

Sur une classe de transformations itérées pour l'ajustement et la simulation stochastique*

A. Bienvenue[†] et D. Rullière[†]

Résumé

Nous proposons une classe paramétrée de distorsions de probabilités qui permettent, itérées, de s'approcher aussi finement que voulu d'une fonction de survie cible. Par construction, la classe proposée respecte plusieurs propriétés que nous détaillons, et en particulier l'inversibilité analytique et la stabilité dont nous présentons l'intérêt pour certaines études actuarielles. Nous étudions l'impact asymptotique de ces distorsions sur le taux de hasard. Nous formulons les conditions sous lesquelles distorsions proposées constituent la base d'un indicateur de risque. Nous établissons la forme des compositions de distorsions ainsi que la convergence de la loi initiale déformée vers la loi cible. Une méthodologie d'estimation et des valeurs d'initialisations sont proposées. Des applications à l'analyse de la mortalité fournissent des résultats pour la modélisation d'évènements catastrophiques, ainsi que des représentations paramétrées très fidèles de l'évolution des lois de mortalité au fil des ans. Une analyse de l'impact de la paramétrisation choisie est également conduite. Nous suggérons finalement un modèle de simulation de mortalité prospective qui découle des constructions précédentes.

Mots-Clés : Distorsions de probabilités, mortalité, composées itérées, déformation hyperbolique, mesure de risque, transformée de fonction de survie, fonction d'altération.

*Ce travail a bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-08-BLAN-0314-01.

[†]Université de Lyon, F-69622, Lyon, France; Université Lyon 1, Laboratoire SAF, EA 2429, Institut de Science Financière et d'Assurances, 50 Avenue Tony Garnier, F-69007 Lyon, France. Alexis.Bienvenue@univ-lyon1.fr, Didier.Rulliere@univ-lyon1.fr

1 Introduction

Lorsque l'on s'intéresse à l'analyse de la mortalité au sein d'une compagnie d'assurances, plusieurs problèmes surgissent : d'une part, il peut être nécessaire de faire appel à une table de mortalité de référence, par exemple nationale, notamment lorsque certaines tranches d'âges sont peu représentées dans la compagnie, et que la construction d'une table complète non paramétrique est exclue. Or, la table de référence porte sur une population soumise à un risque distinct de celui de la compagnie d'assurances. Ces différences sur la population soumise au risque nécessitent parfois une adaptation d'une table en fonction d'une autre, une déformation paramétrée. D'autre part, une analyse fine de la mortalité en fonction de l'âge montre l'existence de phénomènes localisés, d'un taux de hasard non nécessairement monotone, nécessitant une forme paramétrée parfois relativement complexe. Enfin, l'analyse de l'évolution de la mortalité conduit à la recherche d'un modèle paramétrique qui puisse continuer de représenter les données fidèlement au fil des ans.

Ces problèmes ont fait l'objet d'une vaste littérature. Parmi les modèles les plus célèbres, on trouve par exemple les suivants : pour adapter des données de mortalité d'une compagnie d'assurances en fonction d'une table nationale de mortalité, des transformations à hasard proportionnel ou des transformées de Wang peuvent être utilisées (voir Wang, 1996). Heligman, Pollard (1980) vont s'intéresser précisément à la structure de la mortalité en fonction de l'âge. Lee, Carter (1992) vont décrire l'évolution de la mortalité au fil du temps, et de très nombreux autres auteurs présentent différentes techniques paramétriques pour représenter la mortalité et son évolution. Une revue d'autres modèles d'analyse des durées de survie dans un contexte dynamique peut être trouvée dans Pitacco (2004).

Pourtant, ces solutions paramétriques classiques se heurtent à plusieurs écueils :

- D'une part, ces solutions ne permettent pas d'affiner l'adéquation aux données, et l'ajout de paramètres à cette fin est relativement délicate. Ainsi, pour les transformées de Wang (1996), l'application de plusieurs transformations successives ne permet pas d'étendre la classe des fonctions de survie transformées, l'adaptation d'une table à l'aide d'un seul paramètre peut s'avérer insuffisamment précise, et l'ajout de paramètres conduit à dénaturer la transformation en question. Parmi les autres modèles, tels ceux de Heligman, Pollard (1980) ou Lee, Carter (1992), les extensions envisageables peuvent conduire à des formes très différentes selon le nombre de paramètres que l'on désire ajouter, et la convergence des formes obtenues lorsque le nombre de paramètres augmente reste alors à prouver.
- D'autre part, l'usage de nombreux paramètres pour mieux coller aux données conduit à des problèmes importants d'estimation, cette dernière ne pouvant être menée qu'en présence de valeurs d'initialisation suffisamment proches de la solution pour les paramètres. L'ajout de paramètres ou l'introduction d'un cadre prospectif requièrent donc la connaissance de valeurs d'initialisations qui sont parfois délicates à obtenir.
- Enfin, des simulations de dates de décès sont parfois menées en pratique à partir de lois de survie facilement inversibles afin d'accélérer les simulations. Ce choix conduit à s'écarter des modèles précités et à privilégier des lois simples, facilement inversibles. Là encore la bonne représentation des tables est alors limitée par l'usage de lois dont le nombre de paramètres est très réduit, telles celles de Gompertz. Ainsi, des fonctions de répartition inverses de nature paramétrique sont parfois utilisées afin d'opérer des simulations stochastiques mais l'adéquation à un ensemble de tables de mortalité ne pourra pas alors dépasser une précision fixée.

La question initiale à laquelle nous nous efforçons de répondre est celle du rapprochement de ces différentes problématiques au sein d'un même modèle. Si différentes formes paramétriques

ont ainsi pu être proposées pour traiter chacun des problèmes présentés, ces dernières revêtent des aspects très variés, et il semble intéressant de rechercher une forme paramétrique commune, utilisable pour les distorsions, pour les modélisations de tables statiques ou prospectives, pour les fonctions de répartition inverses destinées à la simulation. Par ailleurs, selon la précision désirée, l'ajustement du nombre de paramètres utilisés, sans modifier la nature de l'ajustement, reste une question particulièrement importante, et relativement délicate à traiter à l'aide des ajustements précédemment évoqués.

Pour tenter de fournir un outil utile à l'ensemble des problématiques évoquées, il apparaît naturel de proposer l'exploitation de distorsions de probabilités, et d'envisager la composition de ces distorsions. Nous allons voir que ce choix peut également servir à de nombreuses autres problématiques, comme par exemple la tarification ou la mesure du risque. Le détail de ces autres applications possibles ne sera néanmoins pas l'objet de cet article.

Comme nous l'évoquerons, les déformations de probabilités sont exploitées depuis le dix-huitième siècle pour répondre à des problèmes généraux d'économie, d'assurance ou de finance. Les objectifs poursuivis lors de l'exploitation de ces distorsions de probabilité sont variés, nous en fournissons ici un aperçu :

- Obtenir une forme paramétrique pour une quantité d'intérêt, en améliorant ainsi une référence par rapport aux données réelles (ajuster une table de mortalité officielle à des données d'entreprise, une sinistralité sur un segment par rapport à une sinistralité globale).
- Expliquer un phénomène au moyen de la déformation considérée, l'écart paramétrique à la référence étant ici le centre d'intérêt principal (expliquer l'évolution d'un phénomène au cours du temps, ou l'incidence d'un facteur quantitatif).
- Appliquer une règle prudentielle conduisant à charger davantage la queue d'une distribution, ou d'une façon plus générale tenir compte de phénomènes non observés dans les données (opérer un chargement préservant une tarification par tranche, fournir une marge de solvabilité, mesurer un risque).

Sans doute les premières déformations de probabilité peuvent-elles être attribuées à D'Alembert, J. Le Rond (1768). Quelques années après la déformation de montants au moyen d'une fonction d'utilité logarithme apparaissant dans le traité de Bernoulli, D. (1731), D'Alembert propose de déformer les probabilités elles-mêmes. D'Alembert modifie par exemple, pour un lancer de pièce de monnaie, le temps d'atteinte N d'un événement pile qui passe d'une probabilité d'atteinte au n^e coup 2^{-n} à l'un des choix suivants (avec des notations actualisées, reprise de Pradier (1998)) : $2^{-n} \frac{1}{1+\beta n^n}$, ou $2^{-n(1+\alpha)}$ ou $2^{-n(1+\frac{B}{\sqrt{K-n^q}})}$ ou encore $2^{-n(1+\frac{B}{(K-n)^{q/2}})}$, les différentes constantes n'étant pas ici détaillées. Ironiquement, le propos n'est pas alors d'intégrer une contrainte prudentielle, mais au contraire de diminuer la survenance d'événements rares tels l'obtention d'un grand nombre de tirages face, afin de répondre au fameux paradoxe de Saint-Petersbourg, stipulant qu'une loterie garantissant le gain de 2^N pour une première atteinte de pile au temps N est d'espérance infinie, alors que des individus sont prêts à accepter une somme finie pour offrir cette garantie.

Plus récemment, les distorsions de probabilités ont connu un regain d'intérêt. Ainsi, dans le champ économique, avec notamment la théorie duale de Yaari (1987) : là où les fonctions d'utilité modifient la perception d'un montant de richesse et laissent inchangées les probabilités, les distorsions laissent les montants de richesse inchangés et modifient les probabilités. Ces différences de points de vue peuvent être perçues comme les dignes héritières des vues antagonistes jadis présentées par D'Alembert et Daniel Bernoulli. Une illustration de l'application à l'assurance de la distorsion des probabilités perçues par les individus est donnée par exemple dans Bleichrodt, Eeckhoudt (2006).

Dans le champs actuariel, les distorsions de probabilités ont également été popularisées par les travaux de Wang (1996), qui a exploité différentes transformations pour la fixation de primes d'assurances, ainsi que, plus généralement, pour la mesure du risque. On pourra se référer par exemple, à Wirch, Hardy (1999) pour la présentation des distorsions de probabilités et leur exploitation à la mesure du risque.

L'évaluation et la mesure de risque pour des actifs financiers sont bien sûr concernées par les distorsions de probabilités, et l'on pourra en trouver des illustrations dans Wang (2000) ou Hamada, Sherris (2003). Des contraintes peuvent néanmoins s'exercer sur de telles évaluations, comme le détaille par exemple Pelsser (2007).

D'une façon générale, la mesure de risque, et en particulier la tarification ou l'évaluation d'actifs financiers, vont dépendre de nombreux principes ou axiomes qui seront autant de contraintes potentielles sur les distorsions de probabilité. Sans entrer dans les détails, on peut par exemple se référer à Bühlman (1980), Artzner, Delbaen, Eber, Heath (1997), Landsman, Sherris (2001), Goovaerts et al. (2004), et à de très nombreux autres auteurs qu'il serait malheureusement difficile de citer tant la littérature sur les mesures de risque est vaste. Ainsi, beaucoup de distorsions sont proposées dans une optique prudentielle ou d'évaluation du risque, suivant une axiomatique et un ensemble de contraintes propres à ce champ d'investigation.

Lorsqu'il s'agit de présenter des distorsions susceptibles de s'adapter aussi précisément que l'on souhaite à des données observées, et de permettre par exemple de préserver certaines propriétés d'inversibilité analytique de fonctions de survie, on se heurte à des contraintes bien différentes. Quelques auteurs utilisent par exemple des distorsions pour modéliser l'évolution temporelle de certains risques, telle la mortalité, avec notamment l'article de De Jong, Marshall (2007), basé sur l'évolution de paramètres de transformées de Wang afin d'opérer des projections de mortalité. Néanmoins, certaines propriétés que nous jugeons utiles ne sont pas vérifiées par les transformées utilisées, comme en particulier la capacité d'une transformation de pouvoir être itérée de façon à être aussi fidèle aux données qu'on le souhaite.

Ce seront ces contraintes de fidélité et les déformations induites que nous chercherons à détailler plus précisément, l'objectif étant principalement l'obtention d'une forme paramétrique inversible pour une quantité d'intérêt. L'exigence d'inversibilité analytique émane ici du souci pratique de faciliter la simulation des variables aléatoires continues modélisées, conditionnellement à leur appartenance à un ensemble donné. Dans cet article, les déformations seront données par de simples fonctions réelles appliquées aux fonctions déformées, et l'objectif poursuivi est également d'aborder le problème de la composition de ces fonctions. Dans l'idéal, le résultat recherché est la représentation d'une fonction de survie par la composition de multiples fonctions paramétrées. Ce but se rapproche à terme de l'un de ceux de la décomposition en ondelettes d'une fonction : obtenir une classe de fonctions assez générale pour engendrer (ici par composition) les fonctions cibles dans plusieurs types de problèmes, assez pertinente pour qu'un nombre restreint de paramètres suffise en pratique à bien les représenter, tout en préservant certaines propriétés susceptibles d'être utiles pour le traitement de certains problèmes actuariels. Face à cet objectif ambitieux, nous nous contenterons ici de présenter quelques fonctions dont les propriétés de composition et d'inversibilité nous ont paru intéressantes, et qui se sont révélées efficaces dans nos applications numériques.

2 Contraintes sur les transformations

Il est sans doute inutile de s'étendre sur l'immense variété des choix qui peuvent être faits pour déformer une probabilité, et dont même les premiers travaux de D'Alembert fournissent une illustration. Aussi, au travers de cet article, nous allons nous attacher à restreindre cet

ensemble de choix, en proposant un ensemble de contraintes qui nous ont semblé pertinentes au regard de plusieurs problématiques actuarielles.

Afin de détailler ces contraintes de façon un peu plus formelle, nous brosons ici un cadre minimaliste, qui sera détaillé par la suite, de l'application de transformations itérées de fonctions de survie. Nous nous intéresserons à une classe \mathcal{T} de déformations, qui s'appliqueront sur des fonctions de survie d'une classe \mathcal{S} , de sorte que les transformées soient elles-mêmes des fonctions de survie :

$$\forall T \in \mathcal{T}, \forall S \in \mathcal{S}, T \circ S \in \mathcal{S}.$$

La classe des déformations \mathcal{T} sera constituée d'un ensemble fonctions réelles T_θ , pour des vecteurs de paramètres $\theta \in \Theta$, $\Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}^*$:

$$\mathcal{T} = \{T_\theta : [0, 1] \rightarrow [0, 1]\}_{\theta \in \Theta}.$$

Nous essaierons de déterminer une déformation possédant un nombre limité de paramètres et une expression analytique susceptible d'être calculée simplement à l'aide des logiciels ou langages usuels. Nous nous fixerons cinq contraintes pour ces transformations, détaillées dans les paragraphes suivants.

C1. Inversibilité Les techniques de simulations étant très communément utilisées dans les travaux actuariels, la préservation du caractère inversible d'une fonction de survie déformée passe par la connaissance analytique de l'inverse de la déformation T_θ considérée : cette connaissance permet alors, par fonction de survie inverse d'une variable aléatoire uniforme sur un intervalle inclus dans $]0, 1[$, de simuler aisément une variable aléatoire de loi déformée, conditionnellement à l'appartenance de la variable aléatoire à un intervalle donné. Le choix du travail sur les fonctions de survie s'explique quant-à-lui par la présence, tant en assurance-vie qu'en assurance dommages, de conditionnements sur le dépassement de seuils par les variables aléatoires considérées. Plus formellement, cette contrainte se traduira par :

$$\forall u \in]0, 1[, \forall T_\theta \in \mathcal{T}, \exists !v \in]0, 1[, T_\theta(v) = u. \quad (1)$$

C2. Stabilité Dans l'idéal, nous chercherons également à préserver l'intérêt intuitif de pouvoir déformer une fonction dans un sens comme dans l'autre, en exigeant que les déformations inverses soient de même nature que les déformations elles-mêmes. Cela facilitera l'obtention de propriétés de symétrie, ainsi que l'implémentation informatique des déformations et déformations inverses. De surcroît, sous cette contrainte, l'interversion entre fonction cible et fonction initiale modifiera les paramètres des déformations, mais ne modifiera pas la nature des déformations utilisées, ce qui peut se justifier en l'absence d'informations particulières a priori sur la forme de la cible ou des fonctions approchées. Plus formellement, cette contrainte se traduira par :

$$\forall T_\theta \in \mathcal{T}, \exists \theta' \in \Theta, T_\theta^{-1} = T_{\theta'} \in \mathcal{T}. \quad (2)$$

C3. Régularité Une contrainte très pragmatique est de pouvoir interpréter la déformation, et pouvoir en estimer ses paramètres. Nous chercherons par exemple à déterminer l'impact de chacun des paramètres sur certaines quantités communément utilisées (espérances, primes stop-loss...), à cerner les conséquences de la fixation des valeurs minimales et maximales admissibles pour chaque paramètres. Cela nous conduira à établir des contraintes sur les paramètres, à savoir les composantes des vecteurs θ , $\theta \in \Theta$. Afin de bénéficier d'arguments quantitatifs lorsqu'un paramètre varie, et par souci de simplicité, nous préférons imposer que l'ensemble des valeurs

de paramètres soit un pavé ouvert de \mathbb{R}^p . Selon la même logique, l'interprétation de l'impact d'un paramètre sur la déformation ne devrait pas conduire à séparer l'analyse en différents cas, et doit pouvoir être interprétée logiquement, ce qui nous amènera à formuler des conditions de dérivabilité et de continuité : une variation d'un paramètre conduit naturellement à une variation de la déformation, et nous supposons que cette variation doit être continue. Cela nous conduira à supposer que toutes les dérivées partielles de T_θ par rapport aux composantes de θ sont continues, de sorte que T_θ sera continûment différentiable.

