

Document de recherche

Le mécanisme mathématique du vieillessement biologique

**Direction de la recherche –
Commission sur la recherche universitaire**

Octobre 2019

Document 219106

*This document is available in English
© 2019 Institut canadien des actuaires*

Le mécanisme mathématique du vieillissement biologique

Boquan Cheng, Bruce Jones, Xiaoming Liu et Jiandong Ren

Département de la statistique et de l'actuariat

Université Western Ontario

London, Ontario N6A 5B7 Canada

Bien que le vieillissement soit un phénomène biologique universel et omniprésent, la description du mécanisme du vieillissement en termes mathématiques précis, en particulier la façon d'en établir un modèle et de quantifier le taux de vieillissement, représente un défi non résolu depuis des siècles. Dans le présent document, nous proposons un modèle markovien de type coxien pouvant fournir une description quantitative des caractéristiques bien connues du vieillissement, un processus génétiquement déterminé, progressif et essentiellement irréversible. Notre modèle présente une structure unique, y compris un taux de transition constant pour le processus du vieillissement, ainsi qu'une forme fonctionnelle pour la relation entre le vieillissement et la mort, assortie d'un paramètre de forme pour saisir l'effet de la détérioration biologique dû au vieillissement. La force de passage d'un état à un autre dans le processus markovien indique la force biologique intrinsèque de vieillissement. Le taux de sortie croissant qui s'y rattache saisit la force externe de stress attribuable au risque de mortalité qui s'exerce sur un organisme vivant.

L'idée de ce document provient de Lin et Liu (2007). Une grande différence tient au fait que, dans le présent document, notre modèle utilise une forme fonctionnelle pour les paramètres du modèle, ce qui permet une représentation parcimonieuse, mais flexible pour divers modèles de vieillissement. Le cadre mathématique que nous proposons peut servir à classer le modèle de vieillissement et les paramètres clés du modèle peuvent être utilisés pour mesurer et comparer la façon dont le vieillissement humain évolue au fil du temps et entre les populations.

1. Introduction

[Traduction libre] *La vie humaine est une affaire très personnelle. C'est votre vie, la mienne, celle de notre voisin. Chaque vie est une entité propre et distincte, mais il y a une empreinte commune sur tous... Bien des vies sont écourtées... un accident ou une maladie aiguë peut éteindre l'ampoule avant qu'elle ne soit brûlée. Mais de toute manière elle s'éteindra même si de tels événements ne se produisent pas.*

Dublin et coll. (1949)

Dans la phrase ci-dessus, Dublin et coll. a bien décrit le fait que, même si les durées de vie individuelles peuvent être influencées par de nombreux facteurs personnels et aléatoires, il y a ce processus de vieillissement omniprésent qui sous-tend la vie de tous les êtres humains.

Alors, qu'est-ce que le « vieillissement »?

Pour commencer, citons la définition du « vieillissement » tirée de Jones (1956) : [traduction libre] « *Le vieillissement, tel qu'il est appliqué aux organismes vivants, représente la diminution génétiquement déterminée, progressive et essentiellement irréversible avec le temps de la capacité d'un organisme ou d'une de ses parties de s'adapter à son environnement, qui se manifeste par une réduction de sa capacité à résister aux stress auxquels il est soumis (c.-à-d. l'augmentation de la sensibilité à certaines maladies avec l'âge), et qui prend fin par la mort de l'organisme.* »

La définition de Jones fournit certaines caractéristiques du processus de vieillissement : détermination génétique, progressivité et irréversibilité incontournable. Cependant, dans cette définition du vieillissement, on ne sait trop quelles capacités diminuent avec le temps, et comment. À l'instar d'une grande partie des recherches effectuées à l'époque (avant le milieu du XX^e siècle), les études sur le processus de vieillissement étaient essentiellement descriptives et n'étaient pas accessibles aux fins d'évaluation en raison de l'absence de données directement observables sur le vieillissement. C'est pourquoi les théories antérieures sur le vieillissement tentaient d'associer l'impact du vieillissement à l'augmentation des taux de décès pour justifier quantitativement leur hypothèse. En d'autres termes, il était courant par le passé que le modèle de mortalité à croissance exponentielle (la loi de Gompertz sur la mortalité, en raison de son bon ajustement à une grande partie des taux de mortalité chez les adultes) agissait comme « manifeste » du processus de vieillissement « indirectement » observé.

Toutefois, nous aimerions clarifier la relation entre le vieillissement et le décès : bien que le taux de mortalité soit fortement corrélé au taux du vieillissement, les deux ne sont pas tout à fait équivalents. De nombreux facteurs externes peuvent causer un décès prématuré et(ou) même modifier le processus de vieillissement. Pour n'en citer que quelques-uns, ces facteurs comprennent les antécédents familiaux, l'éducation, le revenu, le mode de vie, l'état civil, les accidents et les maladies aiguës. Selon Herskind et coll. (1995), une composante génétique additive explique environ 2 % de la variabilité de la durée de vie, ce qui indique que des facteurs non génétiques contribuent de façon importante à la durée de vie. Plutôt que d'envisager le vieillissement et le décès comme une relation de cause à effet, on croit maintenant que le profil d'âge dans le risque de mortalité découle de l'interaction entre le processus du vieillissement individuel et les stress externes, comme l'indique Yashin et coll. (2012) et de nombreux renvois dans cet ouvrage.

Un autre changement important est survenu au cours des dernières décennies en ce qui concerne les études sur le vieillissement, soit que de nombreux chercheurs ont commencé à recueillir des données longitudinales sur les facteurs liés au vieillissement. Parmi ces facteurs, mentionnons l'indice de masse corporelle, la tension artérielle diastolique, la tension artérielle, la fréquence cardiaque, le taux de glycémie, l'hématocrite et le cholestérol sérique; encore une fois, voir Yashin et coll. (2012) et des références à ces données pour les recherches récentes en ce sens. Contrairement à de nombreuses études antérieures, les observations issues de ces

études proviennent de jeunes personnes en santé dans le but de comprendre comment les processus biologiques se développent chez les humains vieillissants avant même l'apparition des maladies chroniques. Dans Yashin et coll. (2012), on constate que différents facteurs affichent des tendances différentes de l'âge moyen et des propriétés dynamiques différentes. Certains facteurs suivent des changements monotones avec l'âge; d'autres sont non monotones (p. ex., ils augmentent, puis diminuent). Ces constatations révèlent également que lorsque les indices physiologiques d'une personne s'écartent de ses valeurs « optimales », le risque de mortalité est supérieur à la moyenne. De notre point de vue, leurs recherches pourraient être plus utiles pour comprendre pourquoi certaines personnes peuvent vivre plus longtemps que d'autres. Toutefois, si nous voulons élaborer un outil quantitatif global pour décrire l'effet du vieillissement de la population, nous pourrions devoir nous concentrer sur des statistiques plus générales.

