nuancé dans la mesure où certaines catastrophes naturelles peuvent entraîner autant de décès qu'une épidémie sévère. A titre d'exemple, on peut citer les inondations de 1933 en Chine qui ont entraîné le décès de plus de trois millions de personnes (cf. Tableau 1) alors que la « grippe asiatique » de 1958 a entraîné un million de décès d'après Linfoot (2007). Sans compter que de plus, le nombre de catastrophes naturelles est en forte augmentation : entre 1997 et 2006, elles ont augmenté de 60 % d'après la Fédération Internationale de la Croix-Rouge. Le nombre de morts a même doublé par rapport à la période 1987-1996, passant de 600 000 à 1,2 million. Le nombre de personnes touchées en moyenne chaque année est passé de 230 millions à 270 millions entre les deux décennies. En 2007, le nombre de catastrophes a bondi de 20 % par rapport à 2006, dépassant pour la première fois la barre annuelle des 500 catastrophes, toujours selon la Fédération Internationale de la Croix-Rouge. Il semblerait donc que laisser de côté les catastrophes naturelles dans le risque de souscription vie ne reflète pas exactement l'ampleur réelle du risque de catastrophe.

Ensuite, on remarquera que les modèles utilisés pour déterminer des scénarios catastrophes font appel à de multiples hypothèses sur le taux de contamination de la pandémie, le taux de létalité, le taux d'hospitalisation, les avancées dans le domaine médical, la fréquence de la catastrophe... On peut d'ailleurs noter concernant ce dernier

point que dans le modèle repris par la Mutualité Française, les hypothèses de fréquence retenues varient entre 1/30 ans et 1/600 ans. La Mutualité Française s'est basée sur l'hypothèse haute d'une fréquence de 1/600 ans pour déterminer le coefficient de 7 % pour l'assurance santé au premier euro, affirmant alors que cela est cohérent avec le coefficient de 6,5 % retenu par le CEIOPS. Cependant, le CEIOPS a déterminé ce taux de 6,5 % sur la base d'une pandémie extrême ayant une fréquence de survenance de 1/200 ans. Ainsi, en prétendant comparer les 7 % de son étude et les 6,5 % du CEIOPS, la Mutualité Française oublie que les scénarios qui ont conduit à ces résultats ne sont pas comparables. La multitude d'hypothèses utilisées dans ces études et ces scénarios conduit donc à faire des comparaisons qui n'ont pas de sens.

Les incohérences observées entre les différentes branches dans le QIS4 suggèrent de réfléchir à un scénario commun de catastrophe (qui devrait inclure à la fois le risque de pandémie mais aussi celui de catastrophes naturelles et technologiques) pour l'ensemble de l'activité d'assurance de personnes et qui pourrait être modélisé à part entière dans un module spécifique de la formule standard. Cette réflexion pourrait également être particulièrement utile pour les compagnies qui choisiront d'appliquer le modèle interne. En effet, la prise en compte des catastrophes y est aujourd'hui très peu évoquée. Dans cette optique, nous proposerons dans la troisième partie de ce mémoire un modèle de simulation de chocs de mortalité.

Cette réflexion doit être sérieusement entamée par les compagnies d'assurance dans la mesure où la réassurance traditionnelle présente aujourd'hui certaines limites. En effet, en ce qui concerne le risque de pandémie, celui est considéré par les réassureurs comme étant un phénomène de masse qui n'est pas de nature à activer une couverture XL par évènement. De plus, la plupart des traités de réassurance par tête ou par évènement excluent désormais systématiquement les risques de terrorisme, guerre et autre accident technologique.

Nous allons maintenant étudier dans une seconde partie les deux approches principales de modélisation du risque catastrophe en mortalité.

## ***PARTIE II :***

### ***LES PRINCIPALES APPROCHES DE MODELISATION DE LA MORTALITE EXTREME***

Les travaux concernant l'étude de l'impact d'une catastrophe sur la mortalité sont très variés par leur nombre et par les méthodes employées. On peut cependant discerner deux types d'approche à savoir l'approche épidémiologique et l'approche actuarielle. L'approche épidémiologique se base avant tout sur des avis d'expert et des hypothèses de politique préventive ou curative. Elle accorde généralement peu d'importance aux événements passés. Les modèles actuariels quant à eux se basent sur des modèles de mortalité stochastique. Les données historiques dans cette approche ont une importance toute particulière pour calibrer les modèles.

## CHAPITRE 1. L'APPROCHE EPIDEMIOLOGIQUE

Dans ce chapitre, on s'intéressera aux modèles dits épidémiologiques mis en place pour quantifier le risque catastrophe en cas de pandémie.

### Section 1.1. Présentation générale

Les modèles de mortalité extrême qui suivent cette approche se basent sur des recherches d'experts, des données épidémiologiques et dans une moindre mesure sur des données historiques. Généralement, ce type de modèle exclut les événements du genre catastrophes naturelles ou accident technologique. Les paramètres qui déterminent des taux d'infection et de létalité sont généralement fixés sur la base de recherches scientifiques.

Le modèle servant de base est le modèle dit « sain-malade-mort » qui utilise des chaînes de Markov. On considère simplement que pour un individu  $i$ , trois états sont possibles :

- S= sain ;
- m = malade ;
- M=mort.

L'évolution d'un état à l'autre peut être représentée par le diagramme suivant :

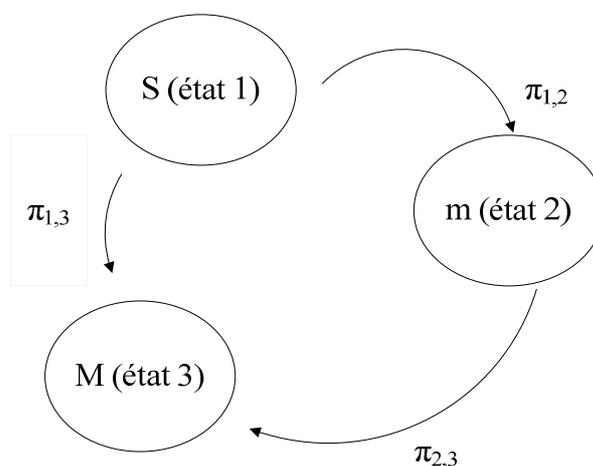


Schéma 2 - Diagramme classique d'un modèle "sain - malade - mort"

Une épidémie suppose bien entendu une dynamique temporelle, le but étant de mesurer l'effet de cette épidémie au fur et à mesure de sa diffusion progressive. Les probabilités de transition ne sont donc pas constantes et varient suivant la période. Elles sont généralement déterminées de la manière suivante :

- $\pi_{1,3}$  : ces probabilités sont déterminées par les  $q_x$  de la table de mortalité de référence ;
- $\pi_{1,2}$  : cette probabilité est proportionnelle à l'effectif contaminé :

$$\pi_{1,2} = \alpha \frac{|C|}{|I|}$$

où  $|X|$  désigne le cardinal de l'ensemble  $X$ .  $C$  représente l'ensemble des personnes contaminées et  $I$  l'ensemble de la population ;

- $\pi_{2,3}$  : le taux de décès pour les individus contaminés est déterminé à partir du taux de base, aggravé de la manière suivante :  $q_x^c = q_x(1 + \lambda)$ .

En plus des taux de survenance de base par âge, ce modèle nécessite la détermination des paramètres suivants :

- $\alpha$  qui s'interprète comme étant le coefficient de diffusion de l'épidémie ;
- la liste des individus initialement contaminée et qui permet d'initialiser le modèle ;
- $\lambda$  le taux de sur-sinistralité des personnes contaminées.

Les taux de sur-sinistralité et les coefficients de diffusion ne sont généralement pas estimés historiquement mais se basent sur des avis d'experts. Les variantes de ce modèle sont très nombreuses. On peut rajouter des états suivant le degré de maladie, inclure un état de guérison partiel ou total... Pour une présentation plus détaillée de l'utilisation des chaînes de Markov en biologie et dans les modèles « proies-prédateurs », on pourra consulter Karlin (1969).

Tous les modèles ont leur particularité, mais le schéma reste globalement le même. Plusieurs scénarios sont simulés concernant l'impact démographique, le lieu où l'épidémie apparaît, sa propagation géographique, son cycle de vie, la production et l'efficacité des vaccins, les mesures de prévention prises par les institutions... A partir de tous ces scénarios, on obtient une distribution de taux de surmortalité à partir de laquelle il est possible de calculer des quantiles. Cette approche se retrouve dans l'étude de la Mutualité Française (2008) et de l'INVS (2005) dont nous avons déjà parlé précédemment. Elle se retrouve aussi dans le modèle développé par l'entreprise *Risk Management Solutions* (un des modèles les plus utilisés pour modéliser des catastrophes avec ceux de AIR et d'Equecat) qui eut un rôle de conseil lors de l'émission de CAT Mortality bonds en 2006 (Vita III) ou encore dans l'étude Swiss Re de juin 2007 : *Pandemic Influenza : a 21st century model for mortality shocks* que nous allons présenter maintenant.

## Section 1.2. L'étude Swiss Re

Cette étude est présentée de manière détaillée dans cette section dans la mesure où c'est l'étude la plus récente et la plus complète existante à ce jour. Cette section reste en grande partie descriptive, la compagnie Swiss Re n'étant pas entrée dans les détails de ses modélisations, ce qui est compréhensible.

L'étude de Swiss Re se décompose en six parties.

La première partie est consacrée à la description de la grippe aviaire. Il semblerait que le H5N1 soit le virus le plus à même d'avoir des conséquences pour l'homme mais l'étude rappelle que d'autres souches de grippe aviaire existent et qu'historiquement, certaines ont contaminé des populations humaines bien plus facilement que l'a fait le virus H5N1. Celui-ci devra d'ailleurs subir de nombreuses mutations pour être réellement transmissible à l'homme et dangereux. C'est le cas également du virus H1N1 qui ne constituera pas une réelle menace pour l'homme tant qu'il n'aura pas muté.

La seconde partie s'intéresse aux effets de la grippe aviaire sur l'homme. Il en ressort qu'au milieu du XX<sup>ème</sup> siècle, les traitements inadaptés par antibiotiques ont eu pour conséquence une augmentation dramatique des taux de décès dus à la grippe et à la pneumonie. Aujourd'hui, la seule mesure préventive efficace connue pour éviter une contamination chez les personnes à risque est le vaccin.

La troisième partie se propose de recenser historiquement les grandes pandémies connues qui ont eu lieu dans le monde. Dans les trois cents dernières années, dix à treize pandémies graves ont eu lieu, à des intervalles de temps très irréguliers. Sur la base de fréquences historiques, il apparaît que la probabilité qu'une pandémie grave survienne dans une année est comprise entre 3 % et 4 %. La sévérité des pandémies passées a été plutôt régulière, à l'exception de la grippe espagnole de 1918 qui fut exceptionnelle. En effet, la dernière pandémie qui semble avoir eu le même impact sur la mortalité remonte à 1580 d'après les descriptions faites. Le tableau suivant recense les dernières pandémies connues entre 1700 et 2000. Il est construit à partir de l'étude Swiss Re, elle-même basée sur *Pandemic Influenza 1700-1900: a study in historical epidemiology* de Patterson (1986) :

Années	Nombre d'années depuis la dernière pandémie	Origine ou lieu du premier dépistage	Type de virus	Nombre de décès estimés	Taux de surmortalité additif (population américaine)
1729-1730	?	Russie ?			
1732-1733	2	Russie			
1781-1782	48	Russie, Chine ?			
1788-1789 (?)	6	Russie			
1830-1831	41-48	Russie, Chine			
1833	2	Russie			
1836-1837	3	Russie ?			
1889-1890	52-56	Russie	H2		
1899-1900	9		H3		
1918-1919 (grippe espagnole)	18	France, Etats-Unis	H1N1	40-50 millions	0,53 %
1957-1958 (grippe asiatique)	38	Chine	H2N2	1-2 millions	0,041 %
1968-1969 (grippe de Hong-Kong)	10	Chine	H3N2	1 million	0,017 %

Tableau 7 - Caractéristiques des pandémies survenues entre 1700 et 2000

Dans une quatrième partie, l'étude présente la situation actuelle et la façon dont se préparent les institutions nationales et les organisations internationales. Le monde est aujourd'hui plus préparé que jamais à une pandémie de grippe et depuis le début du siècle, les organisations et les institutions élaborent des stratégies et des plans de réaction. De plus, les avancées médicales et l'amélioration de l'hygiène ont éliminé un nombre important de causes historiques des maladies infectieuses. Cependant, la capacité de production des vaccins serait insuffisante en cas de nécessité de vaccination massive en début de pandémie.

La cinquième partie est celle qui nous intéresse le plus puisqu'elle présente en détails le modèle utilisé par Swiss Re. Chaque simulation de scénario de pandémie se déroule de la manière suivante. Tout d'abord, deux réalisations des variables aléatoires « létalité du virus » et « capacité à se répandre du virus » sont générées. On appelle « létalité du virus » la proportion de personnes décédées à cause du virus parmi les personnes contaminées par le virus. On appelle « capacité à se répandre du virus » le nombre moyen de personnes qui pourraient être contaminées par une autre personne porteuse du virus au début de la pandémie. Ce paramètre, noté  $R_0$  dans les études d'épidémiologie, est un paramètre clé. A titre d'exemple, pour une rougeole, ce paramètre vaut entre 15 et 20, pour une grippe il est généralement compris entre 1,8 et 2,5 et il se situe aux alentours de 1,1 pour une hépatite B. Les deux distributions de ces variables aléatoires ont été estimées sur la base de données historiques pour la population des Etats-Unis. Lors de l'estimation de ces deux distributions, le poids accordé à la grippe espagnole a été réduit par rapport à celui des autres pandémies. En effet, l'étude souligne qu'une pandémie similaire à celle de 1918 n'aurait pas le même impact aujourd'hui sur la population. En effet, la guerre de 14-18 avait probablement aggravé la létalité de la pandémie en raison du fort nombre de personnes atteintes de la tuberculose et des personnels soignants manquants car mobilisés pour soigner les blessés sur le front. Sans compter que depuis, le virus de la grippe a été découvert ainsi que son vaccin. Swiss Re a finalement retenu une distribution de Pareto généralisée pour la distribution de la variable aléatoire « capacité à se répandre du virus » et une distribution lognormale pour la variable aléatoire « létalité du virus ».

Ces deux réalisations ainsi générées après simulation servent de base au scénario qui sera développé par la suite. Ils permettent de déterminer un grand nombre de paramètres que l'on peut regrouper en quatre classes. Le calcul détaillé de ces paramètres n'est pas exposé dans l'étude de Swiss Re :

- les caractéristiques épidémiologiques du virus :
  - o profil de mortalité : à partir de la valeur simulée de la variable « létalité du virus », le modèle crée un profil de mortalité par âge ;
  - o taux d'infection par âge : déterminé en fonction des caractéristiques des deux maladies principales entraînées par le virus : pneumonie bactérienne ou pneumonie virale ;
  - o période d'incubation : elle n'est pas explicitement simulée mais est plutôt déterminée de manière implicite en fonction de l'âge et du taux d'infection ;
  - o degré de contagion, comportement des individus, réceptivité au virus par tranches d'âge : on trouve très peu de précisions sur la manière dont sont simulées ces composantes dans l'étude ;
- les caractéristiques démographiques :
  - o la répartition de la population par âges : déterminée à partir de statistiques produites par le US Census Bureau en 2006 ;

- les contacts humains : déterminés à partir du nombre moyen de contacts qu'un individu a avec des individus de classe d'âge différente quotidiennement d'après une étude de l'*American Journal of Epidemiology* (Wallinga et al. (2006)) ;
- l'immunité préexistante, les déplacements de la population : là encore peu d'informations sont données pour expliquer sur quelle base sont déterminées ces caractéristiques ;
- les interventions non médicales suite à la pandémie : la restriction des contacts humains (déterminée en fonction du nombre de décès atteint), la restriction des déplacements de la population ;
- les interventions médicales suite à la pandémie :
  - l'efficacité des antibiotiques : Swiss Re précise qu'un système de scoring a été développé sur la base de plusieurs études d'experts ;
  - l'efficacité des antiviraux : Swiss Re suppose qu'ils seraient inefficaces pour une pandémie sur quatre ;
  - l'efficacité des vaccins et la capacité de production : Swiss Re suppose qu'ils seraient efficaces dans 90 % des cas ;
  - les ressources médicales : disponibilité des médecins et des infirmières estimée sur la base d'une étude de la World Health Organization *Core Health Indicator* (2006).

Le schéma suivant tiré de l'étude de Swiss Re permet de résumer comment sont utilisées les deux nombres simulés au début du scénario pour obtenir les variables qui permettront de produire les résultats finaux :

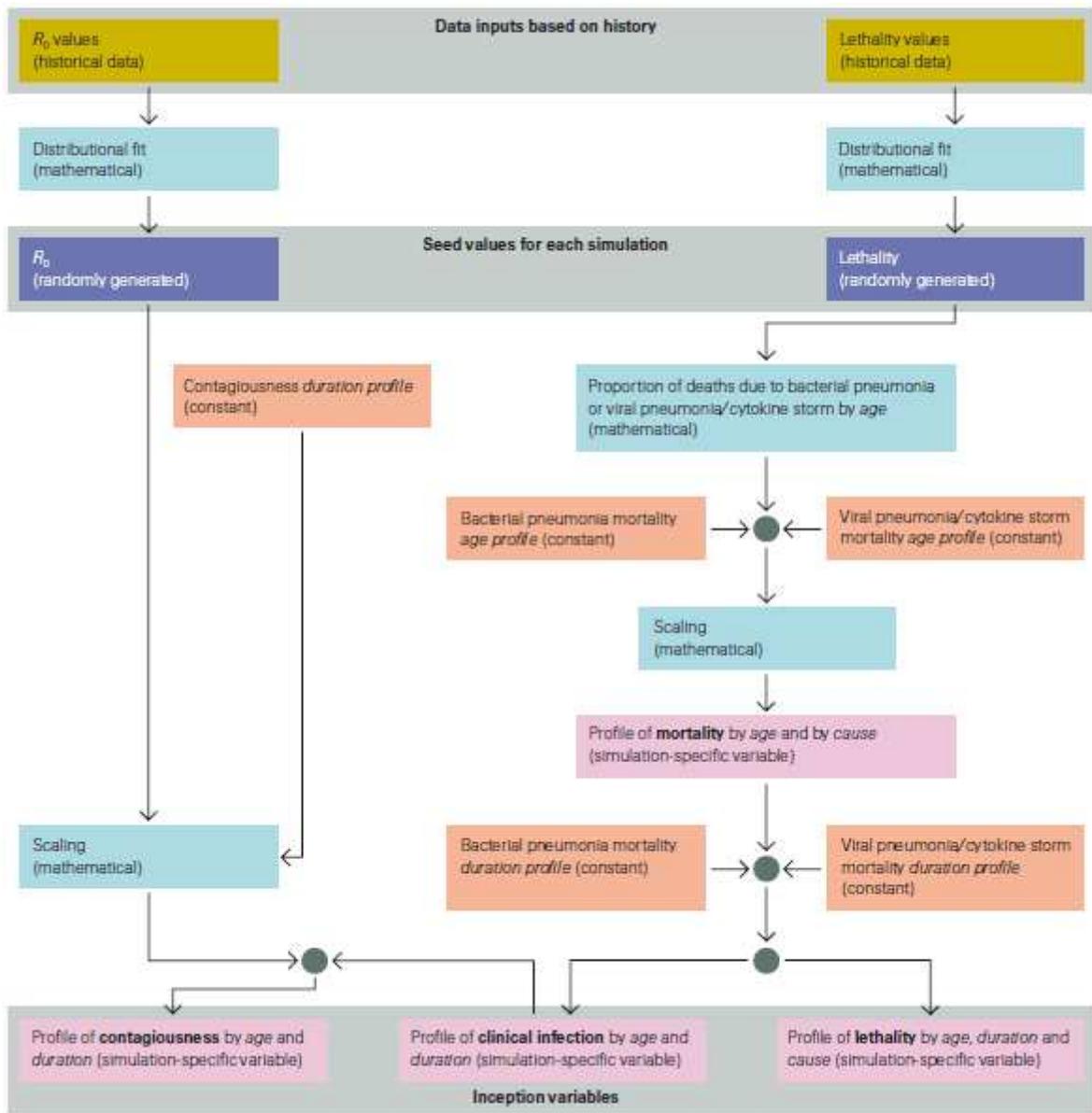


Schéma 3 - Fonctionnement global du modèle de Swiss Re

Chaque nouvelle simulation créera donc un nouveau jeu d'hypothèses à partir des deux paramètres d'origine et permettra d'estimer le niveau de surmortalité suite à un tel scénario. Plusieurs simulations sont effectuées afin d'obtenir une distribution de taux de surmortalité et des quantiles peuvent ensuite être calculés pour différents niveaux. Comme nous l'avons déjà mentionné, pour un évènement de fréquence de survenance d'une fois tous les deux cents ans, le taux de surmortalité attendu suite à une pandémie de grippe serait compris entre 0,1 % et 0,15 % dans la plupart des pays développés tandis qu'il serait compris entre 0,15 % et 0,4 % pour les pays en voie de développement.

La dernière partie de l'étude de Swiss Re illustre pourquoi la modélisation de chocs de mortalité est indispensable d'un point de vue économique pour les assureurs, les agences de notation et plus globalement pour le projet de Solvabilité 2.

On notera que le modèle présenté par Swiss Re a toutefois certaines limites. Tout d'abord, il est déterministe dans le sens où il ne prend pas en compte le caractère très aléatoire d'un début de pandémie, c'est-à-dire entre autres le comportement du virus lorsqu'il arrive dans un nouveau pays. En effet, le modèle est basé sur l'hypothèse que tout nouveau virus entrant sur le territoire donne suite à une pandémie. Ensuite, il ne prend pas en compte la dispersion du virus entre les pays bien que le modèle ait été développé pour trente sept pays différents. Enfin, le fait que certains paramètres soient déterminés sur la base d'études conduit un certain degré d'incertitude sur la valeur de ces paramètres comme le montrent les tests de sensibilité qui ont été effectués par Swiss Re (page 66-68 du rapport) et qui démontrent une forte sensibilité des paramètres aux hypothèses.

Après avoir décrit l'approche épidémiologique pour la modélisation de la mortalité extrême, intéressons-nous à l'approche « actuarielle ».

## CHAPITRE 2. L'APPROCHE « ACTUARIELLE »

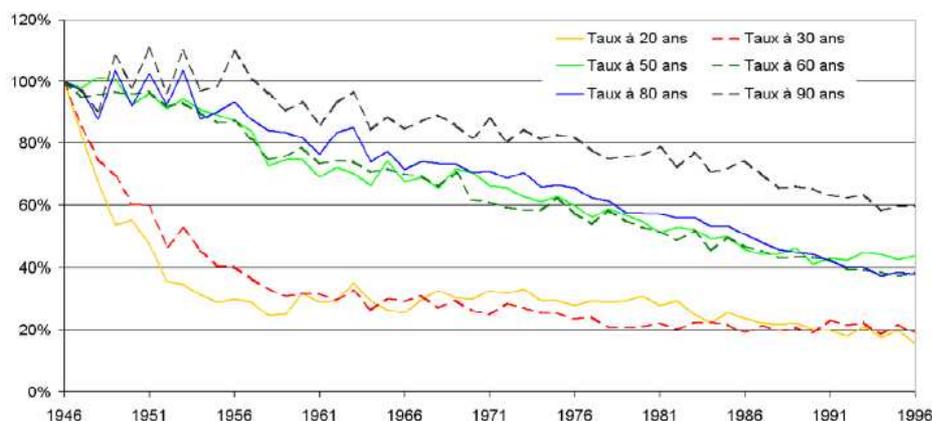
Dans ce chapitre, nous allons présenter une approche alternative aux modèles épidémiologiques basée sur l'utilisation de modèles de mortalité stochastique. Nous étudierons plus en détails l'approche de Bauer et Kramer (2007) à laquelle nous apporterons quelques améliorations notamment concernant la modélisation de la tendance.

### Section 2.1. Présentation générale

De manière classique, pour modéliser la mortalité, on spécifie généralement un taux de hasard  $\mu(x,t)$  déterministe en fonction de l'âge  $x$  et de l'année courante  $t$ . Différents risques pèsent sur une telle méthode :

- un risque de fluctuation d'échantillonnage, qui, lorsque le taux de hasard est correctement spécifié, se mutualise notamment par l'application de la loi des grands nombres (les fluctuations s'amointrissent lorsque la taille du portefeuille augmente). De plus, ces fluctuations peuvent être quantifiées grâce au théorème central limite ;
- un risque de modèle : si la mortalité observée dans le futur est différente de celle prédite par le modèle, l'écart ne sera pas mutualisable car il s'appliquera à toutes les têtes ;
- un risque systématique : dans le cas où le taux de hasard est défini de manière paramétrique, l'estimation des paramètres peut conduire elle aussi à des incertitudes. Si l'estimateur est déterminé à l'aide de la méthode du « maximum de vraisemblance », on peut cependant déterminer des intervalles de confiance (la loi asymptotique étant connue) pour quantifier la part de risque non mutualisable. Un cas particulier de l'erreur d'estimation peut-être le risque qui pèse sur l'estimation de la tendance.

L'utilisation d'un taux de hasard déterministe pour modéliser la mortalité conduit à des surfaces de mortalité régulières. Cependant, certaines études montrent que sur le long terme, on voit bien se dégager une tendance pour l'évolution de ces taux, mais qu'il subsiste une variabilité autour de cette tendance. Cette remarque est confirmée par des données de l'Institut National d'Etudes Démographiques :



Graphique 1 - Evolution des taux de décès au cours du temps (données INED)

Ces variations autour de la tendance ne se mutualisent pas puisqu'elles affectent l'ensemble des individus. Le but est donc de trouver un modèle de mortalité qui soit capable de prendre en compte ces variations autour de la tendance. Pour cela, on a recours aux modèles stochastiques de mortalité.

Dans ce type de modèle, le taux de hasard n'est plus déterministe mais stochastique ce qui permet de quantifier le risque non diversifiable. Les modèles existants sont nombreux et peuvent être de type :

- abattement ou majoration des taux conditionnels de mortalité ;
- modèles de mortalité (Gompertz-Makeham par exemple) dont les paramètres sont rendus aléatoires ;
- modèles de type Lee-Carter ou log-Poisson (cas particuliers de modèles stochastiques même s'ils ont été élaborés à la base pour construire des extrapolations de surfaces de mortalité) ;
- modèles d'inspiration financière (modèles de taux du type retour vers la moyenne par exemple).

Nous n'allons pas développer ici en détails comment fonctionnent ces modèles, le lecteur intéressé pourra consulter Planchet et Thérond (2006) ou encore Cairns et al. (2006). Nous présenterons dans la suite un exemple de modèle de mortalité stochastique.

Les modèles qui suivent cette approche estiment l'évolution future de la mortalité extrême à l'aide de méthodes statistiques et actuarielles basées sur des données historiques. Généralement, le modèle est subdivisé en plusieurs composantes :

- une composante régulière qui modélise les variations des taux de mortalité sans tenir compte des événements catastrophiques. Nous en avons déjà parlé précédemment avec les modèles stochastiques ;

- une composante maladie qui capture l'excès de mortalité suite à une pandémie. La fréquence et la sévérité de la pandémie sont modélisées séparément sur la base de données historiques d'anciennes pandémies ;
- une composante terrorisme qui effectue des simulations de chocs de mortalité consécutifs à une attaque terroriste.

A partir d'un modèle estimant la mortalité « naturelle » sans tenir compte des événements extrêmes, on peut se demander s'il est préférable de modifier ce modèle de manière multiplicative ou additive pour introduire l'effet de pandémies par exemple. Dans Lin et Cox (2006), les événements catastrophiques sont modélisés sur la base de chocs multiplicatifs (c'est-à-dire que plus l'âge est élevé, plus le choc sera important sur la mortalité). On peut penser que cette approche est raisonnable dans la mesure où les personnes âgées ont un système immunitaire plus faible que les jeunes et qu'elles seront donc plus affectées par une pandémie. Or, ceci n'est pas totalement vrai dans la mesure où pour une catastrophe autre qu'une pandémie (une attaque terroriste par exemple), il n'y a aucune raison que les personnes âgées soient plus touchées que les personnes plus jeunes. De plus, si l'on reste dans le domaine de la pandémie, des « tempêtes de cytokine » peuvent avoir lieu. On appelle « tempête de cytokine » un phénomène qui touche des personnes ayant un très bon système immunitaire qui réagit de manière disproportionnée à un virus, entraînant la mort de l'individu. Ce phénomène est une des explications qui a aussi été avancée pour justifier l'importance du nombre de jeunes adultes touchés par la grippe espagnole de 1918. Il est donc en général préférable d'opter pour un modèle de surmortalité additive.

Dans la suite, nous allons présenter un tel modèle introduit par Bauer et Kramer (2007).

## **Section 2.2. L'approche de Bauer et Kramer**

L'approche de Bauer et Kramer reprend en partie celle développée par Miltersen et Persson (2005) qui propose une approche dite « par intensité » pour déterminer des taux de mortalité. Cette approche est courante en finance et plus particulièrement en risque de crédit pour valoriser des CDO par exemple (voir Le Page (2000) pour une présentation).

### *2.2.1. Le modèle utilisé*

Donnons nous un horizon  $T^*$ , un espace probabilisé  $(\Omega, \mathcal{F}, P)$  et une filtration  $\mathbf{F}$  telle que  $\mathbf{F} = (F_t)_{0 \leq t \leq T^*}$ . Nous nous donnons aussi une population où chaque individu a un âge  $x_0$  en  $t = 0$ , un processus stochastique  $X = (X_t)_{0 \leq t \leq T^*}$   $\mathbf{F}$ -adapté, de dimension  $d$ , positif, continue à droite avec limite à gauche et donnons nous enfin une fonction  $\mu(.,.) : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  continue.

On définit l'instant de décès d'un individu  $i$  d'âge  $x_0$  comme étant le premier instant de saut d'un processus de Cox d'intensité  $\mu(x_0 + t, X_t)$  c'est-à-dire :

$$\tau_{x_0} = \inf \left\{ t : \int_0^t \mu(x_0 + s, X_s) ds \geq E_i \right\} \quad (1)$$

où  $E_i$  est une variable aléatoire suivant une loi exponentielle de paramètre 1, indépendante de  $X$  et où  $E_i$  et  $E_j$  sont supposées être indépendantes pour tout  $i \neq j$ .

Avant d'aller plus loin, arrêtons-nous un instant sur les processus de Cox. Les processus de Cox sont des processus de Poisson doublement stochastiques, c'est-à-dire que l'intensité du processus de Poisson est elle-même aléatoire. Un processus de Cox est souvent utilisé pour modéliser l'arrivée des catastrophes. Une façon de caractériser ce processus de Poisson doublement stochastique est de le définir par sa fonction caractéristique. En effet, soit  $(\Omega, A, P)$  un espace probabilisé,  $(N_t)_t$  un processus ponctuel adapté à une filtration  $\mathbf{F} = (F_t)_{0 \leq t \leq T^*}$  et  $(\lambda_t)_t$  un processus positif et mesurable.  $(N_t)_t$  est un processus de Poisson doublement stochastique si  $\forall 0 \leq s \leq t, \forall u \in \mathbb{R}, E \left[ e^{iu(N_t - N_s)} | F_s \right] = \exp \left[ (e^{iu} - 1) \int_s^t \lambda_v dv \right]$ . Dans ce cas,  $(N_t)_t$  lui-même n'est un processus de Poisson que conditionnellement à  $(\lambda_t)_t$ .

Intéressons nous maintenant à une seule tête prise au hasard dans notre population initiale. On définit les filtrations  $G_t = \sigma \{ X_s, 0 \leq s \leq t \}$ ,  $H_t = \sigma \{ 1_{\tau_{x_0} < s}, 0 \leq s \leq t \}$  et  $F_t = \max(G_t, H_t)$ . A partir de l'équation (1), on détermine la probabilité pour qu'un individu d'âge  $x_t$  en  $t$  soit encore en vie à la date  $T$  par la formule suivante :

$${}_{T-t}P_{x_t}^{(t)} = E \left[ 1_{\tau_{x_t} > T} | G_t \right] = E \left[ \exp \left\{ - \int_t^T \mu(x_0 + s, X_s) ds \right\} | G_t \right]$$

Il est maintenant temps de spécifier la fonction  $\mu$  et le processus stochastique  $X$ . C'est là que Bauer et Kramer apportent leur contribution en proposant de retenir un taux de mortalité instantané stochastique sous la forme :

$$\mu(x_t, t) = \mu(x_t, Y_t, \Gamma_t) = Y_t \cdot e^{b \cdot x_t + c} + \Gamma_t$$

où :

- $b$  et  $c$  sont des constantes réelles ;
- $Y_t = (Y_t)_t$  un processus stochastique de retour vers la moyenne qui suit l'équation différentielle stochastique suivante :

$$dY_t = \alpha(\beta - Y_t)dt + \sigma\sqrt{Y_t}dW_t$$

avec  $Y_0 > 0$ ,  $W_t = (W_t)_t$  un mouvement brownien standard et  $\alpha$ ,  $\beta$  et  $\sigma$  des constantes positives. Ce processus est un processus de type Cox-Ingersoll-Ross utilisé en finance pour modéliser l'évolution du taux instantané ;

- $\Gamma_t = (\Gamma_t)_t$  un processus de saut de la forme Gamma Ornstein-Uhlenbeck qui suit l'équation différentielle stochastique suivante :

$$d\Gamma_t = -\kappa\Gamma_t \cdot dt + dJ_t$$

avec  $\Gamma_0 = 0$ ,  $\kappa$  une constante positive qui représente la « vitesse » de retour à la normale après une catastrophe et  $J$  un processus de Poisson composé d'intensité  $\lambda$  dont les sauts sont modélisés via une loi exponentielle de paramètre  $\zeta$ . C'est donc cette composante additive qui modélise le risque de catastrophe. On suppose que  $\Gamma$  est indépendante de  $Y$ .

Après quelques calculs (menés en Annexe), on trouve une expression simplifiée pour  ${}_{T-t}P_{x_t}^{(t)}$  :

$$\begin{aligned} {}_{T-t}P_{x_t}^{(t)} = & \exp\left\{u(T-t) + v(T-t)Y_t e^{b(x_0+t)+c}\right\} \exp\left\{-\frac{\Gamma_t}{\kappa}\left(1 - e^{-\kappa(T-t)}\right) - \frac{\lambda(T-t)}{\zeta\kappa+1}\right\} \\ & \exp\left\{\frac{\lambda\zeta}{\zeta\kappa+1} \log\left[1 + \frac{1}{\zeta\kappa}\left(1 - e^{-\kappa(T-t)}\right)\right]\right\} \end{aligned} \quad (2)$$

où  $u$  et  $v$  satisfont les équations différentielles suivantes :

$$\begin{aligned} u'(x) &= \alpha\beta v(x) e^{b(x_0+T-x)+c}, u(0) = 0 \\ v'(x) &= -1 - (\alpha - b)v(x) + \frac{1}{2}\sigma^2 v^2(x) e^{b(x_0+T-x)+c}, v(0) = 0. \end{aligned}$$

Les calculs qui mènent à cette expression pour  ${}_{T-t}P_{x_t}^{(t)}$  sont développés en « Annexe 1. Calcul du taux de survie du modèle de Bauer ». Les équations différentielles satisfaites par  $u$  et  $v$  sont des équations différentielles ordinaires dites de Riccati. Leur résolution (cf. « Annexe 2. Résolution d'une équation différentielle ordinaire de Riccati ») nous permet d'obtenir une expression explicite pour  ${}_{T-t}P_{x_t}^{(t)}$ .

### 2.2.2. Le calibrage du modèle : la composante catastrophe

Dans cette partie, nous nous intéresserons au calibrage de la composante additive catastrophe  $\Gamma$  seulement. En effet, l'indépendance de la composante continue et de la composante à sauts nous permet de procéder à l'ajustement séparément. Pour ce faire, nous allons ajuster le modèle sur la base de données historiques. Nous allons utiliser les données

de Linfoot (2007) qui présentent la fréquence et la sévérité des dernières épidémies de grippe pour la population américaine. Linfoot (2007) recense 31 épidémies de grippe sur les 420 dernières années et a pu répertorier la surmortalité de cinq d'entre elles :

Epidémie	Excès de mortalité multiplicatif sur le taux de mortalité
1918-1920	26,55 %
1957	6,92 %
1968	1,78 %
1977	1,08 %
2003	3,23 %

Tableau 8 - Excès de mortalité des cinq dernières pandémies selon Linfoot (2007)

On rappelle que la composante catastrophe  $\Gamma$  est modélisée de la manière suivante :  $\Gamma$  est un processus de saut de la forme Gamma Ornstein-Uhlenbeck qui suit l'équation différentielle stochastique suivante :

$$d\Gamma_t = -\kappa \Gamma_t dt + dJ_t$$

avec  $\Gamma_0 = 0$ ,  $\kappa$  une constante positive qui représente la vitesse de retour à la normale après une catastrophe et  $J$  un processus de Poisson composé d'intensité  $\lambda$  dont les sauts sont modélisés via une loi exponentielle de paramètre  $\zeta$ , c'est-à-dire que l'on a :

$$J_t = \sum_i^{N(t)} e_i$$

où les  $e_i$  sont des réalisations d'une loi exponentielle de paramètre  $\zeta$ .