Il faut noter que l'estimation des paramètres découlant en pratique d'une optimisation, ces contraintes de continue différentiabilité s'avèreront parfaitement naturelles, comme cela sera détaillé dans un paragraphe d'estimation. Finalement, la contrainte de régularité sera formulée comme suit :

$$\Theta \text{ pavé ouvert de } \mathbb{R}^p, \quad (3)$$

$$\forall x \in [0, 1], \theta \mapsto T_\theta(x) \text{ continûment différentiable,} \quad (4)$$

$$\forall \theta \in \Theta, x \mapsto T_\theta(x) \text{ continûment différentiable.} \quad (5)$$

C4. Convergence Afin notamment de répondre au besoin d'approcher au mieux un jeu de données à partir d'une fonction de survie de référence, nous fixerons une contrainte de convergence : l'application itérée de déformation doit permettre de réduire une distance spécifiée entre n'importe quelle fonction de survie cible et n'importe quelle fonction de survie initiale, de façon à ce que les transformées itérées de la référence initiale convergent, dans un sens qui sera spécifié, vers la fonction de survie cible. Nous supposons de surcroît que lorsque la fonction de survie initiale est identique à la cible, la transformation ne déforme pas cette cible : la fonction identité fera donc partie de la classe de déformation considérée. Plus formellement la contrainte de convergence se traduira par :

$$\text{Id} \in \mathcal{T}, \quad (6)$$

$$\forall S_0, S_1 \in \mathcal{S}, \exists \text{ une suite } (T_i)_{i \in \mathbb{N}} \text{ d'éléments de } \mathcal{T}, T_n \circ \dots \circ T_1(S_0) \xrightarrow{n \rightarrow +\infty} S_1. \quad (7)$$

C5. Paramétrisation Il est possible de modifier la paramétrisation d'une déformation au moyen d'une bijection \mathcal{H} de l'ensemble Θ des paramètres vers un nouvel ensemble $\tilde{\Theta}$. On peut ainsi remplacer une déformation T_θ par $\tilde{T}_\theta = T_{\mathcal{H}(\theta)}$. Si l'ensemble des déformations reste le même, l'interprétation donnée aux paramètres, les contraintes s'imposant sur ceux-ci, ainsi que la facilité d'estimation pourront être modifiés. Nous avons ici choisi de privilégier certaines paramétrisations. Nous préférons les paramètres d'une transformation inverse lorsqu'ils s'expriment simplement en fonction des paramètres de la transformation initiale. Parmi ces paramétrisations préférées, dont la définition reste subjective, nous distinguerons une classe particulière qui peut être décrite plus formellement : de par l'axiome C2, il existe une bijection I_T qui à tout $\theta \in \Theta$ associe un $\theta' \in \Theta$ tel que $T_\theta^{-1} = T_{\theta'}$, et nous présenterons des paramétrisations conduisant à

$$I_T(\theta) = D_T \cdot \theta, \quad (8)$$

$$D_T \text{ matrice diagonale de diagonale } \vec{d} = (d_1, \dots, d_p), d_1, \dots, d_p \in \{-1, 1\}. \quad (9)$$

Nous qualifierons une telle paramétrisation de symétrique. Lors du passage à la transformation inverse, un paramètre numéro i sera inchangé si $d_i = 1$, nous dirons alors qu'il s'agit d'un paramètre de position. Il changera au contraire de signe si $d_i = -1$, nous dirons qu'il s'agit alors d'un paramètre d'éloignement. Outre la facilité d'implémentation des transformations inverses,

ce choix conduira pour les paramètres d'éloignement à une valeur nulle pour la transformation identité, et à une gestion parfaitement symétrique des transformations et de leur inverse. Néanmoins, l'ensemble Θ_0 des paramètres conduisant à l'identité, $\Theta_0 = \{\theta, T_\theta = \text{Id}\}$, contiendra tous les vecteurs paramètres pour lesquels les composantes d'éloignement sont nulles, quelles que soient les valeurs des composantes de position. Il pourra ainsi s'avérer utile de réduire le nombre de paramètres de position non directement fixés avant l'estimation. En présence d'un paramètre d'éloignement θ_e , il est toujours possible remplacer un paramètre θ_p de position par $\theta_e \theta_p$ ou $\theta_p(1 - e^{\theta_e})/(1 + e^{\theta_e})$, par exemple, qui sont des paramètres d'éloignement. En l'absence de paramètre de position, les paramètres bénéficieront de l'absence de contrainte de positivité, et un vecteur paramètre nul conduira à la transformation identité. La paramétrisation sera dite complètement symétrique dans la situation où

$$\Theta = \mathbb{R}^p, \quad (10)$$

$$\forall \theta \in \Theta, \quad T_\theta^{-1} = T_{-\theta}, \quad (11)$$

ce qui implique en particulier $T_{\vec{0}} = \text{Id}$ et peut ainsi faciliter l'interprétation de l'évolution de chaque paramètre lors de l'estimation. Un avantage des paramétrisations complètement symétriques est qu'elles offrent également la possibilité de supprimer un paramètre tout en conservant les transformations inverses dans la même classe, en choisissant simplement 0 pour la valeur du paramètre étudié.

3 Transformations étudiées

3.1 Définitions

On placera notre transformation dans l'échelle des logit, qui a montré son intérêt dans de multiples modèles. Nous restreindrons ici le champ d'application des déformations aux fonctions de survie de variables aléatoires réelles. Soit \mathcal{S} l'ensemble des fonctions de survie des variables aléatoires réelles d'espérance finie, c'est-à-dire l'ensemble des fonctions S décroissantes et càd-làg de \mathbb{R} dans $[0, 1]$ telles que $S(x) = 1$ pour tout $x \leq 0$ et $\int_0^{+\infty} S(t) dt < \infty$. Pour $S \in \mathcal{S}$ et f une fonction bijective croissante de \mathbb{R} dans \mathbb{R} , on note T_f la fonction de $[0, 1]$ dans $[0, 1]$ définie par

$$T_f(u) = \begin{cases} 0 & \text{si } u = 0, \\ \text{logit}^{-1}(f(\text{logit}(u))) & \text{si } 0 < u < 1, \\ 1 & \text{si } u = 1. \end{cases}$$

et $\mathbb{T}_f(S)$ la fonction de survie définie par

$$\mathbb{T}_f(S)(x) = T_f(S(x)).$$

Nous appellerons f la fonction d'altération de la transformation T_f . La fonction logit et son inverse, $\text{logit}(x) = \ln(x/(1-x))$ et $\text{logit}^{-1}(x) = 1/(1+e^{-x})$, sont ici utilisés de façon très classique afin de garantir que, quelle que soit la transformation f utilisée, la transformée reste bien comprise entre 0 et 1. Ce choix n'est pas crucial, la déformation de la fonction de survie reposant essentiellement sur la fonction f . Il faut noter que l'on aurait pu pour ce choix prendre n'importe quelle fonction de répartition Φ et son inverse. Toutefois, le choix de la fonction d'altération permet aisément de passer d'une situation à l'autre :

$$\text{logit}^{-1}(f(\text{logit}(u))) = \Phi\left(\tilde{f}(\Phi^{-1}(u))\right),$$

avec le choix

$$f(u) = \text{logit}(\Phi(\tilde{f}(\Phi^{-1}(\text{logit}^{-1}(u))))).$$

En particulier, même si les objectifs poursuivis sont ici différents, la transformation spécifiée peut permettre de retrouver la transformée de Wang (1996), lorsque Φ désigne la fonction de répartition d'une loi normale, et $\tilde{f}(x) = x + \lambda$, $\lambda \in \mathbb{R}$. L'avantage premier de la fonction logit par rapport à d'autres fonctions qui auraient pu être choisies est sa facilité d'inversion analytique. Il faut également noter sa simplicité de calcul, les fonctions exponentielles et logarithmes étant directement calculables par le coprocesseur arithmétique de la plupart des ordinateurs actuels, à l'inverse des fonctions de répartition de lois normales. On remarquera enfin que les transformations T se composent de la manière suivante :

$$\mathbb{T}_g \circ \mathbb{T}_f = \mathbb{T}_{g \circ f},$$

et l'on a en conséquence

$$(\mathbb{T}_f)^{-1} = \mathbb{T}_{(f^{-1})}.$$

3.2 Impact sur les caractéristiques de la variable aléatoire

On considère une variable aléatoire réelle X de fonction de survie S , et une variable aléatoire \hat{X} de fonction de survie $\hat{S} = \mathbb{T}_f(S)$. Nous allons observer dans cette section la façon dont plusieurs caractéristiques de X ont été modifiées par cette transformation.

Variable aléatoire transformée

Il est naturel de s'interroger sur l'existence d'une fonction réelle qui permettrait de passer directement de X à \hat{X} . De par les propriétés élémentaires des transformations étudiées, on établit facilement le résultat suivant :

Proposition 1 Si l'on suppose la fonction de survie S inversible, et si l'on définit la fonction g_S par

$$g_S(t) = S^{-1} \circ \text{logit}^{-1} \circ f^{-1} \circ \text{logit} \circ S(t), \quad t \in \mathbb{R},$$

alors la fonction g_S est telle que pour tout $t \in \mathbb{R}$,

$$\mathbb{T}_f S(t) = \text{P} [g_S(X) > t].$$

En particulier, \hat{X} a même loi que $g_S(X)$ et cette représentation de la transformation fournit un lien direct avec les modèles à temps de défaut accéléré (Accelerated Failure Times, AFT), dont on trouvera une présentation et des extensions dans Bagdonavicius, Nikulin (2002).

Taux de hasard

La taux de hasard de X est la fonction μ définie sur \mathbb{R} par

$$\mu(t) = -\frac{S'(t)}{S(t)}.$$

Notons $\hat{\mu}$ le taux de hasard de \hat{X} . Comme $\hat{S} = \mathbb{T}_f(S)$,

$$\frac{\hat{S}'(t)}{\hat{S}(t)} = \frac{T_f'(S(t))S'(t)}{T_f(S(t))}.$$

Avec la définition de T_f , on obtient rapidement

$$\frac{T'_f(u)}{T_f(u)} = (1 - T_f(u)) \frac{f'(\text{logit}(u))}{u(1-u)},$$

de sorte que, lorsque $S(t) < 1$,

$$\frac{\widehat{\mu}(t)}{1 - \widehat{S}(t)} = f'(\text{logit}(S(t))) \frac{\mu(t)}{1 - S(t)}. \quad (12)$$

En particulier, on peut établir le résultat suivant :

Proposition 2 Si $\mu(t)$ et $\widehat{\mu}(t)$ désignent les taux de hasard respectifs d'une variable aléatoire et de sa transformée, alors quand t tend vers l'infini,

$$\frac{\widehat{\mu}(t)}{\mu(t)} \sim f'(\text{logit}(S(t))). \quad (13)$$

Ainsi, quand f admet une direction asymptotique $f'(t) \rightarrow a$, le taux de hasard est multiplié par a à l'infini.

Fonctions de survie à variations régulières

On s'intéresse ici au cas où S est à variations régulières d'exposant $\rho \leq 0$, c'est-à-dire :

$$\frac{S(tx)}{S(t)} \xrightarrow[t \rightarrow +\infty]{} x^\rho \quad \forall x > 0.$$

Dans ce cas,

$$\ln(S(tx)) - \ln(S(t)) \xrightarrow[t \rightarrow +\infty]{} \rho \ln(x),$$

et, comme $\ln(1 - S(tx))$ et $\ln(1 - S(t))$ tendent tous deux vers 0 quand t tend vers l'infini,

$$\text{logit}(S(tx)) - \text{logit}(S(t)) \xrightarrow[t \rightarrow +\infty]{} \rho \ln(x).$$

On écrit maintenant

$$\begin{aligned} \frac{\widehat{S}(tx)}{\widehat{S}(t)} &= \frac{1 + e^{-f(\text{logit}(S(t)))}}{1 + e^{-f(\text{logit}(S(tx)))}} \\ &= \frac{e^{f(\text{logit}(S(tx)))} + e^{f(\text{logit}(S(tx))) - f(\text{logit}(S(t)))}}{e^{f(\text{logit}(S(tx)))} + 1} \end{aligned}$$

Dans le cas où f admet une asymptote en $-\infty$:

$$f(u) - (au + b) \xrightarrow[u \rightarrow -\infty]{} 0, \quad (14)$$

les exposants $f(\text{logit}(S(tx)))$ et $f(\text{logit}(S(t)))$ tendent vers $-\infty$ quand t tend vers l'infini, et la différence $f(\text{logit}(S(tx))) - f(\text{logit}(S(t)))$ tend vers $a\rho \ln(x)$, de sorte que, finalement,

$$\frac{\widehat{S}(tx)}{\widehat{S}(t)} \xrightarrow[t \rightarrow +\infty]{} x^{a\rho},$$

ce qui signifie que \widehat{S} est à variations régulières d'exposant $a\rho$: la pente a de l'asymptote de f en $-\infty$ joue le rôle de coefficient multiplicatif pour l'exposant de variations régulières de S . Notons maintenant (pour les valeurs de p où l'intégrale converge)

$$Z_p^*(x) = \int_x^{+\infty} t^p S(t) dt = \frac{1}{p+1} \mathbb{E} \left[(X^{p+1} - x^{p+1})_+ \right].$$

En particulier, si X est un montant de sinistre, $Z_0^*(x) = \mathbb{E}(X - x)_+$ représentera l'espérance des sinistres à la charge d'un réassureur en cas de traité Stop-loss de priorité x . Dans le cas où S est à variation lente, le théorème 1, p. 281, du livre de Feller (voir Feller, 1968), nous fournit l'équivalent suivant quand $x \rightarrow +\infty$ pour $\rho + p + 1 < 0$:

$$Z_p^*(x) \sim \frac{x^{p+1} S(x)}{-(\rho + p + 1)}.$$

Pour les primes Stop-Loss, on dispose donc de l'équivalent suivant lorsque $\rho < -1$:

$$Z_0^*(x) \sim \frac{xS(x)}{-(\rho + 1)}.$$

Toujours avec l'hypothèse (14), on obtient

$$\frac{\widehat{S}(x)}{e^b S(x)^a} = \frac{1}{e^b S(x)^a + e^{a \ln(1-S(x))+o(x)}}$$

de sorte que, si $a\rho + p + 1 < 0$,

$$\widehat{Z}_p^*(x) \sim \frac{x^{p+1} e^b S(x)^a}{-(a\rho + p + 1)}.$$

Ces résultats permettent en particulier de lier les primes Stop-Loss avant et après transformation :

Proposition 3 Soient $Z_0^*(x) = \mathbb{E}(X - x)_+$ l'espérance des sinistres à la charge d'un réassureur en cas de traité Stop-loss de priorité x , et $\widehat{Z}_0^*(x)$ la même espérance sur la variable aléatoire modifiée. Si le charge initiale des sinistres est à variations régulières d'exposant $\rho \leq 0$, et si f admet une asymptote de pente a en $+\infty$, alors

$$\widehat{Z}_0^*(x) \sim e^b \frac{(-(\rho + 1))^a}{-(a\rho + 1)} x^{1-a} Z_0^*(x)^a \quad \text{quand } x \rightarrow +\infty.$$

3.3 Conditions de cohérence pour mesure du risque

Comme nous l'avons évoqué en introduction, l'objet des distorsions présentées n'est pas initialement la mesure du risque. Néanmoins, il est bien connu que l'on peut construire des mesures de risque à partir de distorsions de probabilité. Ces mesures de risque sont vues comme des fonctionnelles de la loi de probabilité de la variable aléatoire considérée : on peut définir une mesure R_m du risque porté par une variable aléatoire positive X , de fonction de survie S , comme :

$$R_f(S) = \int_0^{+\infty} T_f(S(x)) dx.$$

où T_f désigne une distorsion de probabilité. Comme cela est mentionné dans l'article de Wirch, Hardy (1999), de par les travaux de Denneberg (1994), il est acquis qu'une telle mesure est cohérente au sens de Artzner, Delbaen, Eber, Heath (1997) si T_f est concave. Par souci de simplicité, nous dirons qu'une fonction d'altération est cohérente si elle conduit à une mesure R_f cohérente.