Par « statistiques plus générales », nous entendons des concepts semblables à la charge allostatique définie par Crimmins et coll. (2003) ou au biomarqueur défini par Belsky et coll. (2015). En résumé, ces indices sont établis à partir de diverses fonctions physiologiques dans plusieurs systèmes de l'organisme (p. ex., les fonctions pulmonaire, parodontale, cardiovasculaire, rénale, hépatique et immunitaire), afin de mesurer le fardeau biologique cumulatif qui s'est développé au fil du temps ou de marquer le degré de détérioration d'un corps vieillissant. Malgré l'intensification des études au cours des dernières décennies et l'accès à de meilleures données longitudinales, la recherche dans ce domaine en est encore à son stade embryonnaire. Cette situation est en grande partie attribuable à la complexité du processus de vie et de décès. De notre point de vue, il semble impossible d'atteindre un indice de vieillissement biologique communément accepté dans un proche avenir.

Par conséquent, plutôt que d'adopter une approche absolue selon laquelle il faut choisir un ou plusieurs indices représentatifs du vieillissement, nous proposons un modèle markovien de type coxien qui peut fournir une description quantitative des caractéristiques bien connues du vieillissement — le processus « génétiquement déterminé, progressif et essentiellement irréversible ». La structure de notre modèle est unique. Elle comporte entre autres un taux de transition constant pour le processus du vieillissement, et une forme fonctionnelle pour la relation entre le vieillissement et la mort, assortis d'un paramètre de forme pour saisir l'effet de détérioration biologique dû au vieillissement. Le modèle proposé renferme deux forces distinctes, mais connexes dans ce cadre markovien : la force du vieillissement (c.-à-d. le processus biologique intrinsèque) et la force de la mort (c.-à-d. les forces externes découlant du stress environnemental, y compris les accidents, l'accessibilité à la nutrition et aux soins médicaux, etc.).

L'idée de ce document provient de Lin et Liu (2007). Une grande différence tient au fait que, dans le présent document, notre modèle utilise une forme fonctionnelle pour les paramètres du modèle, ce qui permet une représentation parcimonieuse, mais flexible pour divers modèles de vieillissement.

Le modèle proposé appartient à une grande catégorie de distributions : les soi-disant distributions de type phase. Il y a longtemps que l'on utilise des distributions de type phase pour la modélisation des taux de survie dans la catégorie des distributions « d'absorption dans

le temps »; voir Aalen (1995), Asmussen et coll. (1996), Lin et Liu (2007), ainsi que Su et Sherris (2012), pour n'énumérer que quelques sources. Les distributions de type phase peuvent être clairement mises en œuvre lorsque les états du processus sous-jacent sont observables et peuvent être prédéfinis. Autrement, la souplesse de ces distributions pourrait se transformer en un inconvénient, car elle rendrait l'estimation du modèle beaucoup plus difficile en raison de la propriété de non-unicité de la distribution de type phase (Slud et Suntornchost, 2014). Notre processus défini du vieillissement est un processus latent. Toutefois, une forme fonctionnelle soigneusement conçue pour les paramètres du modèle aide à atténuer le problème de non-identifiabilité.

Le cadre mathématique que nous proposons peut servir à spécifier le modèle du vieillissement et les paramètres clés du modèle peuvent être utilisés pour mesurer l'effet de détérioration associé au vieillissement. Notre modèle peut également fournir des comparaisons biologiquement significatives sur l'évolution du vieillissement humain au fil du temps et entre les populations. Le modèle présente les autres avantages suivants :

- Le modèle compte un petit nombre de paramètres qui peuvent être évalués de façon pratique à l'aide de l'estimation du maximum de vraisemblance. La structure parcimonieuse du modèle fait en sorte que celui-ci peut être identifié de façon unique à l'aide d'un certain ensemble de données et que le modèle qui en découle peut fournir une interprétation biologique significative.
- Le modèle est souple quant à la façon dont il représente l'incidence du vieillissement sur la mortalité.
- Le modèle introduit un caractère aléatoire associé au vieillissement et permet l'hétérogénéité quantifiable au plan de l'âge physiologique des individus d'une population. (Remarque : l'âge physiologique et l'âge biologique sont des termes interchangeables utilisés pour saisir le degré de vieillissement biologique d'un individu. Normalement, l'âge physiologique est calibré de manière à ce qu'un individu d'âge chronologique x ait un âge physiologique supérieur à (inférieur à) x si la personne a plus de (moins de) une personne moyenne d'âge chronologique x .)
- Cette hétérogénéité permet de prendre en compte la sélection de la mortalité à des âges plus avancés, c'est-à-dire une tendance de mortalité plus marquée chez les personnes en moins bonne santé à un âge donné, laissant ainsi un groupe de survivants en meilleure santé.

Le présent document est structuré comme suit. À la section 2, nous présentons notre modèle de vieillissement de type phase. Nous commençons par décrire brièvement les distributions de type phase, puis les caractéristiques d'une sous-catégorie appelée distribution coxienne de type phase. Nous présentons ensuite notre modèle, qui est un cas particulier de ces distributions, et nous décrivons la structure des paramètres du modèle. À la section 3, nous discutons de la façon dont le modèle peut être étalonné à l'aide des données sur la durée de vie et nous illustrons ces idées à l'aide d'un petit ensemble de données renseignant sur la durée de vie des résidents d'un village-retraite. Nous démontrons par la suite comment notre modèle peut fournir une interprétation utile pour le processus de vieillissement intégré. À la section 4, nous

présentons une analyse plus approfondie de notre modèle. Nous illustrons comment ce modèle peut saisir la distribution de la durée de vie d'un mécanisme de vieillissement très différent. Au moyen d'une étude par simulation, nous montrons la difficulté à estimer le nombre d'états du modèle avec seulement des données sur la durée de vie et nous démontrons comment le nombre d'états influence la variabilité du vieillissement des sujets selon le modèle. La section 5 termine le document par un résumé et quelques suggestions de recherches supplémentaires.