Posons  $Z_t = \Gamma_t e^{\kappa t}$  et écrivons la formule d'intégration par parties :

$$dZ_t = d\Gamma_t e^{\kappa t} + \Gamma_t d(e^{\kappa t}) + d\langle \Gamma, e^{\kappa \cdot} \rangle_t.$$

Mais  $\langle \Gamma, e^{\kappa \cdot} \rangle_t = 0$  car  $d(e^{\kappa t}) = \kappa e^{\kappa t} dt$ . On en déduit donc que  $dZ_t = e^{\kappa t} dJ_t$  puis que  $\Gamma_t = e^{-\kappa t} \int_0^t e^{\kappa s} dJ_s = e^{-\kappa t} \gamma_t$  où  $\gamma_t$  suit une loi Gamma.

Essayons d'interpréter de manière discrète la forme de ce processus. La forme intégrale  $\Gamma_t = e^{-\kappa t} \int_0^t e^{\kappa s} dJ_s$  est égale à  $e^{-\kappa t} \sum_{T_i < t} e^{\kappa T_i} \Delta J_{T_i}$  où les  $T_i$  sont les instants de saut du

processus de Poisson et où l'on a par définition  $\Delta J_{T_i} = J_{T_{i+1}} - J_{T_i}$ . Les quantités  $\Delta J_{T_i}$  sont donc des réalisations de loi exponentielle.

Le fait que Linfoot (2007) recense 31 pandémies de grippe sur les 420 dernières années nous pousse à choisir la valeur  $\frac{31}{420} = 7,4\% = \lambda$  pour l'intensité du processus de Poisson puisque pour une loi de Poisson, l'estimateur du maximum de vraisemblance de son paramètre n'est rien d'autre que sa moyenne empirique. Si on se place  $\tau$  années après le premier saut (c'est-à-dire en  $T_1 + \tau$ ), on a alors l'expression suivante (forme discrète) :

$$\Gamma_{T_1+\tau} = e^{-\kappa(T_1+\tau)} \left[ e^{\kappa T_1} \Delta_1 + e^{\kappa T_2} \Delta_2 + \dots \right]$$

où  $T_1$  représente l'instant du premier saut (de la première catastrophe) et  $\Delta_1$  l'amplitude de ce saut qui suit une loi exponentielle de paramètre  $\zeta$ . Ainsi, dans le modèle, l'impact d'un évènement catastrophique sur la mortalité après  $\tau$  années est donné par  $\Delta_1 e^{-\kappa \tau}$  où  $\Delta_1$  représente le saut initial. Si l'on suppose qu'un évènement affecte la mortalité sur une année seulement, c'est-à-dire que l'impact sur la mortalité au bout d'une année a diminué de 99 %, on a la relation :

$$\Delta_1 e^{-1 \cdot \kappa} = 0.01 \Delta_1$$

et donc  $\kappa = 4,6052$ .

Pour ajuster paramétriquement la distribution exponentielle des  $\Delta_i$ , on va utiliser les données présentées dans le Tableau 8. Les  $\Delta_i$  étant indépendants et identiquement distribués, on notera dans la suite  $\Delta$ . Cependant, l'excès de mortalité qui est présenté dans ce tableau est un excès multiplicatif sur les taux de mortalité et non additif sur les taux de hasard instantané comme dans le modèle de Bauer. Dans la suite, on note cet excès multiplicatif  $e$ . On recherche donc le lien entre  $\Delta$  et  $e$ . On a les relations :

$$\begin{aligned} (1+e) \tilde{q}_t &= 1 - e^{-\int_0^t \tilde{\mu}(x+s, t+s) + \Delta e^{-\kappa s} ds} \\ &= 1 - e^{-\frac{\Delta}{\kappa} (1-e^{-\kappa})} e^{-\int_0^t \tilde{\mu}(x+s, t+s) ds} \\ &= 1 - e^{-\frac{\Delta}{\kappa} (1-e^{-\kappa})} (1 - \tilde{q}_t) \end{aligned}$$

d'où :

$$\frac{\Delta}{\kappa} (1 - e^{-\kappa}) = \ln \left\{ \frac{1 - \tilde{q}_t}{1 - (1+e) \tilde{q}_t} \right\}$$

$$\Delta = \frac{\kappa}{(1-e^{-\kappa})} \ln \left\{ \frac{1-\tilde{q}_t}{1-(1+e)\tilde{q}_t} \right\} \quad (3)$$

avec  $\tilde{q}_t$  le taux de mortalité moyen (pondéré en fonction du sexe et de l'âge) de la population américaine à la date t, qui ne prend pas en compte l'effet des catastrophes. On notera  $q_t = (1+e)\tilde{q}_t$  le taux de mortalité classique qui tient compte des catastrophes. Le problème maintenant est de déterminer un taux de mortalité moyen pour les épidémies de 1918-1919, 1957, 1968, 1977 et 2003. Pour ces quatre dernières années, nous allons utiliser les données du site Internet de la *Human Mortality Database* qui fournit pour chaque année comprise entre 1933 et 2005 le nombre de décès, l'exposition et la répartition hommes / femmes par tranche d'âge de 5 ans. Pour chaque tranche d'âge, on calcule tout d'abord le taux de décès brut en divisant le nombre de décès par l'exposition. Puis on pondère ces taux par la répartition du nombre d'individus par âge. Une fois cette pondération faite pour les hommes et pour les femmes séparément, on calcule un taux global en pondérant par la répartition hommes / femmes de l'année. On obtient :

Année	Taux de mortalité $q_t$
1957	0,955 %
1968	0,964 %
1977	0,859 %
2003	0,837 %

Tableau 9 - Estimation du taux de mortalité pondéré aux Etats-Unis

Grâce à la formule  $q_t = (1+e)\tilde{q}_t$  et aux données concernant les excès de mortalité du Tableau 8, on obtient les taux de mortalité « hors catastrophe » suivants :

Année	Taux de mortalité $\tilde{q}_t$
1957	0,893 %
1968	0,947 %
1977	0,850 %
2003	0,811 %

Tableau 10 - Estimation du taux de mortalité "hors catastrophe" pondéré aux Etats-Unis

Pour la pandémie de 1918-1919, nous sommes confrontés à une difficulté. Il est très difficile de trouver des données concernant le nombre de décès, les expositions et la

composition de la population en 1918. On sait simplement que la pandémie de 1918-1919 a causé le décès de près de 500 000 personnes aux Etats-Unis d'après le site Internet *Death Toll from Disasters, War & Accident*. A titre d'information, dans un article pour le magazine « L'Histoire » (2003), Pierre Darmon estime le nombre de décès dus à la grippe espagnole à 128 000 pour l'année 1918 et à 52 500 pour l'année 1919 en France. D'après le Tableau 7 répertoriant les caractéristiques des pandémies survenues entre 1700 et 2000 dans l'étude de Swiss Re, on sait que la surmortalité additive sur le taux de mortalité de l'épidémie de 1918-1919 est évaluée à 0,53 % aux Etats-Unis. Le taux de mortalité « hors catastrophe » peut raisonnablement être estimé à 1,31 % en 1918 d'après plusieurs sources (*National Vital Statistics Reports*, vol. 54, no. 20, Aug. 21, 2007 par exemple. Voir aussi le graphique donnée en « Annexe 3. Mortalité aux Etats-Unis et en Europe entre 1918 et 1919 »). On trouve donc  $q_t$  en résolvant l'équation :

$$q_t = \tilde{q}_t + 0.53 \%$$

On a donc :  $q_{1918-1919} = 1,84 \%$  .

On obtient ainsi la valeur du paramètre  $\Delta$  à l'aide de la formule (3) :

Année	Paramètre $\Delta$
1918-1919	1,0878 %
1957	0,2902 %
1968	0,0792 %
1977	0,0431 %
2003	0,1228 %

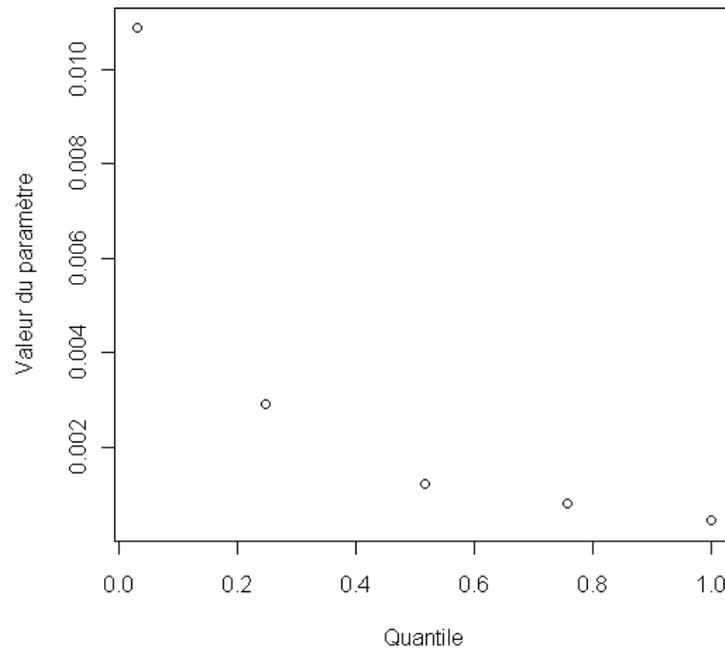
Tableau 11 - Valeur du paramètre  $\Delta$

Nous allons maintenant ajuster ces données aux quantiles d'une loi exponentielle à déterminer. Pour cela, on doit donc déterminer pour chaque pandémie à quel quantile elle correspond. Ces quantiles désignent des quantiles de la vraisemblance, c'est-à-dire le nombre de pandémies pires ou égales à celle considérée rapporté au nombre total de pandémie. Par exemple, la pandémie de grippe espagnole a été la plus grave en 420 ans. Etant donnée la fréquence annuelle de survenance d'une pandémie de 7,4 % qui a déjà été calculée précédemment, on arrive à un quantile à  $\frac{1}{420} \cdot \frac{1}{0.074} = 3,2 \%$  (VaR à 31 ans ou pandémie de période de retour de 31 ans). En fonction de la sévérité des autres pandémies, on détermine ainsi les quantiles suivants :

Année	Quantile	Paramètre $\Delta$
1918-1919	3,2 %	1,0878 %
1957	27,4 %	0,2902 %
1968	75,8 %	0,0792 %
1977	100 %	0,0431 %
2003	51,6 %	0,1228 %

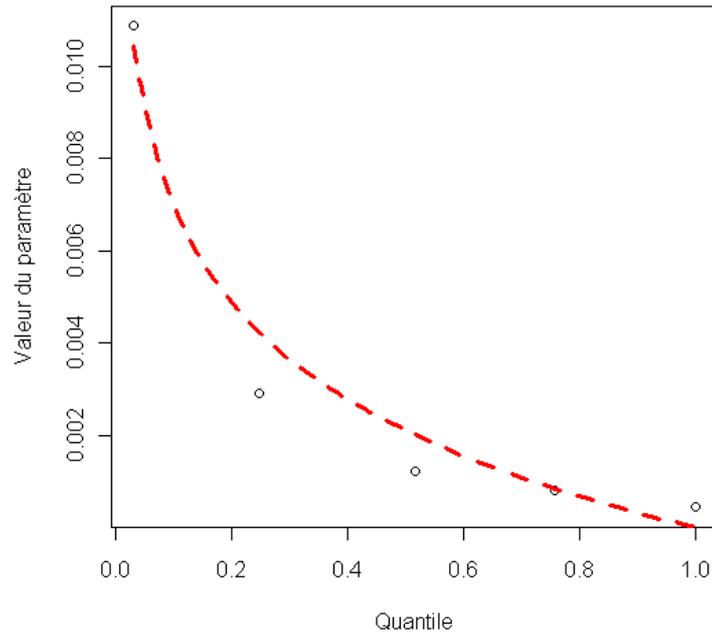
Tableau 12 - Quantiles des pandémies

On utilise le logiciel R pour l'ajustement. Voici tout d'abord la répartition des points dont on va se servir pour déterminer le paramètre de la loi exponentielle :



Graphique 2 – Valeur du paramètre  $\Delta$  en fonction des quantiles de vraisemblance

L'ajustement des données aux quantiles d'une loi exponentielle nous fournit une estimation du paramètre  $\zeta$  égale à 329,5. Voici la courbe ajustée aux données :



Graphique 3 - Valeur du paramètre  $\Delta$  en fonction des quantiles de vraisemblance avec ajustement

On remarque que l'ajustement à une loi exponentielle est plutôt discutable. A titre de comparaison, on rappelle ici les résultats trouvés dans l'article de Bauer et Kramer. Les auteurs ont utilisé une même structure de population pour toutes les pandémies pour pondérer les taux de mortalité à chaque âge. Ils ont choisi la structure de la population américaine de 1959 puis de 2003. Dans notre application en revanche, nous avons utilisé la structure de la population telle qu'elle l'était réellement lors de chaque pandémie prise en compte, ce qui explique en partie la différence des résultats. Par exemple, pour pondérer le taux de mortalité de 1968, nous avons utilisé directement la composition de la population de 1968 et non pas celle de 1959 ou de 2003.

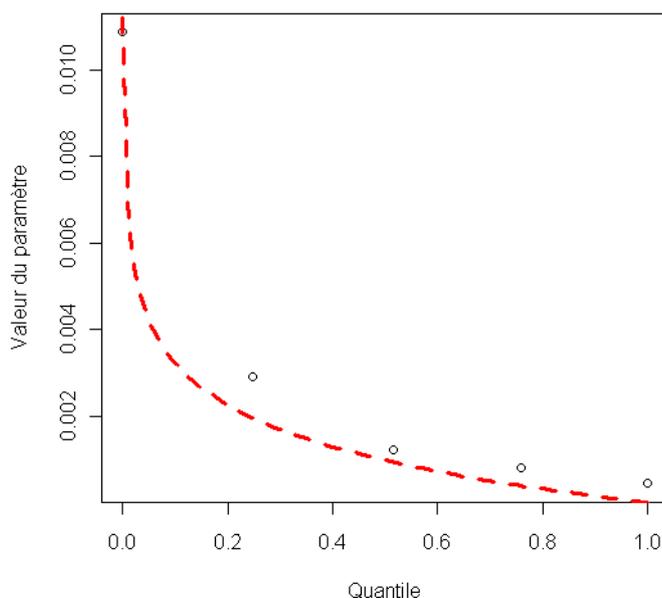
	Résultats de Bauer et Kramer		Rappel du résultat trouvé
	1959	2003	
$\zeta$	317,77	357,31	329,50
Moyenne $\frac{1}{\zeta}$	0,003147	0,002799	0,003035

Tableau 13 - Comparaison des résultats

Etant donné les années de survenance des pandémies utilisées pour calibrer le modèle, il est tout à fait logique que la valeur du paramètre soit plus proche de celle trouvée pour une structure de population de 1959 que de celle trouvée pour une structure de population de 2003. Ces différences de résultat nous amènent à nous demander quelle structure de population choisir. Nous en discuterons dans une prochaine partie. Pour le

moment, on peut revenir aux aspects réglementaires et calculer par exemple le taux de surmortalité additif pour une pandémie de sévérité comparable à une pandémie qui surviendrait une fois tous les deux cents ans. A partir de l'ajustement exponentiel que nous avons fait, on calcule facilement le quantile correspondant. La valeur trouvée pour le paramètre  $\Delta$  est de 1,61 %. Afin de comparer cela avec d'autres données, nous allons transformer cette valeur du paramètre  $\Delta$  en un taux de mortalité additif, toujours à l'aide de la formule (3). On obtient alors un taux de surmortalité additif de 0,32 % pour un taux de mortalité « hors catastrophe » fixé à 0,848 % (taux de mortalité en France en 2008). Ainsi, si une pandémie de sévérité comparable à une pandémie qui surviendrait une fois tous les deux cents ans survenait aujourd'hui en France, le taux de mortalité global sur l'année pourrait être estimé à 1,168 % (au lieu de 0,848 %).

Cette valeur du taux de surmortalité est assez élevée si l'on se souvient que l'étude Swiss Re calcule un taux compris entre 0,1 % et 0,15 %. Elle reste cependant cohérente avec les résultats d'autres études (cf : Tableau 5). Ceci n'est pas étonnant dans la mesure où l'étude Swiss Re a appliqué un retraitement particulier à la pandémie de 1918-1919. En effet, l'étude suggère de comparer la pandémie de 1918-1919 à une pandémie ayant un quantile de vraisemblance de  $\frac{1}{3000} = 0,033\%$  (VaR à 3000 ans. On rappelle que l'examen des fréquences historiques conduit à une VaR à 31 ans). L'argument principal présenté concerne surtout les évolutions médicales qui ont eu lieu entre 1918 et aujourd'hui qui réduisent considérablement le risque qu'une pandémie d'une telle ampleur se reproduise. Nous allons refaire l'estimation du paramètre de la loi exponentielle avec cette nouvelle pondération pour la pandémie de 1918-1919. L'ajustement des données aux quantiles d'une loi exponentielle nous fournit une estimation du paramètre  $\zeta$  égale à 714,2. Voici la courbe ajustée aux données :



Graphique 4 - Valeur du paramètre  $\Delta$  en fonction des quantiles de vraisemblance avec ajustement

On peut calculer de nouveau le taux de surmortalité additif pour une pandémie de sévérité comparable à une pandémie qui surviendrait une fois tous les deux cents ans. A partir de l'ajustement exponentiel on calcule le quantile correspondant. La valeur trouvée pour le paramètre  $\Delta$  est de 0,742 %. On transforme cette valeur en un taux de surmortalité additif et l'on obtient cette fois-ci 0,158 % pour un taux de mortalité « hors catastrophe » fixé à 0,848 % (taux de mortalité en France en 2008). Ainsi, si une pandémie de sévérité comparable à une pandémie qui surviendrait une fois tous les deux cents ans survenait aujourd'hui en France, le taux de mortalité global sur l'année pourrait être estimé à 1,006 % (au lieu de 0,848 %). Ce taux de surmortalité est plus cohérent avec ceux présentés par Swiss Re que le taux calculé précédemment. On remarquera donc que la détermination d'un taux de surmortalité est très sensible au poids que l'on attribue à la pandémie de 1918-1919. Afin de bien mettre ceci en évidence, nous présentons dans le tableau suivant le taux de surmortalité obtenu en fonction du poids accordé à la pandémie de grippe espagnole :

<b>Période de retour de la pandémie de 1918-1919</b>	<b>Taux de surmortalité à 99,5 %</b>
3 000 ans	0,158 %
750 ans	0,189 %
370 ans	0,210 %
75 ans	0,280 %
37 ans	0,328 %
7,4 ans	0,541 %

Tableau 14 - Taux de surmortalité à 99,5 % en fonction de la fréquence d'apparition de la pandémie de 1918-1919

Dans la suite, nous retiendrons plutôt la seconde application avec la valeur du paramètre de la loi exponentielle égale à 714,2 et un taux de surmortalité additif de 0,158 %.

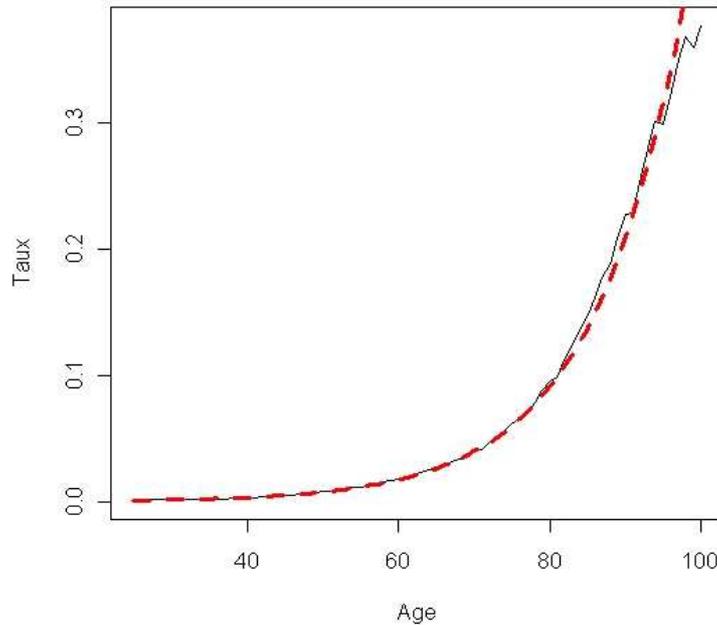
### 2.2.3. Le calibrage du modèle : la composante continue

Afin de comprendre l'influence de la composante à sauts dans la modélisation, nous allons estimer la composante continue. Pour cela, nous allons partir des données du site Internet de la *Human Mortality Database* qui fournit pour chaque année entre 1959 et 2003 aux Etats-Unis des taux de mortalité pour chaque âge :  $m(x'_i, t')$ . A partir de ces taux de mortalité, on en déduit une approximation de l'intensité de mortalité :

$$\begin{aligned}
\tilde{\mu}(x'_t + 0.5, t' + 0.5) &= \tilde{\mu}(x'_t + 0.5, t' + 0.5) \frac{\int_0^1 l(x'_t + s, t' + s) ds}{\int_0^1 l(x'_t + s, t' + s) ds} \\
&\approx \frac{\int_0^1 l(x'_t + s, t' + s) \mu(x'_t + s, t' + s) ds}{\int_0^1 l(x'_t + s, t' + s) ds} \\
&= m(x'_t, t')
\end{aligned}$$

où  $l(x'_t, t')$  représente l'exposition au risque des individus d'âge  $x'_t$  à la date  $t'$ .

Par soucis de simplicité, on posera  $t=0$  pour le 01/07/1959 et on a donc à notre disposition la valeur des intensités  $\tilde{\mu}(x_t, t)$  pour les âges  $x_t \in \{0.5, 1.5, \dots, 100.5\}$  et les années  $t \in \{1959.5, 1960.5, \dots, 2003.5\}$ . Dans la suite, on dira que « l'année 1959.5 » correspond à la période allant du 01/07/1959 au 30/06/1960. On rappelle que si l'on néglige la composante catastrophe à sauts, l'intensité est de la forme  $\mu(x_t, t) = Y_t \cdot e^{b \cdot x_t + c}$  et on choisit  $Y_0 = 1$ . On a alors  $\mu(x_0, t) = e^{b \cdot x_0 + c}$  qui n'est rien d'autre qu'un modèle de Gompertz. Nous allons donc tout d'abord déterminer les coefficients  $b$  et  $c$ . Nous nous restreindrons aux âges  $x_t \in \{25.5, 26.5, \dots, 100.5\}$  afin d'éviter que l'évolution particulière de la mortalité aux faibles âges n'altère trop l'estimation des paramètres d'un ajustement de type Gompertz. Pour estimer les coefficients, nous utilisons une régression linéaire du logarithme des taux de décès. On obtient les résultats suivants :  $b = 0,08290$  et  $c = -9,0300$ . Graphiquement l'ajustement donne :



Graphique 5 - Ajustement des taux de décès à une loi de Gompertz (année 1959.5)

Les coefficients  $b$  et  $c$  étant indépendants du temps, on a pour chaque année  $t$  la relation suivante :

$$\log[\tilde{\mu}(x, t)] - bx_t - c = \log(Y_t).$$

On peut alors estimer  $\log(Y_t)$  pour l'année  $t$  comme étant la moyenne des  $\log[\tilde{\mu}(x, t)] - bx_t - c$  pour tout  $x_t \in \{25.5, 26.5, \dots, 100.5\}$ . On trouve alors la forme de l'ajustement de Gompertz pour toutes les années. Par exemple, si l'on souhaite obtenir l'ajustement pour l'année 40 (1999.5), on calcule tout d'abord :

$$\exp\left\{\frac{1}{76} \sum_{i=25}^{100} \log[\tilde{\mu}(i+0.5, 40)] - b(i+0.5) - c\right\} = Y_{40} = 0,6684$$

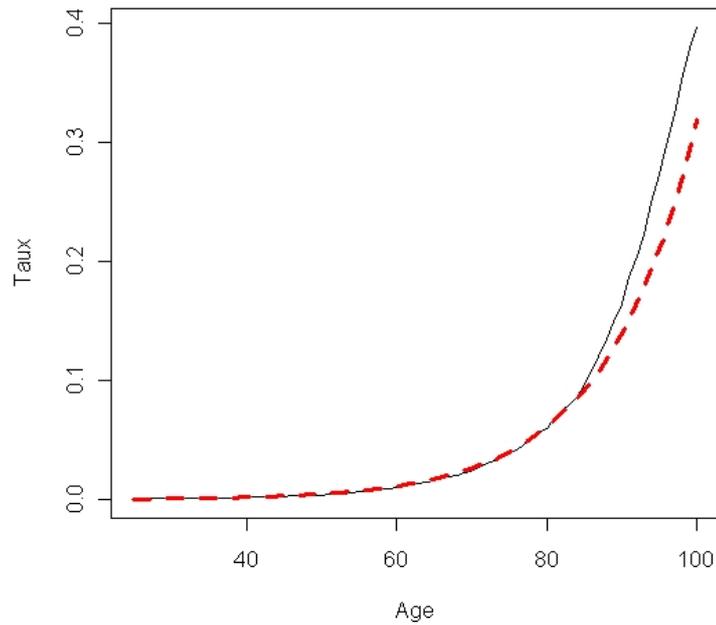
et on obtient donc facilement l'ajustement :

$$\mu(x, 40) = Y_{40} \exp(bx + c)$$

et plus généralement :

$$\mu(x, t) = Y_t \exp(bx + c)$$

Voici ce que l'on obtient graphiquement :

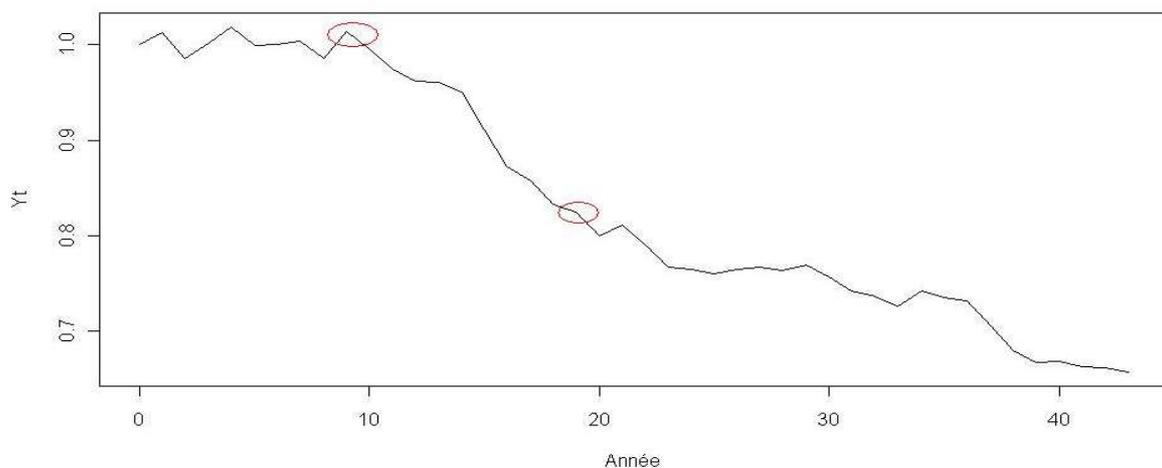


Graphique 6 - Ajustement des taux de décès à une loi de Gompertz (année 1999.5)

On remarque que l'ajustement reste très satisfaisant pour les âges inférieurs à 85 ans mais que la mortalité est sous-estimée aux âges élevés. Le but maintenant est d'estimer les paramètres de l'équation différentielle suivante :

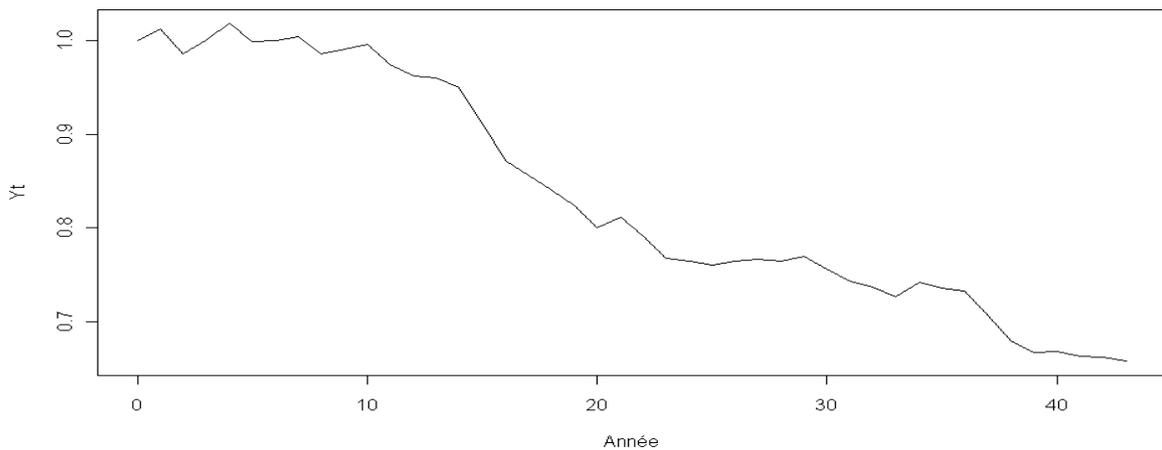
$$dY_t = \alpha(\beta - Y_t)dt + \sigma\sqrt{Y_t}dW_t$$

Pour cela, on va tout d'abord chercher à obtenir la série temporelle des  $Y_t$  pour  $t \in \{0, \dots, 43\}$  à l'aide de la procédure que nous avons décrite précédemment. On obtient le graphique suivant :



Graphique 7 - Série temporelle des  $Y_t$

On sait qu'entre 1959 et 2003, il y a eu deux pandémies « importantes » : la grippe de Hong-Kong de 1968 et la grippe « Russe » de 1978 (entourées sur le graphique précédent). Etant donné que nous souhaitons modéliser à part les pandémies, celles-ci ne doivent pas être prises en compte dans la composante continue. Nous devons donc supprimer ces deux pics de notre série temporelle. Pour cela, nous allons remplacer les deux points en question par le résultat d'une interpolation linéaire entre les deux points encadrant l'année 9 et l'année 18. Le choix de l'interpolation linéaire se fait naturellement étant donné la forme linéaire de la série temporelle sur un nombre réduit de points. On obtient la série ajustée suivante :



Graphique 8 - Série temporelle des  $Y_t$  ajustée

A partir de cette série, nous allons pouvoir déterminer les coefficients de l'équation  $dY_t = \alpha(\beta - Y_t)dt + \sigma\sqrt{Y_t}dW_t$  que l'on discrétise à l'aide d'un schéma d'Euler :

$$Y_{t+1} - Y_t = \alpha(\beta - Y_t) + \sigma\sqrt{Y_t}N_t$$

où  $N_t$ ,  $t = 0, \dots, T-1 = 42$  sont des lois normales centrées réduites indépendantes (loi des accroissements d'un mouvement brownien standard).

En retravaillant un peu cette équation, on obtient :

$$\sigma N_t = \frac{Y_{t+1}}{\sqrt{Y_t}} - \sqrt{Y_t} + \alpha\sqrt{Y_t} - \frac{\alpha\beta}{\sqrt{Y_t}}$$

qui suit une loi normale centrée de variance  $\sigma^2$ . Les trois premiers moments de cette loi sont 0,  $\sigma^2$  et 0. En prenant les estimateurs empiriques de ces trois moments, on doit résoudre le système suivant afin de déterminer les estimateurs des moments des paramètres  $\alpha$ ,  $\beta$  et  $\sigma$  :

$$\begin{cases} \frac{1}{T} \sum_{t=0}^{T-1} \left( \frac{Y_{t+1}}{\sqrt{Y_t}} - \sqrt{Y_t} + \alpha \sqrt{Y_t} - \frac{\alpha\beta}{\sqrt{Y_t}} \right) = 0 \\ \frac{1}{T-1} \sum_{t=0}^{T-1} \left( \frac{Y_{t+1}}{\sqrt{Y_t}} - \sqrt{Y_t} + \alpha \sqrt{Y_t} - \frac{\alpha\beta}{\sqrt{Y_t}} \right)^2 = \sigma^2 \\ \frac{1}{T} \sum_{t=0}^{T-1} \left( \frac{Y_{t+1}}{\sqrt{Y_t}} - \sqrt{Y_t} + \alpha \sqrt{Y_t} - \frac{\alpha\beta}{\sqrt{Y_t}} \right)^3 = 0 \end{cases}$$

On trouve les valeurs suivantes pour les estimateurs des moments :

$$\hat{\alpha} = 0,0102383, \hat{\beta} = 0,0504190, \hat{\sigma}^2 = 0,0105278$$

Nous allons utiliser ces valeurs comme valeurs de départ pour la détermination des paramètres à l'aide d'une méthode de maximisation de la log-vraisemblance. En effet, on connaît la loi de  $Y_t$  où  $Y_t$  suit un processus de Cox-Ingersoll-Ross. Cette variable aléatoire suit une loi du khi-deux décentré avec  $\delta = \frac{4\alpha\beta}{\sigma^2}$  degrés de liberté et un paramètre de non

centralité égal à  $\pi = \frac{Y_0}{\frac{\sigma^2}{4\alpha}(e^{\alpha} - 1)}$  (voir par exemple Planchet (2009)). On procède à la

maximisation de la log-vraisemblance et on obtient les valeurs suivantes des paramètres :

$$\hat{\alpha}^{EMV} = 3,8523 \times 10^{-9}, \hat{\beta}^{EMV} = 4,9580 \times 10^{-2}, \hat{\sigma}^{2EMV} = 2,5510 \times 10^{-3}$$

Afin de pouvoir mettre en évidence l'influence de la composante à sauts, nous allons estimer de nouveau ces coefficients sans lisser la série temporelle ci-dessus, c'est-à-dire en conservant les deux évènements catastrophiques identifiés. On trouve alors les valeurs suivantes pour les estimateurs des moindres carrés :

$$\hat{\alpha} = 0,0102824, \hat{\beta} = 0,0528465, \hat{\sigma}^2 = 0,0093859$$

puis pour les estimateurs du maximum de vraisemblance :

$$\hat{\alpha}^{EMV} = 4,3052 \times 10^{-9}, \hat{\beta}^{EMV} = 7,8174 \times 10^{-2}, \hat{\sigma}^{2EMV} = 2,5510 \times 10^{-3}$$

#### 2.2.4. Test du modèle et mise en évidence de l'utilité de la composante à sauts

Maintenant que nous avons en notre possession tous les paramètres estimés, nous pouvons simuler quelques probabilités de survie afin de tester le modèle. On rappelle que l'on a une expression « simple » pour  ${}_{T-t}p_x^{(t)}$  qui représente la probabilité pour qu'un individu d'âge  $x_t$  en  $t$  soit encore en vie à la date  $T$  :

$${}_{T-t}p_{x_t}^{(t)} = \exp\left\{u(T-t) + v(T-t)Y_t e^{b(x_0+t)+c}\right\} \exp\left\{-\frac{\Gamma_t}{\kappa}\left(1 - e^{-\kappa(T-t)}\right) - \frac{\lambda(T-t)}{\zeta\kappa+1}\right\}$$

$$\exp\left\{\frac{\lambda\zeta}{\zeta\kappa+1} \log\left[1 + \frac{1}{\zeta\kappa}\left(1 - e^{-\kappa(T-t)}\right)\right]\right\}$$

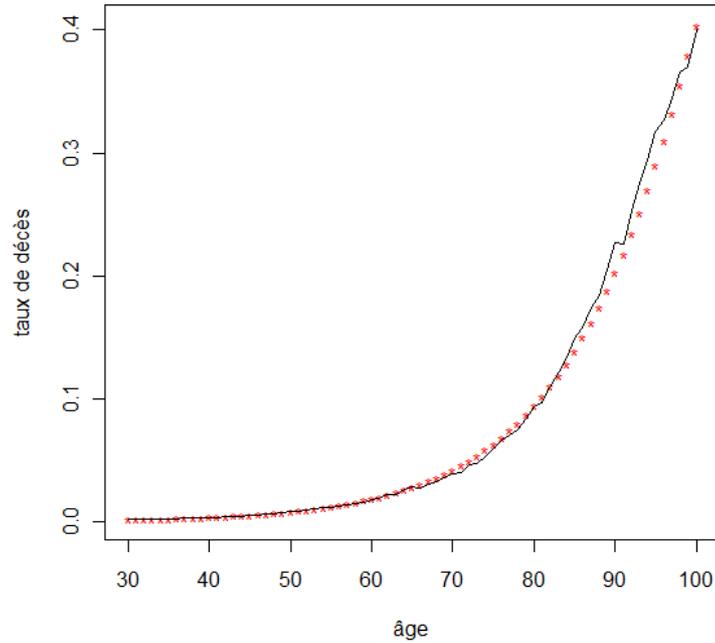
où  $u$  et  $v$  satisfont les équations différentielles suivantes qui sont résolues en annexe :

$$u'(x) = \alpha\beta v(x) e^{b(x_0+T-x)+c}, u(0) = 0$$

$$v'(x) = -1 - (\alpha - b)v(x) + \frac{1}{2}\sigma^2 v^2(x) e^{b(x_0+T-x)+c}, v(0) = 0.$$

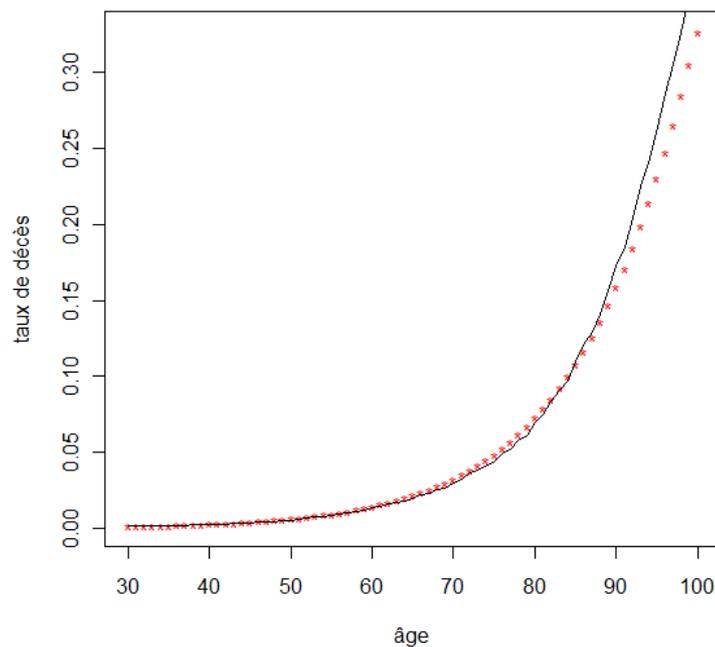
La quantité  ${}_{T-t}p_{x_t}^{(t)}$  est une variable aléatoire, nous allons donc simuler plusieurs réalisations et en prendre la moyenne pour approcher l'espérance de cette variable aléatoire. Pour pouvoir comparer la quantité obtenue directement avec des courbes de mortalité par âge et par année, on choisira  $t$  et  $T$  tels que  $T - t = 1$  pour avoir la probabilité de survie (et donc facilement la probabilité de décès) dans l'année. En ce qui concerne les âges, on choisira la plage d'âge  $x_t \in \{25.5, 26.5, \dots, 100.5\}$  étant donné que les coefficients du modèle de Makeham ont été estimés sur cette plage. Cela signifie qu'en pratique, nous serons assez limités. En effet, si l'on choisit la valeur minimale possible pour  $x_0$ , c'est-à-dire 25.5 ans (ce qui signifie que le 01/07/1959, l'individu en question a 25.5 ans), nous pourrions obtenir au plus une estimation de la probabilité de survie d'un individu âgé de 100.5 ans au 01/07/2034. Au 01/07/2030, nous pourrions obtenir des probabilités de survie pour les individus âgés de 96.5 à 100.5 ans et ainsi de suite. Pour obtenir des plages d'âge plus larges pour ces années, il faudrait prendre un  $t = 0$  à une date plus récente et refaire le même raisonnement que celui que nous avons présenté pour estimer les paramètres de la composante continue. Cependant, cela conduirait à utiliser moins de données pour produire les estimations.