Un des intérêts des distorsions de probabilité cohérentes est de permettre un principe de chargement de primes qui soit compatible avec l'axiomatique précitée de mesure du risque et compatible avec une tarification par tranche, en proposant par exemple comme prime chargée pour la variable aléatoire $(\min(X, b) - a) \mathbb{1}_{X > a}$, $0 < a < b$:

$$\Pi_{[a,b]} = \int_a^b T_f(S(x)) dx .$$

Pour une fonction d'altération f , on montre que les dérivées première et seconde de T_f au point $x \in]0, 1[$ sont respectivement égales à :

$$\begin{aligned} T'_f(x) &= \frac{\hat{x}(1 - \hat{x})}{x(1 - x)} f'(\text{logit } x) , \\ T''_f(x) &= \frac{\hat{x}(1 - \hat{x})}{x^2(1 - x)^2} [(1 - 2\hat{x})f'(\text{logit } x)^2 + f''(\text{logit } x) + (2x - 1)f'(\text{logit } x)] , \end{aligned}$$

où $\hat{x} = T_f(x)$, f' et f'' désignant les dérivées première et seconde de f . Il s'ensuit que $T''_f(x)$ a le signe de $s_f(x) = (1 - 2\hat{x})f'(\text{logit } x)^2 + f''(\text{logit } x) + (2x - 1)f'(\text{logit } x)$. On en déduit facilement le résultat ci-après.

Proposition 4 Si la fonction d'altération f est telle que pour tout $x \in]0, 1[$

$$f''(\text{logit } x) \leq 2f'(\text{logit } x)^2(\hat{x} - x) + (1 - 2x)(1 - f'(\text{logit } x))f'(\text{logit } x) \quad (15)$$

alors la mesure de risque

$$R_f(S) = \int_0^{+\infty} T_f(S(x)) dx ,$$

est cohérente au sens de Artzner, Delbaen, Eber, Heath (1997).

Si la classe des $\{T_f\}_{\theta \in \Theta}$ contient des fonctions concaves, elle contiendra également, de par l'axiome de stabilité, les fonctions inverses qui ne sont pas concaves. Si l'on recherche des fonctions f destinées à produire des mesures cohérentes de risque, et si l'on choisit d'utiliser l'axiomatique présentée précédemment, des conditions sur les paramètres seront impératives pour indiquer si la mesure associée est cohérente.

Si T_f est concave, avec $T_f(0) = 0$ et $T_f(1) = 1$, il est clair que pour tout $x \in]0, 1[$, $T_f(x) > x$ et par suite

$$\forall y \in \mathbb{R}, f(y) \geq y .$$

En particulier, il est impossible que f soit strictement concave sur tout \mathbb{R} . En posant $y = \text{logit } x$, on établit la condition pour qu'une fonction d'altération soit cohérente, à vérifier pour tout $y \in \mathbb{R}$

$$f''(y) \leq 2f'(y)^2 \sigma_f(y) + \frac{e^{-y} - 1}{e^{-y} + 1} (1 - f'(y))f'(y) ,$$

avec

$$\sigma_f(y) = \text{logit}^{-1}(f(y)) - \text{logit}^{-1}(y) .$$

Une réécriture de σ_f conduit à $\sigma_f(y) = \frac{e^{-y} - e^{-f(y)}}{(1 + e^{-y})(1 + e^{-f(y)})}$, puis au résultat suivant :

Proposition 5 Soit f une fonction d'altération telle que pour tout $y \in \mathbb{R}$,

$$f''(y) \leq f'(y)^2(h \circ f)(y) - f'(y)h(y),$$

avec $h(y) = \frac{1 - e^{-y}}{1 + e^{-y}}$ fonction croissante de y

alors la mesure de risque

$$R_f(S) = \int_0^{+\infty} T_f(S(x))dx,$$

est cohérente au sens de Artzner, Delbaen, Eber, Heath (1997).

L'objet du présent article n'est pas spécifiquement de fournir des mesures de risque : nous allons voir toutefois qu'il est possible de trouver des fonctions d'altération conduisant à une mesure de risque cohérente. Considérons en particulier les fonctions d'altérations du type $f(y) = ay + b$, avec $a > 0$. Si $a \neq 1$, on vérifie facilement que $s_f(0^+)s_f(1^-) < 0$, la mesure de risque associée R_f ne pourra pas être cohérente, ce qui est relativement logique puisque dans ce cas, il existe des abscisses pour lesquelles $T_f(x) < x$. Par contre pour $a = 1$, lorsque $f(y) = y + b$, la mesure R_f est cohérente dès lors que $b > 0$, ce qui conduit bien à aggraver la loi de sinistralité S . Ce premier exemple fournit une déformation cohérente qui s'opère dans le même esprit que la transformée de Wang (1996), seule l'échelle changeant, la fonction logit remplaçant ici la fonction de répartition d'une loi normale.

3.4 Classes des fonctions d'altération considérées

Les fonctions que nous considérerons pour f sont les suivantes :

Les fonctions affines Ces fonctions sont définies par deux paramètres $p > 0$ et m :

$$D_{p,m}(x) = p(x - m) + m.$$

Elles sont bien entendu inversibles, avec $(D_{p,m})^{-1} = D_{1/p,m}$. p est la pente de la droite, et m le seuil pour lequel $D_{p,m}(m) = m$. La transformation n'aura pas le même impact avant et après ce seuil. Le choix de paramètres $\rho = \ln p$ et de m conduit à un paramètre d'éloignement et un paramètre de position.

On peut remarquer que pour ces fonctions, la transformation induite correspond à celle suggérée par Brass (voir par exemple Brass (1969) ou Brass (1974)).

Le choix de $h = m(1 - p)/(1 + p)$ en remplacement de m permet la paramétrisation complètement symétrique suivante :

$$\begin{aligned} \bar{D}_{\rho,h}(x) &= D_{e^\rho, m(1+p)/(1-p)}(x), \\ \bar{D}_{\rho,h}(x) &= e^\rho(x + h) + h, h \in \mathbb{R}, \rho \in \mathbb{R}, \\ \bar{D}_{\rho,h}^{-1} &= \bar{D}_{-\rho, -h}. \end{aligned}$$

ρ est la pente de la droite en échelle logarithmique, et h est la hauteur d'intersection de la droite avec celle d'équation $y = -x$.

Les fonctions « angle » Elles ont quatre paramètres : la position du sommet de l'angle (x_0, y_0) , ainsi que les deux pentes $p_1 > 0$ et $p_2 > 0$ de l'angle. Elles s'expriment de la façon suivante :

$$A_{x_0, y_0, p_1, p_2}(x) = \begin{cases} y_0 + p_1(x - x_0) & \text{si } x \leq x_0, \\ y_0 + p_2(x - x_0) & \text{si } x \geq x_0. \end{cases}$$

Ces fonctions sont bijectives, avec

$$(A_{x_0, y_0, p_1, p_2})^{-1} = A_{y_0, x_0, 1/p_1, 1/p_2}.$$

En remplaçant (x_0, y_0) par $(m, h_1) = (\frac{x_0+y_0}{2}, \frac{y_0-x_0}{2})$, m devient un paramètre de position et h_1 devient un paramètre d'éloignement (projections du sommet de l'angle sur les droites d'équations respectives $y = x$ et $y = -x$). On peut alors remplacer m par le paramètre d'éloignement $h_2 = sm$, où s est un paramètre d'éloignement, par exemple $s = \text{signe}(p_1 - p_2)$, choisi de façon à préserver la symétrie de l'angle, et qui n'est nul que dans des situations dégénérées (comprenant bien entendu la fonction identité). Une paramétrisation complètement symétrique des fonctions angles est alors la suivante :

$$\bar{A}_{\rho_1, \rho_2, h_1, h_2}(x) = A_{x_0, y_0, p_1, p_2},$$

avec $p_1 = e^{\rho_1}$, $p_2 = e^{\rho_2}$, $x_0 = sh_2 - h_1$, $y_0 = sh_2 + h_1$, où $s = \text{signe}(\rho_1 - \rho_2)$. On obtient alors

$$\begin{aligned} \bar{A}_{\rho_1, \rho_2, h_1, h_2}^{-1} &= \bar{A}_{-\rho_1, -\rho_2, -h_1, -h_2} \\ (\rho_1, \rho_2, h_1, h_2) &\in \mathbb{R}^4, \\ \bar{A}_{0,0,0,0} &= \text{Id}. \end{aligned}$$

le paramètre h_1 désigne l'écartement du sommet par rapport à la première diagonale d'équation $y = x$, le paramètre h_2 est quant-à-lui nul lorsque le sommet de l'angle se situe sur la deuxième diagonale d'équation $y = -x$. Dans le repère (O, \vec{i}, \vec{j}) , avec $\vec{i} = (1, 1)$ et $\vec{j} = (-1, 1)$, h_1 mesure la position verticale du sommet de l'angle, et h_2 sa position horizontale. Par la suite, nous noterons également $A_{p_1, p_2}^0 = A_{0,0,p_1,p_2}$ l'angle élémentaire ayant son sommet en $(0, 0)$, de sorte que $A_{x_0, y_0, p_1, p_2} = y_0 + A_{p_1, p_2}^0(x - x_0)$. Cela nous permettra en particulier d'utiliser les relations pour $\alpha > 0$:

$$\begin{aligned} A_{\alpha p_1, \alpha p_2}^0(x/\alpha) &= A_{p_1, p_2}^0(x), \\ \alpha A_{p_1, p_2}^0(x) &= A_{\alpha p_1, \alpha p_2}^0(x) = A_{p_1, p_2}^0(\alpha x), \\ A_{p_1, p_2}^0 \circ A_{p'_1, p'_2}^0 &= A_{p_1 p'_1, p_2 p'_2}^0. \end{aligned}$$

Les fonctions hyperboliques qui représentent un adoucissement des précédentes. Elles ont cinq paramètres : la position du sommet de l'angle (x_0, y_0) , les pentes respectives des deux asymptotes p_1, p_2 , et l'écartement ϵ (on prendra par convention $\text{signe}(0) = 0$) :

$$\begin{aligned} H_{x_0, y_0, p_1, p_2, \epsilon}(x) &= y_0 + \frac{p_1 + p_2}{2}(x - x_0) - \text{signe}(p_1 - p_2) \sqrt{\left(\frac{p_1 - p_2}{2}(x - x_0)\right)^2 + \sqrt{p_1 p_2} \epsilon^2}, \\ H_{x_0, y_0, p_1, p_2, \epsilon}^{-1} &= H_{y_0, x_0, 1/p_1, 1/p_2, \epsilon}. \end{aligned}$$

On remarquera que

$$H_{x_0, y_0, p_1, p_2, 0} = A_{x_0, y_0, p_1, p_2}. \quad (16)$$

On peut également proposer une paramétrisation complètement symétrique de l'hyperbole, qui pourra avoir un intérêt lors de la phase d'estimation des paramètres de l'hyperbole :

$$\bar{H}_{\rho_1, \rho_2, h_1, h_2, e} = H_{x_0, y_0, p_1, p_2, \epsilon},$$

avec $p_1 = e^{\rho_1}$, $p_2 = e^{\rho_2}$, $x_0 = sh_2 - h_1$, $y_0 = sh_2 + h_1$, $\epsilon = se$, où $s = \text{signe}(\rho_1 - \rho_2)$. On a alors :

$$\begin{aligned} \bar{H}_{\rho_1, \rho_2, h_1, h_2, e}^{-1}(x) &= \bar{H}_{-\rho_1, -\rho_2, -h_1, -h_2, -e}(x), \\ (\rho_1, \rho_2, h_1, h_2, e) &\in \mathbb{R}^5, \\ \bar{H}_{0,0,0,0,0} &= \text{Id}. \end{aligned}$$

Les fonctions composées de deux angles Il pourra être utile de partir de fonctions déjà composées pour plusieurs raisons :

- lors d'un procédé d'estimation itératif, il est possible qu'il faille s'éloigner tout d'abord de la cible pour pouvoir ensuite s'en approcher davantage, et l'utilisation de fonctions déjà composées peu permettre d'éviter cet écueil.
- La composition de fonctions peut rendre caduque l'intérêt de certains paramètres. Ainsi, la composée de n fonctions angles est entièrement déterminée par $2n + 2$ paramètres, et non n fois les 4 paramètres d'un angle : connaître la forme d'une composée peut permettre de réduire ainsi le nombre de paramètres utilisés.
- Une connaissance particulière (par exemple asymptotes en $y = x$ pour une transformation locale) peut nous amener à simplifier l'expression de fonctions composées et à réduire encore le nombre de paramètres utilisés.
- La connaissance de la fonction composée peut permettre de définir un ensemble de paramètres dont l'effet sur la courbe résultante et la signification sera plus claire par rapport à l'agrégation des paramètres de chaque fonction composée.

La composée de deux angles possède six degrés de libertés, et l'on constate rapidement que la transformation $A_{x_0, y_0, p_1, p_2} \circ A_{x'_0, y'_0, p'_1, p'_2}$ ne dépend que des six valeurs $p_1 p'_1$, $p_2 p'_2$, p_2/p_1 , $p_1(x_0 - y'_0)$, x'_0 , y_0 . En exploitant la paramétrisation symétrique précédemment présentée pour l'angle, on voit que la composée de $\bar{A}_{\rho_1, \rho_2, h_1, h_2} \circ \bar{A}_{\rho'_1, \rho'_2, h'_1, h'_2}$, ne dépend donc que de $v_1 = \rho_1 + \rho'_1$, $v_2 = \rho_2 + \rho'_2$, $v_3 = \rho_2 - \rho_1$, $v_4 = e^{\rho_1}[(s'h'_2 + h'_1) - (sh_2 - h_1)]$, $v_5 = s'h'_2 - h'_1$, $v_6 = sh_2 + h_1$. On peut en premier lieu s'interroger sur les paramétrisations susceptibles de présenter une composition de deux angles, nécessitant 6 degrés de libertés, comme composée de deux angles à k et k' paramètres, avec $k + k' = 6$.