2. Le modèle proposé du vieillissement de type phase (MVP)

2.1. Mathématiques de base appliquées aux distributions de type phase

Dans cette section, nous définissons la distribution de type phase et nous en décrivons les propriétés de base. Nous avons également tenté de structurer le document de manière à faciliter la discussion sur les défis liés à l'utilisation de ce type de modèle en pratique. Pour la théorie générale des distributions de type phase, voir, par exemple, Neuts (1982) et Asmussen (1989).

Définition : Supposons un processus de Markov homogène dans le temps Y_t défini dans un espace-état fini $S = E \cup \Delta = \{1, 2, \dots, m\} \cup \Delta$, où Δ est absorbant et les états dans E sont transitoires. Supposons que Y_t a une distribution initiale $(\alpha^T, 0)$ (exprimée sous forme de vecteur ligne) et un générateur infinitésimal

$$\begin{pmatrix} \Lambda & \mathbf{q} \\ \mathbf{0} & 0 \end{pmatrix} \quad (1)$$

où $\mathbf{q} = -\Lambda \mathbf{e}$ et \mathbf{e} est le vecteur des colonnes.

Supposons que T désigne le temps précédant l'absorption ou le temps précédant le décès dans un contexte de vie humaine. Nous disons alors que T suit une distribution de type phase avec représentation (α, Λ) .

Un aspect très surprenant des modèles de type phase réside dans leur interprétation intuitive du mécanisme biologique et mathématique sous-jacent du système qui débouche sur la distribution du temps de survie. En langage simple, le temps de survie dans un tel système correspond à la somme des séjours dans tous les états que le processus a traversés avant d'être absorbé. La probabilité de survie à tout moment t équivaut à la probabilité totale que le processus se trouve dans un état autre que celui de l'absorption au moment t .

Pour $s \geq 0, t \geq 0, i, j \in E$, supposons que $P_{ij}(s, s+t)$ est la probabilité de passage de l'état i à l'état j au cours de la période $(s, s+t]$:

$$P_{ij}(s, s+t) = \Pr(Y_{s+t} = j | Y_s = i). \quad (2)$$

Étant donné que Y_t est un processus de Markov homogène dans le temps, la probabilité $P_{ij}(s, s+t)$ ne dépend pas de s ; nous formulons donc $P_{ij}(t) = P_{ij}(s, s+t)$ pour tous les $s \geq 0$.

Soit $\mathbf{P}(t) = \{P_{ij}(t)\}_{i,j \in E}$ la matrice de probabilité de transition (parmi les états transitoires) de Y_t dans l'intervalle $[0, t]$. Elle respecte donc l'équation prospective de Kolmogorov,

$$\frac{d}{dt} \mathbf{P}(t) = \mathbf{P}(t) \mathbf{\Lambda}, \quad (3)$$

avec la condition initiale $\mathbf{P}(0) = \mathbf{I}$, où \mathbf{I} représente la matrice d'identité.

L'équation prospective de Kolmogorov présente la solution unique

$$\mathbf{P}(t) = \exp(\mathbf{\Lambda}t), \quad (4)$$

où la fonction exponentielle de la matrice $\exp(\mathbf{\Lambda}t)$ est définie par

$$\exp(\mathbf{\Lambda}t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{\Lambda}^n.$$

Par conséquent, la fonction de survie de la variable aléatoire de la période menant à l'absorption T peut être exprimée sous la forme

$$S(t) = \boldsymbol{\alpha}^T \exp(\mathbf{\Lambda}t) \mathbf{e}, \quad t > 0. \quad (5)$$

Il convient de noter que le i^{e} élément du vecteur

$$\mathbf{p}(t) = \boldsymbol{\alpha}^T \exp(\mathbf{\Lambda}t) \quad (6)$$

débouche sur la probabilité que $Y_t = i$, et $\mathbf{p}(t) \mathbf{e}$ produit la probabilité que $Y_t \in E$.

Les distributions de type phase ont des expressions de matrice simples pour leur fonction de survie, leur fonction de densité de la probabilité, leur transformation de Laplace et les moments ordinaires. En outre, la catégorie de distributions de type phase est dense dans l'espace de toutes les distributions positives continues, c'est-à-dire qu'une distribution définie sur $(0, \infty)$ peut, du moins en principe, être approximée de près au moyen d'une distribution par étapes. Elle a donc été appliquée efficacement dans bien des domaines pour obtenir des solutions d'analyse explicites. Toutefois, l'obtention de solutions numériques en résolvant l'équation différentielle (4) ou en calculant la fonction exponentielle de la matrice (5) n'est pas évidente, plus particulièrement lorsque la matrice $\mathbf{\Lambda}$ (c.-à-d. le nombre d'états, m) est de grande dimension.

De plus, une distribution générale de type phase comporte de nombreux paramètres et, parfois, la façon de paramétriser un modèle n'est pas unique (Asmussen et coll. (1996)). Par conséquent, nous savons qu'il est difficile d'établir l'estimation des paramètres des distributions de type phase. Comme nous le verrons dans la seconde partie du présent document, le problème de non-unicité de la spécification des paramètres se confond généralement avec la difficulté d'estimer les paramètres dans le modèle, surtout lorsque celui-ci renferme un grand nombre d'états et que ses paramètres sont proches les uns des autres.

Nous soulignons que le présent document envisage l'application de la modélisation de type phase au processus du vieillissement, à l'instar de Aalen (1995), Lin et Liu (2007) et Su et Sherris (2012). Dans de telles applications, le modèle peut comprendre un grand nombre d'états, mais la matrice d'intensité est clairesemée et respecte certaines structures.