Afin de vérifier la qualité du modèle, nous allons comparer les valeurs prédites des taux de mortalité de ce modèle aux taux de mortalité réellement observés. Pour avoir une plage d'âge assez importante, on choisira l'année d'observation allant du 01/07/1964 au 30/06/1965 afin de pouvoir observer des individus âgés entre 30.5 et 100.5 ans. Nous comparerons ces résultats obtenus aux taux de mortalité de l'année 1965. Voici le graphique que l'on obtient :

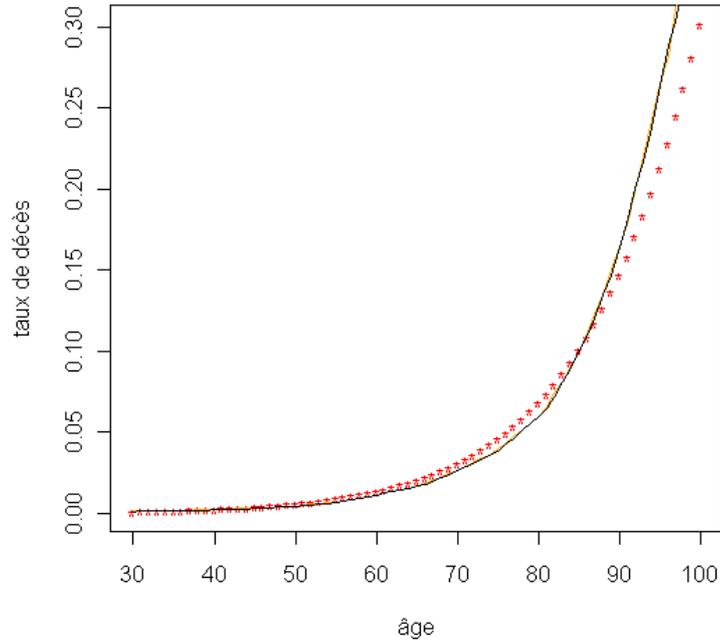


Graphique 9 - Comparaison taux de décès observés et théoriques 1964-1965

On a représenté en noir la courbe correspondant à l'année 1965 et avec des étoiles la courbe obtenue à l'aide du modèle qui intègre la composante catastrophe pour l'année d'observation allant du 01/07/1964 au 30/06/1965. A première vue, il semblerait que le modèle soit assez proche des taux réels, avec une légère surestimation pour la tranche d'âge 60-80 ans et une sous-estimation pour la tranche d'âge 85-100 ans. On peut tracer ces mêmes courbes pour d'autres années :

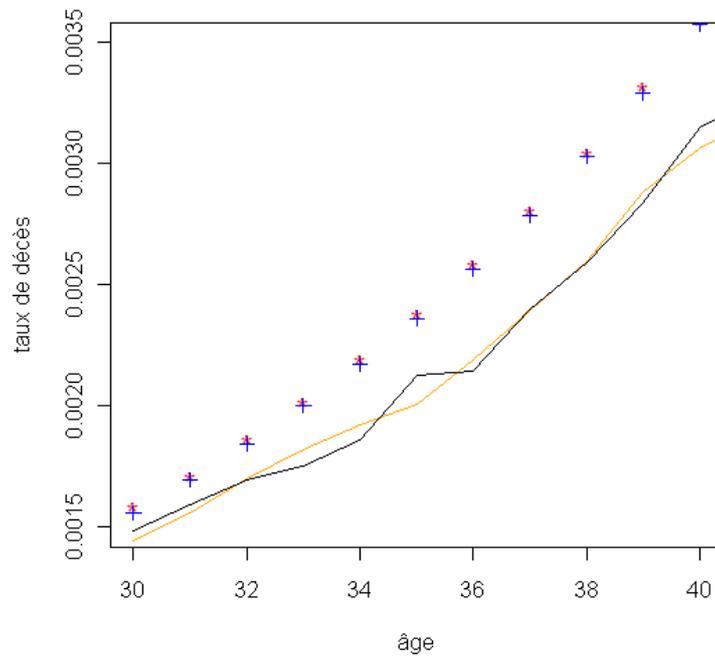


Graphique 10 - Comparaison taux de décès observés et théoriques 1985-1986



Graphique 11 - Comparaison taux de décès observés et théoriques 1995-1996

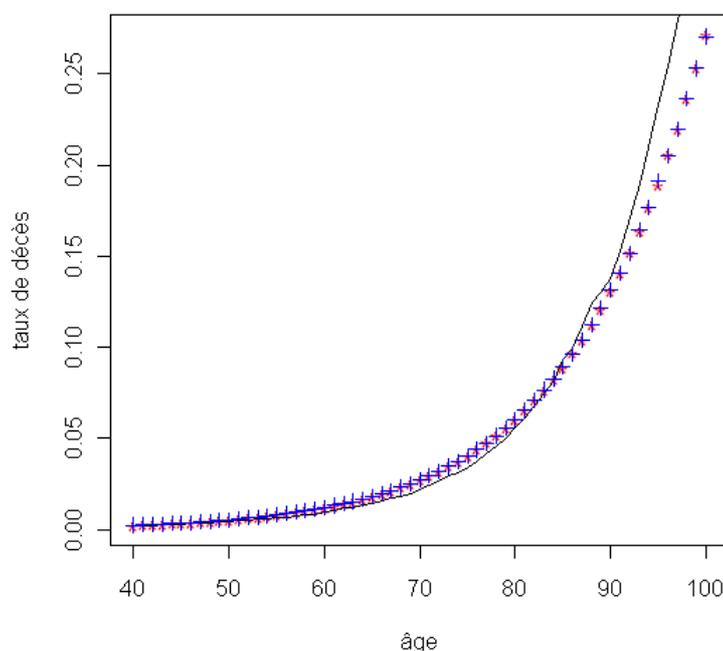
Comme nous l'avons déjà remarqué, plus on avance dans le temps, moins le modèle est adapté aux âges élevés. En effet, le modèle de Makeham n'est pas un modèle adéquat pour rendre compte du phénomène d'orthogonalisation des tables. Afin de mettre en évidence le rôle de la composante catastrophe, nous allons regarder de plus près la tranche d'âge 30-40 ans en revenant à notre premier exemple pour l'année 1964-1965 :



Graphique 12 - Comparaison taux décès observés et théoriques sur la plage d'âge 30-40 ans (1964-1965)

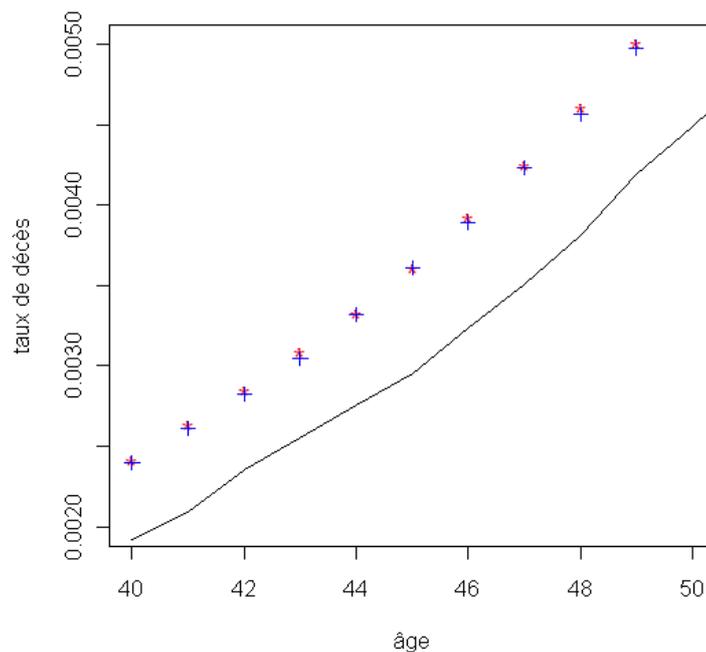
On a toujours en noir la courbe correspondant à l'année 1965 et l'on a rajouté en orange la courbe correspondant à l'année 1964. Les croix représentent la courbe obtenue à l'aide du modèle sans la composante catastrophe pour l'année d'observation allant du 01/07/1964 au 30/06/1965 et les étoiles la courbe obtenue à l'aide du modèle avec la composante catastrophe. Ce graphique nous permet de soulever un premier inconvénient : l'écart entre la courbe incluant la composante « catastrophe » et la courbe « hors catastrophe » est très faible en comparaison avec l'écart observé avec les taux réels. En effet, on calcule un écart relatif de l'ordre de 0,20 % entre les taux de décès théoriques avec et sans la composante catastrophe et un écart relatif de l'ordre de 4,9 % entre les taux de décès théoriques sans la composante catastrophe et les taux de décès réellement observés en 1964.

On rappelle que les données historiques utilisées pour calculer les paramètres du modèle couvrent la période 1959-2003. Nous pouvons donc tester les capacités prédictives du modèle en simulant des taux de décès pour la période allant du 01/07/2005 au 30/06/2006 et comparer les résultats avec les taux effectivement observés en 2005 :



Graphique 13 - Comparaison taux de décès observés et théoriques 2005

On a représenté avec une courbe les taux de décès réellement observés en 2005. Les croix représentent les taux de décès théoriques sans la composante catastrophe et les étoiles les taux de décès théoriques avec la composante catastrophe pour l'année d'observation allant du 01/07/2005 au 30/06/2006. Ces deux dernières composantes sont très proches si bien qu'il est difficile de les distinguer à cette échelle. On peut les discerner si l'on regarde de plus près la tranche d'âge 40-50 ans par exemple :



Graphique 14 - Comparaison des taux de décès observés et théoriques sur la plage d'âge 40-50 ans (2005)

Les écarts entre les taux théoriques et réels sont encore plus marqués pour cette année d'observation. On calcule un écart relatif de l'ordre de 0,12 % entre les taux de décès théoriques avec et sans la composante catastrophe et un écart relatif de l'ordre de 9,1 % entre les taux de décès théoriques sans la composante catastrophe et les taux de décès réellement observés en 1964.

### 2.2.5. Critique du modèle

Dans ce paragraphe, nous allons dresser un bilan des difficultés qu'entraîne l'utilisation d'un tel modèle.

Commençons tout d'abord par le calibrage de la composante de base. Pour estimer les paramètres de la composante de base de  $\mu(x,t)$  à savoir  $Y_t \cdot e^{b \cdot x + c}$ , on s'appuie sur des données historiques qui incluent nécessairement les événements catastrophiques. Or ceux-ci doivent être exclus puisqu'ils seront estimés dans une composante à part. Un travail préliminaire doit donc être fait pour que les événements exceptionnels ne perturbent pas l'estimation des paramètres de la série temporelle  $Y_t$ . Ce retraitement s'accompagne nécessairement de décisions arbitraires pour lisser la série temporelle  $Y_t$  (dans notre cas nous avons décidé d'utiliser une interpolation linéaire).

Un autre inconvénient que l'on peut citer réside dans le choix de  $Y_t = (Y_t)_t$ . En effet en pratique, on n'observe pas de retour vers la moyenne dans la série temporelle  $Y_t$ . Ceci pose le problème du choix de la structure du modèle. De plus, d'après une étude de Luciano (2005), il est acquis que les processus de type retour vers la moyenne ne rendent pas correctement compte du phénomène d'orthogonalisation des tables observé pour les

générations les plus récentes qui traduit le fait que de plus en plus de personnes décèdent à un âge élevé.

En ce qui concerne le calibrage de la composante catastrophe, plusieurs autres problèmes se posent. Tout d'abord, le choix du nombre de catastrophes qui sont utilisées pour calibrer le modèle. Dans notre cas, on a utilisé cinq points ce qui est très peu pour une estimation. Nous avons d'ailleurs bien remarqué graphiquement que l'ajustement à une loi exponentielle est plutôt mauvais avec un nombre si réduit de points. On ne peut cependant pas obtenir plus de points par manque de données sur les pandémies antérieures à la grippe espagnole de 1918-1919.

De plus, comme nous l'avons déjà mentionné, Linfoot (2007) répertorie 31 pandémies dans les 420 dernières années ce qui conduit à une fréquence annuelle de 7,4 %. On rappelle que dans l'étude Swiss Re que nous avons présentée précédemment, 10 à 13 pandémies ont été répertoriées pour les 300 dernières années ce qui conduit à une fréquence annuelle comprise entre 3 % et 4 %. Toujours pour l'estimation des paramètres, nous avons dû faire certaines hypothèses assez arbitraires telles que par exemple l'hypothèse qu'une catastrophe n'affecte la mortalité que pendant une année. Nous avons aussi soulevé le problème du choix de la structure de la population pour choisir le taux de surmortalité additif. Choisir l'estimation obtenue avec une structure de population de 2003 serait plutôt cohérent compte de tenu de la structure de la population actuelle. Mais on peut aussi penser que choisir une structure de population semblable à celle de 1959 serait plus cohérent avec la structure de la population lors de la survenance d'une pandémie.

Nous avons aussi observé que le poids accordé à l'épidémie de grippe espagnole de 1918-1919 change considérablement les résultats obtenus. En effet, l'étude de Swiss Re propose une période de retour de 3000 ans tandis que Bauer et Kramer n'appliquent aucun retraitement particulier à la pandémie de grippe espagnole et conservent une période de retour de 31 ans. Le taux de surmortalité passe successivement de 0,158 % à 0,32 % dans ces deux cas de figure. Or cette fréquence d'apparition ne peut être déterminée avec certitude puisqu'elle est basée notamment sur des hypothèses de progrès médicaux.

Lors de la simulation de taux de mortalité à l'aide du modèle, nous avons remarqué que l'effet de la composante catastrophe est assez faible en comparaison avec l'erreur commise sur la tendance si bien que l'utilité de cette composante peut être remise en question face au problème plus important de l'estimation de la composante de base.

On pourrait proposer de modéliser cette tendance d'une manière différente de ce qui a été fait par Bauer et Kramer, à l'aide d'un modèle de Lee-Carter par exemple. C'est ce que nous allons faire dans la partie suivante.

### Section 2.3. Amélioration du modèle de Bauer et Kramer : tendance de type « Lee-Carter »

Dans cette partie, nous allons reprendre l'idée principale du modèle de Bauer et Kramer mais allons apporter une légère modification en estimant la tendance par un modèle de Lee-Carter. Cette méthode est classique pour déterminer des taux instantané de mortalité, nous en rappelons les grandes lignes.

#### 2.3.1. Modèle de Lee-Carter

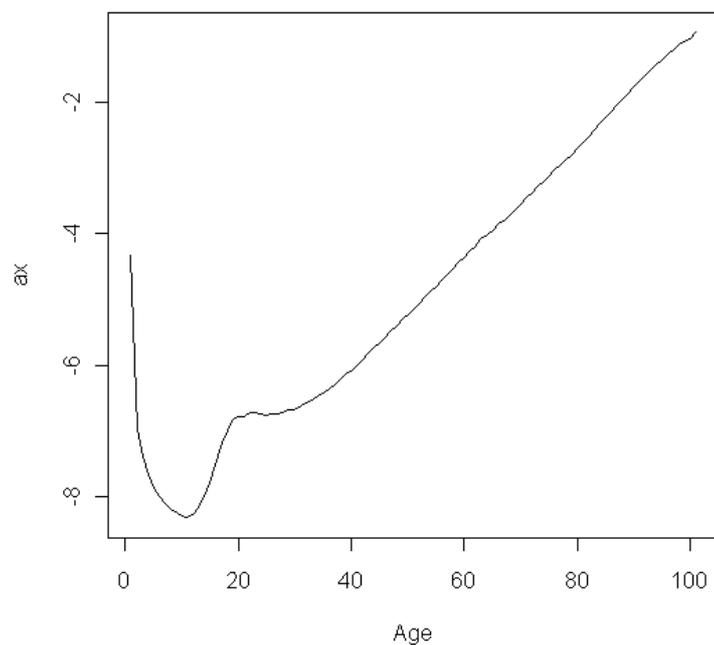
La modélisation retenue pour le taux instantané de mortalité est la suivante :

$$\ln \mu_{xt} = \alpha_x + \beta_x k_t + \varepsilon_{xt}$$

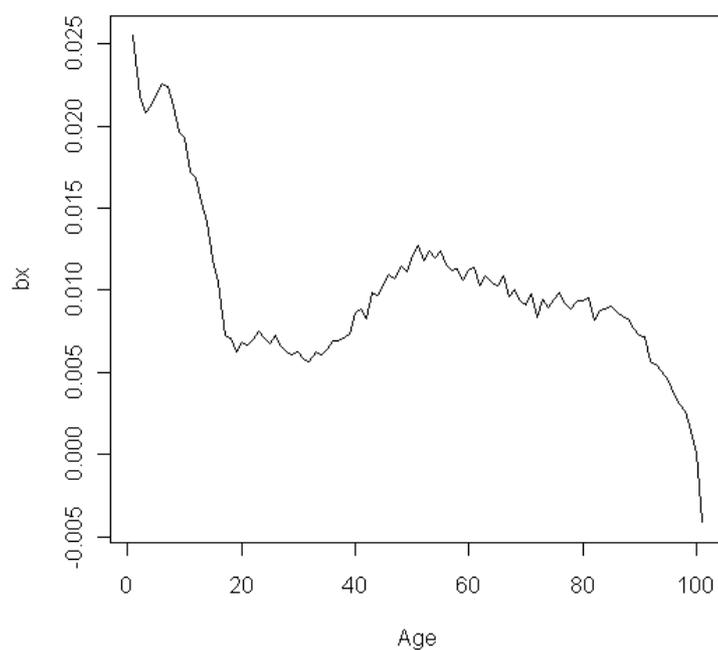
où les  $\varepsilon_{xt}$  sont des variables aléatoires qui suivent une loi  $N(0, \sigma^2)$  indépendantes et identiquement distribuées. Le but est donc d'ajuster la série temporelle du logarithme des taux instantanés de décès indicée par l'âge  $x$  et le temps  $t$  à une structure paramétrique déterministe à laquelle s'ajoute un phénomène aléatoire. Nous ne reviendrons pas sur la question de l'ajustement des paramètres du modèle, le lecteur intéressé pourra consulter notamment Planchet et Thérond (2006). Le modèle de Lee-Carter est implicitement décrit comme étant un modèle en temps continu. Cependant, nous disposons de données discrètes, il convient donc d'adapter le modèle. Pour cela, l'hypothèse classique consiste à dire que la fonction de hasard est constante sur chaque carré du diagramme de Lexis afin d'obtenir la formule simple  $\mu_{xt} = -\ln(1 - q_{xt})$ . Cependant, cette hypothèse n'est en pratique pas vérifiée pour les âges élevés. Nous allons donc décrire le modèle directement en temps discret en utilisant les logits des taux de décès  $\ln\left(\frac{q_{xt}}{1 - q_{xt}}\right)$  (les logits pour des taux de mortalité faibles sont très peu différents de la variable  $\ln(\mu_{xt})$ ). On trouve alors facilement une expression du taux de décès à l'aide de la transformation inverse de la fonction logistique  $y \rightarrow \frac{e^y}{1 + e^y}$ .

On part toujours de nos données de la population américaine observée entre 1959 et 2003. On va utiliser la fonction *lca* du package *demography* de R pour estimer les paramètres du modèle de Lee-Carter. Avant tout comme précédemment, nous devons supprimer l'effet des pandémies de 1968 et de 1978 de nos données afin d'obtenir un modèle « hors catastrophe » auquel on rajoutera par la suite la même composante catastrophe que celle que nous avons utilisée dans la première version de notre modèle. On utilise toujours une interpolation linéaire entre les taux encadrant les deux années pour lesquelles on observe une catastrophe.

On obtient les résultats suivants :



Graphique 15 - Valeurs de alpha en fonction de l'âge par la méthode de Lee-Carter

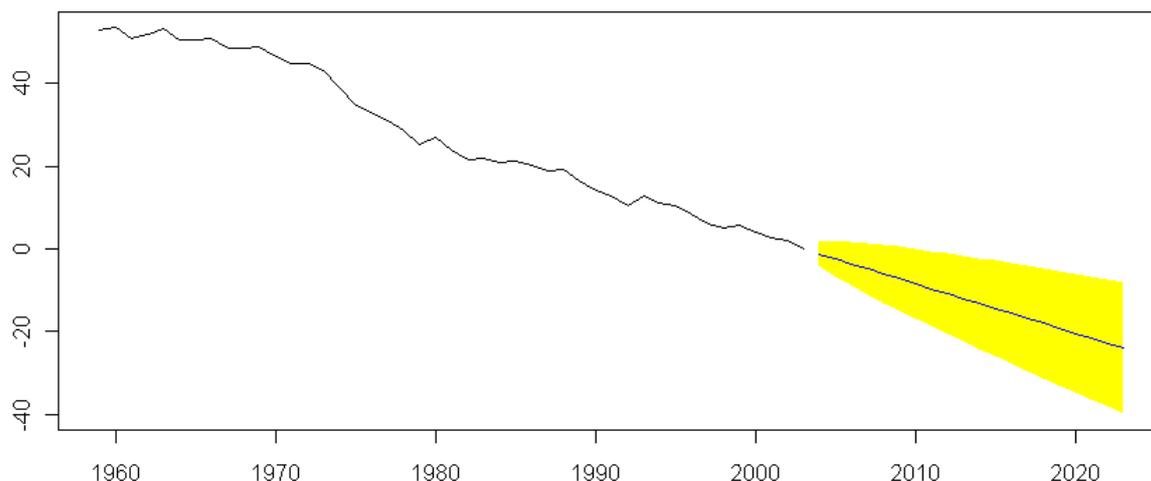


Graphique 16 - Valeurs de beta en fonction de l'âge par la méthode de Lee-Carter

En ce qui concerne la composante temporelle ( $k_t$ ), il est nécessaire de la modéliser afin de pouvoir extrapoler les taux futurs. On se propose généralement d'utiliser une simple régression linéaire de type :

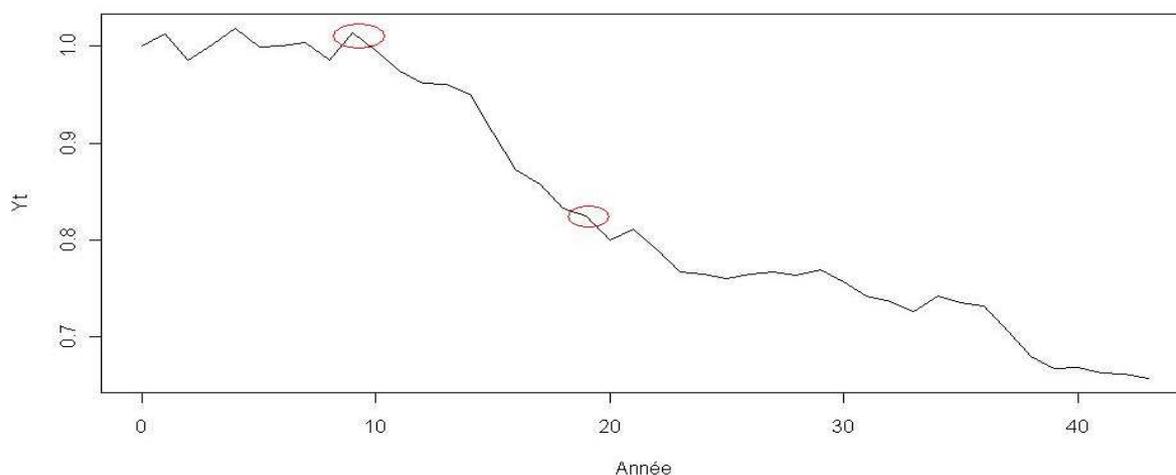
$$k_t = at + b + \varepsilon_t$$

où  $(\varepsilon_t)$  est un bruit blanc gaussien. On obtient alors le graphique suivant :



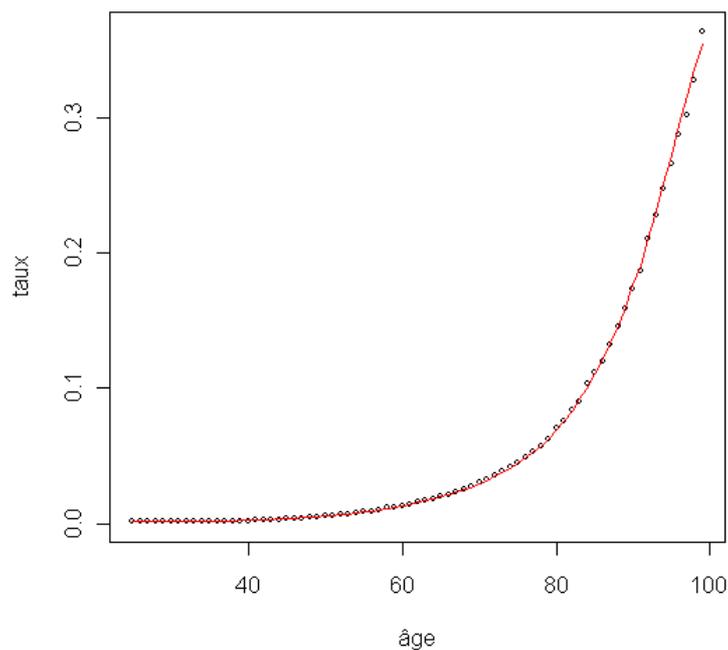
Graphique 17 - Valeurs de  $k$  en fonction de l'année par la méthode de Lee-Carter et extrapolation

La zone en jaune correspond à l'intervalle de confiance à 95 %. On remarquera la très grande similitude des séries  $k_t$  du modèle de Lee-Carter et  $Y_t$  du modèle de Cox-Ingersoll-Ross que nous avons traité précédemment et que nous reprenons ici pour mémoire :



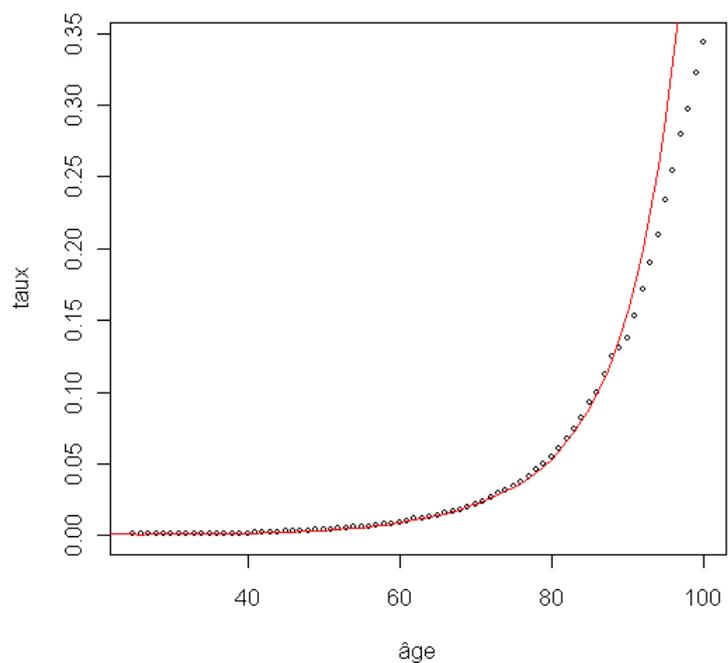
Graphique 18 - Série temporelle des  $Y_t$  du modèle de Cox-Ingersoll-Ross

A partir de l'estimation des paramètres, nous pouvons alors tester la qualité du modèle en traçant les courbes des taux de mortalité théoriques et observés pour différentes années d'observation. Ces taux de mortalité théoriques se retrouvent simplement en utilisant la transformation inverse du logit  $\frac{1}{1+e^{-z}}$  sur les valeurs estimées par le modèle. Pour l'année 1985 par exemple, on a le graphique suivant :



Graphique 19 - Comparaison taux de décès observés et théoriques 1985

On a tracé la courbe des taux théoriques. Les points représentent quant à eux les taux réels observés en 1985. On remarque que contrairement au modèle précédent, le phénomène d'orthogonalisation est bien pris en compte ici. On peut aussi tracer la comparaison pour l'année 2005 qui était particulièrement mauvaise dans le modèle précédent. Pour faire cela ici, on a besoin de la valeur des coefficients  $k_t$  pour les années postérieures à 2003 qui ont été extrapolés. On obtient le graphique suivant :



Graphique 20 - Comparaison taux de décès observés et théoriques 2005

Les résultats sont nettement meilleurs que ceux présentés dans le modèle précédent (cf. Graphique 13) où le modèle s'écartait notablement des valeurs observées à partir de 50 ans. Ici, l'écart ne devient important qu'à partir de 85 ans. Pour quantifier cette amélioration, on peut calculer la déviation entre les décès observés et les décès théoriques à l'aide de la distance classique du khi-deux dans les deux modèles :

$$\frac{(D_{x,t} - D_{x,t}^*)^2}{D_{x,t}^*}$$

où  $D_{x,t}$  représente le nombre de décès observés à l'âge  $x$  et pour l'année  $t$  et  $D_{x,t}^*$  représente les décès théoriques. Cette statistique n'est utilisée ici que comme mesure de l'écart entre les taux ajustés et les taux bruts et non comme une statistique de test. La valeur globale de la déviation pour l'année  $t$  est égale à la somme des khi-deux locaux pour tous les âges. Pour le premier modèle, on trouve une valeur de 39,29 tandis que pour le second modèle que l'on propose, cette valeur devient 5,48. Cette plus faible valeur de la déviation confirme le fait que le second modèle est mieux adapté.

Nous allons maintenant ajouter la composante catastrophe à la composante continue. Ici, l'expression du taux de mortalité sera plus simple que ce qui a été fait dans le modèle précédent. On part toujours de  ${}_{T-t}p_{x_t}^{(t)} = E\left[1_{\tau_{x_t} > T} | G_t\right] = E\left[\exp\left\{-\int_t^T \mu(x_0 + s, X_s) ds\right\} | G_t\right]$  avec  $G_t = \sigma\{X_s, 0 \leq s \leq t\}$  et  $\mu(x_t, t) = \mu(x_t, Y_t, \Gamma_t) = Y_t + \Gamma_t$ . Ici,  $Y = (Y_t)_t = (\mu_{x_t})_t$  du modèle de Lee-Carter et on conserve  $\Gamma = (\Gamma_t)_t$  qui suit l'équation différentielle suivante :

$$d\Gamma_t = -\kappa \Gamma_t dt + dJ_t.$$

L'indépendance des deux composantes  $Y$  et  $\Gamma$  nous permet d'écrire :

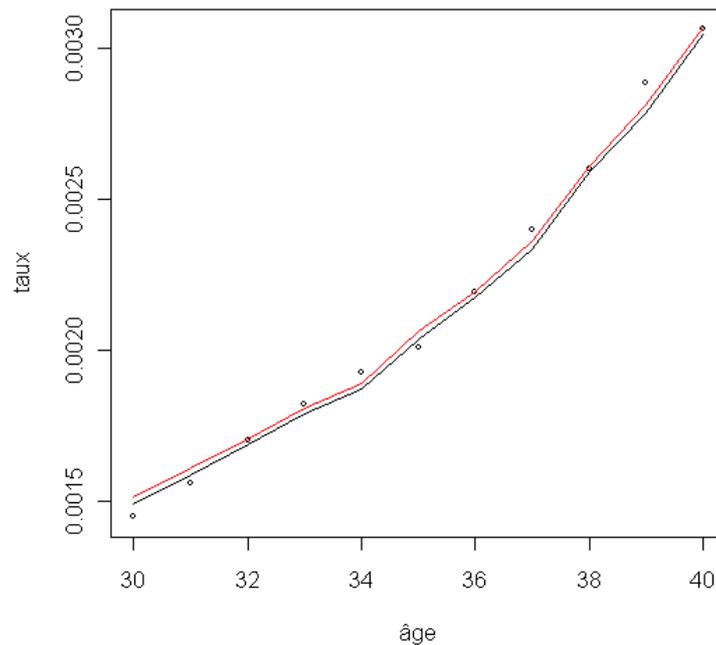
$$\begin{aligned} {}_{T-t}p_{x_t}^{(t)} &= E\left[\exp\left\{-\int_t^T Y_s + \Gamma_s ds\right\} | G_t\right] \\ &= E\left[\exp\left\{-\int_t^T Y_s ds\right\} | G_t\right] E\left[\exp\left\{-\int_t^T \Gamma_s ds\right\} | G_t\right] \end{aligned}$$

Le calcul de la seconde composante est identique à ce qui a été fait précédemment. En ce qui concerne la composante continue, nous recherchons l'expression du taux de décès à partir du logarithme du taux de mortalité instantané extrapolé à l'aide du modèle de Lee Carter. Comme nous l'avons déjà évoqué, on utilisera le fait que les logits des taux de décès sont très proches des logarithmes des taux instantanés de mortalité. Pour nos applications, on impose l'égalité suivante  $T - t = 1$  et on a donc la formule finale suivante :

$$p_{x_t}^{(t)} = \left(1 - q_{x_t}^{(t)(LeeCarter)}\right) \exp\left\{-\frac{\Gamma_t}{\kappa} (1 - e^{-\kappa}) - \frac{\lambda}{\zeta \kappa + 1}\right\} \exp\left\{\frac{\lambda \zeta}{\zeta \kappa + 1} \log\left[1 + \frac{1}{\zeta \kappa} (1 - e^{-\kappa})\right]\right\}$$

Comme nous l'avons fait dans le modèle précédent, nous allons approcher l'espérance de la variables aléatoire  $T-t P_x^{(t)}$  en simulant plusieurs réalisations de cette variable et en prenant la moyenne de ces réalisations.

On peut reproduire le Graphique 12 avec ce nouveau modèle, c'est-à-dire représenter sur le même graphique les taux de mortalité observés, théoriques hors catastrophes et théoriques avec catastrophe pour l'année 1964 et les individus âgés de 30 à 40 ans :



Graphique 21 - Comparaison des taux de décès observés et théoriques en 1964 pour les individus âgés entre 30 et 40 ans

Les ronds correspondent toujours aux données réellement observées en 1964, la courbe supérieure rouge aux taux déterminés par le modèle qui inclut la composante catastrophe et la courbe inférieure bleue les taux déterminés par le modèle sans la composante catastrophe.

Cette fois-ci, l'écart observé entre les taux réels et théoriques est très faible et ne rend donc pas inutile l'utilisation d'une composante catastrophe.