S'agissant du cas $k = k' = 3$, on remarque que si l'on pose $h_1 = h'_1 = 0$ (sommets d'angles sur la diagonale $y = x$), on peut déduire ρ_1 de $v_4 = e^{\rho_1}(v_6 - v_5)$, puis ρ_2 de $v_3 = \rho_2 - \rho_1$, puis ρ'_1 et ρ'_2 de v_1 et v_2 , puis (lorsque les coefficients s et s' sont non nuls) h'_2 de v_5 et h_2 de v_6 . Finalement, quels que soient les angles composés, on peut retrouver le résultat de la composition même si $h_1 = h'_1 = 0$. Ce résultat n'est toutefois valable que si $v_4/(v_6 - v_5) > 0$, c'est-à-dire $\frac{x_0 - y'_0}{y_0 - x'_0} > 0$ (on montre que l'on peut aisément traiter le dernier cas problématique $r = 0$ ou $r' = 0$). On obtient alors le résultat suivant :

Remarque 1 Toute composition de deux angles de sommets (x_0, y_0) et (x'_0, y'_0) , avec $\frac{x_0 - y'_0}{y_0 - x'_0} > 0$, peut s'écrire sous la forme de la composée de deux angles dont les sommets sont sur la droite d'équation $y = x$:

$$\bar{A}_{\rho_1, \rho_2, 0, h_2} \circ \bar{A}_{\rho'_1, \rho'_2, 0, h'_2}.$$

ou

$$A_{m, m, p_1, p_2} \circ A_{m', m', p'_1, p'_2}.$$

Nous pourrions écrire la totalité des compositions d'angles en prenant le premier sommet (x_0, y_0) soit sur la droite d'équation $y = x$, soit sur celle d'équation $y = -x$. Ce résultat est intéressant dans la mesure où il pourra nous permettre de composer des angles à 3 paramètres seulement, ce qui nous permettra explicitement de n'utiliser que six paramètres au lieu de huit. Il permet en outre l'obtention d'une symétrie d'écriture entre la composée de deux angles et l'inverse de cette composée, s'écrivant directement comme composée de deux inverses, et dont l'appartenance à une même classe de fonctions est immédiate. Il peut être facilement vérifié en écrivant

$$A_{x_0, y_0, p_1, p_2} \circ A_{x'_0, y'_0, p'_1, p'_2}(x) = y_0 + A_{\alpha p_1, \alpha p_2}^0 \left(A_{\frac{p'_1}{\alpha}, \frac{p'_2}{\alpha}}^0 (x - x'_0) + x'_0 - \left(x'_0 - \frac{y'_0 - x_0}{\alpha} \right) \right), \quad (17)$$

en posant $\alpha = (x_0 - y'_0)/(y_0 - x'_0)$, lorsque cette quantité est définie et positive. La situation $x_0 = y'_0$ conduit quant-à-elle à confondre les deux sommets d'angle, avec alors $A_{x_0, x_0, p_1, p_2} \circ A_{x'_0, x'_0, p'_1, p'_2}(x) = y_0 + A_{p_1 p'_1, p_2 p'_2}^0 (x - x_0)$. Afin de mieux exploiter la composée de n fonctions, nous abandonnerons néanmoins la composée de deux angles à trois paramètres au profit des représentations suivantes.

S'agissant du cas $k = 2$ et $k' = 4$, afin de gérer les composées successives de fonctions, il peut être intéressant d'écrire une composée de n angles comme un angle à 4 paramètres composé avec $n - 1$ angles à deux paramètres, ce qui permet alors de retrouver les $2n + 2$ degrés de libertés de la composée. Une contrainte supplémentaire que l'on s'impose lors du choix des deux paramètres du premier angle, est que l'inverse d'un tel angle à deux paramètres soit aussi un angle à deux paramètres de la même classe, ce qui exclut des choix comme $A_{0, y_0, 1, p_2}$ dont l'inverse $A_{y_0, 0, 1, 1/p_2}$ ne s'écrit pas $A_{0, y'_0, 1, p'_2}$. Nous verrons par la suite que cette contrainte est fondamentale pour pouvoir composer plus de deux angles. De l'équation 17, on déduit le résultat suivant :

$$A_{x_0, y_0, p_1, p_2} \circ A_{x'_0, y'_0, p'_1, p'_2} = A_{y_0, y_0, 1, \frac{p_2}{p_1}} \circ A_{x'_0, p_1(y'_0 - x_0) + y_0, p_1 p'_1, p_1 p'_2} \quad (18)$$

$$\left(A_{x_0, y_0, p_1, p_2} \circ A_{x'_0, y'_0, p'_1, p'_2} \right)^{-1} = A_{x'_0, x'_0, 1, \frac{p'_1}{p'_2}} \circ A_{y_0, -\frac{1}{p'_1}(y'_0 - x_0) + x'_0, \frac{1}{p_1 p'_1}, \frac{1}{p_1 p'_2}} \quad (19)$$

Si l'on pose $m = \frac{x'_0 + y_0}{2}$, $h = \frac{x'_0 - y_0}{2}$, $w = \sqrt{\frac{p'_1}{p_1}}(y'_0 - x_0)$, $\eta_1 = \ln p_1 p'_1$, $\eta_2 = \ln p_2 p'_2$, $\alpha = \ln p'_2 p_1 - \frac{\ln(p_1 p'_1) + \ln(p_2 p'_2)}{2}$, alors :

$$\begin{aligned} \mathbb{A}_{\eta_1, \eta_2, \alpha, m, h, w} &= A_{x_0, x_0, 1, p_2} \circ A_{x'_0, y'_0, p'_1, p'_2}, \\ (\mathbb{A}_{\eta_1, \eta_2, \alpha, m, h, w})^{-1} &= \mathbb{A}_{-\eta_1, -\eta_2, \alpha, m, -h, -w}, \\ (\eta_1, \eta_2, \alpha, m, h, w) &\in \mathbb{R}^6. \end{aligned}$$

avec : $x_0 = m - h$, $p_2 = e^{\frac{\eta_2 - \eta_1}{2} - \alpha}$ et $x'_0 = m + h$, $y'_0 = e^{\eta_1/2} w + m - h$, $p'_1 = e^{\eta_1}$, $p'_2 = e^{\alpha + \frac{\eta_1 + \eta_2}{2}}$. η_1 et η_2 sont des mesures des pentes des asymptotes de la composée, α est une mesure de l'écart entre la pente centrale et à la pente moyenne des asymptotes. Si $w = 0$, alors $y'_0 - x_0 = 0$ et la composée d'angles n'a plus qu'un seul sommet : w est une mesure de l'écartement entre les deux sommets de la fonction composée. Enfin, m et h peuvent être vues comme des mesures de la position horizontale et verticale de la fonction composée.

les fonctions composées de n angles Notons simplement A_4 un angle à 4 paramètres, et A_2 un angle à deux paramètres (de type $A_{x_0, x_0, 1, p}$). Nous sommes intéressés par la forme que peut revêtir une composition du type $A_4 \circ A'_4 \cdots \circ A''''_4$. Nous avons vu précédemment que toute composée de type $A_4 \circ A'_4$ pouvait s'écrire sous la forme $A_2 \circ A_4$. Un argument de commutativité permet facilement d'étendre ce résultat : On a vu précédemment que

$$\left(\mathbb{A}'_{\eta_1, \eta_2, \alpha, m, h, w} \right)^{-1} = \mathbb{A}'_{-\eta_1, -\eta_2, \alpha, m, -h, -w},$$

et donc que $(\mathbb{A}')^{-1}$ est de la même forme que $\mathbb{A}' = A_2 \circ A_4$. Or, on a aussi $(\mathbb{A}'_{\eta_1, \eta_2, \alpha, m, h, w})^{-1}$ de la forme $A_4^{-1} \circ A_2^{-1}$, et si une bonne paramétrisation permet à A_2^{-1} de revêtir la même forme que A_2 , alors l'on pourra écrire que toute composée d'angle peut s'écrire sous la forme $A_2 \circ A_4$ ou $A'_4 \circ A'_2$. De par notre choix précédent pour A_2 , de la forme $A_{y, y, 1, p}$, on a bien $(A_{y, y, 1, p})^{-1} = A_{y, y, 1, p'}$, avec $p' = 1/p$, et A_2^{-1} appartient à la même classe que A_2 . Or, à partir de l'équation 17, on peut établir :

$$A_{x_0, y_0, p_1, p_2} \circ A_{x'_0, y'_0, p'_1, p'_2} = A_{x'_0 - \frac{y'_0 - x_0}{p'_1}, y_0, p_1 p'_1, p_2 p'_1} \circ A_{x' - 0, x'_0, 1, \frac{p'_2}{p'_1}}.$$

En utilisant ce résultat ou en calculant l'inverse de l'équation 19, on obtient lorsque $y_0 = x_0$ et $p_1 = 1$,

$$A_{x_0, x_0, 1, p_2} \circ A_{x'_0, y'_0, p'_1, p'_2} = A_{-\frac{1}{p'_1}(y'_0 - x_0) + x'_0, x_0, p'_1, p'_1 p_2} \circ A_{x'_0, x'_0, 1, \frac{p'_2}{p'_1}}. \quad (20)$$

Il s'ensuit que l'on peut établir le résultat suivant :

Lemme 1 Toute composition de deux angles peut s'écrire sous la forme de la composée de deux angles à 2 et 4 paramètres,

$$A_4 \circ A'_4 = A''_4 \circ A_2 = A'_2 \circ A'''_4,$$

où les A_2 désignent des angles à deux paramètres, du type $A_{x_0, x_0, 1, p}$, $x_0 \in \mathbb{R}$, $p > 0$, ayant leur sommet sur la diagonale d'équation $y = x$, et leur première pente égale à 1.

Il est aisé d'écrire toute composition de n angles à 4 paramètres sous la forme :

$$(A_4 \circ A'_4) \circ A''_4 \circ \dots = (A_2 \circ A'''_4) \circ A''_4 \circ \dots = A_2 \circ (A'''_4 \circ A''_4) \circ \dots = A_2 \circ (A'_2 \circ A''''_4) \circ \dots$$

Il est ainsi possible de se ramener à une composition d'angles à deux paramètres, à l'exception du dernier, à 4 paramètres. La propriété de commutativité précédemment évoquée permet ensuite de se ramener à une situation où l'angle à 4 paramètres peut occuper n'importe quelle position.

Théorème 1 Toute composée de n angles peut se ramener à une composition de un angle à 4 paramètres avec $n - 1$ angles à 2 paramètres du type $A_{x_0, x_0, 1, p}$, quelle que soit la position choisie pour l'angle à 4 paramètres. En particulier, toute composition d'angle peut s'écrire sous la forme

$$A'_4 \circ A''_4 \circ A'''_4 \circ \dots = A_4 \circ A_2 \circ A'_2 \circ \dots,$$

ou de façon plus générale

$$A_4^{\circ(n)} = A_2^{\circ(k)} \circ A_4 \circ A_2^{\circ(n-1-k)}, \quad 0 \leq k \leq n - 1,$$

où $A_p^{\circ(k)} = A'_p \circ A''_p \circ \dots \circ A_p^{\circ(k)}$ désigne la composée de k angles à p paramètres, $p \in \{2, 4\}$, avec $A_p^{\circ(0)} = \text{Id}$. Les A_2 désignent des angles à deux paramètres, du type $A_{x_0, x_0, 1, p}$, $x_0 \in \mathbb{R}$, $p > 0$, ayant leur sommet sur la diagonale d'équation $y = x$, et leur première pente égale à 1.

Les $2n + 2$ paramètres nécessaires à la description de la fonction composée se décomposent alors sous la forme $4 + 2(n - 1)$, et aucun paramètre n'est inutile. Le choix de la paramétrisation, qui jusqu'ici découlait d'une simple préférence de formulation, est ici important : si l'inverse de l'angle à deux paramètres n'appartenait pas à la même classe, comme pour $A_{0, x_0, 1, p_2}$ par exemple, on ne pourrait plus établir le résultat précédent sans imposer que A_4 soit en dernière

position, même si l'on peut par ailleurs montrer que $A_2 \circ A_4$ peut représenter toute composition de deux angles. Finalement, il s'avère que lors d'un procédé itératif, nous serons amenés à composer entre eux des angles à deux paramètres. Se pencher sur la forme de telles compositions facilite l'interprétation de compositions multiples. Ainsi une composée de type $A_2 \circ A'_2 \circ A''_2 \circ \dots$ peut s'écrire sous la forme :

$$A_{1,p_2}^0 \left(A_{1,p'_2}^0 \left(A_{1,p''_2}^0 (\dots) + x'_0 - x_0 \right) + x'_0 - x_0 \right) + x_0,$$

et les asymptotes en $-\infty$ et $+\infty$ ont pour équations respectives $y = x$ et $y = (p_2 p'_2 p''_2 \dots)x + (x_0 + p_2(x'_0 - x_0) + p_2 p'_2(x''_0 - x'_0) + \dots)$. Ainsi, lors d'un procédé itératif où les asymptotes ont déjà été calées, il sera possible de supprimer deux paramètres en fixant la contrainte d'une asymptote d'équation $y = 1x + 0$ en $+\infty$.

Fonctions Décalage (Shift) Les fonctions Décalage sont des fonctions dont la dérivée va croître puis décroître, afin par exemple de modifier implicitement les taux de hasard, en tâchant d'exploiter l'équation 12. On les définit précisément comme une version lissée de la composée de deux angles à deux paramètres, du type $A_{x_0, x_0, 1, p} \circ \bar{A}_{x'_0, x'_0, 1, p'}$. On impose de surcroît des directions asymptotiques unitaires en $+\infty$ et $-\infty$, et $p' = 1/p$. Finalement, les fonctions Décalage sont définies comme :

$$Z_{m, h, \rho, \epsilon} = H_{m-h, m-h, 1, e^\rho, \epsilon} \circ H_{m+h, m+h, 1, e^{-\rho}, \epsilon}.$$

Si l'on choisit par exemple $h = 0$, alors la fonction sera l'identité, et h peut représenter une mesure de l'écartement des asymptotes de cette fonction, l'asymptote en $+\infty$ ayant pour équation $y = x + h(1 - e^\rho)$. m est un paramètre renseignant sur la position centrale horizontale de la bifurcation, et ρ représente quant-à-lui la rapidité de la bifurcation. On remarque que :

$$(Z_{m, h, \rho, \epsilon})^{-1} = Z_{m, -h, \rho, \epsilon}$$

Ces fonctions agissent dans le même esprit qu'une transformée de Wang (voir Wang, 1996) qui serait opérée progressivement à partir d'une certaine abscisse. Elles peuvent en conséquence revêtir un intérêt particulier en assurance dommages, dans les situations où seule la queue de distribution doit être modifiée.