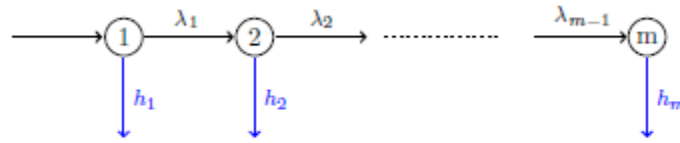
2.2. Distributions coxiennes de type phase

Dans la présente section, nous discuterons de la distribution de type coxien, qui constitue un type spécial de distribution de type phase comportant une distribution initiale $\alpha^T = (1, 0, \dots, 0)$ et une matrice d'intensité

$$\Lambda = \begin{bmatrix} -(\lambda_1 + h_1) & \lambda_1 & & & & \\ & -(\lambda_2 + h_2) & \lambda_2 & & & \\ & & \ddots & & & \\ & & & -(\lambda_{i-1} + h_{m-1}) & \lambda_{i-1} & \\ & & & & & -h_m \end{bmatrix}, \quad (7)$$

où $\lambda_i > 0$ et $h_i > 0$ pour toute la fonction $i \in E$. Le modèle coxien est illustré dans le graphique 1. Puisque le processus ne débute qu'à l'état 1, la fonction de survie du modèle coxien ne dépend que de la première ligne de la matrice de transition $\mathbf{P}(t)$, désignée par $\mathbf{p}_1(t)$. Supposons que le k^e élément de $\mathbf{p}_1(t)$ est désigné $P_{1k}(t)$ ou simplement $P_k(t)$.

Graphique 1 : Diagramme d'un processus markovien de type coxien



Alors, l'équation prospective de Kolmogorov prend la forme suivante :

$$\begin{cases} \frac{dP_1(t)}{dt} = -(\lambda_1 + h_1)P_1(t); \\ \frac{dP_k(t)}{dt} = \lambda_{k-1}P_{k-1}(t) - (\lambda_k + h_k)P_k(t), \quad k = 2, 3, \dots, m. \end{cases} \quad (8)$$

Évidemment, $P_1(t) = e^{-(\lambda_1 + h_1)t}$. En outre, $P_k(t)$ pour $k = 2, \dots, m$ peut être obtenu par itération, ce qui donne

$$P_k(t) = \sum_{j=1}^k \frac{(-1)^{k-1} \lambda_1 \dots \lambda_{k-1}}{\prod_{s=1, s \neq j}^k (\lambda_j + h_j - \lambda_s - h_s)} e^{-(\lambda_j + h_j)t}, \quad k = 2, \dots, m. \quad (9)$$

Enfin, la fonction de survie de T est donnée par

$$S(t) = \sum_{k=1}^m P_k(t). \quad (10)$$

Lorsque la dimension du modèle coxien est petite et que les valeurs paramétriques sont éloignées de sorte que le dénominateur de (9) n'est pas trop petit, l'équation (9) peut être utilisée pour évaluer rapidement la distribution. Toutefois, en modélisation de la durée de vie, le modèle à utiliser peut contenir un grand nombre d'états avec des fonctions λ_i s et h_i s très proches l'une de l'autre. L'utilisation de l'équation (9) engendre des problèmes numériques puisque dans l'ensemble les valeurs de chaque terme sont importantes et présentent des signes qui alternent.

2.3. Une catégorie de distributions coxiennes de type phase – Le MVP que nous proposons

La distribution de type coxien propose une approche naturelle et attrayante pour modéliser le vieillissement. Malheureusement, la distribution générale de type coxien comporte de nombreux paramètres, ce qui rend l'estimation difficile. Nous pouvons largement simplifier le modèle tout en conservant une grande souplesse en imposant une certaine structure. Autrement dit, nous pouvons spécifier des formes mathématiques pour le modèle des valeurs λ_i , qui représentent la force interne du vieillissement, et le modèle des valeurs h_i , qui représentent la force externe de la mort. Si les fonctions mathématiques utilisées à cette fin comprennent un petit nombre de paramètres, notre tâche sur le plan de l'estimation s'en trouve allégée.

Supposons d'abord que le nombre d'états (âges physiologiques) m dans notre modèle est connu. Nous aborderons la façon de déterminer m plus loin dans le présent document.

Supposons maintenant que $\lambda_i = \lambda$, une constante, pour tous les $i = 1, 2, \dots, m - 1$. Ce faisant, nous supposons que le taux de vieillissement est uniforme au fil du temps, de sorte que le taux d'augmentation de l'âge physiologique est uniforme. Ce résultat est logique, car l'âge civil augmente de façon uniforme. Même si le taux de vieillissement que nous supposons est constant, il est variable en périodes de transition. Par conséquent, à n'importe quel âge civil, une personne peut se trouver dans un certain nombre d'états différents qui représentent des âges physiologiques différents.

La valeur de λ doit correspondre à la valeur de m . Le taux de progression dans les états doit être suffisamment important pour que certaines personnes atteignent l'état m , sinon il n'est pas nécessaire d'avoir autant d'états m . Toutefois, nous voulons qu'une petite proportion seulement de personnes survivent à l'état m , sinon le taux de décès modèle s'aplatira aux âges plus avancés. Nous supposons donc que

$$\lambda = m/\psi, \tag{11}$$

où ψ peut être considéré comme la durée de vie des personnes constituant la population d'intérêt. Il n'est pas nécessaire que le paramètre ψ soit un âge limite, car notre modèle ne comporte pas d'âge limite. Toutefois, ψ devrait représenter un âge avancé auquel ne survivra qu'une très faible proportion de personnes. Ce paramètre peut être estimé à partir de données ou choisi en fonction d'une opinion antérieure. Pour les durées de vie humaines, une valeur de ψ entre 100 et 120 peut être raisonnable.

Nous précisons ensuite une forme pour les valeurs h_i . Supposons d'abord que h_1 et h_m sont fixes. Il s'agit des paramètres à estimer. Nous avons ensuite besoin d'un modèle lisse des valeurs de h_i pour i entre 1 et m . Nous y parvenons en supposant

$$h_i^s = \frac{m-i}{m-1} h_1^s + \frac{i-1}{m-1} h_m^s.$$

En d'autres termes, les pouvoirs de h_i sont obtenus à titre d'interpolation linéaire entre les pouvoirs correspondants de h_1 et h_m . Nous avons donc

$$h_i = \left(\frac{m-i}{m-1} h_1^s + \frac{i-1}{m-1} h_m^s \right)^{1/s}.$$

Nous pouvons utiliser cette équation pour n'importe quelle valeur s réelle, sauf lorsque l'expression $s = 0$, n'est pas définie. Toutefois, le cas limite $s \rightarrow 0$ donne