Les modifications apportées au modèle initial de Bauer sur la composante continue améliorent grandement les qualités du modèle. Cependant, bien que la composante continue ait été améliorée, la composante catastrophe présente quelques faiblesses que nous avons déjà évoquées. On notera surtout que dans ce modèle, on ne s'intéresse qu'au risque de catastrophe provoquée par une pandémie. Cette modélisation ne prend pas en compte les autres catastrophes naturelles ou d'origine humaine. De plus, la composante catastrophe est indépendante de l'âge ce qui n'est en réalité pas raisonnable car même s'il est vrai qu'il n'y a aucune raison pour laquelle un tremblement de terre toucherait plus une

tranche d'âge qu'une autre, une pandémie en revanche ne touche pas tous les individus de manière uniforme en fonction de leur âge.

Dans cette partie, nous avons passé en revue les deux approches principales de modélisation de la surmortalité suite à une pandémie. L'approche épidémiologique se base plus particulièrement sur des études scientifiques à caractère médical ou biologique. Elles font donc intervenir des avis d'experts qui apportent un jugement particulièrement précieux dans le cadre de projections. Cependant, les résultats sont très sensibles aux hypothèses retenues. Doyle et al. (2005) dans leur étude pour l'INVS, affirment eux-mêmes que leur modèle épidémiologique « *a été construit sur un ensemble d'hypothèses dont la vraisemblance est inconnue* » et que les résultats sont « *à prendre avec précaution* ».

Les modèles « actuariels » de mortalité stochastique avec composante de sauts du même genre que Bauer et Kramer accordent plus d'importance aux données historiques. Nous avons vu que pour que la composante de choc soit utile, la tendance doit être particulièrement bien estimée afin de ne pas rendre obsolète cette composante de chocs. Concernant cette composante, nous avons aussi remarqué que le calibrage est assez peu robuste étant donné le faible nombre de pandémies pour lesquelles on dispose de suffisamment d'informations et la sensibilité des résultats au poids accordé à la grippe espagnole. On rappelle que pour une période de retour supposée de 3000 ans pour la grippe espagnole (préconisée par Swiss Re), on trouve un taux de surmortalité de 0,158 %. En revanche, si l'on se base sur les fréquences historiques de survenance des pandémies, la période de retour de la pandémie de grippe espagnole de 1918-1919 est de 31 ans ce qui conduit à un taux de surmortalité de 0,32 %. En ce qui concerne le choix de modélisation de la tendance, il est lui aussi délicat. Choisir un processus de retour vers la moyenne comme dans le modèle de Bauer et Kramer n'est pas réellement adapté étant donné que l'on n'observe pas de retour vers la moyenne des taux de mortalité en pratique. De manière générale, l'utilisation de modèles affines pour estimer la mortalité est extrêmement délicate (voir par exemple Gourieroux et Monfort (2007)). Le calibrage est également difficile dans la mesure où les catastrophes doivent être supprimées des données d'origine pour estimer les paramètres de la tendance. Dans la présentation de son modèle lors du *Mortality Workshop* du 1<sup>er</sup> février 2008 organisé par AXA, le CREST et la Fondation du Risque, Bauer a lui-même affirmé que « *due to high parameter uncertainty it's hard to judge which results are "more accurate"* » et a invité les acteurs de l'assurance à une précaution toute particulière dans l'interprétation des résultats.

Dans la partie suivante, nous proposerons un modèle qui évite une majorité de ces inconvénients et qui permet de prendre en compte tous les types de catastrophes sans se restreindre aux pandémies.

### ***PARTIE III :***

#### ***UNE APPROCHE ORIGINALE : UN MODELE DE LISSAGE A CHOCS***

Comme nous l'avons vu dans la partie précédente, une approche épidémiologique ou stochastique de type Bauer soulève beaucoup de difficultés pour modéliser l'effet des catastrophes sur la mortalité. Dans cette partie, nous allons proposer une autre approche basée sur des techniques de lissage. Le but est de trouver à partir d'une surface brute de mortalité une surface sous-jacente lisse qui reflète l'évolution « normale » de la mortalité et une composante supplémentaire qui permet de refléter les chocs identifiés sur cette surface. A partir de cette surface de surmortalité, nous pourrions calculer le taux correspondant à une catastrophe de période de retour de 200 ans. Pour effectuer ce lissage par splines, on recourra à une modélisation probabiliste qui conduira naturellement à un modèle linéaire généralisé dont la résolution sera facilitée par l'utilisation de modèles linéaires généralisés matriciels.

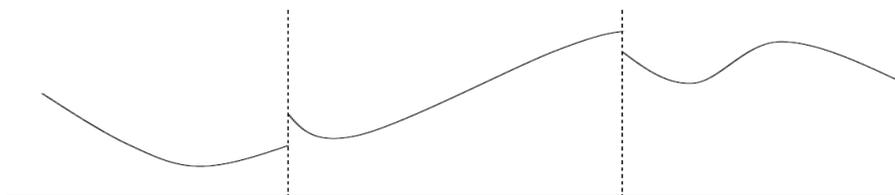
## CHAPITRE 1. LE LISSAGE PAR SPLINES

Dans ce chapitre, nous allons aborder plus en détail les techniques de lissage par splines. Pour une synthèse des méthodes statistiques de lissage dans les espaces fonctionnels, le lecteur intéressé pourra consulter Besse et Cardot (2001). Et pour une application à l'estimation de lois pour le maintien en arrêt de travail, le lecteur peut également consulter Planchet et Winter (2006).

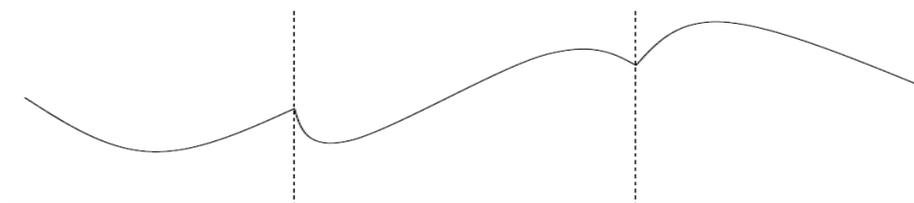
Lors de l'estimation de taux bruts, on observe des irrégularités qui ne reflètent pas exactement le phénomène que l'on souhaite observer et qui sont dues principalement aux conditions de l'expérience et aux fluctuations d'échantillonnage qui viennent perturber les résultats. Un lissage permet donc de gommer ces irrégularités et de représenter de manière plus fidèle le phénomène que l'on souhaite étudier.

### Section 1.1. Quelques généralités

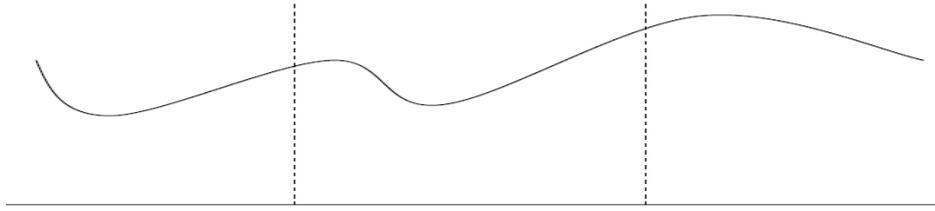
L'idée principale d'un lissage par splines est d'ajuster sur chaque sous-intervalle de la fonction à lisser une fonction simple et de raccorder ces différentes fonctions sur chaque sous-intervalle en imposant des conditions de régularité aux bornes. En pratique, on utilise souvent des polynômes. On aura donc sur chaque sous-intervalle des arcs à raccorder aux points de jonction (appelés aussi « nœuds »). Pour un polynôme de degré 3, on veillera à ce que les raccordements aux points de jonction conservent la continuité, l'égalité des pentes et des courbures. Voici différentes fonctions splines possibles :



Graphique 22 - Fonction spline discontinue au niveau des nœuds



Graphique 23 - Fonction spline continue



Graphique 24 - Fonction spline de dérivée première et seconde continues

Nous nous intéresserons dans la suite à ce dernier cas. Prenons l'exemple d'un spline cubique (polynôme de degré 3) à deux arcs. On suppose que l'on découpe notre intervalle initial  $[x_0; x_2]$  en deux sous intervalles  $[x_0; x_1]$  et  $[x_1; x_2]$ . Notre but est de trouver le spline  $q_x$  qui lisse « le mieux » les taux bruts à notre disposition. Pour cela, on suppose que  $q_x$  s'écrit :

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

où  $p_0$  et  $p_1$  sont des polynômes de degré 3. On impose les conditions suivantes au point de jonction  $x_1$  :

$$\begin{cases} p_0(x_1) = p_1(x_1) \\ \frac{d}{dx} p_0(x_1) = \frac{d}{dx} p_1(x_1) \\ \frac{d^2}{dx^2} p_0(x_1) = \frac{d^2}{dx^2} p_1(x_1) \end{cases}$$

On posera donc  $p_0(x) = c_1 + c_2x + c_3x^2 + c_4x^3$  et  $p_1(x) = p_0(x) + c_5(x - x_1)^3$  où  $c_1, c_2, c_3, c_4, c_5$  sont des coefficients à déterminer. Ces coefficients peuvent être déterminés à l'aide d'un estimateur des moindres carrés pondéré. Nous y reviendront par la suite.

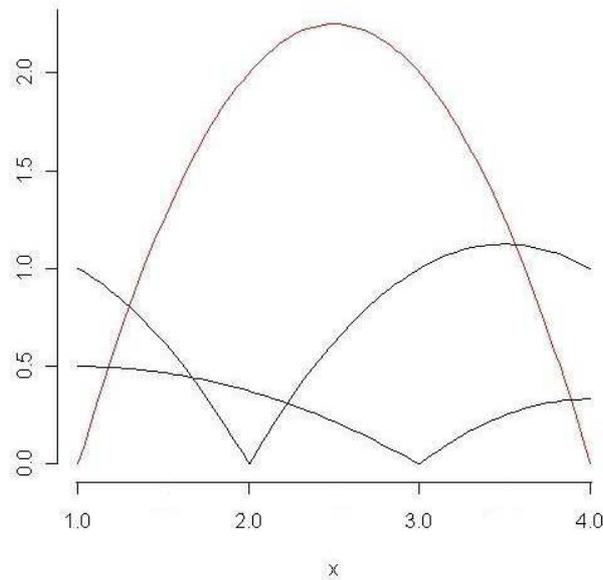
## Section 1.2. Définition d'un B-spline

Un B-spline est une combinaison de polynômes raccordés entre eux d'une manière spécifique. Dans le cadre que nous avons présenté, un B-spline sera donc le résultat du lissage de notre fonction d'origine.

Un B-spline de degré  $q$  est composé de  $q+1$  polynômes, chacun de degré  $q$ . On compte donc  $q$  nœuds auxquels se rejoignent les polynômes et à chacun de ces nœuds, les dérivées successives (jusqu'à la  $q-1^{\text{ème}}$ ) sont continues. Un B-spline est positif sur un domaine couvert par  $q+2$  nœuds. Partout ailleurs, il est nul. Excepté aux bornes de l'intervalle, un B-spline recouvre  $2q$  polynômes des B-splines voisins. Pour un  $x$  donné différent des nœuds, on a  $q+1$  B-splines non nuls.

Prenons un intervalle ayant pour bornes  $x_{\min}$  et  $x_{\max}$  et divisons le en  $n'$  intervalles de même amplitude à l'aide de  $n'+1$  nœuds. Chaque sous intervalle sera donc couvert par  $q+1$  B-splines de degré  $q$ .

Voici une illustration d'un B-spline de degré 2 sur l'intervalle  $[1; 4]$  accompagné de parties de B-splines voisins ayant pour frontière un nœud du B-spline principal :



Graphique 25 - Exemple de B-spline

On a bien 2 nœuds aux points 2 et 3 et le B-spline est positif entre 1 et 4. Pour un  $x$  donné différent de 1, 2, 3 ou 4, on a bien 3 B-splines non nuls. Dans la suite, on exigera de plus qu'à chaque nœud, la somme des valeurs prises par les B-splines qui couvrent ce nœud soit égale à 1. En dehors des nœuds, cette somme peut-être différente de 1.

On notera désormais  $B_j(x; q)$  la valeur au point  $x$  du  $j^{\text{ème}}$  B-spline de degré  $q$  pour une répartition équidistante de nœuds. La courbe lissée  $\hat{y}$  des données  $(x_i; y_i)$  est la combinaison linéaire  $\hat{y}(x) = \sum_{j=1}^n \hat{a}_j B_j(x; q)$ . L'indice de sommation nous permet de relier un B-spline à un nœud. Pour le B-spline  $B_j(x; q)$ , on associe le nœud  $x_j$  qui est le point à partir duquel le B-spline devient non nul.

De Boor (1978) nous donne une formule simple pour dériver des B-splines :

$$\begin{aligned} h \sum_j a_j B_j'(x; q) &= \sum_j a_j B_j(x; q-1) - \sum_j a_{j+1} B_{j+1}(x; q-1) \\ &= - \sum_j \Delta a_{j+1} B_j(x; q-1) \end{aligned}$$

où  $h$  est la distance entre les nœuds et  $\Delta a_j = a_j - a_{j-1}$ .

Pour des dérivées de second ordre, on a la relation :

$$h^2 \sum_j a_j B_j''(x; q) = \sum_j \Delta^2 a_j B_j(x; q-2)$$

avec  $\Delta^2 a_j = \Delta \Delta a_j = a_j - 2a_{j-1} + a_{j-2}$ .

### Section 1.3. Pénalités

En pratique, on est confronté au problème du choix des nœuds. Un nombre de nœuds trop important conduirait à un « sur-ajustement » et à l'inverse un nombre de nœuds trop faible pourrait conduire à un « sous-ajustement ». Ce problème est difficile à résoudre et aujourd'hui, aucun algorithme ne permet d'optimiser le nombre et l'emplacement des nœuds. O'Sullivan (1986, 1988) propose d'utiliser initialement un très grand nombre de nœuds et, afin d'éviter un « sur-ajustement », préconise l'utilisation de pénalités sur la dérivée seconde afin de restreindre la souplesse de la courbe ajustée.

Supposons que l'on ait  $m$  points  $(x_i; y_i)$  et que l'on souhaite lisser la courbe à l'aide de  $n$  B-splines. On détermine les coefficients  $a_j$  à l'aide d'un critère des moindres carrés. La fonction objectif à minimiser est donc :

$$S = \sum_{i=1}^m \left\{ y_i - \sum_{j=1}^n a_j B_j(x_i) \right\}^2 .$$

Les pénalités sont directement introduites sur la fonction objectif. O'Sullivan (1986, 1988) propose d'introduire une pénalité du type :

$$S = \sum_{i=1}^m \left\{ y_i - \sum_{j=1}^n a_j B_j(x_i) \right\}^2 + \lambda \int_{x_{\min}}^{x_{\max}} \left\{ \sum_{j=1}^n a_j B_j''(x) \right\}^2 dx .$$

Eilers et Marx (1996) quant à eux proposent de retenir la forme suivante :

$$S = \sum_{i=1}^m \left\{ y_i - \sum_{j=1}^n a_j B_j(x_i) \right\}^2 + \lambda \sum_{j=k+1}^n (\Delta^k a_j)^2$$

où  $\Delta^k$  est l'opérateur de différence d'ordre  $k$ .

On retiendra cette dernière forme qui réduit les dimensions du problème, les faisant passer de  $m$ , le nombre d'observations, à  $n$ , le nombre de B-splines. Le paramètre  $\lambda$  nous permet de garder un contrôle sur le degré de la pénalité, on appellera ce paramètre le

paramètre de pénalité. On cherche donc à minimiser  $S$  à l'aide des  $a_j$  dans une logique de lissage par moindres carrés.

Le système d'équations du modèle retenu nous conduit à la représentation matricielle suivante :

$$\mathbf{B}'\mathbf{y} = (\mathbf{B}'\mathbf{B} + \lambda\mathbf{D}'_k\mathbf{D}_k)\mathbf{a}$$

où  $\mathbf{D}_k$  est la représentation matricielle de l'opérateur de différenciation  $\Delta^k$ ,  $\mathbf{a}' = (a_1, \dots, a_n)$ ,  $\mathbf{y}' = (y_1, \dots, y_m)$  et les éléments de  $\mathbf{B}$  sont les  $b_{ij} = B_j(x_i)$ .

La forme de cette représentation n'est pas anodine, elle ressemble très fortement à la représentation d'un modèle linéaire généralisé, c'est pourquoi on essaye de se ramener à une modélisation probabiliste.

Si l'on se place donc dans le cadre d'un modèle linéaire généralisé, on introduit un prédicteur linéaire  $\eta_i = \sum_{j=1}^n b_{ij}a_j$  et la fonction de lien canonique  $\eta_i = g(\mu_i)$  où  $\mu_i$  est l'espérance de  $y_i$ . On soustrait la pénalité à la log-vraisemblance  $l(\mathbf{y}; \mathbf{a})$  pour obtenir la log-vraisemblance pénalisée :

$$L = l(\mathbf{y}; \mathbf{a}) - \frac{\lambda}{2} \sum_{j=k+1}^n (\Delta^k a_j)^2.$$

L'optimisation de  $L$  conduit au système d'équations suivant écrit sous forme matricielle :

$$\mathbf{B}'(\mathbf{y} - \boldsymbol{\mu}) = \lambda\mathbf{D}'_k\mathbf{D}_k\mathbf{a}.$$

On reconnaît ici l'écriture matricielle que l'on retrouve dans les modèles linéaires généralisés avec un terme supplémentaire jouant le rôle de pénalité et la matrice  $\mathbf{B}$  jouant le rôle de la matrice de régression  $\mathbf{X}$ .

Dans le chapitre suivant, nous allons revenir rapidement sur les modèles linéaires généralisés et sur leur résolution avant de présenter les avantages de l'utilisation des modèles linéaires généralisés matriciels.

## CHAPITRE 2. RAPPELS SUR LES MODELES LINEAIRES ET EXTENSION

Dans ce paragraphe, nous allons présenter quelques rappels concernant les modèles linéaires généralisés ainsi qu'une variante très utile à savoir les modèles linéaires généralisés matriciels.

### Section 2.1. Le modèle linéaire généralisé (GLM)

Un modèle linéaire généralisé (GLM) est défini par :

- un prédicteur linéaire  $\eta_i = \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle$  formé à partir des variables explicatives regroupées dans le vecteur  $\mathbf{x}_i$  et du vecteur des paramètres  $\boldsymbol{\theta}$  (pour un individu  $i$ ) ;
- une loi de la famille exponentielle qui sera la loi de la variable réponse  $Y_i$ . On rappelle qu'une densité associée aux paramètres  $\gamma$  (localisation) et  $\phi$  (dispersion) appartient à la famille exponentielle si elle s'écrit sous la forme :

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

Pour cette famille de loi, on rappelle sans démonstration (voir Saporta et al. (2005)) que l'on a les relations suivantes :  $b'(\gamma_i) = \mu_i = EY_i$  et  $Var(Y_i) = b''(\gamma_i)a(\phi)$  ;

- une fonction lien  $g$ , monotone et dérivable telle que  $\mu_i = EY_i$ , avec la notation  $g(\mu_i) = \eta_i$ . La fonction de lien donne la forme de la relation entre l'espérance de la variable expliquée et les variables explicatives. Ceci donne plus de souplesse au modèle et permet aussi de faciliter le calcul de la vraisemblance en choisissant correctement cette fonction compte tenu de la loi de probabilité qui a été attribuée à la variable expliquée. Si  $\eta_i = \gamma_i$ , on dit que  $g$  est le lien canonique de la loi utilisée.

Les paramètres sont ensuite estimés à l'aide de la log-vraisemblance  $\ln L = \sum_{i=1}^n l(\gamma_i, \phi, y_i)$ . Dans le cas du lien canonique avec  $\eta_i = \gamma_i$ , on obtient :

$$\frac{\partial l_i}{\partial \theta_j} = \frac{\partial l_i}{\partial \gamma_i} \frac{\partial \eta_i}{\partial \theta_j}$$

$$\frac{\partial l_i}{\partial \gamma_i} = \frac{y_i - b'(\gamma_i)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)}$$

$$\frac{\partial \eta_i}{\partial \theta_j} = x_i^{(j)} \quad \forall j.$$

D'où les équations de vraisemblance (sous forme matricielle) :

$$\mathbf{X}'(\mathbf{y} - \boldsymbol{\mu}) = 0.$$

Si l'on ne se situe pas dans le cadre d'un lien canonique, on obtient :

$$\frac{\partial l_i}{\partial \theta_j} = \frac{\partial l_i}{\partial \gamma_i} \frac{\partial \gamma_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \theta_j}.$$

Or :

$$\frac{\partial \mu_i}{\partial \gamma_i} = b''(\gamma_i) = \frac{\text{Var}(Y_i)}{a(\phi)}$$

Et on obtient donc les équations de vraisemblance suivantes :

$$\sum_{i=1}^n \frac{(y_i - \mu_i) x_i^{(j)}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad \forall j.$$

L'écriture matricielle de ce système d'équations est la suivante :

$$\mathbf{X}'\mathbf{W}_\delta \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}}(\mathbf{y} - \boldsymbol{\mu}) = 0$$

où  $\mathbf{W}_\delta$  est une matrice diagonale avec les éléments suivants sur sa diagonale :

$$w_{ii} = \left( \frac{\partial \mu_i}{\partial \eta_i} \right)^2 \text{Var}(Y_i)^{-1}$$

et  $\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\mu}}$  est aussi une matrice diagonale avec les éléments  $\frac{\partial \eta_i}{\partial \mu_i}$  sur sa diagonale.

Pour la plupart des modèles linéaires généralisés, les équations qui déterminent les paramètres au sens du maximum de vraisemblance sont non linéaires et les estimateurs n'ont pas d'autres expressions formulables que comme solutions de ces équations. Les logiciels calculent les estimations en utilisant un algorithme itératif pour la résolution d'équations non linéaires. Cet algorithme est généralement l'algorithme des scores de Fisher (voir par exemple Lavergne et Trottier (2000)).

Le système  $\mathbf{X}'\mathbf{W}_\delta \frac{\partial \eta}{\partial \boldsymbol{\mu}}(\mathbf{y} - \boldsymbol{\mu}) = 0$  n'est bien entendu pas linéaire en  $\boldsymbol{\theta}$  (qui intervient dans  $\boldsymbol{\mu}$ ) et peut donc être résolu par la procédure itérative des scores de Fisher. Cet algorithme peut être décrit comme la résolution itérative des équations normales :

$$\mathbf{X}'\mathbf{W}_\delta(\mathbf{z} - \mathbf{X}\boldsymbol{\theta}) = 0 \quad (1)$$

$$\text{où } \mathbf{z} = \boldsymbol{\eta} + \frac{\partial \eta}{\partial \boldsymbol{\mu}}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\theta} + \frac{\partial \eta}{\partial \boldsymbol{\mu}}(\mathbf{y} - \boldsymbol{\mu}).$$

A  $\boldsymbol{\theta}$  fixé,  $\mathbf{z}$  est considéré comme un nouveau vecteur de données,  $\mathbf{W}_\delta^{-1}$  comme une nouvelle matrice de poids et on reconnaît alors dans le système (1) les équations classiques des moindres carrés généralisés associées au modèle  $\mathbf{Z} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$  où  $E(\boldsymbol{\varepsilon}) = 0$ . La méthode d'itération des scores de Fisher procède aux itérations :

$$\boldsymbol{\theta}^{[k+1]} = (\mathbf{X}'\mathbf{W}_\delta\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}_\delta\mathbf{z}$$

$$\text{avec } \mathbf{z} = \mathbf{X}\boldsymbol{\theta}^{[k]} + \frac{\partial \eta^{[k]}}{\partial \boldsymbol{\mu}}(\mathbf{y} - \boldsymbol{\mu}^{[k]}).$$

On arrête l'algorithme lorsque l'écart entre  $\boldsymbol{\theta}^{[k+1]}$  et  $\boldsymbol{\theta}^{[k]}$  est jugé suffisamment faible.

Enfin, on rappelle que l'on peut être amené à utiliser les variances de  $\mathbf{X}\hat{\boldsymbol{\theta}}$  qui correspondent aux éléments diagonaux de  $\mathbf{X}(\mathbf{X}'\mathbf{W}_\delta\mathbf{X})^{-1} \mathbf{X}'$ .

## Section 2.2. Le modèle linéaire généralisé matriciel (*GLAM*)

Pour l'étude de certains phénomènes, les données sont disponibles « naturellement » sous forme de matrices. Par exemple pour des études de santé, on peut avoir à notre disposition des déclarations de maladie « rangées » par année, par âge ou par mois dans des tableaux. Dans le cas de l'étude de la mortalité, on a souvent à notre disposition des matrices qui contiennent le nombre de décès par âge et par année et les expositions par âge et par année. Les modèles linéaires généralisés matriciels permettent de tirer profit de la structure en tableaux des données. Ce type de modèle n'est en réalité rien d'autre qu'un moyen technique pour résoudre des *GLM* et n'apporte pas de nouveauté conceptuelle. Dans la suite, nous présenterons quelles sont les conditions requises par le *GLM* pour être résolu à l'aide de techniques de type « modèle linéaire généralisé matriciel » ainsi que les différents aspects techniques qui facilitent la détermination des coefficients du modèle. Pour une présentation plus détaillée de ce type de modèle, le lecteur intéressé pourra consulter Currie et al. (2006).

On parlera de modèles linéaires généralisés matriciels (*Generalized linear array model : GLAM*) lorsque les données sont présentées sous forme de tableaux et lorsque la

matrice du modèle (matrice des variables explicatives encore appelée matrice de régression) peut-être écrite comme un produit de Kronecker. On rappelle que le produit de Kronecker de deux matrices ( $\mathbf{A}$  de dimension  $m \times n$  et  $\mathbf{B}$  de dimension  $p \times q$ ) s'écrit (voir Rotella et al. (1995)) :

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} \cdots a_{1n}\mathbf{B} \\ \vdots \quad \ddots \quad \vdots \\ a_{m1}\mathbf{B} \cdots a_{mn}\mathbf{B} \end{pmatrix}.$$

Ce type de modèle est particulièrement adapté pour le lissage multidimensionnel. D'un point de vue computationnel, il permet aussi une utilisation réduite de la mémoire de l'ordinateur et est beaucoup plus rapide que la résolution du *GLM* pour des modèles ayant un nombre élevé de paramètres et d'observations.

Afin de se familiariser avec ces concepts, nous allons détailler le cas de la dimension 1 et de la dimension 2.

### 2.2.1. Le cas unidimensionnel

On se place toujours dans le même cadre que les modèles linéaires généralisés avec la donnée d'un prédicteur linéaire, d'une loi de la famille exponentielle pour la variable expliquée et d'une fonction de lien.

On soulignera tout d'abord le fait qu'un *GLAM* unidimensionnel n'est pas exactement la même chose qu'un *GLM* ordinaire bien que ces deux modèles partagent le même objectif, à savoir estimer la fonction linéaire  $\mathbf{X}\boldsymbol{\theta}$  (transformation d'un vecteur à l'aide d'une unique matrice). On notera cependant deux différences importantes : la matrice de poids  $\mathbf{W}_\delta$  dans le *GLM* devient un vecteur de poids  $\mathbf{w}$  dans le *GLAM* et les calculs de  $\mathbf{X}(\mathbf{X}'\mathbf{W}_\delta\mathbf{X})^{-1}\mathbf{X}'$  et de  $\mathbf{X}'\mathbf{W}_\delta\mathbf{X}$  sont simplifiés et peuvent se réduire à la transformation d'un vecteur par une matrice unique dans le *GLAM*. Établissons l'écriture alternative de  $\mathbf{X}'\mathbf{W}_\delta\mathbf{X}$ . Soit  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_c)$  une matrice de dimension  $n \times c$  et  $\mathbf{w}$  le vecteur de poids qui correspond aux éléments diagonaux de  $\mathbf{W}_\delta$ . Le  $(j, k)$ <sup>ème</sup> élément de  $\mathbf{X}'\mathbf{W}_\delta\mathbf{X}$  est donné par :

$$\begin{aligned} (\mathbf{X}'\mathbf{W}_\delta\mathbf{X})_{jk} &= \sum_{i=1}^n w_i x_{ij} x_{ik} \\ &= (x_{1j}x_{1k}, \dots, x_{nj}x_{nk}) \mathbf{w} \quad (1) \\ &= (\mathbf{x}_j * \mathbf{x}_k)' \mathbf{w} \end{aligned}$$

où l'opérateur  $*$  désigne le produit élément par élément. L'élément  $(\mathbf{x}_j * \mathbf{x}_k)'$  correspond au produit des colonnes  $j$  et  $k$  de  $\mathbf{X}$ . Par soucis de simplicité, nous allons introduire un

nouvel opérateur qui effectue cette opération que nous appellerons le tenseur colonne. Le tenseur colonne d'une matrice  $\mathbf{X}$  possédant  $c$  colonnes est défini de la manière suivante :

$$G(\mathbf{X}) = (\mathbf{X} \otimes \mathbf{1}') * (\mathbf{1}' \otimes \mathbf{X})$$

où  $\mathbf{1}$  est un vecteur de 1 de longueur  $c$ .  $G(\mathbf{X})$  possède donc  $c^2$  colonnes et le même nombre de lignes que  $\mathbf{X}$ . A titre d'exemple, si l'on prend la matrice  $\mathbf{M} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$ , alors :

$$G(\mathbf{M}) = \begin{pmatrix} a_{11}^2 & a_{12}a_{11} & a_{13}a_{11} & a_{12}^2 & a_{11}a_{12} & a_{13}a_{12} & a_{13}^2 & a_{11}a_{13} & a_{12}a_{13} \\ a_{21}^2 & a_{22}a_{21} & a_{23}a_{21} & a_{22}^2 & a_{21}a_{22} & a_{23}a_{22} & a_{23}^2 & a_{21}a_{23} & a_{22}a_{23} \end{pmatrix}.$$

En d'autres termes, cet opérateur multiplie deux à deux toutes les colonnes de la matrice d'origine.

Ainsi, on a facilement l'expression suivante d'après le calcul (1) :

$$\mathbf{X}'\mathbf{W}_\delta\mathbf{X}, c \times c \equiv G(\mathbf{X})' \mathbf{w}, c^2 \times 1 \quad (2)$$

(l'opérateur  $\equiv$  signifie que les membres de gauche et de droite ont les mêmes éléments mais ont des dimensions différentes). Le vecteur  $G(\mathbf{X})' \mathbf{w}$  n'est en réalité rien d'autre qu'un empilement des colonnes de  $\mathbf{X}'\mathbf{W}_\delta\mathbf{X}$ .

Dans la suite, on utilisera souvent la notation  $\mathbf{m} = \text{Vec}(\mathbf{M})$  le vecteur des données associé à la matrice  $\mathbf{M}$ . Ce vecteur est de dimension  $n = n_1 n_2$  et la première dimension de  $\mathbf{M}$  varie plus rapidement que la seconde. A titre d'exemple, si l'on a la matrice  $\mathbf{M}$  suivante :

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

avec  $n_1 = 2$  et  $n_2 = 3$ , le vecteur  $\mathbf{m}'$  correspond au vecteur suivant :

$$\mathbf{m}' = (a_{11} \ a_{21} \ a_{12} \ a_{22} \ a_{13} \ a_{23}).$$

Dans notre cas, d'après l'identité (2), on a donc l'égalité  $\text{Vec}(\mathbf{X}'\mathbf{W}_\delta\mathbf{X}) = G(\mathbf{X})' \mathbf{w}$ .

Enfin, on terminera par le calcul des variances de  $\mathbf{X}\hat{\boldsymbol{\theta}}$ . On introduit le vecteur  $\mathbf{S}_m = \text{Var}(\hat{\boldsymbol{\theta}}) = (\mathbf{X}'\hat{\mathbf{W}}_\delta\mathbf{X})^{-1}$ ,  $c \times c$ . On obtient alors le résultat suivant de manière similaire à ce qui a été fait pour l'équation (2) :

$$\begin{aligned} \text{diag} \left[ \text{Var}(\mathbf{X}\hat{\boldsymbol{\theta}}) \right] &= \text{diag}(\mathbf{X}\mathbf{S}_m\mathbf{X}') \\ &= G(\mathbf{X})\mathbf{s} \end{aligned} \quad (3)$$

où  $\mathbf{s} = \text{Vec}(\mathbf{S}_m)$ ,  $c^2 \times 1$ .

### 2.2.2. Le cas bidimensionnel

Désormais, on ne s'intéressera plus qu'à la dimension 2. Plus précisément, on note  $\mathbf{Y}$  la matrice de dimension  $n_1 \times n_2$  dans laquelle sont rangées les données (et on note  $\mathbf{y} = \text{Vec}(\mathbf{Y})$  son équivalent vectoriel) et  $\mathbf{X}$  la matrice du modèle qui peut s'écrire sous la forme d'un produit de Kronecker :

$$\mathbf{X} = \mathbf{X}_2 \otimes \mathbf{X}_1$$

où la matrice  $\mathbf{X}_i$  correspond à la matrice de la  $i^{\text{ème}}$  variable. Cette matrice est de dimension  $n_i \times c_i$ . On considère donc un *GLM* pour  $\mathbf{y}$  avec la matrice de régression  $\mathbf{X}$ .

On introduit aussi la matrice des coefficients  $\boldsymbol{\Theta}$  de dimension  $c_1 \times c_2$ ,  $\boldsymbol{\theta} = \text{Vec}(\boldsymbol{\Theta})$  le vecteur des coefficients correspondant de dimension  $c_1 c_2 \times 1$  et une matrice de poids  $\mathbf{W}$ .

Une identité classique du produit de Kronecker nous permet d'écrire le prédicteur linéaire de notre modèle linéaire généralisé matriciel :

$$\mathbf{X}_1 \boldsymbol{\Theta} \mathbf{X}_2', n_1 \times n_2 \equiv (\mathbf{X}_2 \otimes \mathbf{X}_1) \boldsymbol{\theta}, n_1 n_2 \times 1 \quad (4)$$

La forme simplifiée du premier membre de cette égalité a deux propriétés intéressantes. Tout d'abord, elle nous épargne le calcul du produit de Kronecker et donc de la matrice  $\mathbf{X}$  qui peut avoir des dimensions très importantes. Deuxièmement, si l'on suppose que nos deux matrices  $\mathbf{X}_1$  et  $\mathbf{X}_2$  sont de dimension  $10^3 \times 10^2$ , le calcul du membre de droite conduit à effectuer  $10^{10}$  multiplications tandis que le calcul du membre de gauche conduit à effectuer  $10^8$  multiplications. De manière générale, cette forme d'écriture du produit de Kronecker permet d'optimiser le temps de calcul.

Comme nous l'avons vu dans le paragraphe précédent pour la résolution des modèles linéaires généralisés classiques, la solution se trouve à l'aide d'évaluations itératives de l'algorithme du score :

$$\mathbf{X}'\tilde{\mathbf{W}}_\delta \mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}'\tilde{\mathbf{W}}_\delta \tilde{\mathbf{z}}$$

où le tilde représente une solution approximative et le chapeau une meilleure approximation.

Calculons tout d'abord la quantité  $\mathbf{X}'\mathbf{W}_\delta\mathbf{X}$  avec  $\mathbf{W}$  la matrice de dimension  $n_1 \times n_2$  telle que  $\mathbf{w} = \text{Vec}(\mathbf{W}) = \text{diag}(\mathbf{W}_\delta)$ .

Ainsi :

$$(\mathbf{X}_2 \otimes \mathbf{X}_1)' \mathbf{W}_\delta (\mathbf{X}_2 \otimes \mathbf{X}_1), c_1 c_2 \times c_1 c_2 \equiv G(\mathbf{X}_1)' \mathbf{W} G(\mathbf{X}_2), c_1^2 \times c_2^2.$$

En effet :

$$\begin{aligned} (\mathbf{X}_2 \otimes \mathbf{X}_1)' \mathbf{W}_\delta (\mathbf{X}_2 \otimes \mathbf{X}_1) &\equiv G(\mathbf{X}_2 \otimes \mathbf{X}_1)' \mathbf{w} && \text{d'après l'équation (2)} \\ &\equiv (G(\mathbf{X}_2) \otimes G(\mathbf{X}_1))' \mathbf{w} && \text{les colonnes de } G(\mathbf{X}_2 \otimes \mathbf{X}_1) \text{ sont effectivement} \\ &&& \text{les mêmes que celles de } G(\mathbf{X}_2) \otimes G(\mathbf{X}_1) \text{ mais} \\ &&& \text{dans un ordre différent} \\ &\equiv G(\mathbf{X}_1)' \mathbf{W} G(\mathbf{X}_2) && \text{d'après l'équation (4)} \end{aligned}$$

Enfin, les variances sont calculées en utilisant la version bidimensionnelle de l'équation (3) :

$$\begin{aligned} \text{diag} \left[ \text{Var}(\mathbf{X}\hat{\boldsymbol{\theta}}) \right] &= \text{diag}(\mathbf{X}\mathbf{S}_m\mathbf{X}'), n_1 n_2 \times 1 \\ &\equiv G(\mathbf{X}_1)\mathbf{S}G(\mathbf{X}_2)', n_1 \times n_2 \end{aligned}$$

où les éléments de la matrice  $\mathbf{S}_m = (\mathbf{X}'\hat{\mathbf{W}}_\delta\mathbf{X})^{-1}$ ,  $c_1 c_2 \times c_1 c_2$  sont réorganisés dans une matrice  $\mathbf{S}$ ,  $c_1^2 \times c_2^2$  (nous ne reviendrons pas en détail sur la réorganisation de la matrice  $\mathbf{S}$  qui alourdirait cette présentation. Pour plus de détails, consulter Currie (2006)).