Fonctions Bosse (Bump) Ces fonctions peuvent s'obtenir comme une version lissée de la composition de trois angles de type A_2 , en imposant des asymptotes d'équation $y = x$ en $-\infty$ et $+\infty$, de sorte que l'ajustement soit exclusivement local, si l'on prend l'hypothèse que les asymptotes ont déjà été ajustées (ou le seront, en exploitant le théorème 1). De façon non lissée, ces bosses correspondent donc à :

$$\begin{aligned} B_{x_0, p_2, x'_0, p'_2}^0 &= A_{x_0, x_0, 1, p_2} \circ A_{x'_0, x'_0, 1, p'_2} \circ A_{x''_0, x''_0, 1, p_2}, \\ (B_{x_0, p_2, x'_0, p'_2}^0)^{-1} &= B_{x''_0, 1/p'_2, x_0, 1/p_2}^0, \\ \text{avec } p'_2 &= \frac{1}{p_2 p_2''}, \\ \text{et } x'_0 &= \frac{x''_0(p_2'' - 1) + x_0 p_2''(p_2 - 1)}{p_2 p_2'' - 1}. \end{aligned}$$

Si l'on choisit $m = \frac{x_0+x_0''}{2}$, $h = \frac{x_0-x_0''}{2}$, $\rho = \frac{\ln p_2 + \ln p_2''}{2}$ et $\gamma = \frac{\ln p_2 - \ln p_2''}{2}$, alors les versions lissées et non lissées de la fonction bosse peuvent s'écrire :

$$\begin{aligned}\bar{B}_{m,h,\rho,\gamma}^0 &= A_{m+h,m+h,1,e^{\rho+\gamma}} \circ A_{x_0',x_0',1,e^{-2\rho}} \circ A_{m-h,m-h,1,e^{\rho-\gamma}}, \\ B_{m,h,\rho,\gamma,\epsilon} &= H_{m+h,m+h,1,e^{\rho+\gamma,\epsilon}} \circ H_{x_0',x_0',1,e^{-2\rho,\epsilon}} \circ H_{m-h,m-h,1,e^{\rho-\gamma,\epsilon}}, \\ \text{avec } x_0' &= m + h \left(\frac{e^{2\rho} - 2e^{\rho-\gamma} + 1}{e^{2\rho} - 1} \right), \quad \rho \neq 0.\end{aligned}$$

On a alors lorsque $\rho \neq 0$:

$$B_{m,h,\rho,\gamma,\epsilon}^{-1} = B_{m,-h,-\rho,\gamma,\epsilon}.$$

Le cas $\rho = 0$ correspond à la situation dégénérée où la bosse est confondue avec l'identité (ce cas peut être géré en remplaçant par exemple γ par un paramètre d'asymétrie δ tel que $e^{-\gamma} = 1 - \frac{e^\rho - 1}{e^\rho + 1} \delta$). m désigne la position horizontale de la bosse, h sa hauteur (on obtient la fonction identité pour $h = 0$). Les pentes p_2 et p_2'' influant sur la pente des parties gauche et droite de la bosse, ρ s'interprète comme la rapidité du retour à l'asymptote, et γ comme un coefficient d'asymétrie de la bosse (dans le repère (O, \vec{i}, \vec{j}) , avec $\vec{i} = (1, 1)$ et $\vec{j} = (-1, 1)$). Le choix d'un coefficient d'asymétrie nul conduit à la version lissée suivante de la bosse :

$$B'_{m,h,\rho,\epsilon} = H_{m+h,m+h,1,e^{\rho,\epsilon}} \circ H_{x_0',x_0',1,e^{-2\rho,\epsilon}} \circ H_{m-h,m-h,1,e^{\rho,\epsilon}},$$

avec $x_0' = m + h \frac{1-e^\rho}{1+e^\rho}$, qui est défini même lorsque $\rho = 0$. On a alors :

$$(B_{m,h,\rho,\epsilon})^{-1} = B_{m,-h,-\rho,\epsilon}.$$

Fonctions composées générales d'hyperboles Enfin, il est des situations où l'on cherche à ajouter des paramètres jusqu'à l'obtention d'une forme paramétrique monotone (pour la fonction de survie) qui soit suffisamment fidèle aux données. Dans ces situations, en application du théorème 1, nous composerons des versions lissées d'un angle à 4 paramètres, composé avec des versions lissées de $n - 1$ angles à 2 paramètres. Le paramètre de lissage pourra être commun à toutes ses fonctions, si bien que nous écrirons par exemple :

$$G_{(x_0,y_0,p_1,p_2,e),(a_1,q_1),\dots,(a_n,q_n)} = H_{a_n,a_n,1,q_n,e} \circ \dots \circ H_{a_1,a_1,1,q_1,e} \circ H_{x_0,y_0,p_1,p_2,e}.$$

L'inverse de ces fonctions s'exprime très simplement :

$$(G_{(x_0,y_0,p_1,p_2,e),(a_1,q_1),\dots,(a_n,q_n)})^{-1} = H_{y_0,x_0,\frac{1}{p_1},\frac{1}{p_2},e} \circ H_{a_1,a_1,1,\frac{1}{q_1},e} \circ \dots \circ H_{a_n,a_n,1,\frac{1}{q_n},e}.$$

D'après le théorème 1 ces fonctions englobent la classe des fonctions continues croissantes et affines par morceaux à $n + 1$ points de ruptures, et l'on pourrait choisir n'importe quelle position pour l'hyperbole à 5 paramètres. Nous verrons au paragraphe 4.4 que des valeurs d'initialisation peuvent être aisément obtenues pour chacun des paramètres. Si l'interprétation des paramètres devient plus délicate, ces fonctions seront bien adaptées à la détermination d'une forme paramétrique monotone et inversible sur un jeu de données. Ce type de situation se rencontrera notamment lors de simulations, où le besoin sera de simuler simplement des variables aléatoires de répartition suffisamment proche d'une répartition empirique.

4 Estimation et convergence de l'ajustement itératif

4.1 Méthodologie d'estimation

On souhaite transformer une fonction de survie $S_0 \in \mathcal{S}$ de telle sorte qu'elle s'approche d'une fonction de survie $S \in \mathcal{S}$. La distance que nous utilisons ici sur \mathcal{S} est la distance D définie par

$$D(S, S') = \int_0^{+\infty} |S(t) - S'(t)| dt.$$

Cette distance est bien définie et finie pour tout couple de \mathcal{S}^2 grâce à l'hypothèse de finitude de l'espérance des variables considérés.

Remarque 2 Soient X et X' deux variables aléatoires positives de fonctions de survie respectives S et S' . On a

$$|E X' - E X| \leq D(S, S').$$

La distance D permet donc de contrôler la différence entre espérances de variables aléatoires.

En se restreignant à la famille de transformations $(\mathbb{T}_{f_\theta})_\theta$ obtenue à partir de la famille de fonction $(f_\theta)_\theta$ paramétrisée par le vecteur θ , on obtient :

$$S_1 = \mathbb{T}_{f_{\theta_0^*}}(S_0),$$

où

$$\theta_0^* = \arg \min_{\theta} D(\mathbb{T}_{f_\theta}(S_0), S).$$

Si la fonction S_1 ainsi définie n'est pas suffisamment proche de S , on itère le procédé de façon à définir une deuxième approximation S_2 par

$$S_2 = \mathbb{T}_{f_{\theta_{2,1}^*}}(\mathbb{T}_{f_{\theta_{2,0}^*}}(S_0)),$$

où

$$(\theta_{2,0}^*, \theta_{2,1}^*) = \arg \min_{\theta_0, \theta_1} D(\mathbb{T}_{f_{\theta_1}}(\mathbb{T}_{f_{\theta_0}}(S_0)), S).$$

En continuant de cette manière, on définit une suite $(S_n)_n$ de fonctions de survie.

4.2 Forme des itérées

Nous avons vu que toutes les fonctions d'altération considérées pouvaient être vues comme des versions lissées de composées d'angles. Nous nous intéresserons donc ici en premier lieu à la forme des itérées pour des compositions de fonctions « angle ». Considérons la fonction de survie $S_0^* \in \mathcal{S}$, définie par

$$S_0^*(x) = \begin{cases} 1 & \text{si } x \leq 0, \\ \text{logit}^{-1}(1 - x) & \text{si } x > 0. \end{cases}$$

Nous étudions ici la forme que prennent les transformées itérées de $S_0 = S_0^*$ par T , dans le cas de l'utilisation de la famille $(A_\theta)_\theta$ des fonctions « angle ». Pour toute suite $\Theta = (\theta_n)_n = ((x_n, y_n, \alpha_n, \beta_n))_n$ de \mathbb{R}^4 et $n \in \mathbb{N}^*$, notons $S_n(\theta_0, \dots, \theta_{n-1})$ la fonction de survie définie par récurrence par

$$S_n = \mathbb{T}_{A_{\theta_{n-1}}}(S_{n-1}),$$

et ℓ_n la fonction définie sur \mathbb{R}_+^* par

$$\ell_n(\theta_0, \dots, \theta_{n-1}) = \text{logit}(S_n(\theta_0, \dots, \theta_{n-1})),$$

de sorte que, pour tout $x > 0$,

$$\ell_n(\theta_0, \dots, \theta_{n-1})(x) = A_{\theta_{n-1}} \circ \dots \circ A_{\theta_0}(1 - x).$$

On remarque que ℓ_n est une fonction continue, affine par morceaux, dont les points de rupture de pente sont les x_i , $i = 0, \dots, n-1$. Notons \mathcal{A}_n l'ensemble des fonctions de \mathbb{R}_+^* dans \mathbb{R} continues strictement croissantes et affines par morceaux ayant au plus n points de rupture de pente, et $\mathcal{A} = \cup_n \mathcal{A}_n$. On vient de voir que $\ell_n \in \mathcal{A}_n$. Considérons maintenant un élément f de \mathcal{A}_n . Notons $0 = x_{-1} < x_0 < \dots < x_{n-1}$ ses points de rupture de pente, et, pour $i = -1, \dots, n-1$, $-\alpha_i$ la pente de f sur l'intervalle $]x_i, x_{i+1}[$ (avec la convention $x_n = +\infty$), et $y_i = f(x_i)$. Posons maintenant

$$\theta_0 = (x_0, y_0, \alpha_{-1}, \alpha_0),$$

et, pour $i = 1, \dots, n-1$,

$$\theta_i = (x_i, x_i, 1, \alpha_i/\alpha_{i-1}).$$

On vérifie aisément qu'avec cette valeur des paramètres θ_i , ℓ_i coïncide avec f sur $]1, x_i[$ pour tout i , et en particulier

$$\ell_n(\theta_0, \dots, \theta_{n-1}) = f.$$

Ainsi, finalement, toute fonction de \mathcal{A}_n peut s'écrire comme une transformée itérée n fois de $\text{logit}(S_0)$ par des fonctions « angle », c'est-à-dire que

$$\{\ell_n(\theta_0, \dots, \theta_{n-1}), \theta_0, \dots, \theta_{n-1} \in \mathbb{R}^4\} = \mathcal{A}_n \quad \forall n \in \mathbb{N}^*. \quad (21)$$

4.3 Convergence

Nous allons ici nous attacher à vérifier que les principales fonctions d'altération proposées respectent la contrainte C4 que nous avons posé. Pour $S \in \mathcal{S}$, on notera

$$a^*(S) = \sup\{x, S(x) = 1\} \in \mathbb{R}_+^*, \quad b^*(S) = \inf\{x, S(x) = 0\} \in \mathbb{R}_+^* \cup \{+\infty\},$$

$$M(S) =]a^*(S), b^*(S)[.$$

Théorème 2 Soient $S \in \mathcal{S}$ et $S_0 \in \mathcal{S}$ telles que

1. $a^*(S_0) \leq a^*(S) \leq b^*(S) \leq b^*(S_0)$;
2. S_0 est strictement décroissante sur $]a^*(S_0), b^*(S_0)[$.

Alors, pour les familles $(A_\theta)_\theta$ et $(H_\theta)_\theta$, ainsi que pour les familles construites par composition à partir de celles-ci, comme $(G_\theta)_\theta$,

$$\lim_{n \rightarrow \infty} D(S_n, S) = 0.$$

Remarque 3 La condition (1) est tout à fait naturelle. En effet, si x est tel que $a^*(S) < x < a^*(S_0)$, $S_n(x) = 1$ pour tout n alors que $S(x) < 1$. Donc, si $a^*(S) < a^*(S_0)$,

$$\int_{a^*(S)}^{a^*(S_0)} |S_n(t) - S(t)| dt = \delta > 0,$$

ce qui empêche la convergence de $D(S_n, S)$ vers zéro. Il en est de même pour la condition sur les bornes b^* .

Pour montrer ce résultat, nous utiliserons le lemme suivant d'approximation de fonctions de survie :

Lemme 2 Soit $S \in \mathcal{S}$, $0 \leq a < b < +\infty$, et $\varepsilon > 0$. Alors il existe une suite finie $(t_i)_{0 \leq i \leq n}$ de $[a, b]$ telle que toute fonction de survie $S' \in \mathcal{S}$ qui coïncide avec S sur les t_i vérifie

$$\int_a^b |S'(t) - S(t)| dt < \varepsilon.$$

Preuve du lemme 2:

Soit N un entier vérifiant $(2 + b - a)/N < \varepsilon$. Construisons la suite $(t_i)_i$, ainsi qu'un sous-ensemble J de \mathbb{N} par récurrence de la manière suivante. On pose tout d'abord $t_0 = b$ et $J = \emptyset$, puis :

– si $S(t_i^-) \geq S(t_i) + 1/N$, alors on ajoute i à J , et on pose

$$t_{i+1} = \max(t_i - 1/N^2, a);$$

– sinon, on pose

$$t_{i+1} = \max(\inf\{t, S(t) \leq S(t_i) + 1/N\}, a),$$

de sorte que (du moins si $t_{i+1} > a$)

$$S(t_{i+1}) \leq S(t_i) + 1/N \quad \text{et} \quad S(t_{i+1}^-) \geq S(t_i) + 1/N.$$

On arrête la construction dès qu'un t_i atteint a : on note n cet indice final. La suite $(t_i)_i$ ainsi définie est strictement décroissante et pour tout $i < n - 2$, $S(t_{i+2}) \geq S(t_i) + 1/N$. Cette suite est donc bien finie, de taille au plus $2N$. Considérons maintenant une fonction $S' \in \mathcal{S}$ telle que, pour tout $i \in \{0, \dots, n\}$, $S'(t_i) = S(t_i)$. Comme de plus S' et S sont décroissantes,

$$\begin{aligned} \int_a^b |S'(t) - S(t)| dt &= \sum_{i=1}^n \int_{t_i}^{t_{i-1}} |S'(t) - S(t)| dt \\ &\leq \sum_{i=1}^n \int_{t_i}^{t_{i-1}} (S(t_i) - S(t_{i-1})) dt \\ &\leq \sum_{i \in J} \frac{1}{N^2} + \sum_{i \notin J} \frac{t_{i-1} - t_i}{N} \\ &\leq \frac{2}{N} + \frac{b-a}{N} < \varepsilon. \end{aligned}$$

□

Nous pouvons maintenant démontrer le théorème de convergence :

Preuve du théorème 2:

Commençons par le cas où la famille $(f_\theta)_\theta$ utilisée est la famille des fonctions « angle ». Fixons une valeur de $\varepsilon > 0$. Comme $S \in \mathcal{S}$, $\int_0^{+\infty} S(t) dt$ converge, et il existe donc un réel $b < b^*(S)$ tel que

$$S(b) > 0 \quad \text{et} \quad \int_b^{+\infty} S(t) dt < \varepsilon.$$

Posons par ailleurs $a = a^*(S) + \varepsilon$, de sorte que $S(a) < 1$. Soit $(t_i)_{0 \leq i \leq n}$ la suite définie par le lemme 2 pour S sur l'intervalle $[a, b]$. Soit f l'application croissante affine par morceaux définie sur $I_0 = [\text{logit}(S_0(b)), \text{logit}(S_0(a))]$ par ses points de rupture de pente :

$$f(\text{logit}(S_0(t_i))) = \text{logit}(S(t_i)) \quad \forall i \in \{0, \dots, n\}.$$

Comme $\mathbb{T}_f(S_0)$ coïncide avec S sur les t_i ,

$$\int_a^b |\mathbb{T}_f(S_0)(t) - S(t)| dt < \varepsilon.$$

La fonction f est croissante par construction. Assurons-nous qu'elle est strictement croissante en définissant la fonction affine par morceaux strictement croissante \tilde{f} par

$$\tilde{f}(t) = f(t) + \varepsilon \frac{t - \text{logit}(S_0(b))}{(b-a)(\text{logit}(S_0(a)) - \text{logit}(S_0(b)))} \quad \forall t \in I_0,$$

de telle sorte que

$$|\tilde{f}(t) - f(t)| < \frac{\varepsilon}{b-a}.$$

Comme la fonction

$$t \mapsto \text{logit}^{-1}(t) = \frac{1}{1 + e^{-t}}$$

a une dérivée inférieure à un, on obtient la majoration

$$|\mathbb{T}_{\tilde{f}}(S_0)(t) - \mathbb{T}_f(S_0)(t)| < \frac{\varepsilon}{b-a} \quad \forall t \in [a, b],$$

et ainsi

$$\int_a^b |\mathbb{T}_{\tilde{f}}(S_0)(t) - S(t)| dt < 2\varepsilon. \quad (22)$$

Soit b_0 un réel supérieur à $b + \varepsilon$ pour lequel

$$\int_{b_0}^{+\infty} S_0(t) dt < \varepsilon.$$

Prolongeons maintenant continûment \tilde{f} sur \mathbb{R} en la définissant sur $\mathbb{R} \setminus I_0$ de la manière suivante :

1. \tilde{f} est affine sur $[\text{logit}(S_0(a)), \text{logit}(S_0(a^*(S)))]$, et

$$\tilde{f}(\text{logit}(S_0(a^*(S))) + u) = u + \max \left(\text{logit} \left(\frac{1 - \varepsilon}{a^*(S)} \right), \tilde{f}(\text{logit}(S_0(a))) + 1 \right) \quad \forall u \geq 0.$$