$$h_i = h_1^{\frac{m-i}{m-1}} h_m^{\frac{i-1}{m-1}}.$$

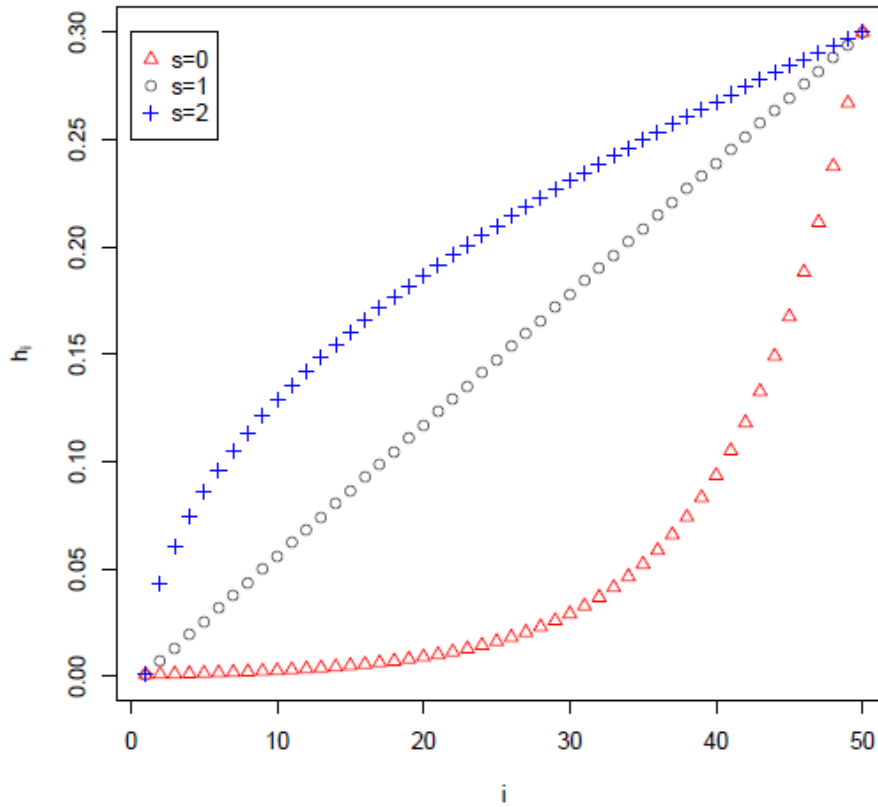
Dans ce cas, $\log h_i$ est obtenu à titre d'interpolation linéaire entre $\log h_1$ et $\log h_m$. Le paramètre s peut donc représenter n'importe quelle valeur réelle. Puis, pour $i = 1, 2, \dots, m$,

$$h_i = \begin{cases} \left(\frac{m-i}{m-1} h_1^s + \frac{i-1}{m-1} h_m^s \right)^{1/s} & s \neq 0, \\ h_1^{\frac{m-i}{m-1}} h_m^{\frac{i-1}{m-1}} & s = 0. \end{cases} \quad (12)$$

Cette structure rappelle la transformation bien connue de Box-Cox introduite par Box et Cox (1964).

Il convient de noter que lorsque $s = 1$, nous avons un modèle linéaire de valeurs h_i avec valeurs i en hausse. Lorsque $s = 0$, nous avons un modèle exponentiel de valeurs h_i . En outre, lorsque $s < 1$, le modèle sera convexe, et quand $s > 1$, le modèle sera concave. Le graphique 2 montre les modèles des valeurs h_i pour $s = 0, 1$ et 2 . Pour le vieillissement humain, nous nous attendrions que $s = 0$ soit la plus pertinente des trois valeurs. La valeur utilisée peut être estimée à partir des données sur la durée de vie.

Graphique 2 : Valeurs de h_i déterminées à l'aide de $h_1 = 0.001$, $h_m = 0.3$ et $s = 0, 1$ et 2 .



En résumé, nous proposons une distribution de type coxien avec cinq paramètres à préciser : m , ψ , h_1 , h_m et s . Nous pourrions être en mesure de choisir m et ψ d'après les connaissances antérieures, ou une combinaison de données sur la durée de vie et de connaissances antérieures. Les valeurs m et ψ ayant été établies, λ est déterminée à partir de (11), et h_1 , h_m et s sont facilement estimées d'après les données sur la durée de vie.

Idéalement, notre modèle de vieillissement serait étalonné à l'aide des données sur le vieillissement, plutôt que simplement les données sur la durée de vie. En d'autres termes, si les observations provenant d'au moins une variable clé du vieillissement sont prises à un ou plusieurs moments pendant la vie de chaque personne en plus des âges au décès, nous pourrions estimer les paramètres du modèle avec moins d'incertitude. Govorun et coll. (2018) décrit une méthode à cette fin. Cependant, cette question dépasse la portée du présent document.

Puisque les états de notre modèle ne sont connectés à aucune variable de vieillissement, ces états ne sont pas observables. Par conséquent, il est impossible d'identifier l'état d'un individu spécifique et d'utiliser cette information dans le calcul des probabilités liées à sa vie future. L'utilité de notre modèle réside davantage dans la capture du mécanisme de vieillissement et dans la compréhension de la variabilité créée par ce mécanisme.

Les taux de mortalité traditionnels des tables de mortalité peuvent être calculés à partir de notre modèle en utilisant

$$q_x = \frac{S(x) - S(x + 1)}{S(x)},$$

où les valeurs de la fonction de survie $S(\cdot)$ sont obtenues de (5). Ces taux de mortalité augmentent avec l'âge. Cependant, il est important de noter qu'aucun des taux de transition de notre modèle n'augmente avec l'âge d'un individu. Ce qui change avec le temps, c'est l'état de l'individu. Aux âges plus avancés, les probabilités associées aux états de vieillissement ultérieurs sont plus élevées. Et comme les taux de mortalité spécifiques à chaque état sont plus élevés dans les états ultérieurs, nous avons des taux de mortalité qui augmentent avec l'âge.

3. Estimation et application des modèles

3.1. Estimation des paramètres

Les paramètres de notre modèle peuvent être calculés à l'aide de l'estimation du maximum de vraisemblance, que nous décrivons ci-dessous.

Supposons que θ est le vecteur paramétrique que nous souhaitons estimer. Si nous avons choisi les valeurs de m et de ψ , donc $\theta = (h_1, h_m, s)$. En outre, supposons que nous souhaitons estimer le vecteur paramétrique à l'aide des données sur la durée de vie seulement. Plus précisément, pour chaque personne observée $j = 1, \dots, n$ nous notons (ℓ_j, t_j, δ_j) , où ℓ_j est l'âge auquel la personne j a été observée pour la première fois, t_j est l'âge auquel cette personne a cessé d'être observée, et $\delta_j = 1$ si cette personne est décédée au moment t_j et $\delta_j = 0$ dans les autres cas.