Cette présentation des modèles *GLAM* a bien mis en évidence l'utilité de cette méthode de résolution des *GLM* pour des modèles avec un nombre important de paramètres et de données. En effet, dans une problématique de lissage, les *GLM* requièrent des capacités de stockage très élevées ainsi qu'un temps de calcul important. L'idée principale d'un *GLAM* consiste en l'utilisation de processus marginaux. En effet, puisque la matrice de régression s'écrit comme le produit de Kronecker de deux matrices marginales  $\mathbf{X}_1$  et  $\mathbf{X}_2$ , la matrice de régression n'est pas stockée comme dans un *GLM* puisqu'un *GLAM* travaille avec  $\mathbf{X}_1$  et  $\mathbf{X}_2$  séparément.

Voyons maintenant dans le chapitre suivant un aspect plus pratique des notions théoriques que nous avons présentées dans ces deux premiers chapitres

### CHAPITRE 3. APPLICATION AU LISSAGE D'UNE COURBE ET D'UNE SURFACE DE MORTALITE

Dans cette partie, nous allons appliquer la théorie présentée à des données réelles pour lisser une courbe et une surface de mortalité.

#### Section 3.1. Présentation des données

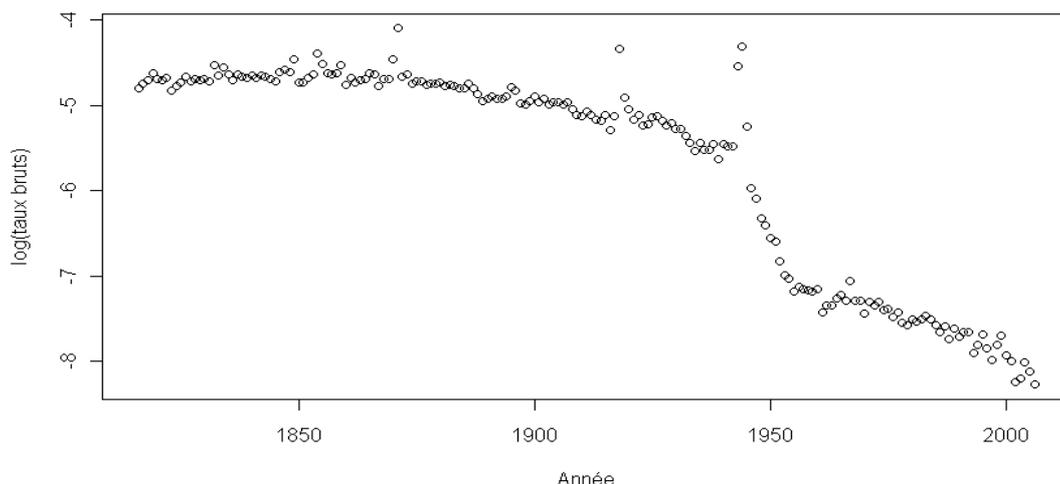
Les données que nous allons utiliser pour cette application ont été obtenues sur le site internet de la *Human Mortality Database* qui référence pour de nombreux pays le nombre de décès, de naissance, l'exposition au risque, la structure démographique de la population et bien d'autres indicateurs utiles pour des études de mortalité. On choisi de s'intéresser à la population française observée entre 1816 et 2006 pour des individus âgés de 0 à 110 ans. On ne s'intéresse qu'à la population des civils et on ne tient pas compte des militaires. Les données sont issues de publications officielles pour les années postérieures à 1900. Pour les années antérieures à 1900, les données ont été récupérées principalement à partir d'une étude du démographe Jean Bourgeois-Pichat (1951) « Evolution générale de la population française depuis le XVIII<sup>ème</sup> siècle ». Les données ont été adaptées en fonction de l'évolution des frontières françaises au cours du temps (Savoie, Alsace-Lorraine) et une correction a été apportée pour les « faux mort-nés » (individus nés vivants mais décédés avant la déclaration de naissance) entre 1899 et 1975.

#### Section 3.2. L'utilisation de P-splines unidimensionnels

Nous avons à notre disposition dans deux matrices distinctes les expositions au risque et les nombre de décès par âge (en ligne) et par année (en colonne). Nous appellerons ces matrices respectivement  $\mathbf{E}$  et  $\mathbf{Y}$  chacune de dimension  $n_a \times n_y$ . On supposera que les années et les âges sont classés par ordre croissant dans les vecteurs  $\mathbf{x}_a$  et  $\mathbf{x}_y$ .

Appliquons la théorie des P-splines (B-splines pénalisés) unidimensionnels que nous avons présentée dans le chapitre précédent. On s'intéresse aux données correspondantes au nombre de décès à 23 ans. On suppose que les décès et les expositions sont regroupés dans deux vecteurs  $\mathbf{d}$  et  $\mathbf{e}$  de longueur  $n_y$ . Nous pouvons déjà tracer le logarithme des taux bruts

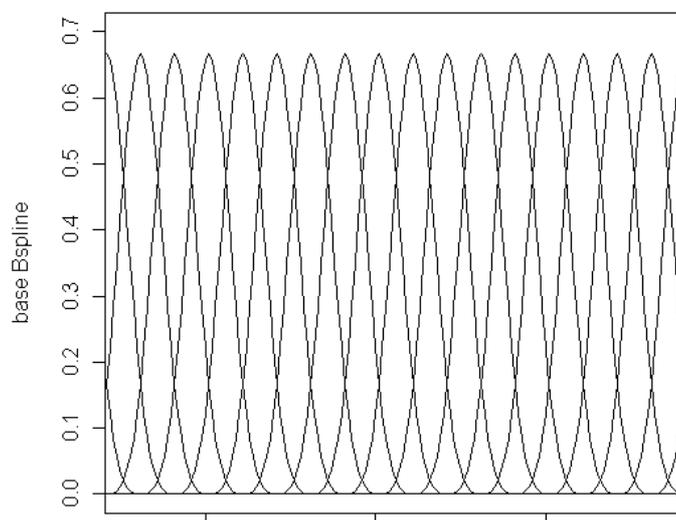
$$\log\left(\frac{\mathbf{d}}{\mathbf{e}}\right) :$$



Graphique 26 - Log des taux bruts pour les individus de 23 ans

Nous allons donc essayer d'estimer la tendance régulière à l'aide de B-splines. Pour cela, nous devons tout d'abord définir une base de B-splines pour la régression. Nous choisirons des B-splines cubiques.

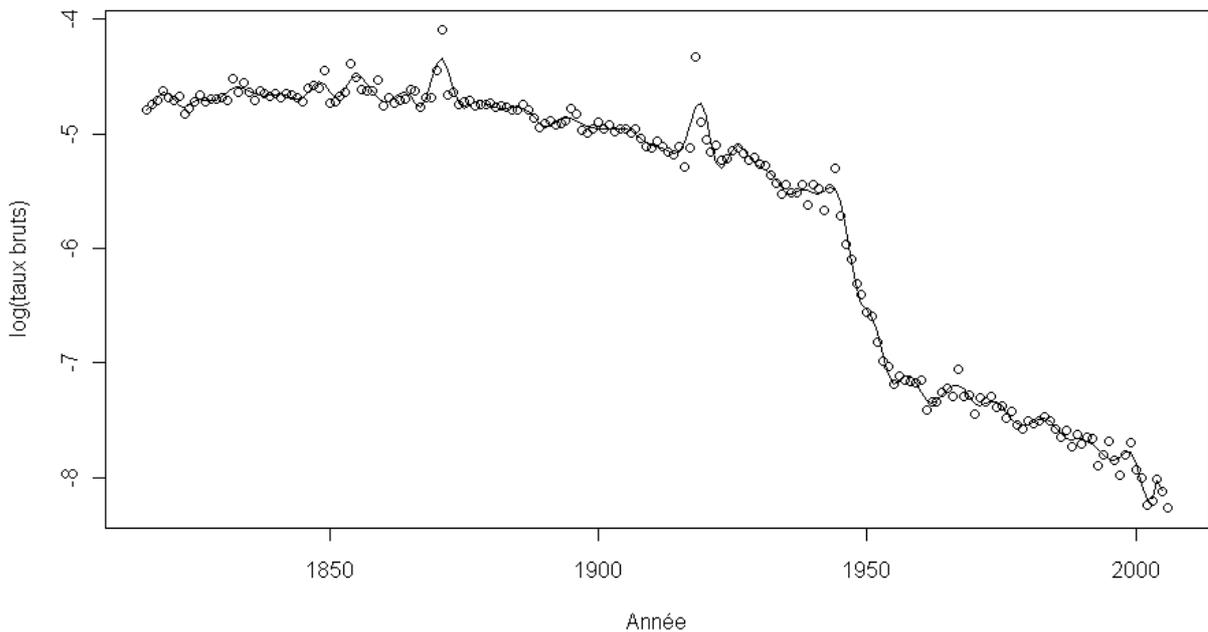
Nous plaçons des nœuds à la valeur minimale et maximale de l'axe des abscisses (ici 1816 et 2006). Nous choisissons ensuite de placer 58 autres nœuds équidistants entre eux pour obtenir un total de 60 nœuds :  $c_y = 60$ . La base de B-splines est alors de la forme suivante :



Graphique 27 - Base de B-splines cubiques

On note  $B_1 \dots B_{c_y}$  cette base de B-splines cubiques. On note ensuite  $\mathbf{B}_y$  la matrice de régression où l'élément  $(i,k)$  correspond à la valeur du B-spline  $B_k$  évalué au  $i^{\text{ème}}$  élément de  $\mathbf{x}_y$ .

Afin de se placer dans le cadre d'un modèle linéaire généralisé, on supposera que les nombres de décès  $d$  ont une distribution poissonnienne de moyenne  $e * \tau$  où  $\tau$  est le taux de mortalité. Si l'on suppose de plus que  $\tau = \exp(\mathbf{B}_y \boldsymbol{\theta})$ , alors nous avons un modèle linéaire généralisé de variables  $d$ , de matrice de régression  $\mathbf{B}_y$ , de variable offset  $\log e$  (une variable offset est une variable qui sert à tarer un modèle si la variable à expliquer dépend linéairement de cette variable offset) avec une erreur de type Poisson et un lien log. Le modèle de type Poisson a été choisi car c'est celui qui est couramment utilisé pour des modèles de comptage (ici comptage du nombre de décès). La régression est effectuée sous R à l'aide de la commande *glm* (sous la forme *glm(matrice\_deces~matrice\_regression-1,family=quasipoisson,offset=log(matrice\_exposition))*). On a alors le résultat suivant :



Graphique 28 - Résultat de la régression avec B-splines

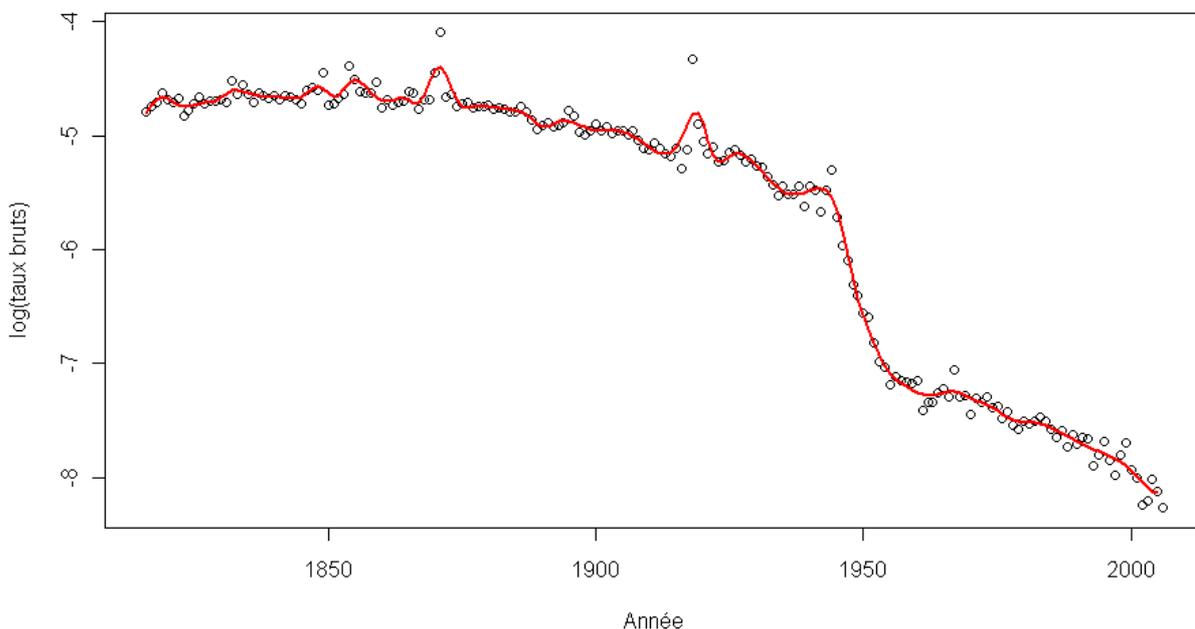
Nous allons maintenant utiliser des P-splines afin de réduire la souplesse de la courbe ajustée c'est-à-dire que nous allons introduire une pénalité dans la log-vraisemblance du modèle. Cette pénalité sera une pénalité de second ordre basée sur les différences finies des coefficients de B-splines adjacents :

$$(\theta_1 - 2\theta_2 + \theta_3)^2 + \dots + (\theta_{c_y-2} - 2\theta_{c_y-1} + \theta_{c_y})^2 = \boldsymbol{\theta}' \mathbf{D}_y' \mathbf{D}_y \boldsymbol{\theta}$$

Comme nous l'avons déjà évoqué, nous introduisons un paramètre  $\lambda$  devant cette pénalité qui permet de garder un contrôle sur la pénalité.

Nous allons appliquer cette pénalité au cas précédent avec un coefficient  $\lambda = 1$  (le choix optimal du paramètre  $\lambda$  n'est pas encore opéré dans cet exemple. On utilisera pour d'autres applications des critères de décision tels que l'AIC ou le BIC). On obtient alors le résultat suivant à l'aide de la commande *gam* du package *mgcv* (sous la forme

`gam(matrice_deces ~ matrice_regression - 1, family = quasipoisson, offset = log(matrice_exposition), paraPen = list(matrice_regression=list(lambda*DtD))):`



Graphique 29 - Résultat de la régression avec P-splines

On remarque que dans ce cas, l'introduction d'une pénalité a permis de réduire le caractère erratique du premier lissage pour les années postérieures à 1945. Pour les années antérieures, l'introduction d'une pénalité n'a visiblement pas eu beaucoup d'impact sur le premier lissage effectué.

Les coefficients estimés se retrouvent en procédant à l'algorithme pénalisé des scores sur les équations :

$$(\mathbf{B}'_y \tilde{\mathbf{W}}_\delta \mathbf{B}_y + \mathbf{P}) \hat{\boldsymbol{\theta}} = \mathbf{B}'_y \tilde{\mathbf{W}}_\delta \tilde{\mathbf{z}}$$

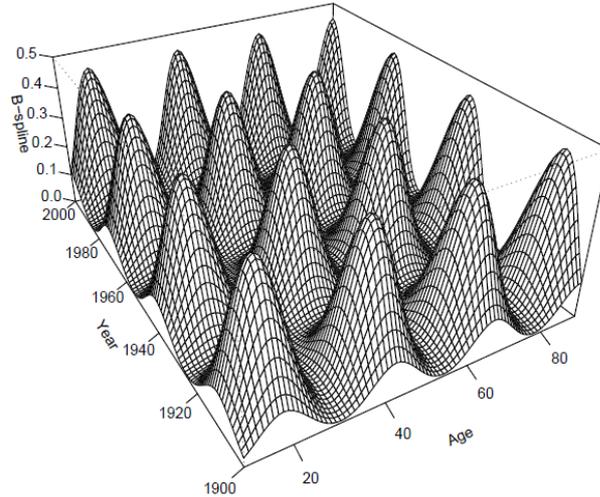
avec  $\mathbf{P} = \lambda_y \mathbf{D}'_y \mathbf{D}_y$ ,  $\lambda_y$  le paramètre de pénalité,  $\tilde{\mathbf{W}}_\delta$  la matrice diagonale des poids avec les éléments  $diag(\tilde{\boldsymbol{\mu}}) = diag(\mathbf{e} * \exp(\mathbf{B}_y \tilde{\boldsymbol{\theta}}))$  sur sa diagonale et  $\tilde{\mathbf{z}} = \mathbf{B}_y \tilde{\boldsymbol{\theta}} + \tilde{\mathbf{W}}_\delta^{-1} (\mathbf{d} - \tilde{\boldsymbol{\mu}})$ . Comme nous l'avons déjà mentionné, le tilde représente une solution approchée et le chapeau une solution améliorée.

### Section 3.3. L'utilisation de P-splines bidimensionnels

Pour une présentation générale des techniques de lissage bidimensionnel avec P-splines, le lecteur intéressé pourra consulter Durban et al. (2002).

Le Graphique 27 ci-dessus présente une base de B-splines cubiques selon les années. De la même manière, nous pouvons créer une telle base selon les âges.

La matrice de régression pour les âges est notée  $\mathbf{B}_a$ . Les matrices  $\mathbf{B}_y$  et  $\mathbf{B}_a$  sont appelées matrices de régression marginales. La matrice de régression qui correspond au cas bidimensionnel est par définition le produit de Kronecker des matrices de régression marginales :  $\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a$  de dimension  $n_a n_y \times c_a c_y$ . Un B-spline bidimensionnel de la base prend donc la forme suivante :



Graphique 30 – Un B-spline de la base bidimensionnelle

De manière identique à ce qui a été fait précédemment, on peut supposer que le nombre de décès  $y_{ij}$  à l'âge  $i$  l'année  $j$  suit une loi de Poisson de moyenne  $\mu_{ij} = e_{ij} \tau_{ij}$ . Nous notons  $\text{Vec}(\mathbf{E})$  et  $\text{Vec}(\mathbf{Y})$  les vecteurs des expositions et des décès. On définit donc un modèle linéaire généralisé matriciel sur la variable  $\text{Vec}(\mathbf{Y})$  avec la matrice de régression  $\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a$ , la variable offset  $\log \text{Vec}(\mathbf{E})$ , le lien  $\log$  et des erreurs de Poisson. A chaque B-spline de la base bidimensionnelle correspond un coefficient de régression  $\theta_{ij}$ ,  $1 \leq i \leq c_a$ ,  $1 \leq j \leq c_y$ . Ces coefficients peuvent être mis dans une matrice  $\Theta$ . Le vecteur des coefficients de régression sera alors noté  $\boldsymbol{\theta} = \text{Vec}(\Theta)$ .

Le modèle linéaire généralisé matriciel pour ce modèle avec la matrice de régression  $\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a$  est :

$$\log \mathbf{M} = \log \mathbf{E} + \mathbf{B}_a \boldsymbol{\Theta} \mathbf{B}_y'$$

où  $\mathbf{M} = \mathbf{E} \mathbf{Y}$ . Cela découle immédiatement de l'identité que nous avons déjà évoquée :

$$\mathbf{B}_a \boldsymbol{\Theta} \mathbf{B}_y' \equiv (\mathbf{B}_y \otimes \mathbf{B}_a) \boldsymbol{\theta}$$

Pour éviter que la surface obtenue soit trop erratique, on peut là aussi pénaliser les lignes et les colonnes de la matrice des coefficients de régression  $\boldsymbol{\theta} = \text{Vec}(\Theta)$  avec la matrice de pénalité :

$$\mathbf{P} = \lambda_a \mathbf{I}_{c_y} \otimes \mathbf{D}'_a \mathbf{D}_a + \lambda_y \mathbf{D}'_y \mathbf{D}_y \otimes \mathbf{I}_{c_a}$$

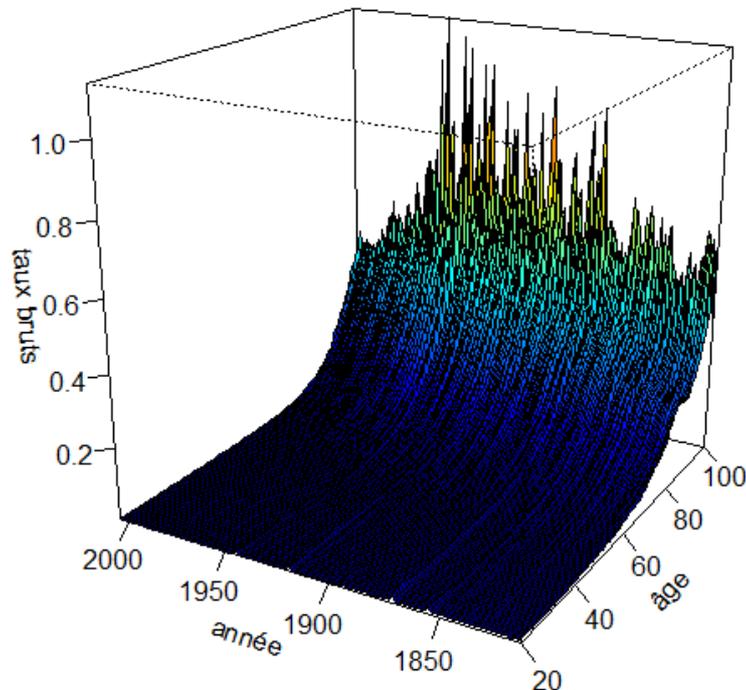
où  $\mathbf{D}_a$ ,  $\lambda_a$ ,  $\mathbf{D}_y$  et  $\lambda_y$  sont les matrices de différence et les paramètres de contrôle de la pénalité pour les âges et les années respectivement.

Les paramètres de pénalité sont déterminés en minimisant le critère BIC. On rappelle que le critère BIC est calculé de la manière suivante :

$$BIC = Deviance + Tr \times \ln(n)$$

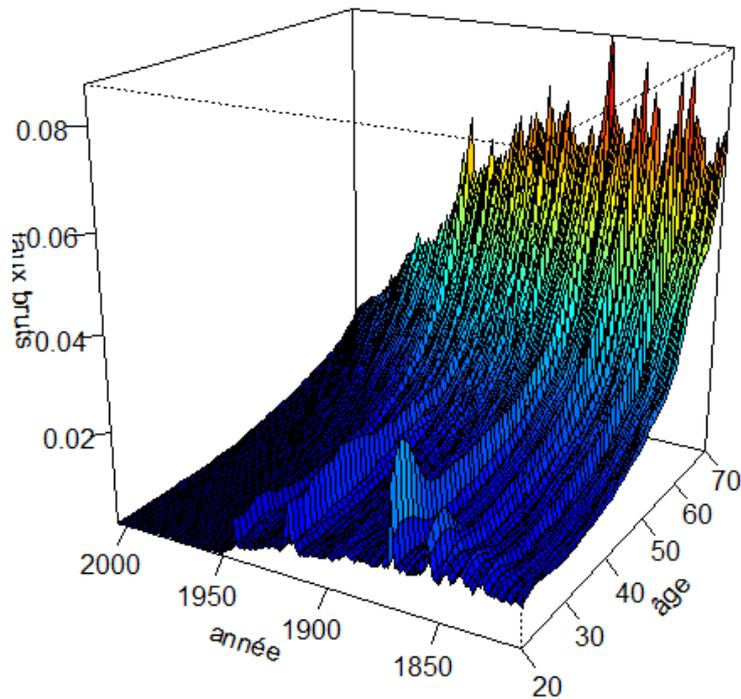
où *Deviance* est l'expression classique de la déviance dans un *GLM* (à savoir  $Deviance = 2 \times (\ln \hat{L} - \ln L)$  où  $\hat{L}$  correspond à la vraisemblance du modèle saturé) et *Tr* correspond à la dimension effective du modèle calculé à partir de la trace de la matrice chapeau  $\mathbf{H} = \mathbf{B}_y (\mathbf{B}'_y \hat{\mathbf{W}}_\delta \mathbf{B}_y + \mathbf{P})^{-1} \mathbf{B}'_y \hat{\mathbf{W}}_\delta$ .

En pratique, nous allons nous intéresser toujours à la même population française que nous avons présentée précédemment. Nous avons à notre disposition les expositions et le nombre de décès par âge et par année sous forme de tableau. On peut tout d'abord tracer la surface des taux bruts de mortalité :



Graphique 31 - Surface des taux bruts de mortalité en France entre 1816 et 2006 (20-100 ans)

Afin d'avoir une vision plus précise du caractère erratique de cette surface, on peut se focaliser sur la tranche d'âge des 20-70 :



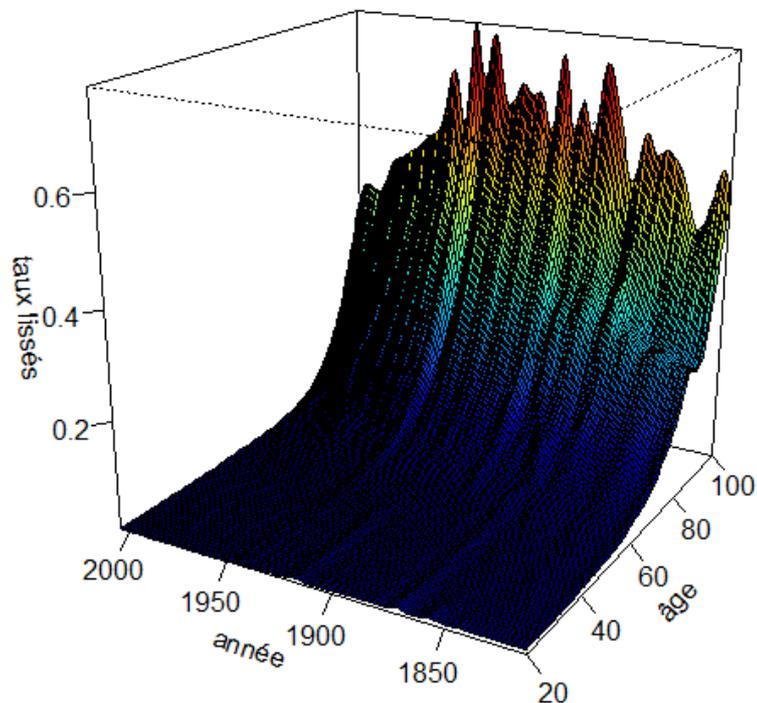
Graphique 32 - Surface des taux bruts de mortalité en France entre 1816 et 2006 (20-70 ans)

On observe sans surprise des pics de mortalité pour les années 1945 et 1918 (conséquence des deux guerres mondiales et de la pandémie de grippe espagnole de 1918) et aussi un pic particulièrement marqué en 1871 (guerre franco-allemande).

Nous allons maintenant appliquer la théorie des *GLAM* et des P-splines bidimensionnels afin de lisser cette surface. D'après le formalisme que nous avons développé précédemment, nous allons devoir :

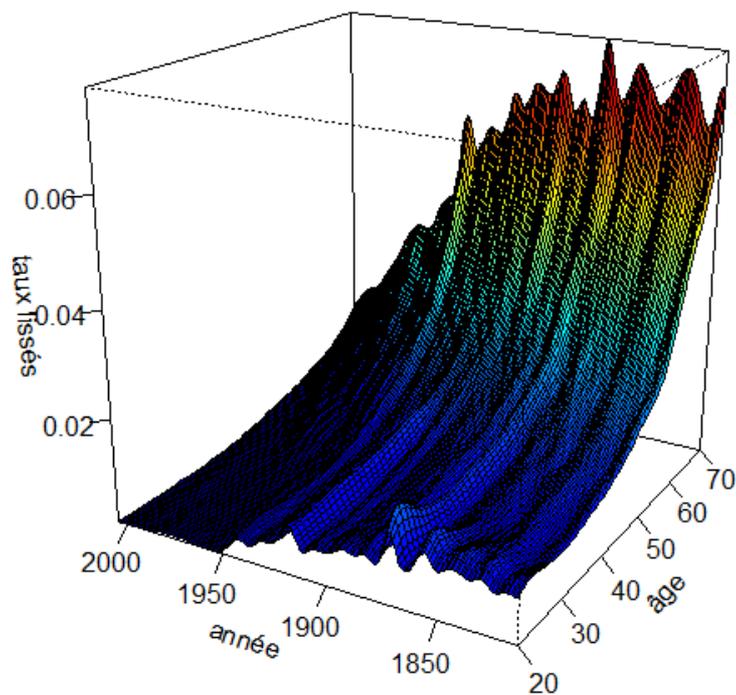
- placer des nœuds pour les B-splines (nous choisissons un intervalle entre chaque nœud de 5 ans à la fois pour les âges et les années d'observation) ;
- construire les matrices marginales  $\mathbf{B}_y$  et  $\mathbf{B}_a$ , leurs tenseurs colonne et la matrice de régression  $\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a$  ;
- construire les matrices de différenciation  $\mathbf{D}_y$  et  $\mathbf{D}_a$  et la matrice de pénalité ;
- optimiser le choix des paramètres de pénalité  $\lambda_a$  et  $\lambda_y$  à l'aide d'une minimisation du critère BIC ;
- résoudre l'algorithme des scores de Fisher avec les paramètres de pénalité déterminés précédemment (l'initialisation des coefficients de régression se fera en résolvant un modèle linéaire simple gaussien non pénalisé :  $\boldsymbol{\theta}^{[0]} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ ). On choisit un seuil de tolérance de  $10^{-6}$  et un nombre d'itérations maximal de 20.

Le lissage donne le résultat suivant :



Graphique 33 - Surface des taux de mortalité lissés en France entre 1816 et 2006 (20-100 ans)

Si l'on se concentre sur la plage d'âge des 20-70 ans, on a la surface lissée suivante :



Graphique 34 - Surface des taux de mortalité lissés en France entre 1816 et 2006 (20-70 ans)

Le critère BIC vaut alors  $BIC = 383400,1$  et les paramètres de pénalité  $\lambda_a = 3,4432$  et  $\lambda_y = 0,001$ .

Malgré le lissage, on repère toujours des chocs qui ne reflètent pas exactement l'évolution « normale » de la mortalité. Le but maintenant sera donc de déterminer dans une composante à part l'effet de ces chocs. Les modèles linéaires généralisés matriciels peuvent être étendus aux modèles linéaires généralisés avec des composantes additives. C'est ce que nous allons faire par la suite en introduisant une composante de chocs additive sous la forme d'un modèle linéaire généralisé matriciel additif.

## CHAPITRE 4. LISSAGE D'UNE SURFACE DE MORTALITE AVEC COMPOSANTE DE CHOCS

Dans ce dernier chapitre, nous allons isoler dans une composante spécifique l'effet des chocs sur la mortalité. La détermination de cette surface de surmortalité sera ensuite extrapolée et utilisée pour calculer des quantiles à 99,5 %.

### Section 4.1. Mise en œuvre

Dans cette partie, on se propose de lisser la surface de mortalité brute en essayant de mettre en évidence les chocs dans une composante séparée. On va donc utiliser un modèle linéaire généralisé additif pénalisé avec deux composantes. La première composante utilisera la matrice de régression définie précédemment à l'aide du produit de Kronecker :  $\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a$  et de la matrice de pénalité  $\mathbf{P} = \lambda_a \mathbf{I}_{c_y} \otimes \mathbf{D}'_a \mathbf{D}_a + \lambda_y \mathbf{D}'_y \mathbf{D}_y \otimes \mathbf{I}_{c_a}$  afin d'obtenir une première surface lissée. En ce qui concerne la composante supplémentaire de chocs, on va introduire une seconde matrice de régression pour cette composante :

$$\mathbf{I}_{n_y} \otimes \tilde{\mathbf{B}}_a$$

où  $\tilde{\mathbf{B}}_a$ , de dimension  $n_a \times c_s$ , est une autre matrice de régression de B-splines pour les âges. Cette seconde base permettra de lisser des courbes pour chaque année selon l'axe des âges. On choisira un nombre de nœuds assez faible afin que les chocs repérés soient très lisses ( $c_s = 11$  par exemple).

Ce modèle additif a donc un prédicteur linéaire de la forme :

$$\log(\text{Vec}(\mathbf{M})) = \log(\text{Vec}(\mathbf{E})) + (\mathbf{B}_y \otimes \mathbf{B}_a) \boldsymbol{\theta} + (\mathbf{I}_{n_y} \otimes \tilde{\mathbf{B}}_a) \tilde{\boldsymbol{\theta}}$$

Nous introduisons une nouvelle pénalité sur les coefficients de la composante de chocs :

$$\tilde{\mathbf{P}} = \lambda_s \mathbf{I}_{n_y} \otimes \mathbf{I}_{c_s} = \lambda_s \mathbf{I}_{n_y c_s}.$$

La matrice de pénalité complète est donc la matrice diagonale par blocs suivante :

$$\begin{pmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{P}} \end{pmatrix}.$$

Nous avons donc maintenant spécifié un *GLM* additif pénalisé avec la variable  $\text{Vec}(\mathbf{Y})$ , des erreurs de type Poisson, un lien log, le prédicteur linéaire et la matrice de régression donnée ci-dessus. Les dimensions du problème sont cette fois-ci très conséquentes puisque l'on a  $n_a n_y = 51 \times 191 = 9741$  observations ( $n_a = 51$  car on ne

s'intéresse qu'à la plage d'âge des 20-70 ans) et  $c_a c_y + c_s n_y = 13 \times 41 + 11 \times 191 = 2634$  variables de régression. Un modèle de type *GLAM* est donc particulièrement adapté ici pour réduire le temps de résolution des équations. Il prend la forme suivante :

$$\log \mathbf{M} = \log \mathbf{E} + \mathbf{B}_a \Theta \mathbf{B}_y' + \tilde{\mathbf{B}}_a \tilde{\Theta}$$

où  $\tilde{\Theta}$ ,  $c_s \times n_y$ , et  $\Theta$ ,  $c_s \times c_y$  sont les matrices des coefficients de régression. Les avantages du *GLAM* sont ici particulièrement flagrants. A titre d'exemple, le calcul de  $(\mathbf{I}_{n_y} \otimes \tilde{\mathbf{B}}_a) \tilde{\theta}$  requiert le stockage en mémoire de  $(\mathbf{I}_{n_y} \otimes \tilde{\mathbf{B}}_a) \tilde{\theta}$  ainsi que  $n_a n_y^2 c_s$  multiplications alors que le calcul de  $\tilde{\mathbf{B}}_a \tilde{\Theta}$  ne nécessite pas le stockage en mémoire de  $(\mathbf{I}_{n_y} \otimes \tilde{\mathbf{B}}_a) \tilde{\theta}$  et requiert seulement  $n_a n_y c_s$  multiplications. Le gain n'est pas négligeable puisque l'on va utiliser une centaine d'années d'observation  $n_y = 191$ .

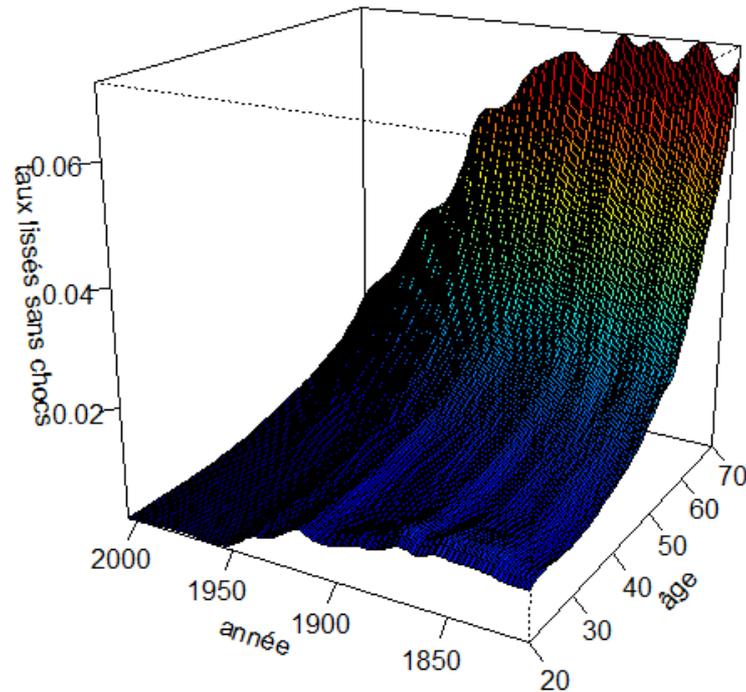
Là encore, nous allons devoir :

- placer des nœuds pour les B-splines (nous choisissons un intervalle entre chaque nœud de 5 ans à la fois pour les âges et les années d'observation et un intervalle de 5 ans également pour les chocs entre les années) ;
- construire les matrices marginales  $\mathbf{B}_y$ ,  $\mathbf{B}_a$  et  $\tilde{\mathbf{B}}_a$ , leurs tenseurs colonne et les matrices de régression  $\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a$  et  $\mathbf{I}_{n_y} \otimes \tilde{\mathbf{B}}_a$  ;
- construire les matrices de différenciation  $\mathbf{D}_y$  et  $\mathbf{D}_a$  et la matrice de pénalité ;
- optimiser le choix des paramètres de pénalité  $\lambda_a$ ,  $\lambda_y$  et  $\lambda_s$  à l'aide d'une minimisation du critère BIC ;  $\theta^{[0]} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$
- résoudre l'algorithme des scores de Fisher avec les paramètres de lissage déterminés précédemment (l'initialisation des coefficients de régression se fera en résolvant un modèle linéaire simple gaussien non pénalisé :  $\theta^{[0]} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ ). On choisit un seuil de tolérance de  $10^{-8}$  et un nombre d'itérations maximal de 20.