On obtient ainsi

$$\mathbb{T}_{\tilde{f}}(S_0)(t) \geq \frac{1 - \varepsilon}{a^*(S)} \quad \forall t \geq a^*(S),$$

de sorte que

$$\int_0^{a^*(S)} (1 - \mathbb{T}_{\tilde{f}}(S_0)(t)) dt \leq \varepsilon. \quad (23)$$

2. \tilde{f} est affine sur $[\text{logit}(S_0(b + \varepsilon)), \text{logit}(S_0(b))]$, et, pour $t \leq \text{logit}(S_0(b + \varepsilon))$,

$$\tilde{f}(t) = t + \min \left(0, \text{logit} \frac{\varepsilon}{b' - (b + \varepsilon)} - \text{logit}(S_0(b + \varepsilon)), \tilde{f}(\text{logit}(S_0(b))) - 1 \right).$$

On obtient ainsi $\tilde{f}(t) \leq t$ pour $t \leq \text{logit}(S_0(b + \varepsilon))$, de sorte que

$$\mathbb{T}_{\tilde{f}}(S_0)(t) \leq S_0(t) \quad \forall t > b + \varepsilon,$$

et ainsi

$$\int_{b_0}^{+\infty} \mathbb{T}_{\tilde{f}}(S_0)(t) dt \leq \int_{b_0}^{+\infty} S_0(t) dt < \varepsilon. \quad (24)$$

Par ailleurs,

$$\tilde{f}(t) \leq \text{logit} \frac{\varepsilon}{b' - (b + \varepsilon)} \quad \forall t \leq \text{logit}(S_0(b + \varepsilon)),$$

de sorte que

$$\mathbb{T}_{\tilde{f}}(S_0)(t) \leq \frac{\varepsilon}{b' - (b + \varepsilon)} \quad \forall t > b + \varepsilon,$$

et ainsi

$$\int_{b+\varepsilon}^{b_0} \mathbb{T}_{\tilde{f}}(S_0)(t) dt < \varepsilon. \quad (25)$$

Enfin, comme $a = a^*(S) + \varepsilon$ et $\mathbb{T}_{\tilde{f}}(S_0)(t)$ et $S(t)$ appartiennent à $[0, 1]$,

$$\int_{a^*(S)}^a |\mathbb{T}_{\tilde{f}}(S_0)(t) - S(t)| dt < \varepsilon \quad \text{et} \quad \int_b^{b+\varepsilon} |\mathbb{T}_{\tilde{f}}(S_0)(t) - S(t)| dt < \varepsilon. \quad (26)$$

En regroupant les inégalités (22), (23), (24), (25) et (26), on voit que la fonction $\tilde{f} \in \mathcal{A}$ permet de transformer S_0 en une bonne approximation de S :

$$D(\mathbb{T}_{\tilde{f}}(S_0), S) < 7\varepsilon.$$

Ainsi, si N est un entier tel que $\tilde{f} \in \mathcal{A}_N$, la relation (21) implique que

$$D(S_n, S) < 7\varepsilon \quad \forall n \geq N.$$

On a ainsi montré que

$$\lim_{n \rightarrow \infty} D(S_n, S) = 0$$

dans le cas des fonctions « angle ». Le cas des fonctions hyperboliques H_θ s'en déduit immédiatement car les fonctions A_θ sont des cas particuliers des fonctions H_θ , il en va de même pour les composées d'hyperboles et pour toutes les classes de fonctions d'altération englobant les angles ou les hyperboles. \square

4.4 Valeurs d'initialisation

Nous avons proposé dans les sections précédentes d'utiliser comme fonctions d'altération soit des composées de fonctions angles, soit des versions lissées de celles-ci, à savoir des composées d'hyperboles. Lors de l'estimation des paramètres, il est nécessaire de disposer d'une valeur initiale pour l'estimation du vecteur paramètre. Il est envisageable de procéder de plusieurs façons :

- Le vecteur paramètre initial peut correspondre à une fonction d'altération identité, si la fonction de survie source n'est pas trop éloignée de la fonction de survie cible. Ce choix peut néanmoins conduire de nombreux algorithmes d'optimisation à une solution éloignée de la solution optimale.
- Pour les composées de plusieurs fonctions, il est plus aisé d'estimer séparément les paramètres optimaux pour chacune des déformations, ce qui peut conduire à des valeurs initiales pour les paramètres agrégés de la fonction composée. Ce choix se heurte néanmoins aux situations où deux effets antagonistes des composées se compensent, par exemple lorsqu'une première fonction d'altération cause un éloignement de la cible pour permettre à une deuxième fonction d'altération de mieux s'en rapprocher ensuite.
- Enfin, l'estimation simultanée de tous les paramètres d'une fonction composée est la solution susceptible de fournir les meilleurs résultats, pour peu que le vecteur paramètre initial ne soit pas trop éloigné du vecteur optimal. Nous nous pencherons sur ce dernier cas en essayant de proposer des valeurs d'initialisation conduisant à une première approximation correcte de la fonction cible.

Les fonctions d'altération considérées sont toutes des composées d'hyperboles, pour lesquelles le paramètre de lissage n'est vraisemblablement pas le plus délicat à estimer. Nous nous focaliserons donc sur l'estimation de composées d'angles à deux et à quatre paramètres. En pratique, nous disposons d'un ensemble fini d'abscisses, $\{x_i\}_{i \in I} \in I$, avec $I = \{1..p\}$, pour lesquels sont données les valeurs de la fonction de survie cible et son logit $l_i = \text{logit } S(x_i)$, ainsi que le logit de la fonction de survie en cours à déformer $\alpha_i = \text{logit } \widehat{S}(x_i)$, pour $i \in I$. Ce nuage de point est un ensemble fini d'une courbe que nous noterons $l(\alpha)$. Nous cherchons f_θ tel que les points $\{(l_i, f_\theta(\alpha_i))\}_{i \in I}$ soient aussi proches que possible de la diagonale Δ d'équation $y = x$. Une possibilité est de rechercher une application \widehat{f}_θ qui associerait l_i à certains des α_i , pour un sous ensemble de points $i \in I_k$, $I_k \subset I$. Pour ces points, nous serions alors assurés d'avoir $\{(l_i, f_\theta(\alpha_i))\}_{i \in I_k}$ inclus dans Δ . Peut-on trouver une composée d'angles proche d'une fonction passant par un sous-ensemble $\{(l_i, \alpha_i)\}_{i \in I_k}$ de points, $I_k \subset I$? Nous allons voir que cela est relativement simple, grâce à la proposition suivante.

Proposition 6 Soit un ensemble de points successifs $\{(u_i, v_i)\}_{i \in \{1, \dots, 3+k\}}$ d'une courbe croissante, avec $u_1 \leq \dots \leq u_{3+k}$ et $v_1 \leq \dots \leq v_{3+k}$, $k \geq 0$. La composée d'angle de type

$$\begin{aligned} G_\theta^{(k)} &= A_{a_k, a_k, 1, q_k} \circ \dots \circ A_{a_1, a_1, 1, q_1} \circ A_{x_0, y_0, p_1, p_2}, \text{ si } k \geq 1, \\ G_\theta^{(k)} &= A_{x_0, y_0, p_1, p_2}, \text{ si } k = 0, \end{aligned}$$

est telle que :

$$G_\theta^{(k)}(u_i) = v_i, \text{ pour tout } i \in \{1, \dots, 3+k\},$$

dès lors que :

$$x_0 = u_2, y_0 = v_2, p_1 = (v_2 - v_1)/(u_2 - u_1), p_2 = (v_3 - v_2)/(u_3 - u_2),$$

et que :

$$a_k = v_{2+k} \text{ et } q_k = \left(\frac{v_{3+k} - v_{2+k}}{u_{3+k} - u_{2+k}} \right) \left(\frac{u_{2+k} - u_{1+k}}{v_{2+k} - v_{1+k}} \right), \quad k \geq 1.$$

Preuve :

On vérifie facilement que $A_{x_0, y_0, p_1, p_2}(u_1) = v_1$, $A_{x_0, y_0, p_1, p_2}(u_2) = v_2$ et $A_{x_0, y_0, p_1, p_2}(u_3) = v_3$. On vérifie ensuite par récurrence que pour tout $i \leq k$, $G^{(k)}(u_i) = v_i$ puis que $G^{(k)}(u_{4+k}) =$

$$v_{2+k} \frac{v_{3+k} - v_{2+k}}{u_{3+k} - u_{2+k}} (u_{4+k} - u_{2+k}) \text{ et } G^{(k+1)}(u_{3+k+1}) = G^{(k+1)}(u_{4+k}) = v_{4+k}. \quad \square$$

Nous utiliserons en conséquence le procédé suivant pour l'initialisation du vecteur paramètre : un sous-ensemble (u_i, v_i) de points sera extrait de l'ensemble (l_i, α_i) , en faisant en sorte de choisir des points aussi éloignés que possible les uns des autres : nous prendrons par exemple le point d'abscisse minimum, le point d'abscisse maximum, puis $k + 1$ points intermédiaires régulièrement espacés (soit $k + 2$ intervalles intermédiaires) :

$$\begin{aligned} (u_j, v_j) &= (\alpha_{s(j)}, l_{s(j)}), \\ \text{avec } s(j) &= \inf_{i \leq p} \left\{ L(i) \geq \frac{L(n)}{k+2}(j-1) \right\}, \quad j = 1, \dots, k+3, \\ \text{et } L(i) &= |l_i - l_1|. \end{aligned}$$

Nous en déduisons par la proposition précédente des valeurs d'initialisation pour toutes les combinaisons précédemment présentées de type $A_2 \circ \dots \circ A_2 \circ A_4$. En fixant la valeur initiale du paramètre de lissage ϵ à une valeur faible non nulle, par exemple dans nos applications $\epsilon = 1$, nous obtiendrons des valeurs d'initialisation pour les fonctions d'altération du type $H_2 \circ \dots \circ H_2 \circ H_4$. Le choix d'une valeur d'initialisation non nulle pour ϵ s'explique par des conditions de continue différentiabilité qui facilitent la convergence des principaux algorithmes d'optimisation. Il est également possible d'obtenir très simplement des valeurs d'initialisation pour les paramètres des fonctions de type $H_4 \circ H_2 \circ \dots \circ H_2$: il suffit pour cela d'ajuster la fonction inverse de $l(\alpha)$ en intervertissant les rôles des l_i et des α_i . L'inverse de l'ajustement obtenu, de type $A_2 \circ \dots \circ A_2 \circ A_4$, sera $A_4^{-1} \circ A_2^{-1} \circ \dots \circ A_2^{-1}$ dont les paramètres se déduisent immédiatement, et qui est bien de même nature que $A'_4 \circ A'_2 \circ \dots \circ A'_2$. Nous aurions également pu choisir pour L_i d'autres fonctions croissantes, comme par exemple $|\widehat{S}(x_i) - \widehat{S}(x_1)|$, ou encore comme la part cumulée de l'observation numéro i dans l'erreur globale, $\sum_{k=1}^i |\widehat{S}(x_i) - S(x_i)|$, qui induisent des répartitions différentes des points de contacts initiaux entre \widehat{S} et S . A titre d'illustration, dans le cas d'ajustement local (bosses liées à un évènement catastrophique), nous avons parfois utilisé $\sum_{k=1}^i e^{-|\widehat{S}(x_i) - S(x_i)|}$ lorsque cette répartition donnait de meilleurs résultats.

5 Applications numériques

Il est possible de proposer des applications des transformations itérées de fonctions de survie à de nombreux domaines. Nous avons choisi ici de présenter des applications à l'analyse des durées de survie. Nous nous baserons dans cette section sur des tables de mortalité issues du site Internet Human Mortality Database (2008), tables données par année de décès, pour les pays Etats-Unis et France (tables par tranche d'âge d'une année et tranche d'année de décès de une année, désignées par l'expression 1×1). Nous nommerons ces deux tables respectivement USA et FRANCE. En l'absence de précision concernant le sexe, les tables désignent celles disponibles pour la totalité de la population, hommes et femmes. Dans le cas discret, il est indispensable de se doter d'une mesure de la distance entre la fonction cible et son ajustement qui soit adaptée au problème. Nous disposons ici d'un ensemble de points correspondant aux valeurs des fonctions de survie à l'étape n , $s_i^n = S^n(x_i)$ et de l'ensemble des valeurs de la fonction de survie cible $s_i = S(x_i)$, pour les différentes abscisses x_i , $i \in \{1, \dots, p\}$. Nous utiliserons comme mesure de la qualité de l'ajustement à l'étape numéro n :

$$I_Q^n = -\log_{10} \left(\frac{1}{p} \sum_{i=1}^p |s_i^n - s_i| \right).$$

	France H_5	France $H_2 \circ H_5$	France Gompertz	USA H_5	USA $H_2 \circ H_5$	USA Gompertz
x_0	-94.09334665	-95.78211241		-114.2086702	-102.6014363	
y_0	0.5117481008	-1.504117646		1.3712451764	0.7420644703	
p_1	0.4643066083	0.439815211		2.4981606661	0.727987894	
p_2	0.06252572	0.2108655161		0.0424709347	0.0666009078	
ϵ	5.7654445858	0.9843955044		26.536409759	10.800577777	
a_1	-	0.630550249		-	6.5281238533	
q_1	-	0.324782732		-	0.00078401	
a	-	-	0.082599221	-	-	0.0766073843
b	-	-	-9.595051243	-	-	-9.047939578
I_Q	2.824361353	2.9225066156	1.7229331698	3.0056691098	3.2141510129	1.8248654389

Tab. 1 – ajustements de tables de mortalité pour l’année de décès 2005, par une loi Gompertz ou par déformations hyperboliques itérées d’exponentielle de paramètre 1.

Cet indicateur donne une idée directe du nombre de décimales correctes de l’ajustement utilisé.

5.1 Ajustement de tables de mortalité

Nous chercherons tout d’abord à déterminer la capacité des déformations à servir d’ajustement initial de mortalité, pour les tables USA et FRANCE, pour l’année de décès 2005, pour tous les âges de 0 à 110 ans, et pour les fonctions d’altération de type Hyperbole et Hyperboles composées. Les déformations sont appliquées à un premier ajustement non paramétré par la fonction de survie $S_0(x) = \exp(-x)$, évidemment très éloigné de la mortalité humaine, avec une espérance de vie à la naissance de une année et une absence de vieillissement. En pratique, d’autres choix pour $L(i)$ ou $S_0(x)$ peuvent naturellement conduire à de meilleures qualité d’ajustement. Le propos ici étant de montrer la faisabilité pratique d’un ajustement sans information a priori sur la fonction de survie, nous ne tenterons pas d’améliorer ces paramètres. Nous noterons cependant que la modification de S_0 peut servir la construction de classes beaucoup plus vastes de fonctions de survie, sans modifier en rien la complexité de la procédure. Ces ajustements seront comparés à ceux fournis par un ajustement classique en assurance vie de type Gompertz à deux paramètres (ajustement que l’on peut facilement inverser), avec pour taux instantané de mortalité $\mu_x = e^{ax+b}$, et $S(x) = e^{(e^b - \mu_x)/a}$. L’estimation est opérée pour ce dernier ajustement en exploitant la linéarité de $\ln(-\ln(1 - q_x)) = ax + (b + \ln(e^a - 1) - \ln(a))$. Les ajustements présentés dans le tableau 1 seront des ajustements de type : $H_5 = H_{x_0, y_0, p_1, p_2, \epsilon}$ ou $H_2 \circ H_5 = H_{a_1, a_1, 1, q_1, \epsilon} \circ H_{x_0, y_0, p_1, p_2, \epsilon}$.