La fonction de probabilité est donnée par

$$L(\theta) = \prod_{j=1}^n \frac{f(t_j; \theta)^{\delta_j} S(t_j; \theta)^{1-\delta_j}}{S(\ell_j; \theta)}, \quad (13)$$

où $S(t; \theta)$ est notre fonction de survie modèle et $f(t; \theta)$ est notre fonction de densité de probabilité modèle. Rappelez-vous que nous avons

$$S(t; \theta) = \alpha^T \exp(\Lambda t) e,$$

et

$$f(t; \theta) = \alpha^T \exp(\Lambda t) q.$$

Les valeurs Λ et q sont des fonctions de θ . Par conséquent, pour une valeur θ donnée, nous pouvons calculer la valeur de la probabilité ou la log-vraisemblance, $l(\theta) = \log L(\theta)$. La log-vraisemblance peut alors être optimisée numériquement par rapport à θ pour obtenir les estimateurs du maximum de vraisemblance.

3.2. Application à l'aide des données provenant d'un village-retraite

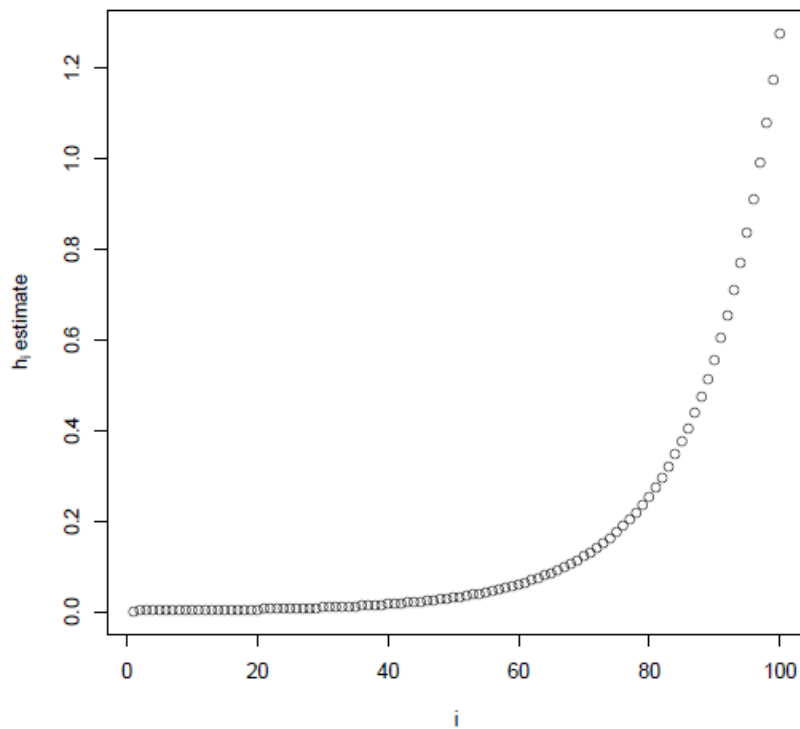
Pour illustrer la façon d'étalonner notre modèle de vieillissement à l'aide de données sur la durée de vie, nous étudions les données de Channing House, un village-retraite de Palo Alto, en Californie. L'ensemble de données comprend l'âge d'entrée et l'âge au décès (ou de fin d'étude) pour 462 personnes (97 hommes et 365 femmes) qui ont habité le village entre janvier 1964 et juillet 1975. Ces résidents participaient à un programme de soins de santé qui facilitait l'accès gratuit aux soins. Il est probable que ce programme a permis d'abaisser la mortalité moyenne. Nous ajustons notre modèle de vieillissement uniquement pour les données sur les femmes. Par conséquent, nous tenons compte des données portant sur un groupe relativement restreint de personnes homogènes – toutes des femmes vivant dans le même village et ayant le même accès aux soins de santé et, fort probablement, des styles de vie semblables. Cette situation est idéale, car une hypothèse d'homogénéité sous-tend notre modèle.

Les dossiers de seulement 362 des 365 femmes ont été conservés, puisque les trois autres avaient des âges d'entrée et de décès identiques. Des 362 femmes, 130 sont décédées pendant leur séjour au village et les 232 autres ont survécu jusqu'à la fin de la période d'observation. La plus jeune participante à l'entrée avait un peu plus de 61 ans et la plus âgée avait un peu moins de 101 ans. Ainsi, les données portent sur le vieillissement compris dans cette fourchette d'âge.

Puisque notre modèle suppose que toutes les personnes commencent à l'état 1, mais que nous supposons une certaine variabilité de l'état à compter de 61 ans, nous supposons par exemple que le processus de vieillissement débute à 50 ans. Cela nous permet d'obtenir une certaine variabilité dans la distribution de l'état à compter de 61 ans. Nous supposons également, quoiqu'un peu arbitrairement, que la durée de vie prend fin à 105 ans; tous les membres du village étaient décédés à cet âge. Donc, $\psi = 105 - 50 = 55$. En outre, selon l'expérience, nous estimons que $m = 100$ états de vieillissement conviennent à la modélisation du vieillissement au-delà de 50 ans, ce qui donne une variabilité raisonnable de l'âge physiologique à divers âges civils, comme l'indique le graphique 6 ci-après. Nous avons ensuite $\lambda = m/\psi = 100/55 = 1,8182$. À l'aide de l'estimation du maximum de vraisemblance, décrite ci-dessus, nous obtenons $\hat{h}_1 = 0,00175471$, $\hat{h}_m = 1,27518$ et $\hat{s} = -0,0734710$. Les estimateurs paramétriques sont très corrélés et les variances sont importantes. Toutefois, les estimations semblent raisonnables et se traduisent par un bon ajustement aux données.

Les estimations de h_i qui découlent de \hat{h}_1 , \hat{h}_m et de \hat{s} sont reproduites dans le graphique 3. Le modèle des valeurs semble raisonnable compte tenu du fait que ces valeurs représentent des taux de mortalité instantanés qui s'appliquent entre 50 et 105 ans.