Avant tout, on notera que si l'on choisit  $\lambda_a = 3,4432$  et  $\lambda_y = 0,001$  avec  $\lambda_s = 10^{12}$ , on obtient  $BIC = 383400,1$ , c'est-à-dire le même résultat que celui que nous avons trouvé dans le paragraphe précédent. En d'autres termes, cela signifie que plus le coefficient  $\lambda_s$  est grand, plus la pénalité sur la composante de chocs est importante et par conséquent le modèle avec la composante de chocs séparée devient équivalent au modèle précédent sans composante de chocs. On souhaite désormais déterminer les paramètres de pénalité à l'aide d'une minimisation du critère BIC. Malgré l'utilisation d'un modèle *GLAM*, l'optimisation

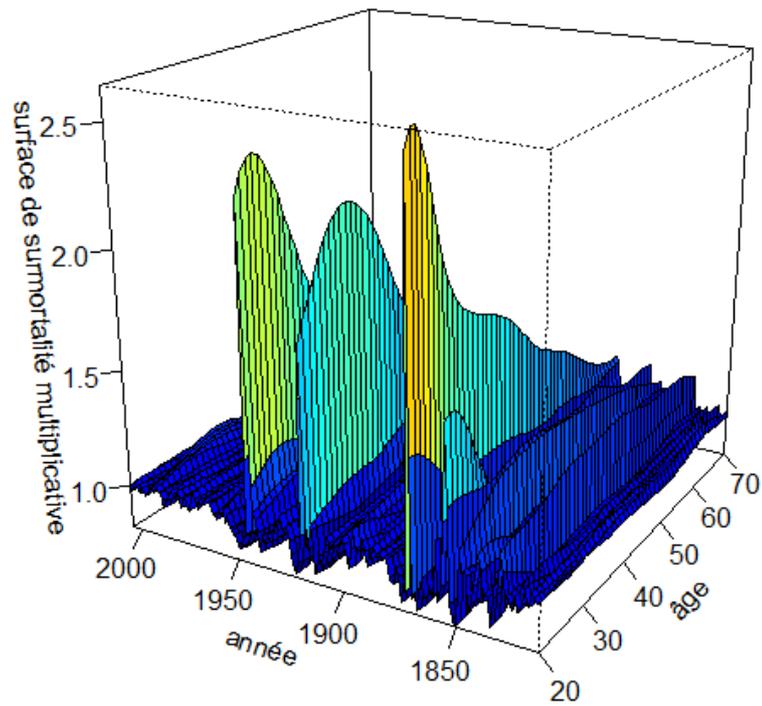
du critère BIC ne peut pas être effectuée dans un temps raisonnable (une journée de calcul n'a pas été suffisante). Nous choisissons donc pour le moment des coefficients de manière arbitraire à savoir  $\lambda_a = 10$ ,  $\lambda_y = 500$  et  $\lambda_s = 800$  ce qui conduit à un  $BIC = 40070,72$ .

On obtient la surface lissée sans chocs suivante en calculant simplement  $\exp(\mathbf{B}_a \Theta \mathbf{B}'_y)$  :



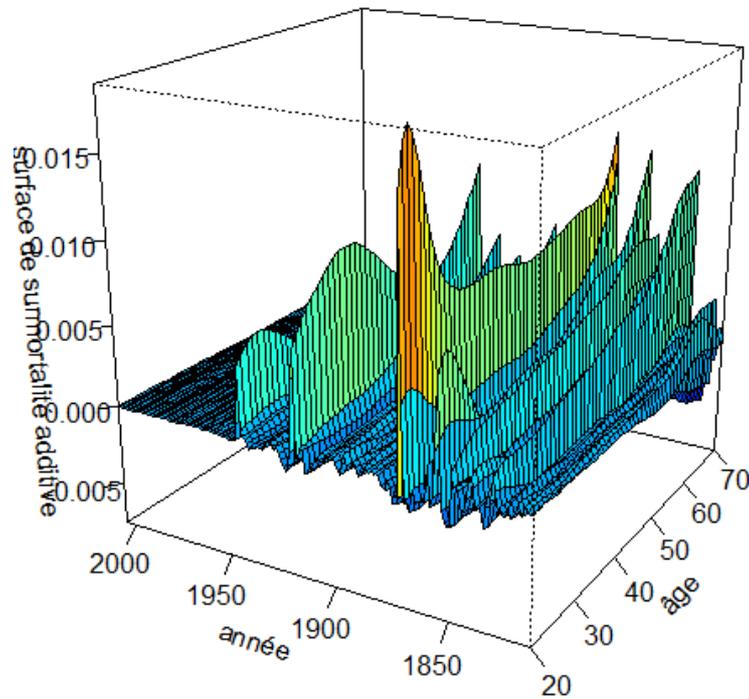
Graphique 35 - Surface des taux lissée sans chocs

De la même manière, on obtient la surface de chocs multiplicative suivante en calculant  $\exp(\check{\mathbf{B}}_a \check{\Theta})$  :



Graphique 36 - Surface de surmortalité multiplicative

On en déduit alors facilement une surface de chocs de mortalité additifs :

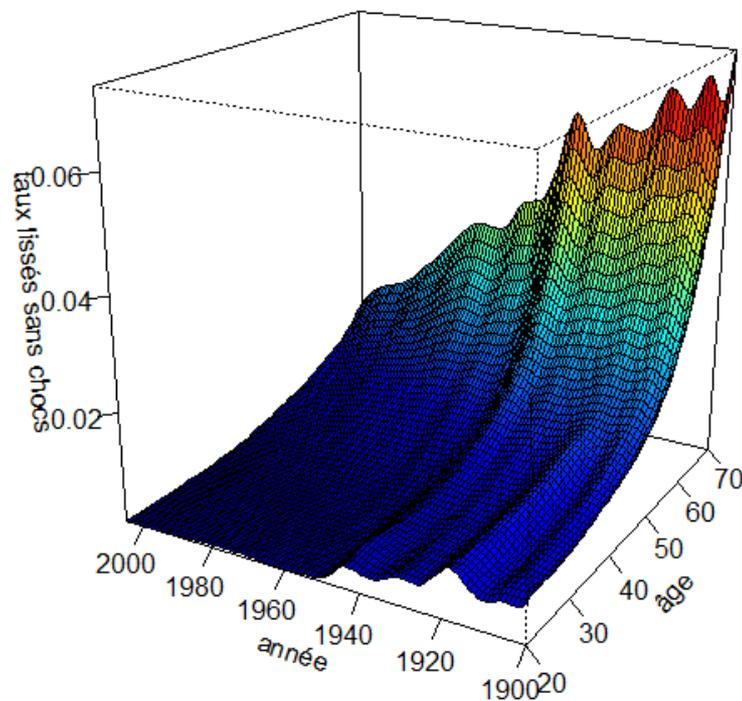


Graphique 37 - Surface de surmortalité additive

On a donc à notre disposition des taux de surmortalité (mais aussi de sous-mortalité) pour tous les âges et pour toutes les années. Ainsi par exemple, on repère que la

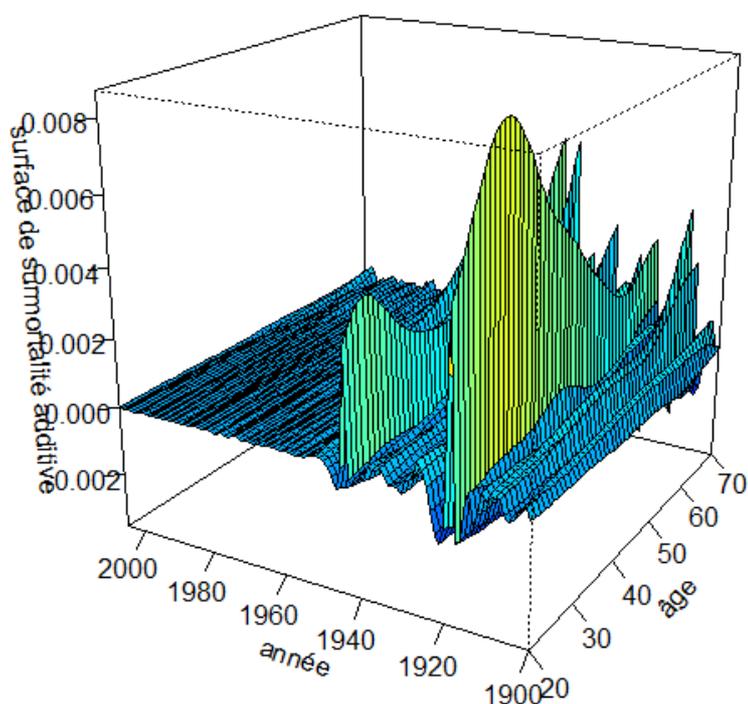
surmortalité additive dépasse 0,5 % pour les individus âgés de 25 ans pour les années 1855 (guerre de Crimée à laquelle s'ajoute une épidémie de choléra), 1871 (guerre franco-allemande et répression de la Commune de Paris), 1918 (fin de la Première Guerre Mondiale et pandémie de grippe espagnole) et 1944 (fin de la Seconde Guerre Mondiale).

Comme nous l'avons évoqué précédemment, l'optimisation du critère BIC n'a pas pu être réalisée dans un temps raisonnable avec autant de données. Nous allons donc nous restreindre à l'observation des individus âgés de 20 à 70 ans pour des années allant de 1900 à 2005 et non plus de 1816 à 2006. Les 106 années que l'on conserve constituent une plage temporelle qui reste assez importante pour pouvoir produire des résultats robustes. Cette fois-ci, l'optimisation du critère BIC est réalisée en un temps acceptable et on obtient les coefficients suivants  $\lambda_a = 8,08$ ,  $\lambda_y = 0,036$  et  $\lambda_s = 1125,9$  ce qui conduit à un  $BIC = 21529,12$ . On obtient cette fois-ci la surface de mortalité suivante :



Graphique 38 - Surface des taux lissée sans chocs

ainsi que la surface de surmortalité additive :

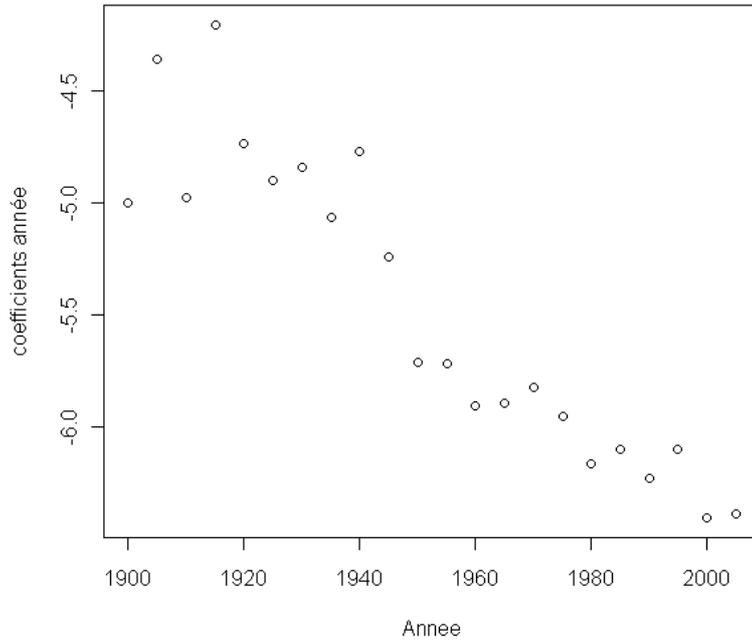


Graphique 39 - Surface de surmortalité additive

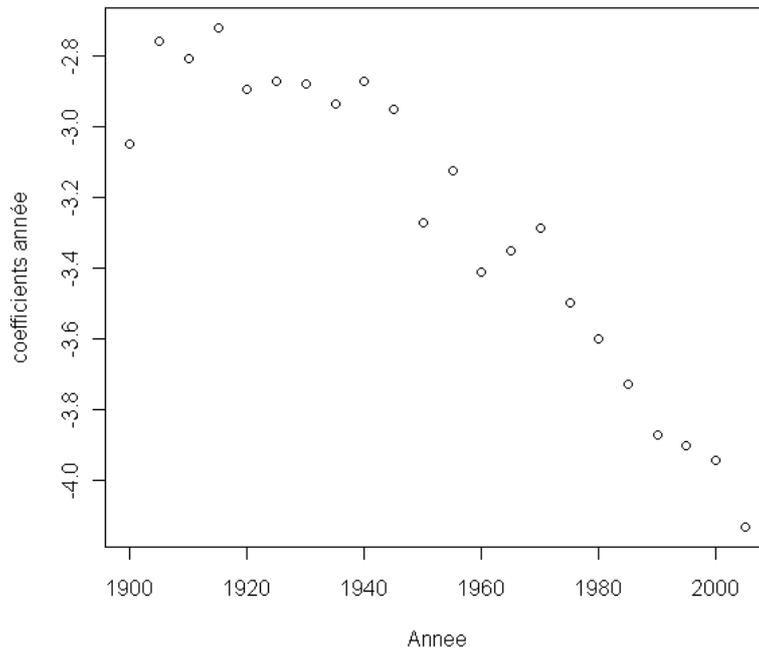
On observe cette fois-ci plus clairement une sous-mortalité additive très prononcée après la guerre de 1914-1918. Ceci peut s'expliquer par le fait que la pandémie de 1918 a entraîné le décès des individus les plus faibles et que par conséquent les individus les plus robustes ont représenté une large majorité au début des années 1920, entraînant ainsi cette sous-mortalité. On remarquera aussi de manière plus discrète un pic de mortalité chez les 60-70 ans en 2003. On peut penser que ce pic traduit la vague de chaleur qui a eu lieu en France cette année et qui a entraîné le décès de plus de 15 000 personnes.

#### Section 4.2. Extrapolation de la surface de mortalité et détermination d'un quantile extrême

A partir de cette surface de mortalité lissée sans chocs, on peut procéder à son extrapolation. En effet, la matrice des coefficients  $\theta$  contient dans ses colonnes les coefficients qui traduisent l'effet « âge » et dans ses lignes les coefficients qui traduisent l'effet « année ». Pour un âge donné, on peut donc déterminer un vecteur de coefficients qui correspondra aux coefficients « année » pour cet âge. La première idée est donc d'extrapoler ces coefficients « année » pour chaque âge. Dans chaque vecteur année (qui correspond à une ligne de  $\theta$ ), nous avons  $c_y$  points (soit 22 points). Voici la forme de quelques séries de coefficients « année » pour un âge fixé :



Graphique 40 - Coefficients "année" pour les individus âgés de 40 ans

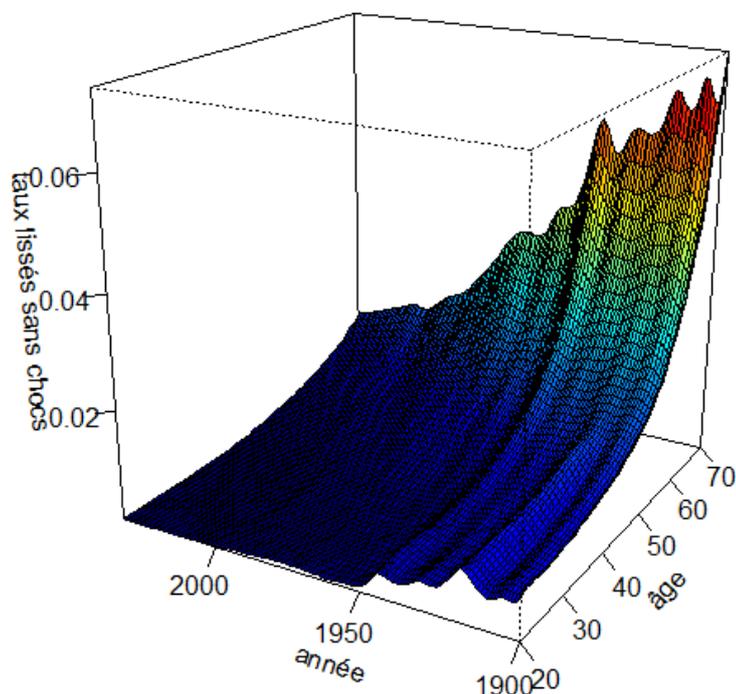


Graphique 41 - Coefficients "année" pour les individus âgés de 70 ans

Au vu de ces graphiques, il est plutôt raisonnable de considérer que les coefficients « année » évoluent linéairement en fonction de l'âge. Pour chaque âge  $x$ , nous allons donc estimer les coefficients d'un modèle linéaire simple du type  $\theta_t^x = at + b + \varepsilon$ .

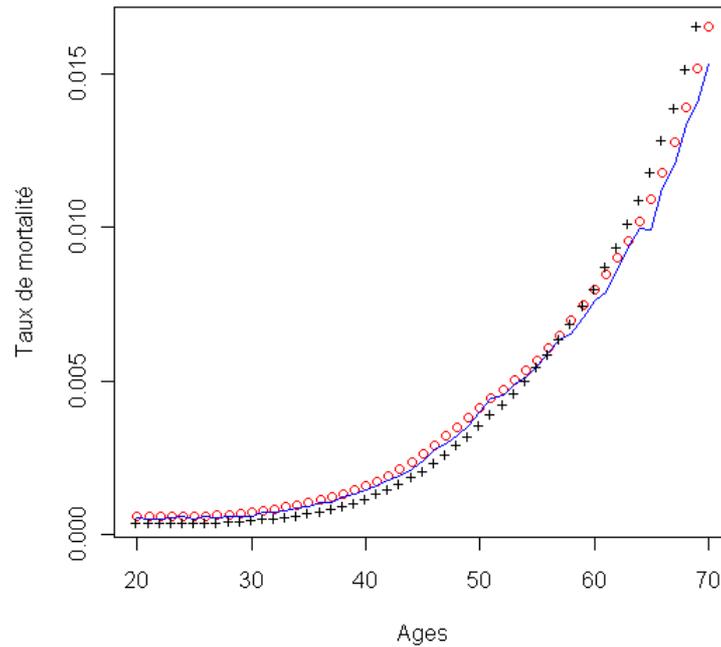
Une fois la régression effectuée pour tous les âges, il est possible de projeter la valeur des coefficients pour des années postérieures à 2005. On obtient donc alors facilement une

surface lissée sans chocs à l'aide de la formule  $\exp(\mathbf{B}_a \Theta^{extrapolés} \mathbf{B}_y^{agrandie})$  où  $\Theta^{extrapolés}$  représente la matrice  $\Theta$  contenant en plus les coefficients extrapolés (une colonne supplémentaire pour une année extrapolée supplémentaire) et où  $\mathbf{B}_y^{agrandie}$  représente la matrice de régression marginale  $\mathbf{B}_y$  avec un nombre de colonnes supplémentaires égal au nombre de colonnes qui ont été rajoutées dans la matrice  $\Theta^{extrapolés}$ . On veille à bien respecter les anciennes caractéristiques de la matrice de régression marginale notamment en ce qui concerne l'espacement des nœuds. La surface de mortalité lissée sans chocs et extrapolée prend la forme suivante (la surface a été extrapolée jusqu'à l'année 2035) :



Graphique 42 - Surface de mortalité sans chocs extrapolée jusqu'en 2035

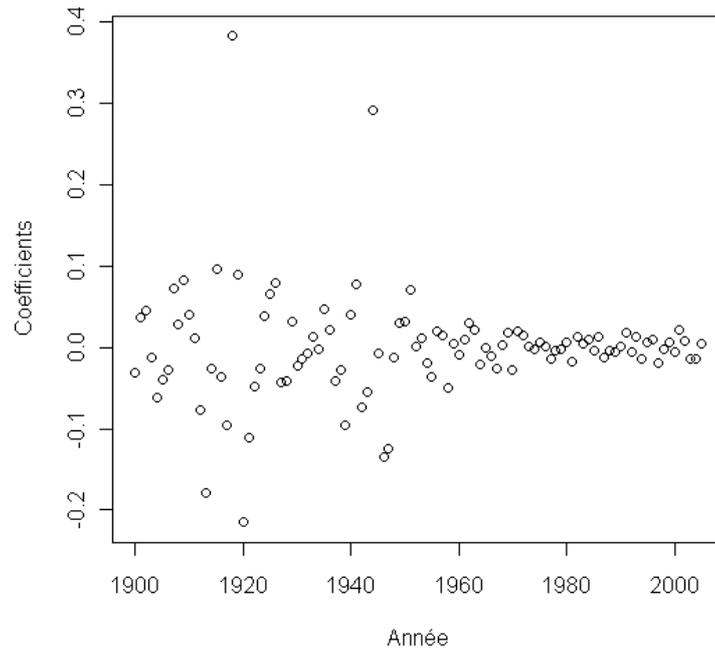
Regardons de plus près quelques résultats. Les données que nous avons utilisées s'arrêtant en 2005, on peut tester la surface extrapolée pour l'année 2006 et la comparer aux données réelles. On peut aussi rajouter le résultat que l'on obtiendrait pour l'année 2006 à l'aide d'un modèle de Lee-Carter ajusté sur les années 1900-2005 et extrapolé lui aussi en 2006. Voici les résultats obtenus :



*Graphique 43 - Comparaison des taux réels, obtenus avec Lee-Carter et obtenus à l'aide de la surface extrapolée pour 2006*

On a représenté avec une ligne les taux de décès réellement observés, avec des croix les taux obtenus à l'aide d'un modèle de Lee-Carter et avec des ronds les taux obtenus à l'aide de la surface extrapolée. Ce graphique montre bien que les taux obtenus à partir de la surface extrapolée sont plus proches des taux réels que ceux obtenus à l'aide du modèle de Lee-Carter.

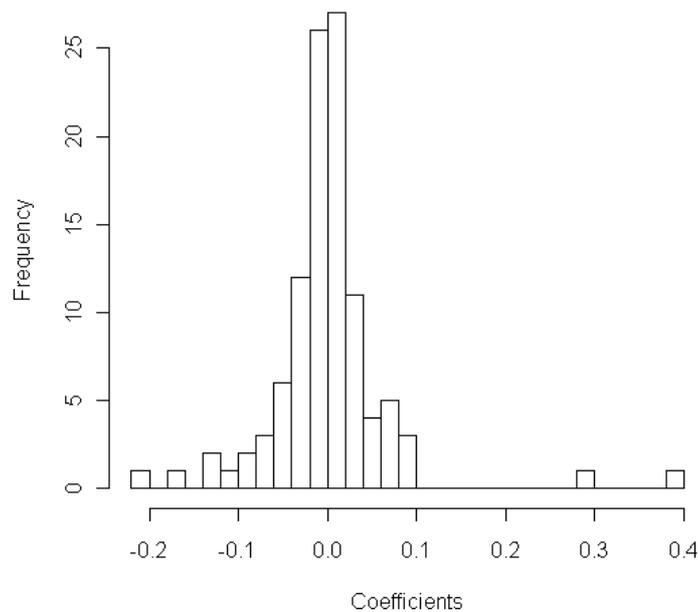
Notre but maintenant est d'extrapoler la surface de chocs que nous avons identifiée. L'idée est légèrement différente de ce que nous avons fait pour l'extrapolation précédente puisque l'objectif n'est plus seulement d'extrapoler la surface mais aussi de calculer des quantiles. En effet ici, nous allons ajuster pour chaque âge la série des coefficients de la composante de choc à une loi paramétrique connue. Une fois l'ajustement opéré, il suffira de simuler cette loi pour simuler un coefficient et par conséquent un choc annuel en termes de taux de mortalité. Voici à titre d'exemple la série des coefficients de la composante de choc pour les individus âgés de 40 ans :



Graphique 44 - Coefficients de choc en fonction de l'année d'observation pour les individus âgés de 40 ans

Plus la valeur du coefficient est élevée, plus le choc est important. On notera donc les deux pics de 1918 et 1914 que nous avons déjà identifiés ainsi que la situation de sous-mortalité après la Première Guerre Mondiale à laquelle nous avons déjà proposé une explication.

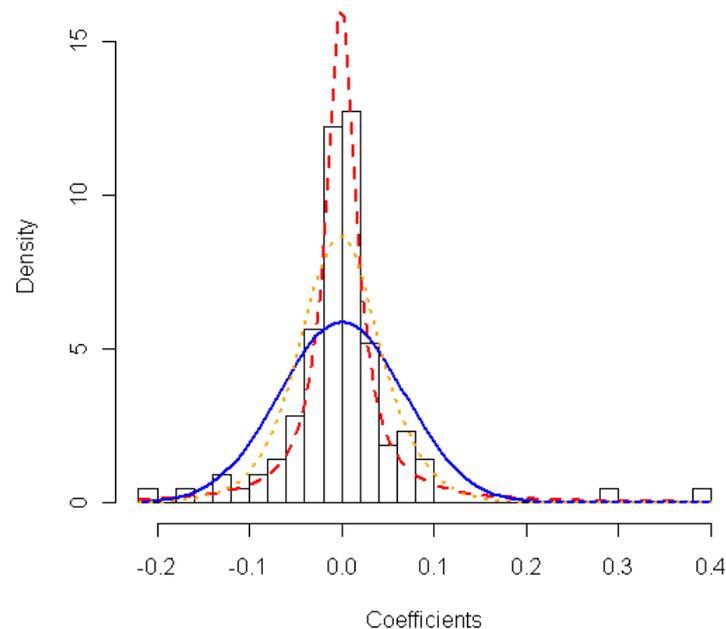
On a l'histogramme correspondant pour la répartition des coefficients :



Graphique 45 - Histogramme des coefficients de choc pour les individus âgés de 40 ans

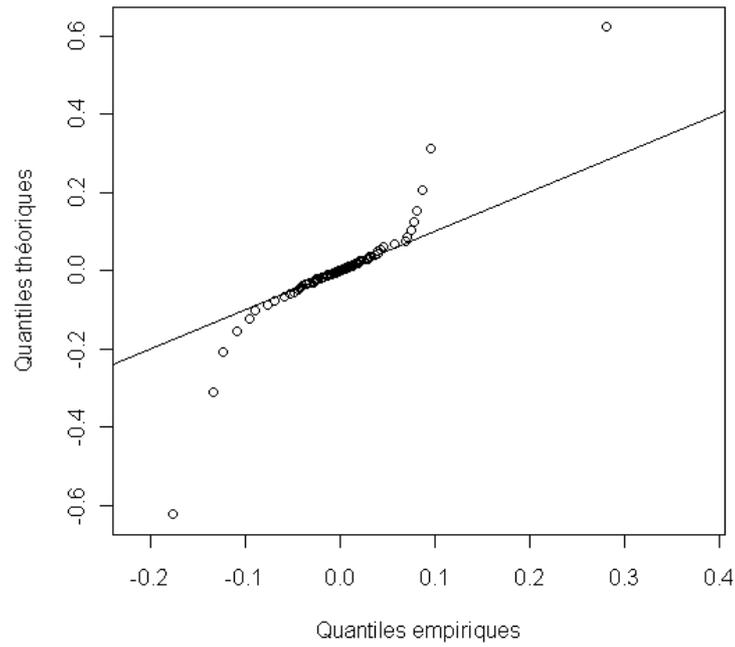
On rappelle que l'on a 11 séries de coefficients. En effet, nous avons choisi 11 nœuds pour les âges sur la plage d'âge 20-70 ans. Chaque nœud est espacé par 5 années. Nous

allons donc devoir ajuster 11 séries de données. Après avoir effectué plusieurs essais avec une trentaine de lois à l'aide de l'utilitaire EasyFit (édité par MathWave Technologies <http://www.mathwave.com/>), on peut raisonnablement penser que pour ce type de distribution, les lois qui conviendront le mieux sont la loi normale, la loi logistique et la loi de Cauchy (les fonctions de densité de ces lois sont rappelées en « Annexe 4. Lois utilisées »). Nous allons expliquer de manière détaillée les étapes de l'ajustement pour la série que nous avons présentée ci-dessus. Pour les dix autres séries, les graphiques détaillés sont présentés en « Annexe 5. Détails des ajustements paramétriques ». Voici la représentation graphique obtenue pour ces trois lois après estimation des paramètres par maximum de vraisemblance :

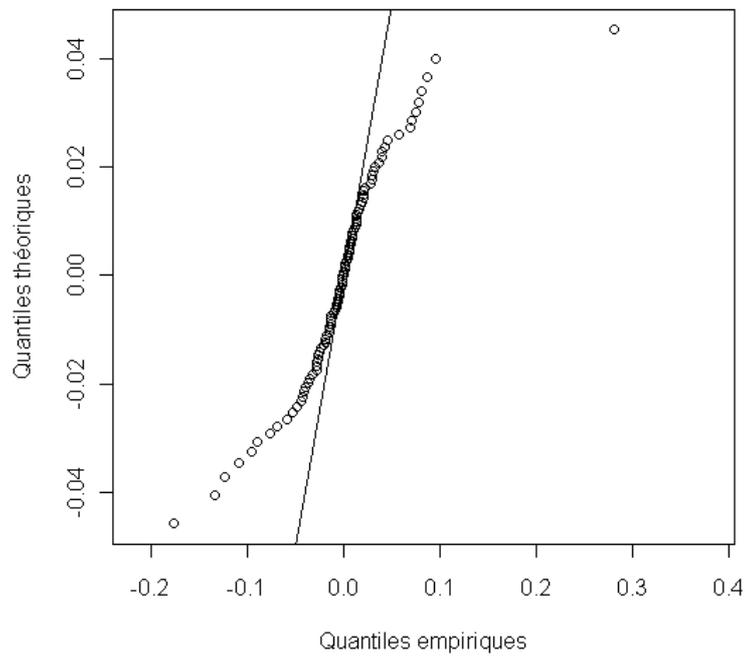


*Graphique 46 - Histogramme et densités théoriques*

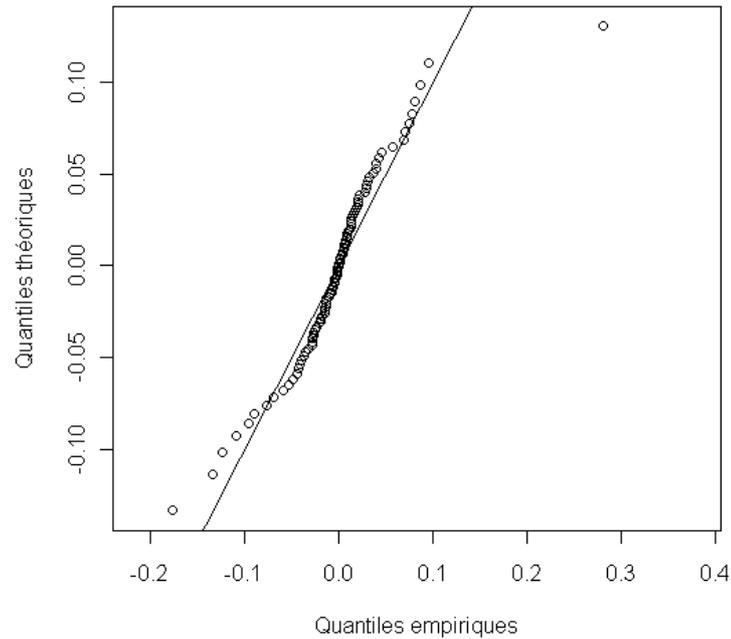
On a représenté avec une ligne la distribution normale, avec des pointillés la distribution de Cauchy et avec des petits pointillés la distribution logistique. Nous pouvons maintenant tracer les QQ-plots pour les trois distributions que nous envisageons :



Graphique 47 - *QQ-plot distribution de Cauchy*



Graphique 48 - *QQ-plot distribution normale*



Graphique 49 - QQ-plot distribution logistique

Il semblerait *a priori* que la loi logistique ne soit pas finalement très bien adaptée aux données. Entre la loi normale et la loi de Cauchy, cette dernière est celle qui présente les meilleurs résultats d'alignement des points du QQ-plot. On va maintenant utiliser un test d'ajustement du khi-deux pour vérifier ces résultats (c'est seul test d'ajustement que l'on peut utiliser puisque nous avons estimé les coefficients de la distribution). La variable de décision est la variable qui prend la valeur :

$$\hat{\chi}^2 = \sum_{i=1}^k \frac{(O_i - C_i)^2}{C_i}$$

où les  $C_i$  correspondent aux effectifs théoriques de la classe  $i$  et les  $O_i$  correspondent aux effectifs réellement observés de la classe  $i$ . Les  $k$  classes sont déterminées de telle sorte que chaque classe contienne au moins 5 observations. On teste les hypothèses suivantes :

- $H_0$  : la distribution observée est conforme à la distribution théorique choisie ;
- $H_1$  : la distribution observée est différente de la distribution théorique.

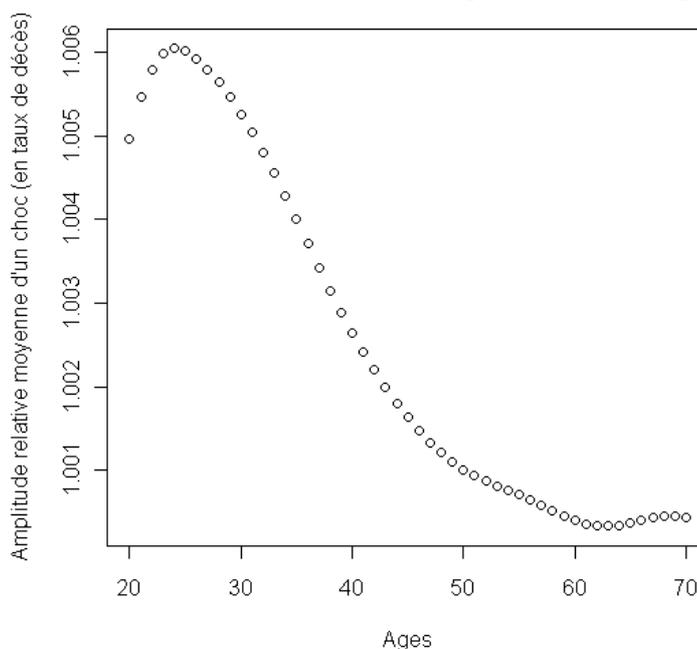
Sous l'hypothèse  $H_0$ , la variable de décision suit une loi du  $\chi^2$  à  $k-1-r$  degrés de liberté où  $r$  est le nombre de paramètres estimés de la loi théorique. On calcule ensuite la  $p$ -valeur du test et on utilise la règle suivante : « si  $p$  est la  $p$ -valeur d'une observation sous l'hypothèse  $H_0$ , on obtient un test de seuil  $\alpha$  par la règle de rejet : « rejet de  $H_0 \Leftrightarrow p < \alpha \gg$  ». Dans la suite, on choisira  $\alpha = 5\%$  .