Au final, nous pouvons observer qu’il est aisé d’obtenir avec 5 ou 7 paramètres des ajustements conduisant à des écarts de l’ordre de 10^{-3} sur la fonction de survie, ce qui, lors d’une inversion, pour un âge de 65 ans par exemple, représente des différences d’âge de l’ordre du mois. L’ajustement de la table USA à l’aide d’une fonction d’altération de type $H_2 \circ H_5$ est illustré dans la figure 1. A titre indicatif, l’indice de qualité relatif à l’ajustement de la table française par la table américaine non transformée (où l’inverse) est de 1.6240037397. Il est parfois possible d’améliorer considérablement l’indice de qualité en choisissant une tranche d’âge d’analyse plus courte, comme cela apparaît dans la table 2. Ainsi, pour la table française, sur la tranche 20-60 ans, nous obtenons un indice de qualité légèrement supérieur à 4 pour une modélisation de type $H_2 \circ H_5$, ce qui représente une erreur absolue moyenne de 10^{-4} entre les fonctions de survie cible et ajustées, avec l’utilisation d’une fonction de survie initiale exponentielle pourtant très éloignée de la réalité. Pour la table américaine, au prix de deux paramètres supplémentaires, nous obtenons un indice de qualité proche de 4.5, ce qui représente une erreur absolue moyenne de l’ordre de $3 \cdot 10^{-5}$.

tranche 20-60 ans				
	France	France	USA	USA
	H_5	$H_2 \circ H_5$	H_5	$H_2 \circ H_2 \circ H_5$
x_0	-41.44818439	-51.9131699	-44.2289858	-47.36803472
y_0	3.6727504339	2.8303184115	3.1120792017	2.8992555156
p_1	0.0793804677	0.0761580284	0.0712430456	0.0719719785
p_2	0.0556743609	0.0919693101	0.0479084682	0.0635147139
ϵ	0.056367023	0.105885775	0.2017659419	0.0818666745
a_1	-	3.5826627318	-	3.3633496933
q_1	-	0.5887705559	-	0.6106991574
a_2	-	-	-	3.9627332679
q_2	-	-	-	2.0129504234
a	-	-	-	-
b	-	-	-	-
I_Q	3.6676370336	4.0456978917	3.6659281404	4.4486252598

Tab. 2 – Ajustements de tables de mortalité pour l'année de décès 2005, sur une tranche d'âge réduite de 20 à 60 ans.

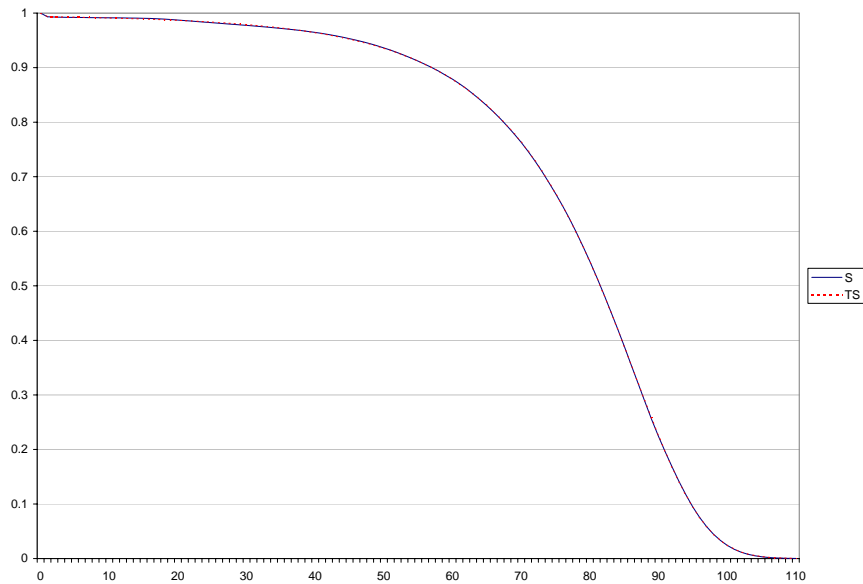


Fig. 1 – Ajustement de la fonction de survie USA pour l'année 2005 par déformation de type $H_2 \circ H_5$ d'une exponentielle de paramètre 1 ($I_Q \simeq 3.2$, courbes pratiquement confondues).

	Bosse sym.	Bosse asym.	H_5	$H_2 \circ H_5$	$H_2 \circ H_2 \circ H_5$
m	-0.131679921	-1.564937408			
h	1.4611733816	-2.67357656			
ρ	-1.609080286	0.8600668251			
γ		-1.187224754			
ϵ	0.405428658	0.2689129359	0.025933307	0.203347333	0.4931553975
x_0			0.9697545225	1.100854581	0.0444691012
y_0			-0.477215771	-0.410166997	-1.019698507
p_1			0.7611150965	0.7457718552	0.9095122669
p_2			3.4347630872	5.4453018072	0.004080047
a_1				1.4366535801	-1.025350693
q_1				0.1736984168	1882.6778835
a_2					1.4179376797
q_2					0.0672308606
I_Q	1.6763399649	2.2496987566	1.9712125335	2.2204988662	2.6610658769

Tab. 3 – Ajustement de la fonction de survie pour l’année 1915 par déformations de la fonction de survie de 1913.

5.2 Modélisation d’un événement catastrophique

Pour cette application nous allons déformer la table française pour l’année de décès 1913, afin de s’approcher de la table française pour l’année de décès 1915, dans les premières années de la première guerre mondiale. La déformation obtenue vise à illustrer la capacité des fonctions d’altérations proposées à modéliser un changement catastrophique sur une table, même lorsque celui-ci affecte de façon très différente les probabilité de survie aux différents âges, et touche ici en particulier les jeunes adultes. Bien qu’il puisse être pertinent d’analyser les différences de déformation selon le sexe, nous avons utilisé ici les mêmes tables que précédemment, tous sexes confondus. Les fonctions de survie utilisées correspondent ici à un produit de probabilités annuelles de survie, comme si la mortalité était continuellement conforme à celle observée au cours de l’année considérée, 1913 ou 1915. Le modèle appliqué ici est le suivant :

$$S^{1915} = \mathbb{T}_f S^{1913} .$$

Il est délicat d’ajuster une déformation locale, et les résultats ne conduisent pas à une aussi bonne précision que lors de l’ajustement d’une table entière. Avec 4 paramètres, nous trouvons un indice de qualité proche de 2 en utilisant une bosse symétrique, ou un peu meilleur avec davantage de paramètres.

Nous proposons ici des ajustements qui minimisent l’écart entre les fonctions de répartitions cible et fonctions de répartition ajustées. Il faut garder à l’esprit que ce choix peut conduire à des différences sur les probabilités annuelles de décès, par exemple. Selon l’exploitation future de la déformation, il pourra être préférable d’utiliser d’autres critères d’optimisations, que nous ne détaillons pas ici.

La tranche d’âge utilisée est de 0 à 104 ans (prendre une tranche plus étendue améliorerait artificiellement l’indice de qualité), et les résultats sont fournis dans la table 3. Pour les ajustements de moins de cinq paramètres, il semble ici préférable d’utiliser les fonctions bosses, qui exploitent le fait que la table est peu modifiée aux âges avancés. La fonction bosse asymétrique permet d’améliorer de façon importante la qualité de l’ajustement, avec l’ajout d’un unique paramètre. Ce dernier ajustement est illustré dans la figure 2. L’indice de qualité pour cette dernière fonction à 5 paramètres est meilleur que celui avec 7 paramètres obtenu par composition d’hyperboles, et la connaissance de propriétés particulières de la fonction d’altération nous a donc ici permis de gagner deux paramètres. La tentative d’amélioration de la précision conduit à l’utilisation de composées d’hyperboles, qui ne permettent toutefois pas de représenter la table de 1915 aussi fidèlement que celle de 2005, du fait de son caractère plus erratique.

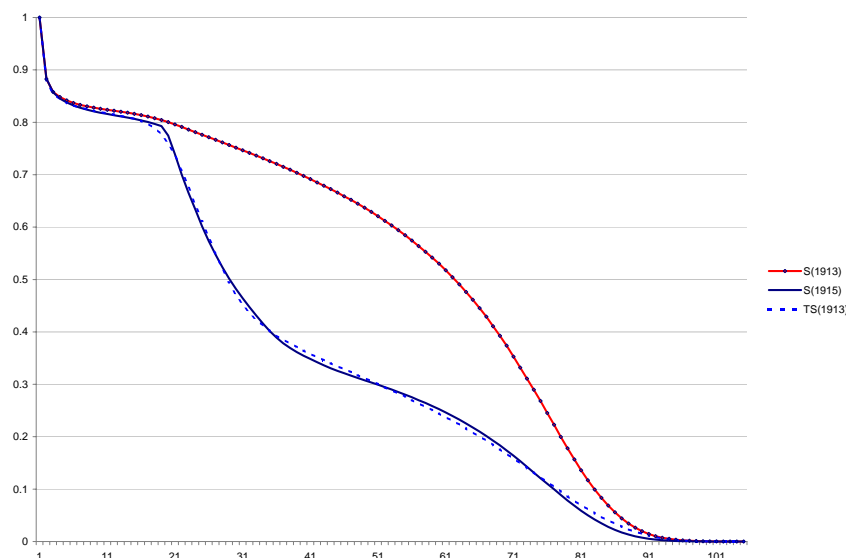


Fig. 2 – Ajustement de la fonction de survie pour l’année 1915 par déformation de la fonction de survie de 1913 par bosse asymétrique ($T_f S^{1913}$ en pointillés, $I_Q \simeq 2.25$).

5.3 Modélisation prospective, déformation statique

Nous tâcherons ici de fournir un ajustement paramétrique de la fonction de survie par âge et par année de décès, afin de tenir compte de l’évolution de l’espérance de vie au fil des ans. Il est d’autant plus délicat d’obtenir un ajustement de bonne qualité sur de longues périodes, l’évolution de la mortalité au fil des ans se résumant difficilement à deux ou trois paramètres. Nous aborderons ici deux modélisations : l’une exploitant une unique déformation de fonctions variant au fil des ans, détaillée ici, l’autre exploitant une déformation par an d’une unique distribution initiale, détaillée au paragraphe suivant.

La première modélisation consiste à approcher les tables prospectives au moyen d’un ajustement rudimentaire de type Gompertz, dont les paramètres évoluent linéairement. Afin d’améliorer la qualité de l’ajustement, une déformation identique pour toutes les années de décès est appliquée. Ce premier modèle se résume donc ainsi, en notant x l’âge considéré et t l’année de décès considérée (par rapport à une année de référence) :

$$S(x, t) = T_{f_\theta}(S_{a_0 + \delta_a t, b_0 + \delta_b t}^{Gompertz}(x)),$$

θ ne dépendant pas de t . Nous opérerons ce type d’ajustement pour les tables françaises et américaines. Afin de cerner au mieux le gain apporté par la déformation T_{f_θ} , nous avons optimisé les paramètres de la déformation de Gompertz en vue de minimiser l’écart entre les fonctions de survie, ce qui permet d’améliorer légèrement l’indice de qualité de l’ajustement initial par rapport aux ajustements de Gompertz des paragraphes précédents. Nous avons choisi d’ajuster les fonctions de survie sur les 30 dernières années de données disponibles, c’est-à-dire sur la période [1975, 2005]. L’année de référence sera ici 1975, et t représentera donc la différence entre l’année de décès considéré et 1975, $t \in \{0, \dots, 30\}$. L’ajustement de Gompertz considéré correspond toujours à un taux instantané de mortalité $\mu_{x,t} = e^{a_t x + b_t}$, et une fonction de survie $S^{Gompertz}(x, t) = e^{(e_t^b - \mu_{x,t})/a_t}$, avec des paramètres a_t et b_t évoluant selon l’année de décès

	France 1975-2005	USA 1975-2005
a_0	-0.491289836	-0.361101767
b_0	70.063157305	44.101327024
δ_a	0.0002907186	0.0002223594
δ_b	-0.040132231	-0.026831193
I_Q moyen 1975-2005	1.8705173846	1.9926931009
I_Q min (année)	1.8351969465 (1991)	1.9177576017 (1975)
I_Q max (année)	1.9166623069 (2005)	2.043199272 (1999)

Tab. 4 – Ajustements simultanés des fonctions de survie pour les années de 1975 à 2005 par une loi de Gompertz avec évolution linéaire des paramètres

	France 1975-2005	USA 1975-2005
x_0	-1.861284225	3.5001285387
y_0	5.1712241962	5.6546689349
p_1	2.3908915996	1.2556625787
p_2	0.2357981499	0.0000250609
ϵ	8.3736456237	34.837517538
I_Q moyen 1975-2005	2.5423560775	2.5848441805
I_Q min (année)	2.2850246327 (2003)	2.2979580277 (1975)
I_Q max (année)	2.725839914 (1995)	2.8075256348 (1990)

Tab. 5 – Ajustements simultanés des fonctions de survie pour les années de 1975 à 2005 par déformation statique d'une loi de Gompertz dont les paramètres évoluent linéairement

considérée :

$$a_t = a_0 + \delta_a t, \quad b_t = b_0 + \delta_b t.$$

Cet ajustement initial compte parmi les plus simples que l'on puisse imaginer pour modéliser des tables prospectives de mortalité. Les résultats obtenus sont indiqués dans la table 4.

L'ajustement est correct, mais pour des besoins de simulation, une erreur de l'ordre de 10^{-2} sur S correspond à un écart de l'ordre d'une année sur x , pour un âge de 60 ans pour l'année de décès 1980, ou encore d'avantage pour les âges plus faibles. Il peut s'avérer important de réduire cette erreur, et nous allons voir que si il est difficile, avec un nombre très limité de paramètres, de gagner une décimale sur S pour chacune des 31 tables de 1975 à 2005, nous pourrions néanmoins améliorer sensiblement la qualité de l'estimation. Afin d'améliorer la précision de l'ajustement, nous allons déformer par une unique déformation statique tous les ajustements Gompertz (dont les paramètres évoluent linéairement). Nous avons choisi ici pour fonction d'altération une hyperbole à 5 paramètres, l'évolution de l'indice de qualité nous ayant pas paru très importante lors de l'ajout de paramètres supplémentaires par composition, les résultats obtenus apparaissent dans la table 5.

Ici, l'erreur a été ramenée sur la majorité des tables entre $2 \cdot 10^{-3}$ et $3 \cdot 10^{-3}$, ce qui représente une amélioration de l'erreur absolue moyenne d'un facteur avoisinant 4, avec simplement 5 paramètres supplémentaires, et en gardant l'inversibilité analytique de la fonction de survie, ce qui pourra être utile pour des simulations.

5.4 Modélisation prospective, déformation dynamique

Une seconde modélisation consiste à représenter chaque table à l'aide d'une hyperbole ou d'une composée d'hyperbole, les paramètres évoluant au fil du temps. Le modèle est le suivant :

$$S(x, t) = T_{f_{\theta(t)}}(S^0(x)).$$

le vecteur $\theta(t)$ dépendant ici de t , et S^0 représentant la fonction de survie d'une loi exponentielle de paramètre 1. Afin d'obtenir un ajustement d'indice de qualité proche de 3, nous utiliserons

	France 1975-2005		USA 1975-2005	
	valeur en t_0	variation annuelle δ_{\dots}	valeur en t_0	variation annuelle δ_{\dots}
x_0	-41.53863841	-0.842434497	-93.31499759	-0.286202887
y_0	28.769734752	0.2211269332	0.2788635868	0.0578831192
p_1	0.5203056857	0.0072216002	0.4488498422	0.0093110481
p_2	0.3563170509	0.017845381	0.1110527388	0.0002171147
ϵ	5.0355587846	0.0676164418	3.7147851699	0.1180810776
a_1	0.5285555707	0.0046795803	4.6840510287	0.0770143004
q_1	0.1191526406	-0.001668731	0.145247124	-0.000537494
I_Q moyen	2.7183640463		2.6993190644	
I_Q min. (année)	2.4151646347 (2004)		2.4670385128 (1975)	
I_Q max. (année)	3.0722898974 (1995)		3.0415419452 (2001)	

Tab. 6 – Ajustements simultanés des fonctions de survie pour les années de décès de 1975 à 2005, par déformations dynamiques d’une loi exponentielle de paramètre 1, avec évolution temporelle linéaire des paramètres des déformations.

comme fonction d’altération la fonction f_{θ_t} de type $H_2 \circ H_5$, qui donnait de bons résultats sur une unique année :

$$f_{\theta(t)} = H_{a_1^t, a_1^t, 1, q_1^t, \epsilon^t} \circ H_{x_0^t, y_0^t, p_1^t, p_2^t, \epsilon^t}.$$

Pour l’évolution du vecteur paramètre $\theta(t)$ au fil du temps, nous ferons le choix très simple d’une évolution linéaire. Il faut noter que de ce fait, les résultats pourront dépendre fortement de la paramétrisation choisie. En particulier, une évolution linéaire sur une pente p n’aurait pas le même impact qu’une évolution linéaire sur le logarithme ρ de cette pente.