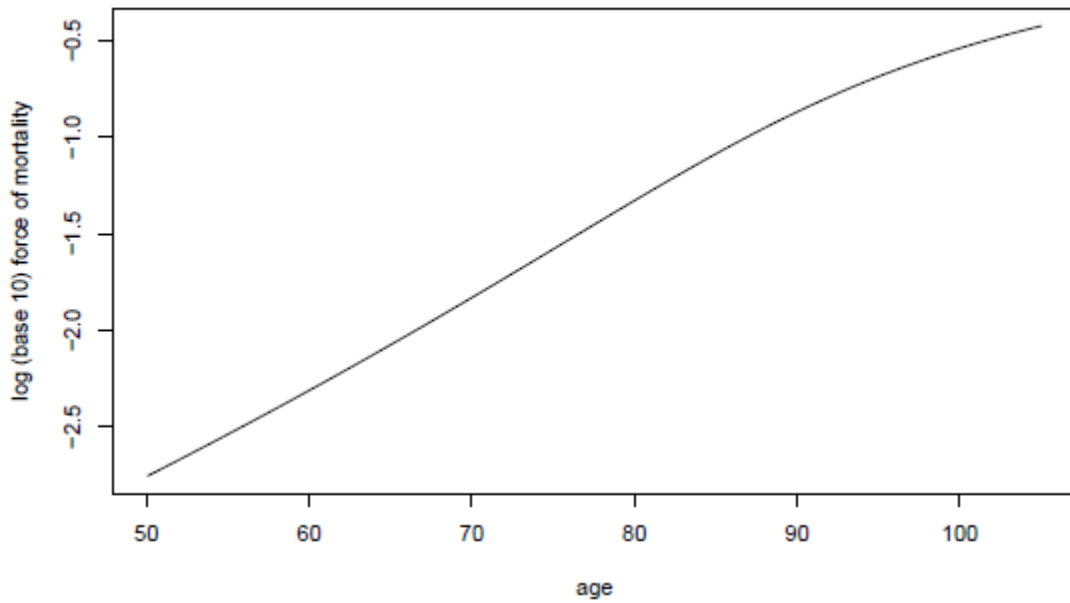
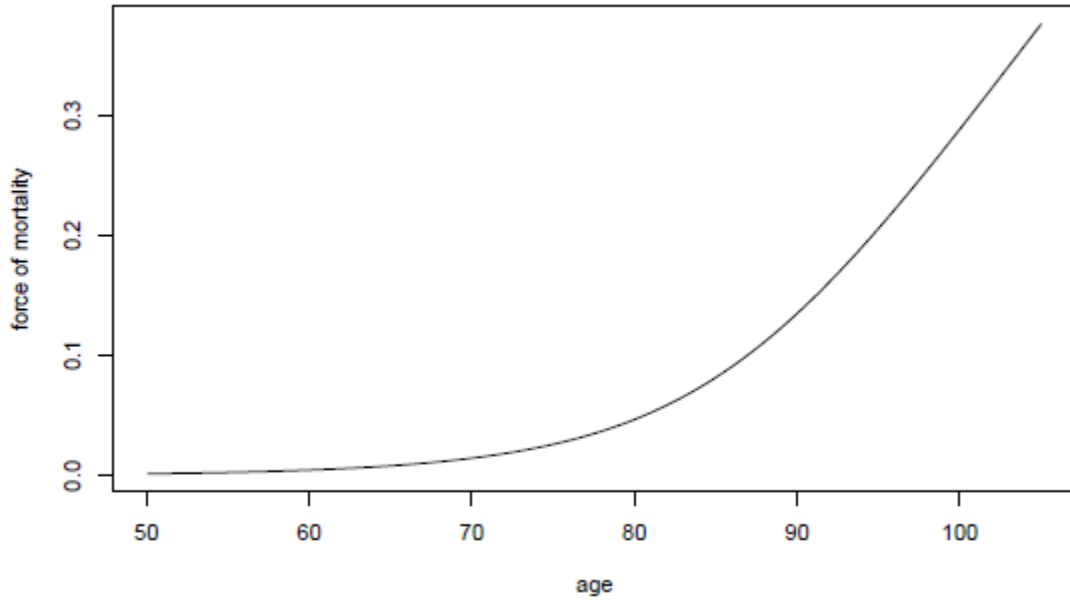
Graphique 3 : Estimations de h_i obtenues par étalonnage du MVP à l'aide des données des femmes résidant à Channing House



*En anglais seulement.

Le graphique 4 montre la force de la mortalité et le logarithme (base 10) de cette force d'après le modèle ajusté. Encore une fois, le comportement est passablement raisonnable.

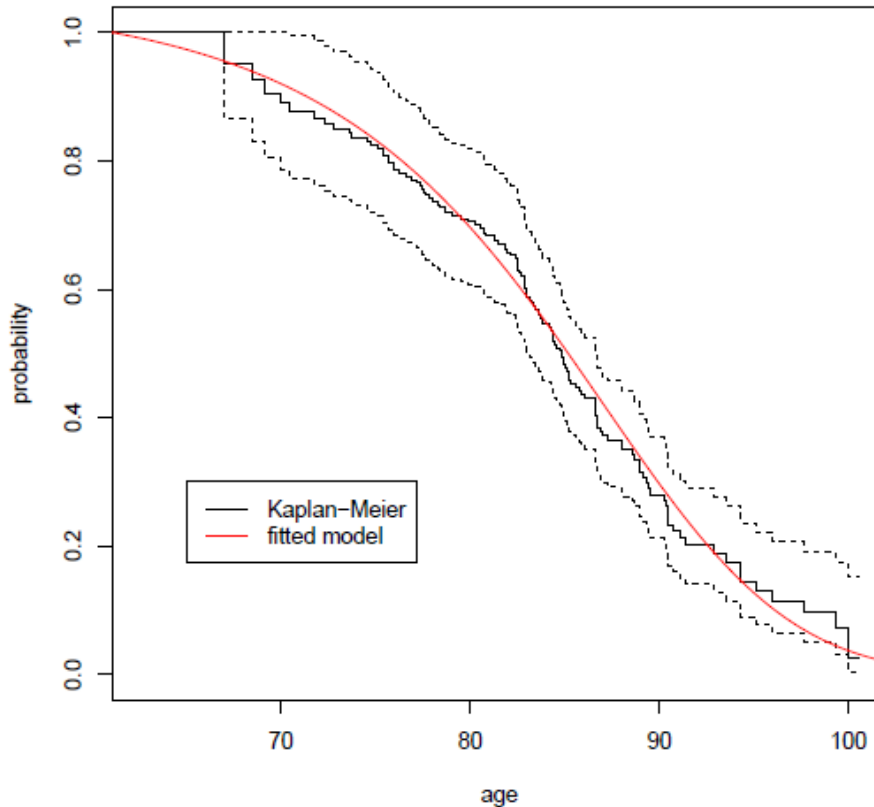
Graphique 4 : Force de la mortalité et logarithme (base 10) de cette force d'après le MVP étalonné à l'aide des données des femmes résidant à Channing House



*En anglais seulement.