On obtient alors les résultats suivants :

	Valeur du khi-deux	<i>p</i> -valeur
<b>Cauchy</b>	7,14	0,9889
<b>Normale</b>	64,77	$3,39.10^{-7}$
<b>Logistique</b>	29,98	0,03765

Tableau 15 - Résultats du test d'adéquation du khi-deux

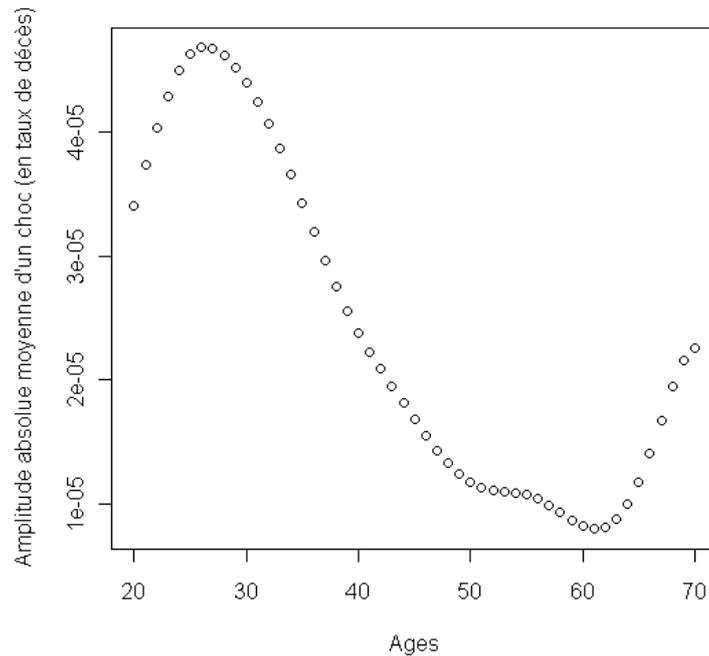
La loi de Cauchy est donc clairement la loi la plus adaptée à cette série. Ceci est particulièrement appréciable dans la mesure où parmi les trois lois que nous avons retenues, la loi de Cauchy est celle qui a l'indice de queue le plus élevé, ce qui va dans le sens de la prudence lorsque nous aurons à calculer des quantiles à 99,5 %. En ce qui concerne les autres séries de coefficients (dont les résultats sont reportés en « Annexe 5. Détails des ajustements paramétriques »), on peut faire la remarque suivante : la loi de probabilité des coefficients n'est pas indépendante de l'âge (et donc de la série) que l'on considère. On note en effet que la structure des coefficients change en fonction de l'âge. Pour les séries allant de 1 à 7 (c'est-à-dire pour les âges allant de 20 à 50 ans), la meilleure distribution ajustée est la distribution de Cauchy comme nous l'avons vu dans l'exemple précédent. A partir de la série 8 et jusqu'à la série 11 (pour des âges allant de 55 à 70 ans), la meilleure distribution ajustée devient la distribution logistique. Comme nous l'avons évoqué, la loi logistique est une loi à queue fine tandis que la loi de Cauchy est une loi à queue épaisse. Cette évolution de la distribution des coefficients de chocs avec l'âge traduit donc le fait que l'amplitude relative d'un choc exprimée en taux de décès décroît avec l'âge. Nous pouvons remarquer cela si l'on trace la moyenne des amplitudes relatives d'un choc que nous avons déterminées au travers de la composante de choc pour chaque âge :



Graphique 50 - Moyenne des amplitudes relatives des chocs par âge (en taux de décès)

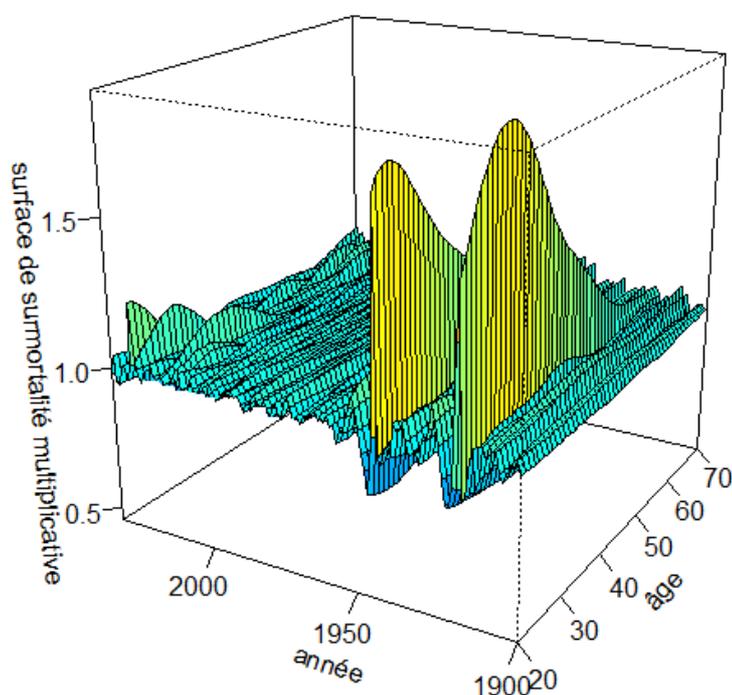
En d'autres termes, cela signifie que les personnes les plus âgées sont relativement moins touchées par les chocs que les personnes les plus jeunes. Etant donnée la différence

de niveau de mortalité « normale », le profil d'amplitude absolue moyenne d'un choc est légèrement différent :



Graphique 51 - Moyenne des amplitudes absolues des chocs par âge (en taux de décès)

Maintenant que nous avons pu déterminer les distributions les plus adéquates et estimé les coefficients de ces distributions, nous allons pouvoir simuler des coefficients et ainsi extrapoler la surface de choc. De manière similaire à ce qui a été fait pour la tendance, pour obtenir une surface de surmortalité / sous-mortalité, il suffit de calculer la quantité  $\exp(\tilde{\mathbf{B}}_a \tilde{\Theta}^{extrapolés})$  où  $\tilde{\Theta}^{extrapolés}$  représente la matrice  $\tilde{\Theta}$  contenant en plus les coefficients simulés pour chaque âge (une colonne supplémentaire pour une année simulée supplémentaire) et où  $\tilde{\mathbf{B}}_a$  représente la matrice de régression non modifiée. La surface de surmortalité multiplicative simulée prend la forme suivante (la surface a été simulée jusqu'à l'année 2035) :



Graphique 52 - Surface de surmortalité multiplicative simulée entre 2005 et 2035

Plaçons nous maintenant dans un contexte assurantiel avec pour objectif de calculer un taux de surmortalité additif correspondant à un choc de sévérité comparable à un choc qui interviendrait une fois tous les deux cents ans (quantile à 99,5 %). Avant tout, il est important de signaler qu'avec nos données précédentes, nous avons pris en compte tous les chocs qui sont survenus dans l'histoire française depuis 1900. Du point de vue d'un assureur, tous les évènements causant des chocs ne sont pas assurables. C'est notamment le cas des guerres que nous avons prises en compte dans notre étude précédente. L'importance de ces pics nous suggère de plus qu'ils doivent avoir une grande influence dans l'estimation des paramètres des distributions des coefficients de chocs. Pour la détermination d'un quantile à 99,5 % dans un contexte d'assurance, il convient donc de travailler sur des données n'incluant pas les décès dus aux deux guerres mondiales.

Il est très difficile de déterminer le nombre de civils tués directement par la guerre. C'est pourquoi au lieu de travailler sur des données de décès et d'expositions modifiées, nous travaillerons directement sur le retraitement des coefficients de choc. Pour ce faire, nous imposerons pour les quatre premières années de la Première Guerre Mondiale des coefficients de choc nuls (on conserve les coefficients de choc déterminés pour les années 1918 et 1919 qui prennent majoritairement en compte la grippe espagnole de 1918-1919). De même, pour les sept années de la Seconde Guerre Mondiale, nous imposerons des coefficients de choc nuls. Après ces retraitements, on applique de nouveau la méthodologie présentée ci-dessus pour estimer les paramètres des lois de distribution auxquelles on ajuste les coefficients.

Désormais, il est possible de déterminer pour chaque distribution ajustée aux coefficients retraités un quantile à 99,5 %. Le coefficient de choc calculé de cette manière correspondra donc au coefficient de choc d'une catastrophe de sévérité comparable à un évènement qui surviendrait une fois tous les 200 ans. Une fois ces coefficients calculés, on les réinjecte dans la matrice des coefficients de choc afin d'obtenir des taux de surmortalité multiplicatifs puis additifs. Dans notre cas, pour l'année 2005, on détermine les taux de surmortalité additifs suivants par âge :

	<b>Taux de surmortalité additif</b>		<b>Taux de surmortalité additif</b>
<b>20</b>	0,125%	<b>46</b>	0,341%
<b>21</b>	0,136%	<b>47</b>	0,352%
<b>22</b>	0,144%	<b>48</b>	0,352%
<b>23</b>	0,148%	<b>49</b>	0,339%
<b>24</b>	0,151%	<b>50</b>	0,314%
<b>25</b>	0,152%	<b>51</b>	0,282%
<b>26</b>	0,152%	<b>52</b>	0,249%
<b>27</b>	0,152%	<b>53</b>	0,220%
<b>28</b>	0,152%	<b>54</b>	0,200%
<b>29</b>	0,153%	<b>55</b>	0,195%
<b>30</b>	0,154%	<b>56</b>	0,209%
<b>31</b>	0,155%	<b>57</b>	0,243%
<b>32</b>	0,158%	<b>58</b>	0,293%
<b>33</b>	0,161%	<b>59</b>	0,357%
<b>34</b>	0,165%	<b>60</b>	0,427%
<b>35</b>	0,170%	<b>61</b>	0,492%
<b>36</b>	0,177%	<b>62</b>	0,540%
<b>37</b>	0,184%	<b>63</b>	0,559%
<b>38</b>	0,193%	<b>64</b>	0,550%
<b>39</b>	0,204%	<b>65</b>	0,519%
<b>40</b>	0,218%	<b>66</b>	0,473%
<b>41</b>	0,234%	<b>67</b>	0,417%
<b>42</b>	0,254%	<b>68</b>	0,356%
<b>43</b>	0,277%	<b>69</b>	0,299%
<b>44</b>	0,301%	<b>70</b>	0,251%
<b>45</b>	0,323%		

Tableau 16 - Taux de surmortalité additifs pour un choc de période de retour de 200 ans

Pour obtenir un taux de surmortalité additif moyen, on peut pondérer ces taux par la composition de la population réelle en 2005 (site internet de l'INSEE). On obtient alors un taux moyen de surmortalité additif de l'ordre de 0,252 %. Ce résultat est tout à fait cohérent avec les études déjà menées auparavant où l'on rappelle que les taux de surmortalité déterminés sont compris entre 0,1 % et 0,61 %.

On peut faire la même chose pour l'année 2009. Cependant cette fois-ci, les données utilisées s'arrêtant en 2005, il est nécessaire avant tout d'extrapoler la tendance comme

nous l'avons déjà fait précédemment et de récupérer la composition de la population en 2009 (toujours dans la catégorie « bilan démographique » du site de l'INSEE). En pondérant de nouveau les taux de surmortalité par âge obtenus à l'aide de la composition de la population, on trouve cette fois-ci un taux de surmortalité moyen de 0,251 %.

### **Section 4.3. Avantages et inconvénients**

Dans cette section, nous allons passer en revue les avantages et les difficultés qui découlent de l'utilisation de cette méthode.

Tout d'abord, on peut noter que ce modèle permet de prendre en compte tous les chocs qui sont survenus dans l'histoire d'un pays. Contrairement aux autres modèles, nous ne nous sommes pas restreints au risque de pandémie. Comme nous avons pu le noter, le choc de 2003 suite à la vague de chaleur a bien été pris en compte et utilisé pour déterminer la surface de surmortalité.

On peut également souligner que l'utilisation des P-splines pour le lissage est particulièrement appréciable dans la mesure où ils offrent une flexibilité qui permet de bien prendre en compte toute les subtilités des données. Cela a pour conséquence une meilleure estimation de la tendance comme nous avons pu le constater.

Le fait que l'on isole à part une composante de chocs permet d'une part de mieux identifier la tendance et d'autre part d'obtenir une surface de surmortalité sur laquelle on peut se baser pour introduire une dose de choc « cohérente » pour les chocs futurs.

D'un point de vue du retraitement des données, cette méthode est aussi beaucoup plus avantageuse que les précédentes dans la mesure où nous n'avons pas eu à effacer les chocs de nos données pour estimer correctement la tendance. Les seuls retraitements que nous avons faits concernent la « suppression » des deux guerres mondiales pour calculer des quantiles à 99,5 %. Pour cela, nous avons simplement remplacé les coefficients de chocs de ces années par des coefficients de chocs nuls.

Cette méthode présente néanmoins deux inconvénients. Tout d'abord nous avons remarqué que la détermination des paramètres de pénalité peut être très lourde en termes de temps de calcul malgré l'utilisation des *GLAM*. Nous avons donc dû restreindre notre plage d'âge et d'années pour obtenir des résultats en un temps raisonnable. Si l'on utilise du matériel informatique plus adapté aux gros calculs, cet inconvénient peut toutefois rapidement disparaître. Le second inconvénient réside dans le fait que nous n'avons pas introduit d'avis d'experts sur la survenance d'une pandémie comme le font d'autres études. Cependant, si un assureur souhaite adapter ce modèle aux particularités de son portefeuille pour déterminer un quantile à 99,5 %, il peut jouer sur le choix de la queue de distribution. Ainsi par exemple, s'il estime que le modèle n'est pas assez prudent pour les individus âgés de 30 ans dans son portefeuille, il peut choisir de modifier la queue de distribution de la distribution de Cauchy qui a été déterminée pour une distribution ayant un indice de queue plus élevé.

## CONCLUSION

---

Dans ce mémoire, nous avons dressé un panorama du risque catastrophe en assurance de personnes et avons présenté des techniques calculatoires pour modéliser des chocs de mortalité.

Nous avons vu que les risques qui pèsent sur la vie humaine sont nombreux et en pleine mutation. Ils peuvent potentiellement faire dévier de manière significative la mortalité anticipée et entraîner des flux de capitaux très importants. D'après Charpentier (2002), « *il faudrait au secteur de l'assurance plusieurs dizaines de milliards de dollars de fonds propres supplémentaires pour offrir une couverture totale de ces risques* ». Face à de tels risques, la réassurance traditionnelle n'offre pas de couverture satisfaisante. Elle ne prend d'ailleurs généralement pas le risque d'assurer une pandémie. Les assureurs ne doivent donc pas négliger ces risques et s'intéresser de plus près à la prise en compte du risque catastrophe dans leurs modèles (d'après Benfield (2008), 15 % des acteurs du marché de l'assurance ne possèdent pas de couverture catastrophe).

D'un point de vue réglementaire et normatif, la question de la prise en compte des catastrophes est encore loin d'être résolue. Les états réglementaires C8 et C9 sont parfois difficilement complétés par les assureurs qui sont confrontés à des problèmes pratiques de simulation des effets d'une pandémie sur leur portefeuille par exemple. Dans le cadre du projet Solvabilité 2, nous avons également mis en évidence les incohérences qui existent actuellement dans le QIS4 entre les différentes branches qui présentent un sous-module catastrophe pour le calcul de la formule standard. Concernant le modèle interne, aucune orientation précise n'a été donnée pour l'instant sur la manière de prendre en compte les catastrophes dans les modèles.

Deux types d'approches principales peuvent être distingués pour modéliser des chocs de mortalité. Nous avons tout d'abord abordé le cas de l'approche épidémiologique en regardant plus en détails l'étude de Swiss Re qui s'écarte nettement des autres études existantes en remettant en cause le poids accordé à la grippe espagnole de 1918. Ce type d'approche se fonde en majorité sur des avis d'experts qui conduisent à des jeux d'hypothèses qui ne sont pas totalement transparents et auxquels sont très sensibles les résultats. Si l'on souhaite accorder plus d'importance aux événements passés pour calibrer un modèle, on se tournera alors vers une approche « actuarielle » basée sur des modèles de mortalité stochastique. L'examen du modèle de Bauer et Kramer met en évidence d'une part le fait que la tendance doit avant tout être particulièrement bien estimée pour que l'utilisation d'une composante catastrophe soit utile et d'autre part les difficultés de calibrage du modèle. En effet, le calibrage de la tendance doit se baser sur des données qui ne tiennent pas compte des chocs alors que ceux-ci y sont obligatoirement présents ce qui nécessite un retraitement délicat et souvent hasardeux. De plus, concernant la composante catastrophe, le calibrage ne s'effectue que sur quelques points ce qui n'est pas suffisant

pour produire des résultats robustes. Nous avons également observé que les résultats sont particulièrement sensibles au poids accordé à la grippe espagnole de 1918. Selon les différentes hypothèses retenues (période de retour estimée entre 31 ans et 3000 ans), le taux de surmortalité à 99,5 % que nous avons déterminé varie entre 0,321 % et 0,158 %. On reprochera également aux deux approches que nous avons envisagées qu'elles ne prennent en compte que le risque de pandémie et ne tiennent pas compte des autres risques.

Devant le manque de transparence des études épidémiologiques et la difficulté de mise en œuvre et de calibrage des modèles de mortalité stochastique, l'objectif du mémoire a également été de proposer une approche originale de modélisation de la surmortalité. L'approche proposée se base sur des techniques de lissage par P-splines et l'utilisation des modèles linéaires généralisés matriciels. L'utilisation de P-splines pour le lissage a pour avantage d'introduire une plus grande flexibilité dans la modélisation de la tendance. Les modèles généralisés matriciels quant à eux permettent de résoudre des modèles linéaires généralisés classiques ayant des propriétés particulières (données sous forme de matrice et matrice de régression s'écrivant comme un produit de Kronecker) de manière plus efficace et rapide avec un grand nombre de paramètres. L'idée est d'identifier dans deux composantes séparées une surface de mortalité lisse jouant le rôle de la tendance et une surface de surmortalité. La tendance ainsi déterminée donne de très bons résultats et son extrapolation reste très simple et se base sur l'extrapolation des coefficients qui correspondent à cette composante. La détermination d'une surface de surmortalité à part entière est également d'un grand intérêt. Grâce au modèle, elle se fait relativement facilement sans se baser sur des jeux d'hypothèses particuliers. Les coefficients de chocs peuvent également être utilisés à des fins d'extrapolation et de calcul de quantiles. Pour cela, l'approche retenue consiste à ajuster sur chaque série de coefficient de choc une distribution paramétrique à partir de laquelle nous pouvons simuler des réalisations et calculer des quantiles à 99,5 %. Pour des âges compris entre 20 et 50 ans, une loi à queue épaisse a été retenue (loi de Cauchy) tandis que pour les individus plus âgés, la loi retenue est une loi à queue fine (loi logistique). L'utilisation de ce modèle appliqué aux données françaises conduit à déterminer un taux de surmortalité de l'ordre de 0,25 % pour une catastrophe ayant une période de retour de deux cents ans, soit un nombre de décès supplémentaire de près de 160 000. Ce taux de surmortalité est cohérent dans la mesure où les études existantes tablent sur un taux compris entre 0,1 % et 0,61 %.

Même si ce modèle présente des défauts et n'est pas le plus complet existant, il a pour mérite d'être assez transparent, de produire des résultats robustes et de proposer une méthode relativement simple à appliquer avec un calibrage qui ne se base pas sur des retraitements de données et des hypothèses discutables. On pourrait également envisager d'appliquer ce modèle à d'autres champs de l'actuariat pour calculer des taux d'incidence d'arrêt de travail ou des taux d'hospitalisation.

## BIBLIOGRAPHIE

---

- ACAM (2004) : Note d'information sur les états C8 et C9 et sur le rapport de réassurance, [www.acam-france.fr](http://www.acam-france.fr)
- AVALOS A., GRANDVALET Y., AMBROISE C. (2005) : Pénalisation  $l_1$  pour les MAG. XXXVII<sup>èmes</sup> Journées de Statistique, Pau.
- BAUER D., KRAMER F. (2007) : Risk and Valuation of Mortality Contingent Catastrophe Bonds.
- BENFIELD (2008) : Renouveau des programmes de réassurance Catastrophe Décès Invalidité en France. <http://www.apref.org/>
- BESSE P., CARDOT H. (2001) : Modélisation statistique de données fonctionnelles. Université de Toulouse, support de cours.
- BIENVENÛE A. (2009) : Modèles linéaires généralisés. Université Lyon I, support de cours.
- BOURGEOIS-PICHAT J. (1951) : Evolution générale de la population française depuis le XVIII<sup>ème</sup> siècle. *Population*, 4 : 635-662.
- BOURGEOIS B., POUGNET G. (2008) : Réassurance et techniques financières de transfert de risque - l'exemple du « mortality bond » : une obligation indexée au risque de (sur)mortalité. *Mémoire d'actuariat CEA Paris*.
- CAIRNS A.J., BLAKE D., DOWD K. (2006) : Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk. *ASTIN Bulletin*, 36:79-120.
- CHARPENTIER A. (2002) : Titrisation des risques catastrophes : les Cat Bonds. FFSA – Direction Etudes, Statistiques & Systèmes d'information.
- CEIOPS (2007) : QIS3. Calibration of the underwriting risk, market risk and MCR. [www.ceiops.org](http://www.ceiops.org).
- CEIOPS (2008) : QIS4 background document. Calibration of SCR, MCR and proxies. [www.ceiops.org](http://www.ceiops.org).
- COMMISSION EUROPEENNE (2008) : QIS4 Technical Specifications. [http://ec.europa.eu/internal\\_market/insurance/index\\_en.htm](http://ec.europa.eu/internal_market/insurance/index_en.htm).
- CURRIE I.D., DURBAN M., EILERS P.H.C. (2006) : Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B* 68, 259-80.
- DARMON P. (2003) : La grippe espagnole submerge la France. *L'Histoire*, n°281, 79-85.

- DE BOOR C. (1978) : *A Practical Guide to Splines*. Springer, Berlin.
- DOYLE A., BONMARIN I., LEVY-BRUHL D., LE STRAT Y., DESENCLOS J-C. (2005) : Estimation de l'impact d'une pandémie grippale et analyse des stratégies. *INVS*.
- DURBAN M., CURRIE I., EILERS P. (2002) : Using P-splines to smooth two-dimensional Poisson data. *Proc 17<sup>th</sup> IWSM*. Chania, Crete. 207-214.
- EILERS P.H.C., MARX B.D. (1996) : Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89-121.
- FEDERATION FRANCAISE DES SOCIETES D'ASSURANCES (2008) : L'assurance de personnes. Données clés 2007.
- FEDERATION INTERNATIONALE DES SOCIETES DE LA CROIX ROUGE ET DU CROISSANT ROUGE (2007) : World Disasters Report, Focus on discrimination.
- GEN RE (2006) : The economic impact of pandemic influenza in the United States: priorities for intervention.
- GOURIEROUX C., MONFORT A. (2007) : Quadratic stochastic intensity and prospective mortality tables. *CREST, University of Toronto*.
- HEALTH AND HUMAN SERVICES (2005) : A potential influenza pandemic: possible macroeconomic effects and policy issues. *Congressional Budget Office*. <http://www.cbo.gov/ftpdocs/69xx/doc6946/12-08-BirdFlu.pdf>
- KARLIN S. (1969) : *Initiation aux processus aléatoires*. Paris : Dunod.
- KIRKBY J.G., CURRIE I.D. (2009) : Smooth model of mortality with period shocks.
- LAVEGNE C., TROTTIER C. (2000) : Sur l'estimation dans les modèles linéaires généralisés à effets aléatoires. *Revue de statistique appliquée*, tome 48 n°1 49-67.
- LE PAGE D. (2000) : Modélisation du risque de défaut : une approche intensité. *Mémoire d'actuariat ENSAE Paris*.
- LIN Y., COX S. (2006) : Securitization of Catastrophe mortality Risks. *Working Paper*, Youngstown State University and Georgia State University.
- LINFOOT A. (2007) : Financing Catastrophic Risk: mortality Bond Case Study. *Présentation de Scottish Re*.
- LUCIANO E., VIGNA E. (2005) : Non mean reverting affine processes for stochastic mortality. *ICER applied Working Paper Series n°4/2005*.
- METZER M.I., COX N.J., FUKUDA K. (1999) : The economic impact of pandemic influenza in the United States: priorities for intervention. *Emerging Infectious Diseases*, 5(5): 659-671.

- MILTERSEN K.R., PERSSON S.A. (2005) : Is mortality dead? Stochastic force of mortality determined by no arbitrage. *Working Paper*, University of Bergen.
- MOODYS (2007) : Bird Flu Risk for US Life Insurers. *Special Comment – Global Credit Research*.
- MUTUALITE FRANCAISE (2008) : Calibrage du risque catastrophe en assurance complémentaire santé.
- O’SULLIVAN F. (1986) : A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science*, 1, 505-527.
- O’SULLIVAN F. (1988) : Fast computation of fully automated log-density and log-hazard estimators. *SIAM J. Sci. Statist. Comput.*, 9, 363-379.
- PATTERSON KD. (1986) : Pandemic Influenza 1700-1900: a study in historical epidemiology. Totowa, New Jersey : Rowan & Littlefield.
- PAUVERT B, (2003) : Accidents collectifs, catastrophes et assurances. *Journal des Accidents et des Catastrophes n°30*. <http://www.jac.cerdacc.uha.fr/>
- PLANCHET F., THEROND P.E. (2006) : *Modèles de durée. Applications actuarielles*. Paris : Economica.
- PLANCHET F., WINTER P. (2006) : L'utilisation des splines bidimensionnels pour l'estimation de lois de maintien en arrêt de travail. *ICA*.
- PLANCHET F. (2009) : Modèles stochastiques de mortalité. *Université de Lyon I, support de cours*.
- ROTELLA F., BORNE P. (1995) : *Théorie et pratique du calcul matriciel*. Paris : Technip.
- SAPORTA G., DROESBEKE J-J., LEJEUNE M. (2005) : *Modèles statistiques pour données qualitatives*. p85, Paris : Technip.
- SAUVET C. (2006) : Solvency II – Quelle modélisation stochastique des provisions techniques prévoyance et non vie ? *Mémoire d’actuariat ISFA Lyon*.
- STRACKE A., HEINEN W. (2006) : Influenza Pandemic: The impact on an Insured Lives Portfolio. *The Actuary*.
- SWISS RE (2007) : Pandemic Influenza: a 21<sup>st</sup> century model for mortality shocks. <http://www.swissre.com/>
- WALLINGA J., TEUNIS P. (2006) : Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164(10) : 936-944.
- ZAJDENWEBER D. (2000) : *Economie des extrêmes*. Paris : Flammarion.

### *Sites Internet*

Le Robert Koch Institute : [www.rki.de](http://www.rki.de)

Le CEIOPS : [www.ceiops.org](http://www.ceiops.org)

La Commission Européenne : [http://ec.europa.eu/internal\\_market/insurance/index\\_en.htm](http://ec.europa.eu/internal_market/insurance/index_en.htm)

Human Mortality Database : <http://www.mortality.org/>

Death Toll from Disasters, War & Accident :  
<http://www.geocities.com/dtmcbride/hist/disasters-war.html#ftnt3>

Bureau de recensement de la population américaine : <http://www.census.gov/popest/>

INSEE : <http://www.insee.fr/>

## ANNEXES

---

### ANNEXE 1. CALCUL DU TAUX DE SURVIE DU MODELE DE BAUER

On part de  ${}_{T-t}P_{x_t}^{(t)} = E\left[1_{\tau_{x_t} > T} | G_t\right] = E\left[\exp\left\{-\int_t^T \mu(x_0 + s, X_s) ds\right\} | G_t\right]$  avec  $G_t = \sigma\{X_s, 0 \leq s \leq t\}$  et  $\mu(x_t, t) = \mu(x_t, Y_t, \Gamma_t) = Y_t \cdot e^{b \cdot x_t + c} + \Gamma_t$ .

$Y = (Y_t)_t$  et  $\Gamma = (\Gamma_t)_t$  suivent les équations différentielles suivantes :

$$dY_t = \alpha(\beta - Y_t)dt + \sigma\sqrt{Y_t}dW_t \text{ et } d\Gamma_t = -\kappa\Gamma_t \cdot dt + dJ_t$$

L'indépendance des deux composantes  $Y$  et  $\Gamma$  nous permet d'écrire :

$$\begin{aligned} {}_{T-t}P_{x_t}^{(t)} &= E\left[\exp\left\{-\int_t^T Y_s e^{b(x_0+s)+c} + \Gamma_s ds\right\} | G_t\right] \\ &= E\left[\exp\left\{-\int_t^T Y_s e^{b(x_0+s)+c} ds\right\} | G_t\right] E\left[\exp\left\{-\int_t^T \Gamma_s ds\right\} | G_t\right] \end{aligned}$$

Intéressons-nous tout d'abord au second membre. On suppose qu'il peut s'écrire :

$$\begin{aligned} E\left[\exp\left\{-\int_0^T \Gamma_s ds\right\} | G_t\right] &= \exp\left\{-\int_0^t \Gamma_s ds\right\} E\left[\exp\left\{-\int_t^T \Gamma_s ds\right\} | G_t\right] \\ &= \exp\left\{-\int_0^t \Gamma_s ds\right\} \exp\{\hat{u}(T-t) + \hat{v}(T-t)\Gamma_t\} \\ &= \hat{f}(t, \Gamma_t) \end{aligned}$$

où  $\hat{u}$  et  $\hat{v}$  sont deux fonctions quelconques telles que  $\hat{u}(0) = 0 = \hat{v}(0)$ .

Une application du lemme d'Itô pour les processus à saut conduit à l'équation suivante :

$$d\hat{f}(t, \Gamma_t) = \hat{f}(t, \Gamma_{t-}) \left( -\Gamma_t (1 + \hat{v}'(T-t) + \hat{v}(T-t)\kappa) - \hat{u}'(T-t) + \lambda \left( \frac{\hat{v}(T-t)}{\zeta - \hat{v}(T-t)} \right) \right) dt + dM_t$$

où  $(M_t)_t$  est une martingale. Puisque  $\hat{f}(t, \Gamma_t)$  devrait être aussi une martingale, on annule la partie en  $dt$  de l'équation précédente pour obtenir le système d'équations différentielles suivant :

$$\begin{aligned}\hat{u}'(x) &= \frac{\lambda \hat{v}(x)}{\zeta - \hat{v}(x)}, \hat{u}(0) = 0 \\ \hat{v}'(x) &= -\kappa \hat{v}(x) - 1, \hat{v}(0) = 0\end{aligned}$$

La seconde équation conduit directement à :

$$\hat{v}(x) = \frac{-1}{\kappa} (1 - e^{-\kappa x}).$$

En remplaçant cette expression dans la première équation et en calculant l'intégrale, on a alors :

$$\hat{u}(x) = \frac{-\lambda x}{\zeta \kappa + 1} + \frac{-\lambda \zeta}{\zeta \kappa + 1} \log \left\{ 1 + \frac{1}{\zeta \kappa} (1 - e^{-\kappa x}) \right\}.$$

On s'intéresse maintenant au premier membre et on suppose là aussi que l'on peut écrire :

$$\begin{aligned}E \left[ \exp \left\{ -\int_0^T Y_s e^{b(x_0+s)+c} ds \right\} | G_t \right] &= \exp \left\{ -\int_0^t Y_s e^{b(x_0+s)+c} ds \right\} E \left[ \exp \left\{ -\int_t^T Y_s e^{b(x_0+s)+c} ds \right\} | G_t \right] \\ &= \exp \left\{ -\int_0^t Y_s e^{b(x_0+s)+c} ds \right\} \exp \left\{ u(T-t) + v(T-t) Y_t e^{b(x_0+t)+c} \right\}\end{aligned}$$

où  $u$  et  $v$  sont deux fonctions quelconques telles que  $u(0) = 0 = v(0)$ . Là aussi, une application du lemme d'Itô nous conduit au système d'équations différentielles suivant :

$$\begin{aligned}u'(x) &= \alpha \beta v(x) e^{b(x_0+T-x)+c}, u(0) = 0 \\ v'(x) &= -1 - (\alpha - b)v(x) + \frac{1}{2} \sigma^2 v^2(x) e^{b(x_0+T-x)+c}, v(0) = 0\end{aligned}$$

Au final, on obtient bien :

$$\begin{aligned}_{T-t} p_{x_t}^{(t)} &= \exp \left\{ u(T-t) + v(T-t) Y_t e^{b(x_0+t)+c} \right\} \exp \left\{ -\frac{\Gamma_t}{\kappa} (1 - e^{-\kappa(T-t)}) - \frac{\lambda(T-t)}{\zeta \kappa + 1} \right\} \\ &\quad \exp \left\{ \frac{\lambda \zeta}{\zeta \kappa + 1} \log \left[ 1 + \frac{1}{\zeta \kappa} (1 - e^{-\kappa(T-t)}) \right] \right\}\end{aligned}$$

## ANNEXE 2. RESOLUTION D'UNE EQUATION DIFFERENTIELLE ORDINAIRE DE RICCATI

Ce sont des équations du type :

$$y' = a + by + cy^2$$

où  $a$ ,  $b$  et  $c$  sont des fonctions qui peuvent dépendre de  $x$ . Dans le cas du modèle de Bauer que nous avons présenté, on rappelle que l'on a une équation de la forme :

$$v'(x) = -1 - (\alpha - b)v(x) + \frac{1}{2}\sigma^2 v^2(x) e^{b(x_0 + T - x) + c}, v(0) = 0$$

où encore :

$$v' = -1 - (\alpha - b)v + Kv^2q(x), v(0) = 0$$

avec  $q(x) = e^{-bx}$ .

En utilisant le changement de variable  $w(x) = Kv(x)q(x)$ , on supprime déjà le coefficient devant le carré :

$$w' = w^2 - Kq - \alpha w, w(0) = 0$$

puis en utilisant le changement de variable  $w = -\frac{z'}{z}$  on obtient :

$$z'' + \alpha z' - Kqz = 0, z(0) = 0.$$

Cette dernière équation n'est pas une équation différentielle linéaire et il n'existe pas de méthode pour trouver une solution à cette équation. Cependant, si l'on peut trouver une solution particulière, alors la méthode de variation de la constante nous donnera une solution générale. Aucune solution particulière évidente ne fonctionnant, on choisit de faire l'approximation  $e^{-bx} = 1 + O(1)$  qui est justifiée par le fait que  $b$  est petit et que  $x$  (c'est-à-dire  $T-t$  dans le cadre de notre modèle) prendra seulement les valeurs 1 ou 2.

On obtient donc l'équation simplifiée suivante :

$$z'' + \alpha z' - Kz = 0, z(0) = 0$$

qui a pour équation caractéristique  $X^2 + \alpha X - K = 0$ .

Les racines réelles de cette équation sont :

$$r_1 = \frac{-\alpha - \sqrt{\alpha^2 + 4K}}{2}, r_2 = \frac{-\alpha + \sqrt{\alpha^2 + 4K}}{2}$$

et les solutions sont donc de la forme :

$$z(x) = \gamma e^{r_1 x} + \mu e^{r_2 x}.$$

La solution de notre équation en  $v$  sera donc :

$$v(x) = -\frac{z(x)'}{Kz(x)q(x)} = -\frac{\gamma r_1 e^{r_1 x} + \mu r_2 e^{r_2 x}}{K e^{-bx} (\gamma e^{r_1 x} + \mu e^{r_2 x})}, v(0) = 0.$$

On rappelle que l'on avait aussi l'équation suivante que l'on doit résoudre :

$$u'(x) = \alpha \beta v(x) e^{b(x_0 + T - x) + c}, u(0) = 0$$

où encore :

$$u'(x) = -\frac{2\alpha\beta}{\sigma^2} \frac{\gamma r_1 e^{r_1 x} + \mu r_2 e^{r_2 x}}{\gamma e^{r_1 x} + \mu e^{r_2 x}}, u(0) = 0$$

que l'on résout facilement :

$$u(x) = -\frac{2\alpha\beta}{\sigma^2} \ln(\gamma e^{r_1 x} + \mu e^{r_2 x}), u(0) = 0.$$

A l'aide des conditions initiales  $v(0) = 0$ ,  $u(0) = 0$ , on obtient le système suivant :

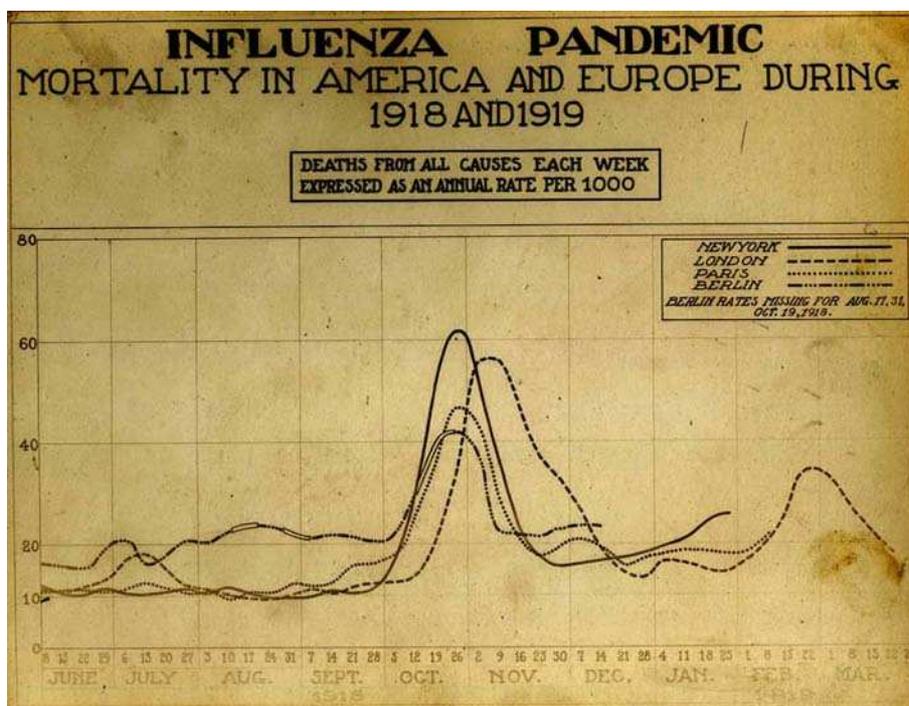
$$\begin{cases} \mu + \gamma = 1 \\ \gamma r_1 + \mu r_2 = 0 \end{cases}$$

qui conduit à :

$$\begin{cases} \mu = \frac{r_1}{r_1 - r_2} \\ \gamma = 1 - \frac{r_1}{r_1 - r_2} \end{cases}$$

et qui achève donc la résolution du système d'équation.

### ANNEXE 3. MORTALITE AUX ETATS-UNIS ET EN EUROPE ENTRE 1918 ET 1919



### ANNEXE 4. LOIS UTILISEES

Nous rappelons ici pour mémoire les densités et les domaines de définition des trois lois que nous avons évoquées pour l'ajustement des coefficients de choc :

Loi normale :

$\mathcal{N}(\mu, \sigma^2)$  de paramètres  $\mu \in \mathbb{R}$  et  $\sigma^2 > 0$  :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \text{ pour } x \in \mathbb{R}$$

Loi de Cauchy (ou loi de Lorentz) :

*Cauchy*(*loc*, *scale*) de paramètres  $loc \in \mathbb{R}$  et  $scale > 0$  :

$$f(x) = \frac{1}{\pi \times scale \times \left[ 1 + \left( \frac{x - loc}{scale} \right)^2 \right]} \text{ pour } x \in \mathbb{R}$$

Cette loi a pour particularité de n'admettre aucun moment.