Les paramètres seront représentés en fonction d’une année de référence. Pour des tables données sur des années de décès 1975-2000, nous prendrons comme année de référence le milieu de la tranche, soit $t_0 = 1990$. La tranche d’âge considérée correspond toujours à l’intégralité des données disponibles, de 0 à 110 ans. Les paramètres de la fonction f_{θ_t} seront supposés évoluer linéairement au fil des années de décès considérées :

$$\theta_t = (x_0^t, y_0^t, p_1^t, p_2^t, \epsilon^t, a_1^t, q_1^t) = (x_0^0 + t\delta_{x_0}, y_0^0 + t\delta_{y_0}, p_1^0 + t\delta_{p_1}, p_2^0 + t\delta_{p_2}, \epsilon^0 + t\delta_{\epsilon}, a_1^0 + t\delta_{a_1}, q_1^0 + t\delta_{q_1}),$$

t représentant ici la différence entre l’année de décès considérée et l’année de référence t_0 .

Les résultats de ces déformations dynamiques sont donnés dans la table 6. On aurait pu craindre ici une plus grande instabilité de la qualité de l’ajustement, mais aucune année de décès ne conduit à un ajustement de qualité inférieure à 2.4 (Le pire ajustement est représenté dans la figure 3). Une majorité d’ajustements ont un indice de qualité supérieur à 2.6, ce qui représente une erreur sur S de l’ordre de $2.5 \cdot 10^{-3}$, la moyenne des ajustements représentant une erreur de l’ordre de $2 \cdot 10^{-3}$.

Enfin, la considération de tranches d’âges moins étendues (ici de 0 à 110 ans), de plage d’années de décès plus restreintes (ici des années 1975 à 2005), conduirait comme précédemment à une amélioration sensible de la qualité de l’ajustement. Le nombre de paramètres utilisé est ici de 14, dont quelques-uns ont une utilité réduite. Ce nombre n’est pas si important au regard du volume de données ajusté. Certains paramètres restent relativement stables au fil du temps. La réflexion pour déterminer quels paramètres conserver n’a pas été développée ici, l’objet de ce paragraphe étant de montrer la faculté d’adaptation des fonctions d’altérations considérées dans un cadre prospectif. Nous aurions pu adopter une autre modélisation, en proposant de modéliser l’ensemble des tables sur la période 1975 – 2005 par une déformation d’une table d’une année de référence, par exemple celle de 1990, en écrivant :

$$S(x, t) = T_{f_{\theta(t)}}(S(x, 1990)).$$

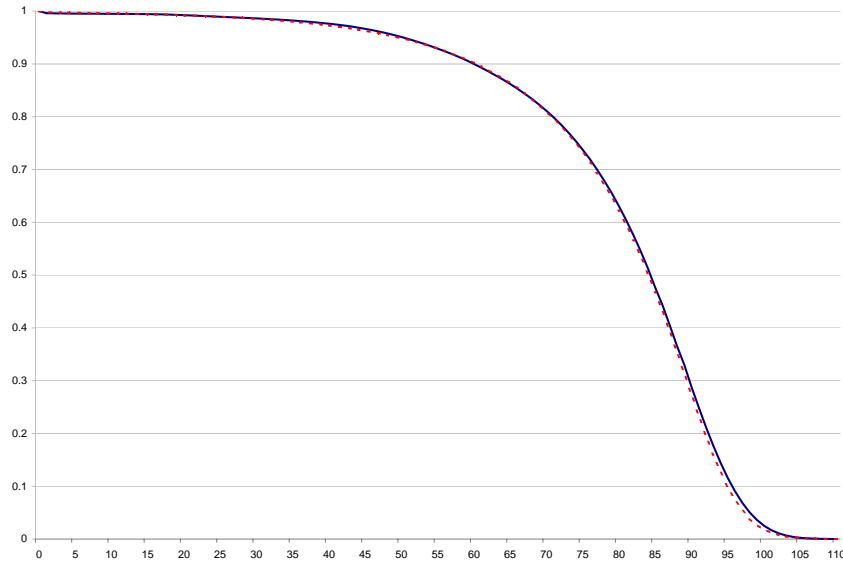


Fig. 3 – Pire ajustement obtenu lors de l’ajustement simultané des fonctions de survie pour les années de 1975 à 2005 (table France, fonction de survie de l’année de décès 2004 en trait plein, $\mathbb{T}_{f_{\theta(t)}} S^0$ en pointillé, $I_Q \simeq 2.4$).

Il semble a priori beaucoup plus simple d’adapter une table de mortalité de l’année 1990 plutôt qu’une exponentielle de paramètre 1 extrêmement inadaptée à la modélisation de la mortalité humaine. A titre d’illustration l’indice de qualité obtenu en ajustant une table de l’année de décès 2005 en fonction de celle de 1990, pour une fonction d’altération de type $H_2 \circ H_5$, est de 3, 13 pour les tables françaises, 3.32 pour les tables américaines. L’amélioration n’est donc que de l’ordre de 0.1 à 0.2 pour l’indice de qualité, par rapport à la déformation d’une exponentielle de paramètre 1. Nous avons en outre pris ici le parti de conserver une expression paramétrique continue dont la fonction inverse est aisée à obtenir de façon analytique.

5.5 Impact de la paramétrisation sur les déformations dynamiques

Enfin, nous avons également essayé d’utiliser une modélisation complètement symétrique (au sens défini dans l’axiome C5) avec une évolution linéaire des paramètres. Dans ce dernier cas, nous utilisons une fonction d’altération de type $\bar{H}_2 \circ \bar{H}_5$, où les paramètres évoluent linéairement. Cela revient notamment à supposer que les logarithmes des pentes des hyperboles évoluent linéairement, plutôt que les pentes elles-mêmes.

$$f_{\theta(t)} = \bar{H}_{0, \rho_2^t, 0, h_2^t, e^t} \circ \bar{H}_{\rho_1^t, \rho_2^t, h_1^t, h_2^t, e^t},$$

avec

$$\theta_t = (\rho_1^t, \rho_2^t, h_1^t, h_2^t, e^t, \rho_2^t, h_2^t) = (\rho_1^0 + t\delta_{\rho_1}, \rho_2^0 + t\delta_{\rho_2}, h_1^0 + t\delta_{h_1}, h_2^0 + t\delta_{h_2}, e^0 + t\delta_e, \rho_2^0 + t\delta_{\rho_2}, h_2^0 + t\delta_{h_2}),$$

t représentant ici la différence entre l’année de décès considérée et l’année de référence t_0 . Les résultats obtenus avec cette modélisation complètement symétrique sont très proches de ceux obtenus avec une paramétrisation non complètement symétrique, avec pour les tables France et USA un même indice de qualité légèrement supérieur à 2.7. Ces résultats apparaissent dans

	France 1975-2005		USA 1975-2005	
	valeur en t_0	variation annuelle δ_{\dots}	valeur en t_0	variation annuelle δ_{\dots}
ρ_1	-0.752373066	0.0010759827	-0.631686318	0.0262924511
ρ_2	-1.230137561	0.0205771249	-1.411088309	0.037514726
h_1	35.012554839	0.2472455037	37.997016123	0.5652152238
h_2	-13.19736708	-0.502650324	-15.2462469	-0.219197073
e	3.9697980803	-0.035415201	4.0934229086	0.0849409362
ρ'_2	-1.948417765	0.0029115404	-1.871937492	-0.017363335
h'_2	0.3459137795	-0.000503276	-0.537955434	-0.022157972
I_Q moyen	2.7060637572		2.7162917298	
I_Q min. (année)	2.4053924319 (2004)		2.5047983294 (1979)	
I_Q max. (année)	3.0432242972 (1995)		2.9357837381 (1996)	

Tab. 7 – Ajustements simultanés des fonctions de survie pour les années de décès de 1975 à 2005, par déformations dynamiques complètement symétriques d’une loi exponentielle de paramètre 1, avec évolution temporelle linéaire des paramètres des déformations.

la table 7. Les pentes évoluant peu dans ces applications relatives à la mortalité humaine, ce résultat n’est pas ici très surprenant : le choix de la paramétrisation pourrait revêtir un impact plus notable dans des applications où l’évolution temporelle a un impact beaucoup plus marqué sur les fonctions de survie.

Là encore, il est possible de réduire notablement le nombre de paramètres. La réduction du nombre de paramètres peut avoir notamment comme intérêt une meilleure maîtrise de l’évolution des fonctions de survie au fil du temps, en limitant les effets antagonistes lorsque plusieurs paramètres évoluent. Elle peut être un préalable à certaines projections, afin de définir une tendance qui ne soit pas noyée par un degré de liberté trop important. A titre d’exemple, une manipulation erronée nous avait conduit à fixer $\delta_{h_1} = \delta_{h_2} = \delta_e = 0$. L’impact avait été relativement faible, conduisant pour les tables France et USA à un même indice de qualité moyen de 2.67. La supposition de tous les paramètres constants à l’exception de h_1^t et h_2^t mène quant-à-elle à un indice de qualité encore très correct, proche de 2.6. Le choix des paramètres significatifs reste naturellement un problème délicat, qui demeure lié au choix de la paramétrisation.

5.6 Simulations stochastiques

Supposons que l’on sache calculer la réciproque de la fonction de survie de départ S . La contrainte d’inversibilité sur les transformations de fonction de survie que nous manipulons permet de simuler facilement (par la méthode d’inversion) la loi obtenue après transformation. En effet, la fonction de répartition F de la loi transformée s’écrit

$$F(x) = 1 - \widehat{S}(x) = 1 - T_f(S(x)),$$

de sorte que

$$F^{-1}(y) = S^{-1}(T_{f^{-1}}(1 - y)).$$

Ainsi, si U est une variable aléatoire de loi uniforme sur $]0, 1[$,

$$X = S^{-1}(T_{f^{-1}}(U))$$

a pour fonction de survie \widehat{S} . Remarquons aussi aussi que, pour tout $u \in \mathbb{R}$, la simulation de la loi conditionnelle de X sachant $X > u$ peut tout aussi bien être effectuée par la méthode d’inversion : la fonction de répartition F_u de cette loi conditionnelle s’écrit

$$F_u(t) = \mathbb{P}[X \leq t | X > u] = \frac{\mathbb{P}[X > u] - \mathbb{P}[X > t]}{\mathbb{P}[X > u]} = 1 - \frac{\widehat{S}(t)}{\widehat{S}(u)},$$

de sorte que

$$F_u^{-1}(y) = \widehat{S}^{-1}((1-y)\widehat{S}(u)) = S^{-1}\left(T_{f^{-1}}\left((1-y)T_f(S(u))\right)\right).$$

En reprenant l'exemple numérique d'ajustement d'une table de mortalité de la section 5.1, la durée de vie résiduelle X_u d'un individu ayant déjà atteint l'âge u pourra être simulée de manière très efficace par la formule suivante :

$$X_u = -\ln\left[T_{f^{-1}}\left(V \cdot T_f(e^{-u})\right)\right] - u,$$

où V est simulée suivant la loi uniforme sur $]0, 1[$ (on utilise simplement la valeur renvoyée par le générateur pseudo-aléatoire). Ainsi, lorsque la fonction d'altération dépend d'un vecteur paramètre explicatif ou bien évolue au fil du temps (comme dans la section 5.4), ce procédé permet de simuler des durées de vie résiduelles conformes à une table de mortalité dépendant d'un ou plusieurs paramètres, par exemple prospective. A titre illustratif, notons f_t la fonction d'altération à appliquer à une fonction de survie de référence S pour modéliser la loi de survie d'un individu de date de naissance t . Pour un individu de cette génération ayant atteint l'âge u , un tirage de la durée de survie résiduelle X_u^t s'obtient à partir d'un tirage V uniforme sur $]0, 1[$, par :

$$X_u^t = S^{-1}\left[T_{f_t^{-1}}\left(V \cdot T_{f_t}(S(u))\right)\right] - u.$$

Références

- Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1997), Thinking Coherently. Risk 10, 68-71.
- Bagdonavicius, V., Nikulin, M. (2002), Accelerated Life Models. Monographs on Statistics and Applied Probability 94.
- Bernoulli, D. (1731), Specimen theoriae novae de mensura sortis. English trad. Exposition of a new theory on the measurement of risk, *Econometrica*, XXI, pp. 223 sqq., 1954.
- Bleichrodt H., Eeckhoudt, L. (2006), Survival risks, intertemporal consumption, and insurance : The case of distorted probabilities. *Insurance : Mathematics and Economics*, 38, 335-346.
- Brass, W. (1969), A generation method for projecting death rates. Population growth and the brain dead. techniques and methods of study. Bechhofer F. editor, Birmingham, Edinburgh University Press : 75-91.
- Brass, W. (1974), Mortality models and their uses in demography. *Transactions of the faculty of actuaries*, 33. 122-133.
- Bühlman, H., (1980) An economic premium principle. *ASTIN Bulletin* 11, 52-60.
- D'Alembert (1768), Vingt-troisième mémoire. V. Sur le calcul des probabilités, *Opuscles Mathématiques*, vol.IV, pp.74-79, Paris, David.
- Denneberg, D. (1994), Non-additive measure and Integral. Kluwer academic Publishers, Dordrecht.
- De Jong, P., Marshall, C., (2007), Mortality projection based on the Wang transform, *ASTIN Bulletin*, 37 (1), 149-161.

- Feller, W. (1968) *An Introduction to Probability Theory and Its Applications, Volume I*, John-Wiley & Sons.
- Goovaerts, M.J., Kaas, R., Dhaene, J., Tang, Q., (2004) Some new classes of consistent risk measures, *Insurance : Mathematics and Economics*, 34, 505-516.
- Hamada, M., Sherris, M. (2003) Contingent claim pricing using probability distortion operators : methods from insurance risk pricing and their relationship to financial theory, *Applied Mathematical Finance*, 10, 19-47
- Heligman, L, Pollard, J.H. (1980) The age pattern of mortality. *Journal of the institute of actuaries*, 107, 49-80.
- Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock (Germany). [http ://www.mortality.org](http://www.mortality.org) or [http ://www.humanmortality.de](http://www.humanmortality.de).
- Landsman, Z., Sherris, M., (2001) Risk measures and insurance premium principles, *Insurance : Mathematics and Economics*, 29 , 103-115.
- Lee, R.D., Carter, L.W. (1992) Modelling and forecasting U.S. mortality (with discussion). *Journal of the American Statistical Association*, 87(419), 659-675.
- Pelsser, A. (2007), On the applicability of the Wang Transform for Pricing Financial Risks, University of Amsterdam and NETSPAR, working paper.
- Pitacco, E. (2004), Survival models in a dynamic context : a survey. *Insurance : Mathematics and Economics*, 35, 2, 279-298.
- Pradier, P.-C. (1998), Concepts et mesures du risque en théorie économique - essai historique et critique, Thèse ENS-Cachan.
- Wang, S.S. (1996), Premium calculation by transforming the layer premium density, *ASTIN Bulletin* 26 (1), 71-92.
- Wang, S.S. (2000), A class of distortion operators for pricing financial and insurance risks, *The Journal of Risk and Insurance* 67(2), 15-36.
- Wirch, J.L., Hardy, M.R. (1999), A synthesis of risk measures for capital adequacy, *Insurance : Mathematics and Economics*, 25, 337-347.
- Yaari, M.E. (1987), The dual theory of choice under risk. *Econometrica* 55 (1), 95-115.