Loi logistique :

*Logistique*(*loc*, *scale*) de paramètres  $loc \in \mathbb{R}$  et  $scale > 0$  :

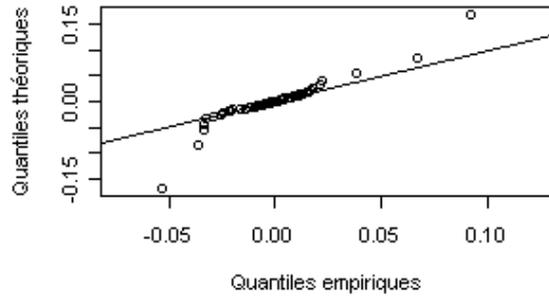
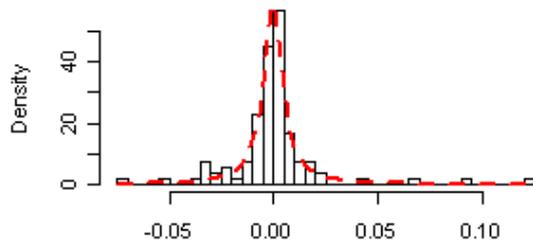
$$f(x) = \frac{\exp\left(-\frac{x - loc}{scale}\right)}{scale \times \left( 1 + \exp\left(-\frac{x - loc}{scale}\right) \right)^2} \text{ pour } x \in \mathbb{R}$$

## **ANNEXE 5. DETAILS DES AJUSTEMENTS PARAMETRIQUES**

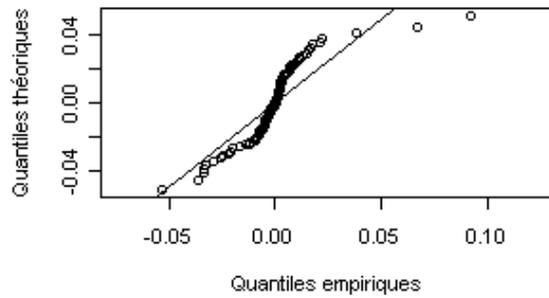
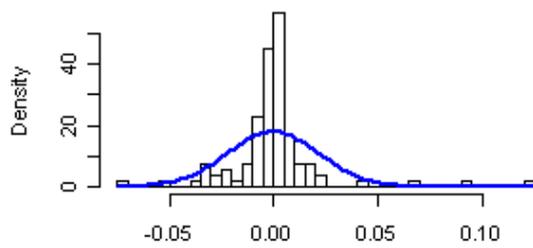
Dans cette Annexe, on reprend tous les résultats des ajustements paramétriques pour les 11 séries de coefficients de chocs. Lorsque les QQ-plots mettent en évidence des points très éloignés de la droite, ceux-ci ont été enlevés de la classe à laquelle ils appartiennent afin de ne pas fausser les résultats du test d'ajustement du khi-deux.

**Série 1**

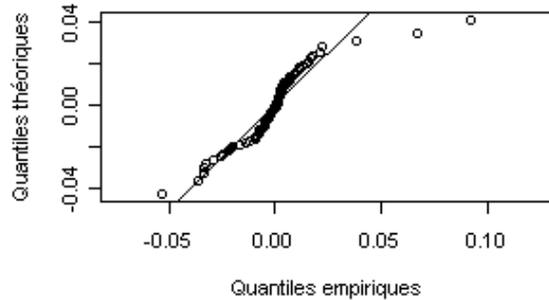
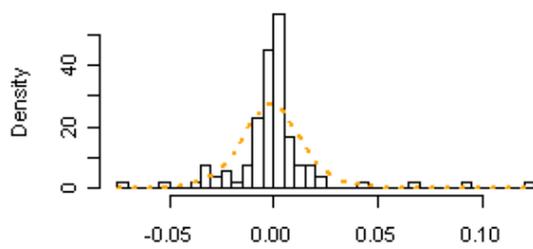
**loi de Cauchy**



**loi Normale**



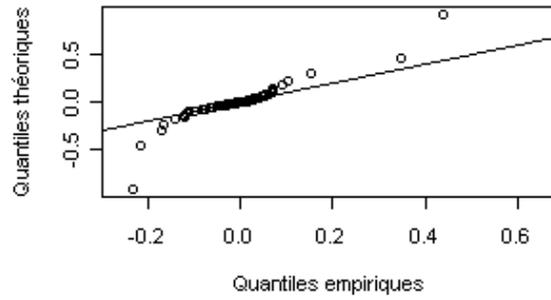
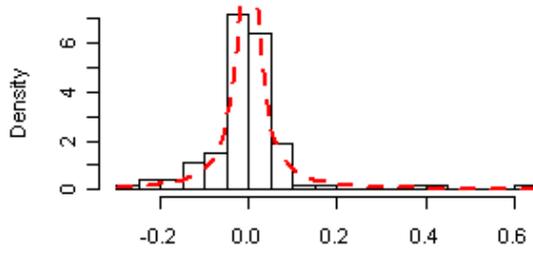
**loi Logistique**



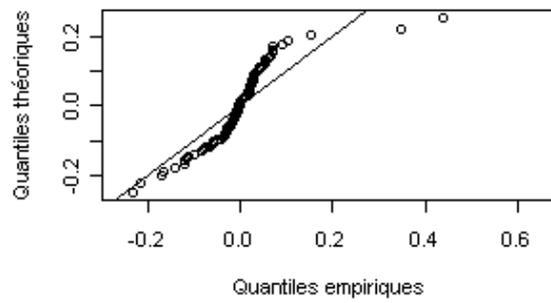
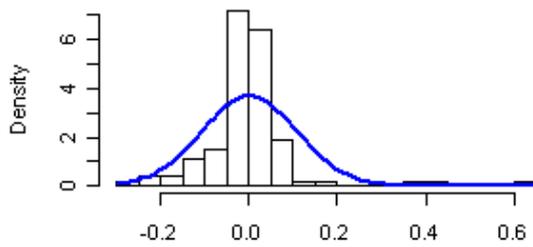
	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,0527
<b>Normale</b>	$1,937 \cdot 10^{-5}$
<b>Logistique</b>	$2,915 \cdot 10^{-4}$

**Série 2**

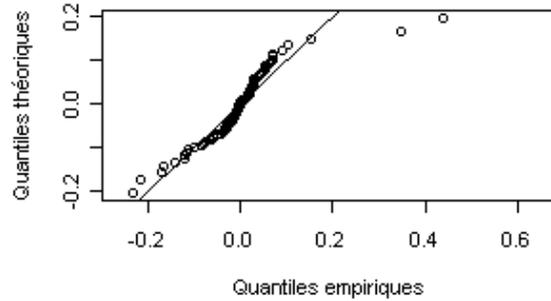
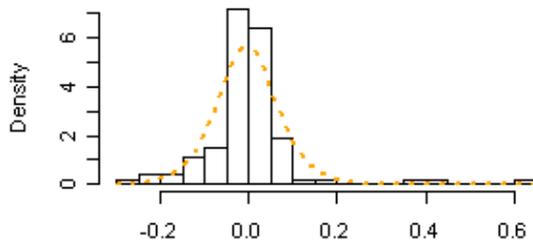
**loi de Cauchy**



**loi Normale**



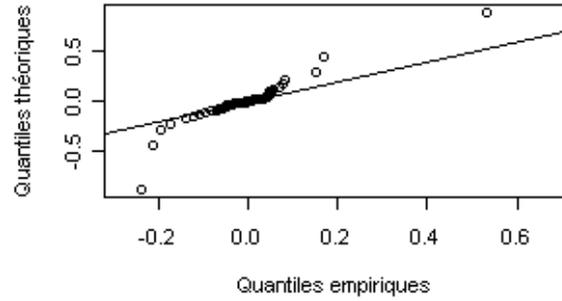
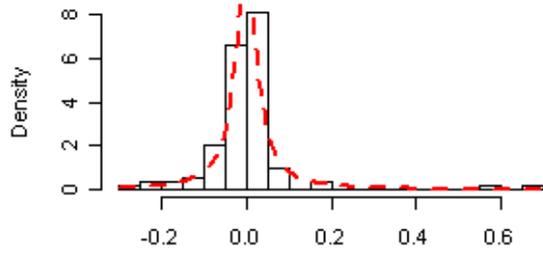
**loi Logistique**



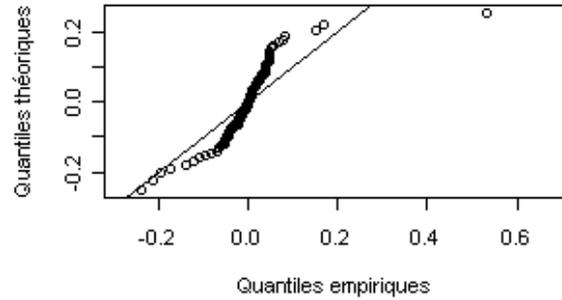
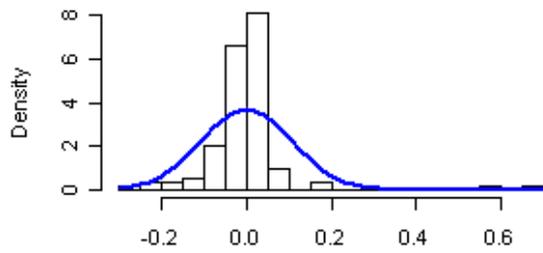
	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,2558
<b>Normale</b>	$8,65 \cdot 10^{-3}$
<b>Logistique</b>	$3,958 \cdot 10^{-2}$

**Série 3**

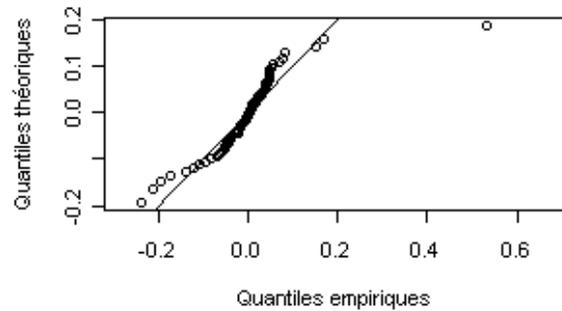
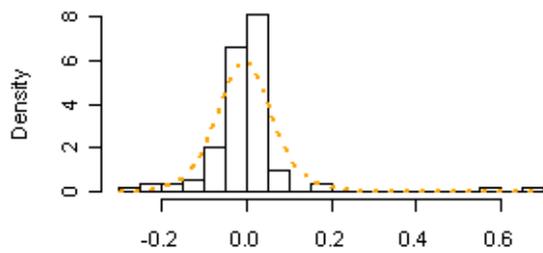
**loi de Cauchy**



**loi Normale**



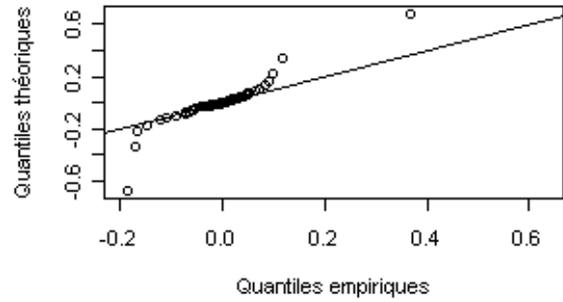
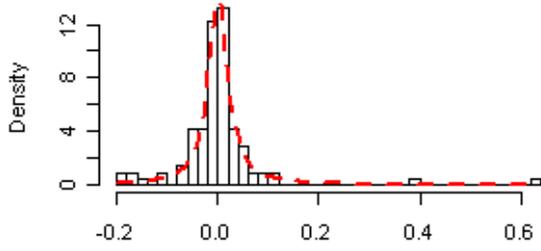
**loi Logistique**



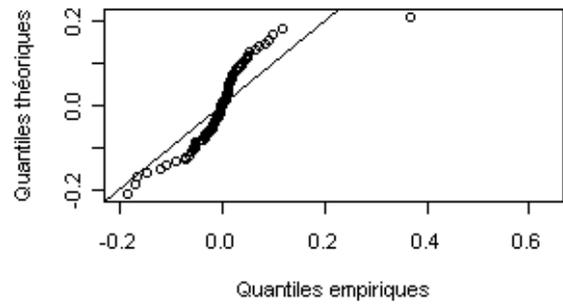
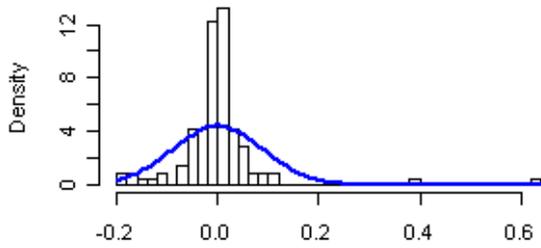
	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,7528
<b>Normale</b>	$6,317 \cdot 10^{-14}$
<b>Logistique</b>	$2,22 \cdot 10^{-3}$

**Série 4**

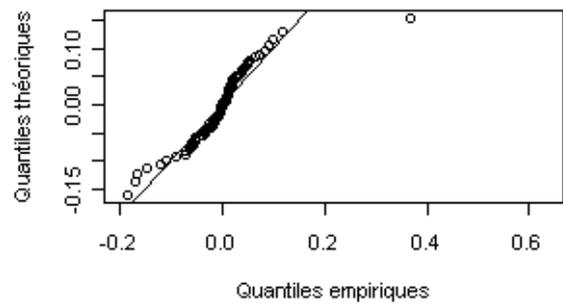
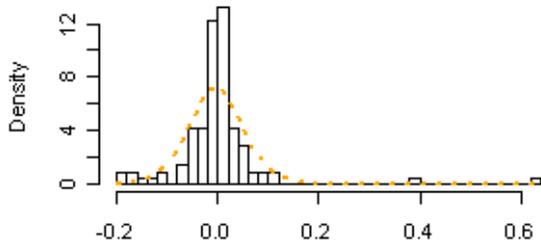
**loi de Cauchy**



**loi Normale**



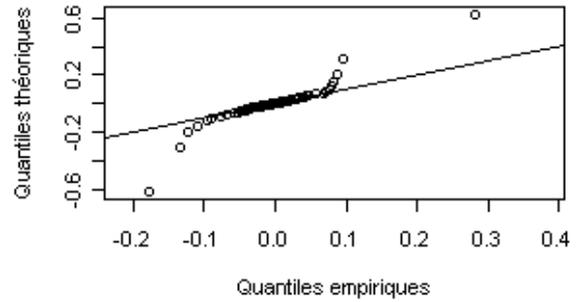
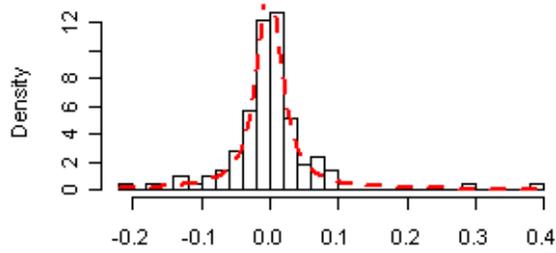
**loi Logistique**



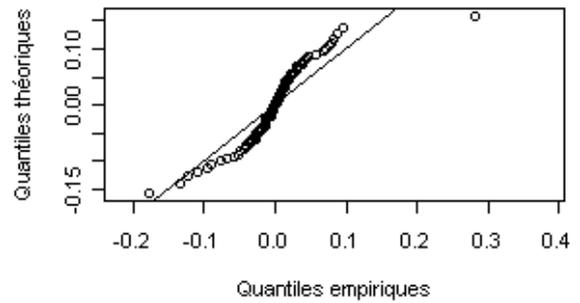
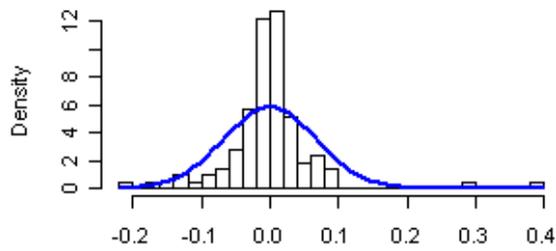
	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,32
<b>Normale</b>	$2,22 \cdot 10^{-16}$
<b>Logistique</b>	$9,62 \cdot 10^{-5}$

Série 5

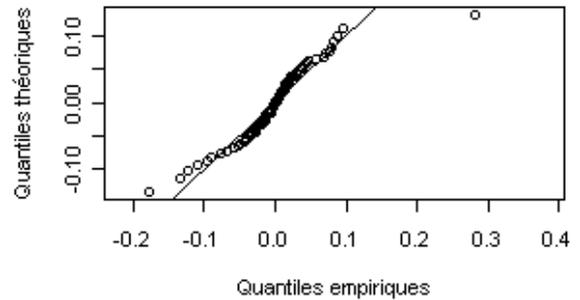
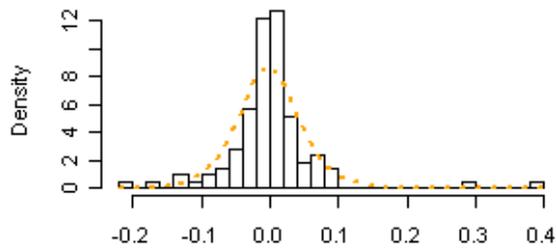
loi de Cauchy



loi Normale



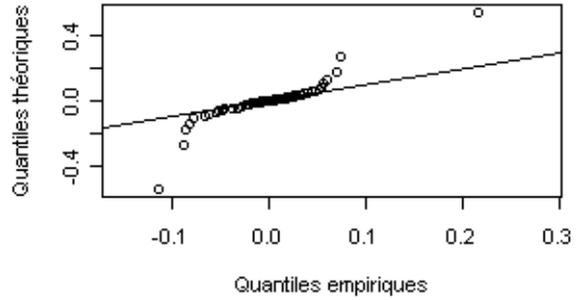
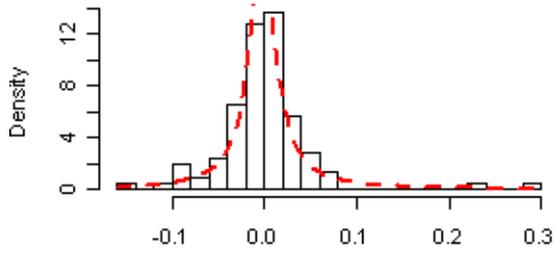
loi Logistique



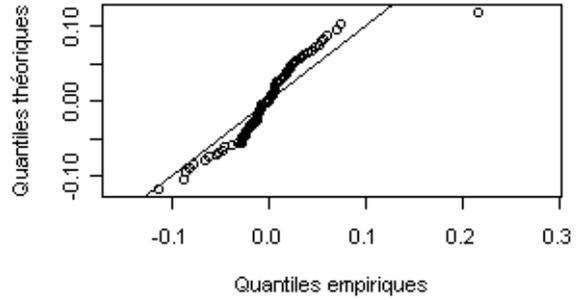
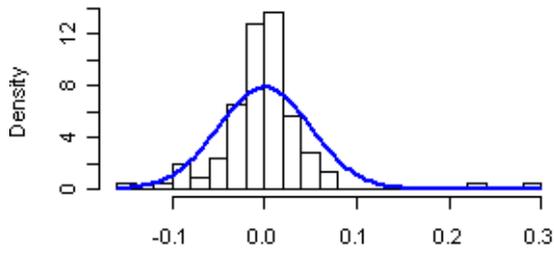
	<i>p</i> -valeur
<b>Cauchy</b>	0,9889
<b>Normale</b>	$3,39 \cdot 10^{-7}$
<b>Logistique</b>	0,03765

Série 6

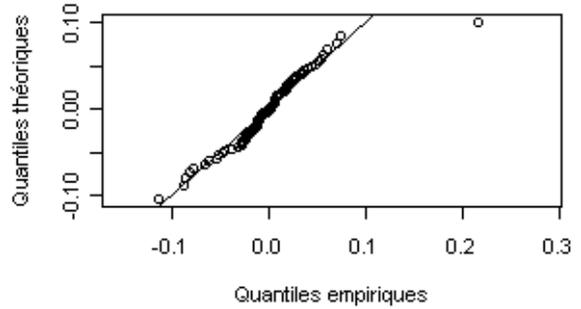
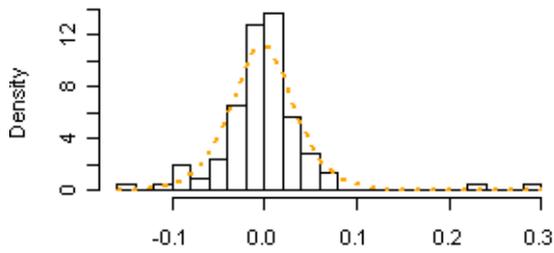
loi de Cauchy



loi Normale



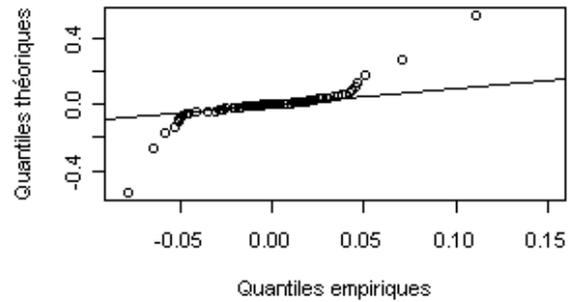
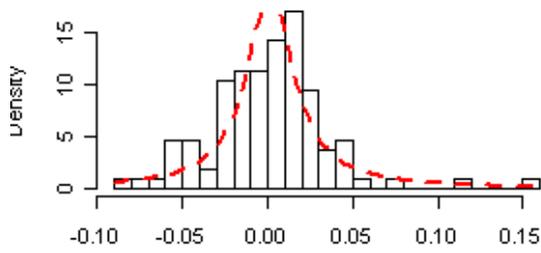
loi Logistique



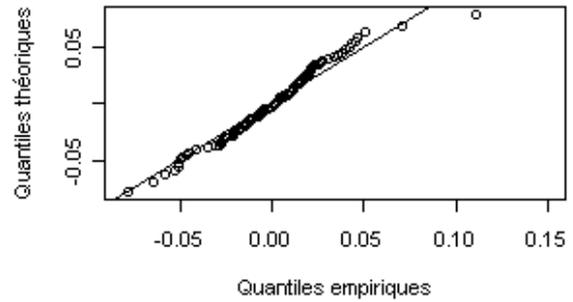
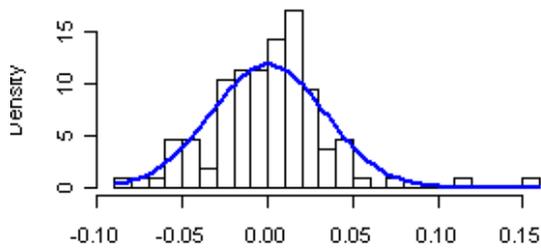
	<i>p</i> -valeur
<b>Cauchy</b>	0,5160
<b>Normale</b>	$3,788 \cdot 10^{-6}$
<b>Logistique</b>	0,1215

**Série 7**

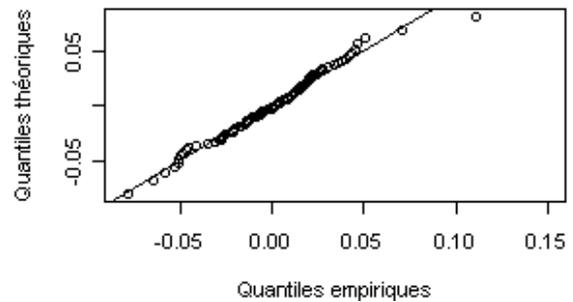
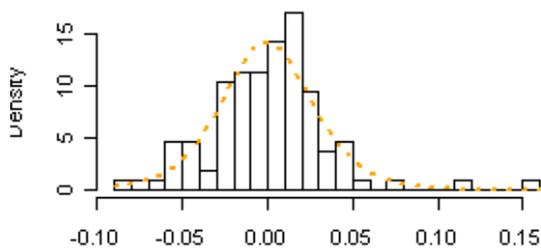
**loi de Cauchy**



**loi Normale**



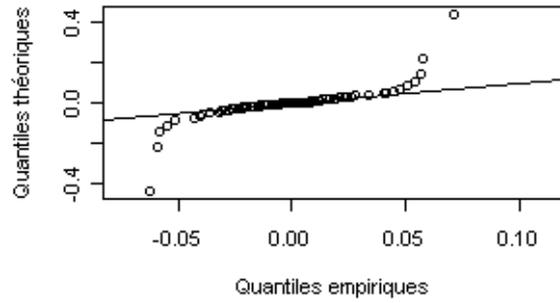
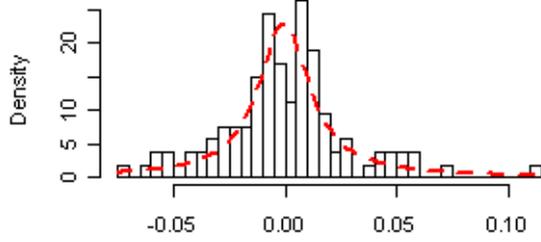
**loi Logistique**



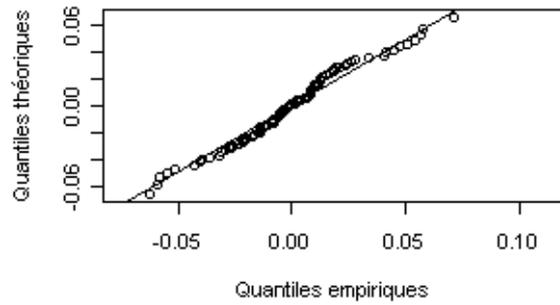
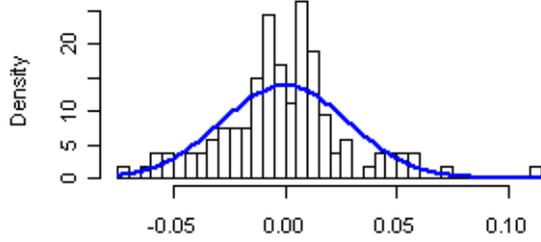
	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,0744
<b>Normale</b>	0,2967
<b>Logistique</b>	0,4956

Série 8

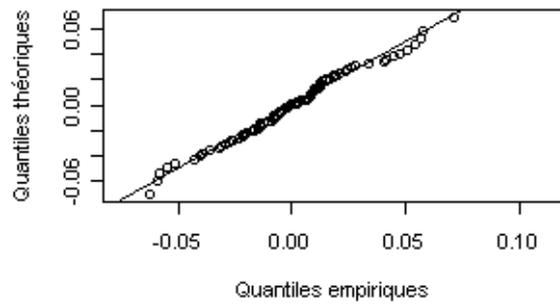
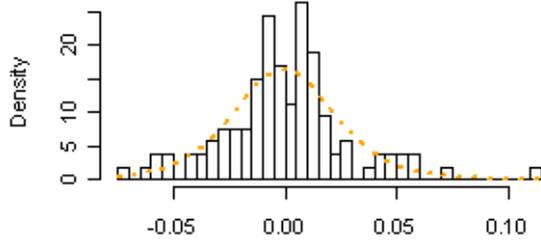
loi de Cauchy



loi Normale



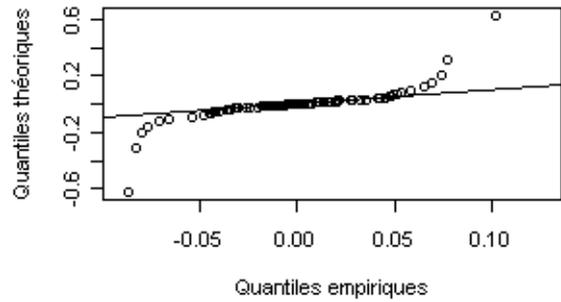
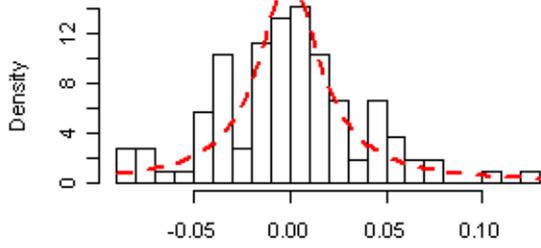
loi Logistique



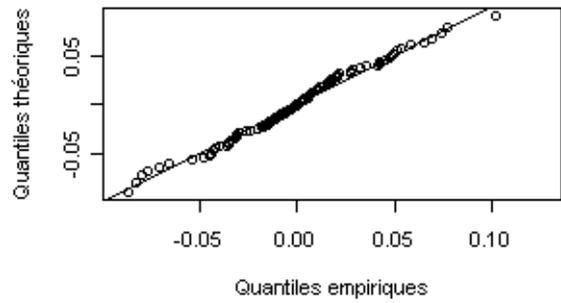
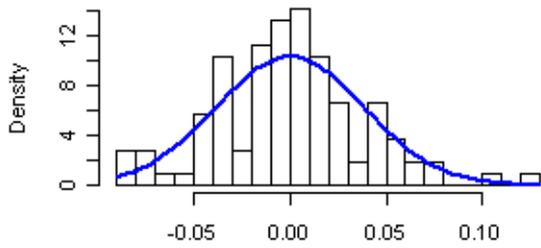
	<i>p</i> -valeur
<b>Cauchy</b>	0,1802
<b>Normale</b>	0,090
<b>Logistique</b>	0,176

**Série 9**

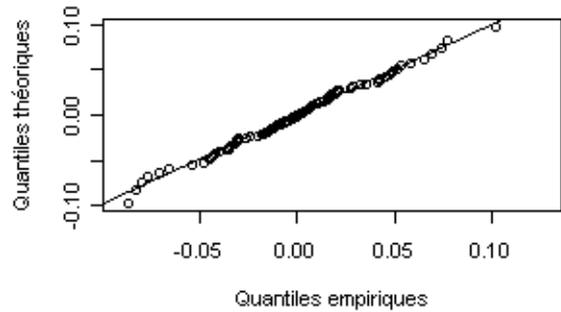
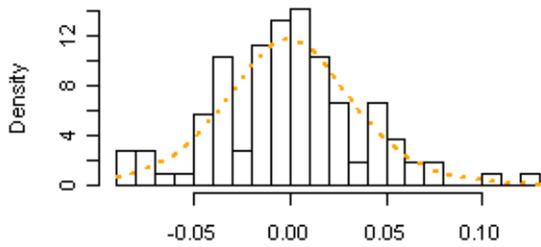
**loi de Cauchy**



**loi Normale**



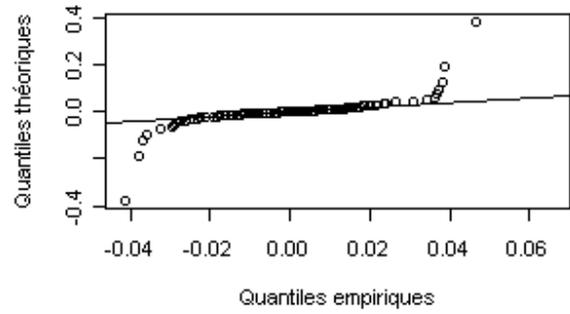
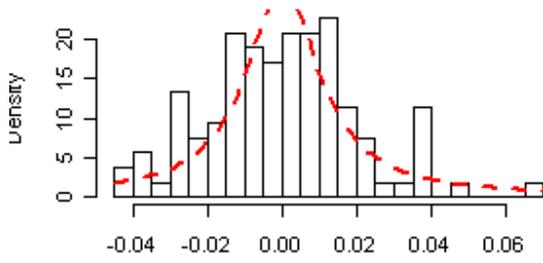
**loi Logistique**



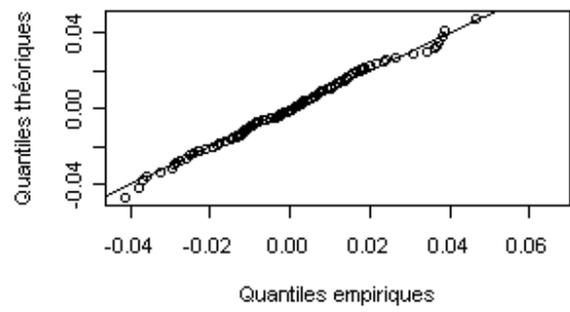
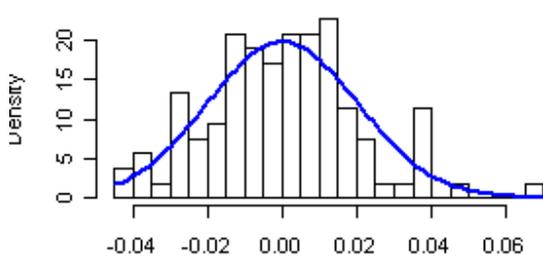
	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,1768
<b>Normale</b>	0,6363
<b>Logistique</b>	0,6374

**Série 10**

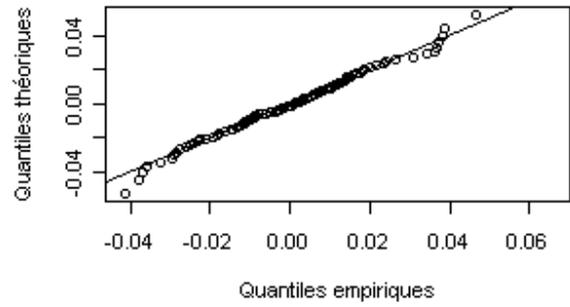
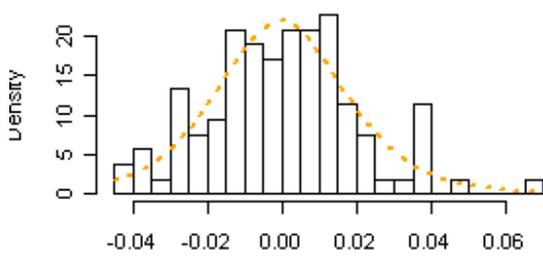
**loi de Cauchy**



**loi Normale**

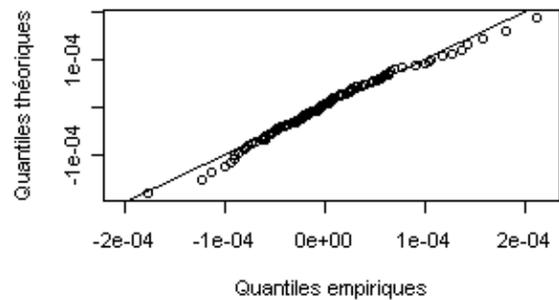
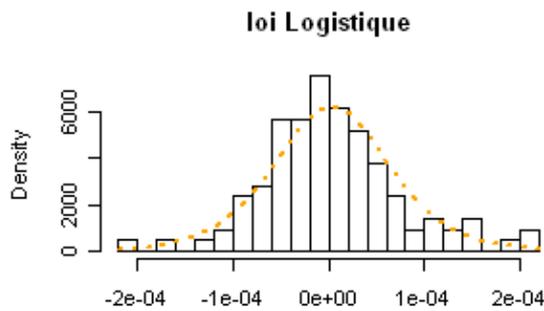
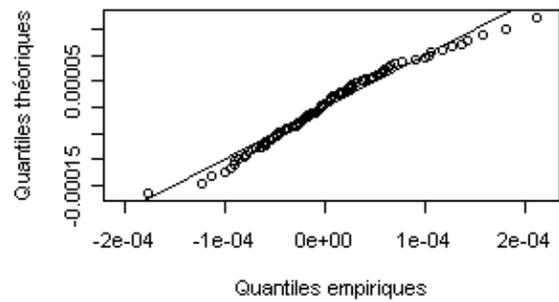
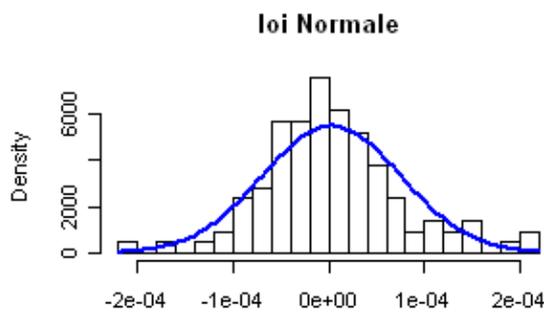
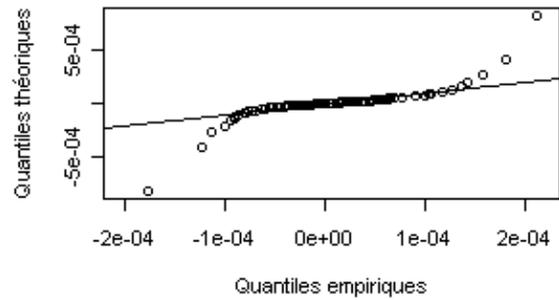
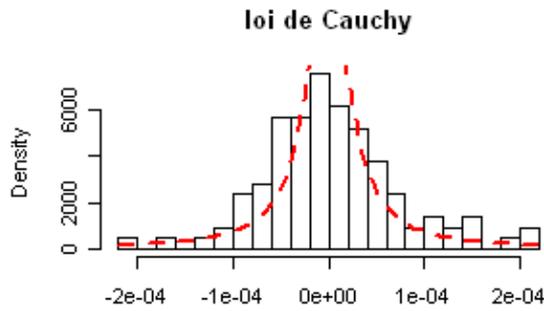


**loi Logistique**



	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,1779
<b>Normale</b>	0,4087
<b>Logistique</b>	0,4455

**Série 11**



	<b><i>p</i>-valeur</b>
<b>Cauchy</b>	0,00388
<b>Normale</b>	0,3898
<b>Logistique</b>	0,5964