Dans le graphique 5, nous illustrons la qualité de l'ajustement de notre modèle aux données de Channing House en traçant la fonction de survie ajustée et les estimations de la fonction de survie non paramétrique de Kaplan-Meier. Nous observons que notre modèle s'ajuste très bien et les estimateurs de notre modèle ajusté respectent les limites de confiance de 95 % d'après l'estimateur de Kaplan-Meier.

Graphique 5 : Fonction de survie d'après le MVP étalonné à l'aide des données sur les femmes de Channing House, de même que les estimations de Kaplan-Meier sur la fonction de survie et les limites correspondantes de confiance de 95 % (pointillés)



*En anglais seulement.

À l'aide de notre modèle ajusté, nous pouvons effectuer une série d'analyses du processus de vieillissement. Par exemple, il est intéressant d'examiner la distribution de l'état d'une personne à des âges différents, compte tenu du fait que la personne est en vie à ces âges. Les probabilités de l'âge physiologique associées à cette distribution, soit $\Pr(Y_t = i | Y_t \in E)$, sont données par le vecteur

$$\mathbf{p}(t)/\mathbf{p}(t)e,$$

où $\mathbf{p}(t)$ peut être calculé à l'aide de l'équation (6). Puisque ces probabilités correspondent à l'âge $50 + t$ dans notre exemple, il est intuitivement intéressant de transformer l'état Y_t en une valeur comparable que nous pouvons interpréter comme l'âge physiologique de la personne.

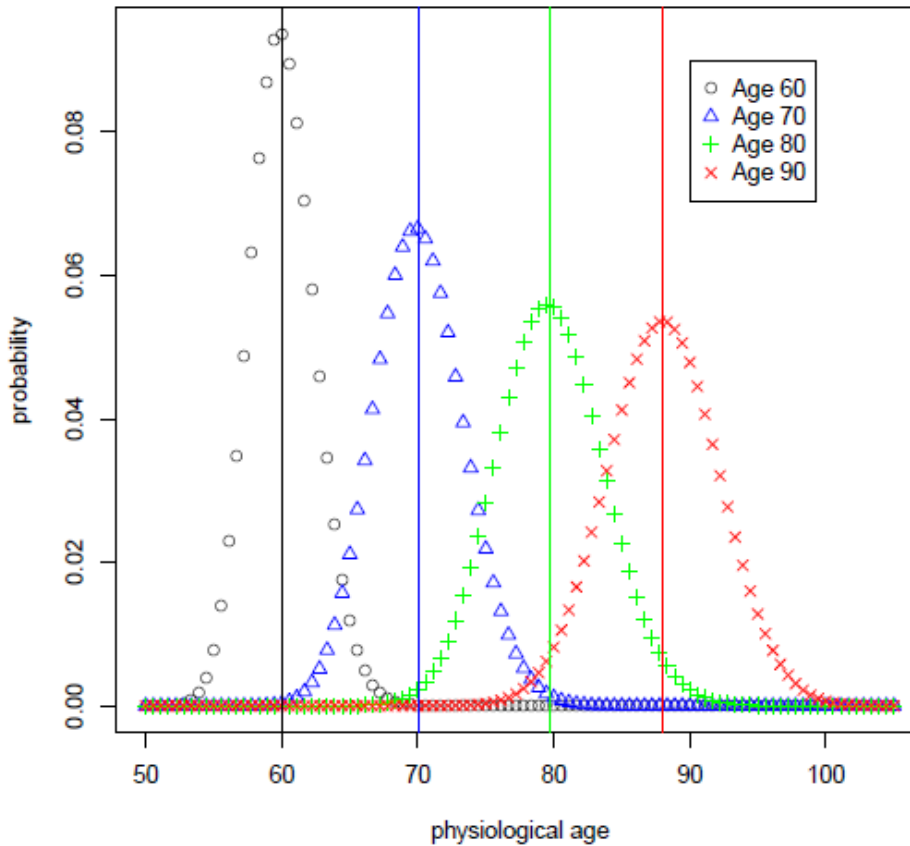
Supposons

$$\hat{\text{Age physiologique}} \text{ à l'âge civil} = 50 + \frac{Y_t - 1}{m - 1} \psi.$$

alors, une personne à l'état 1 a un âge physiologique de 50 ans, et une personne à l'état 100 a un âge physiologique de $50 + \frac{100-1}{99} (55) = 105$. Nous pouvons maintenant déterminer la distribution de l'âge physiologique à n'importe quel âge civil.

Le graphique 6 montre la distribution de l'âge physiologique à 60, 70, 80 et 90 ans pour le MVP étalonné à l'aide des données sur les femmes de Channing House. Nous observons que la distribution de l'âge physiologique est moins variable aux âges moins avancés. Cette situation est principalement attribuable au fait que nous avons amorcé le processus à 50 ans, de sorte qu'à 50 ans, toutes les personnes ont un âge physiologique de 50 ans. Aux âges plus avancés, la variabilité se stabilise. Nous observons également que l'âge physiologique, indiqué par des lignes verticales, est très proche de 60 à 60 ans. Toutefois, aux âges plus avancés, l'âge physiologique moyen diminue nettement par rapport à l'âge civil. Cette situation est imputable à la sélection de la mortalité. Les personnes dont l'âge physiologique est plus avancé présentent un taux de mortalité plus élevé. Par conséquent, les survivants sont davantage ceux dont l'âge physiologique est moins élevé.

Graphique 6 : Distribution de l'âge physiologique à 60, 70, 80 et 90 ans d'après le MVP étalonné à l'aide des données sur les femmes de Channing House. Les lignes verticales indiquent la moyenne des distributions

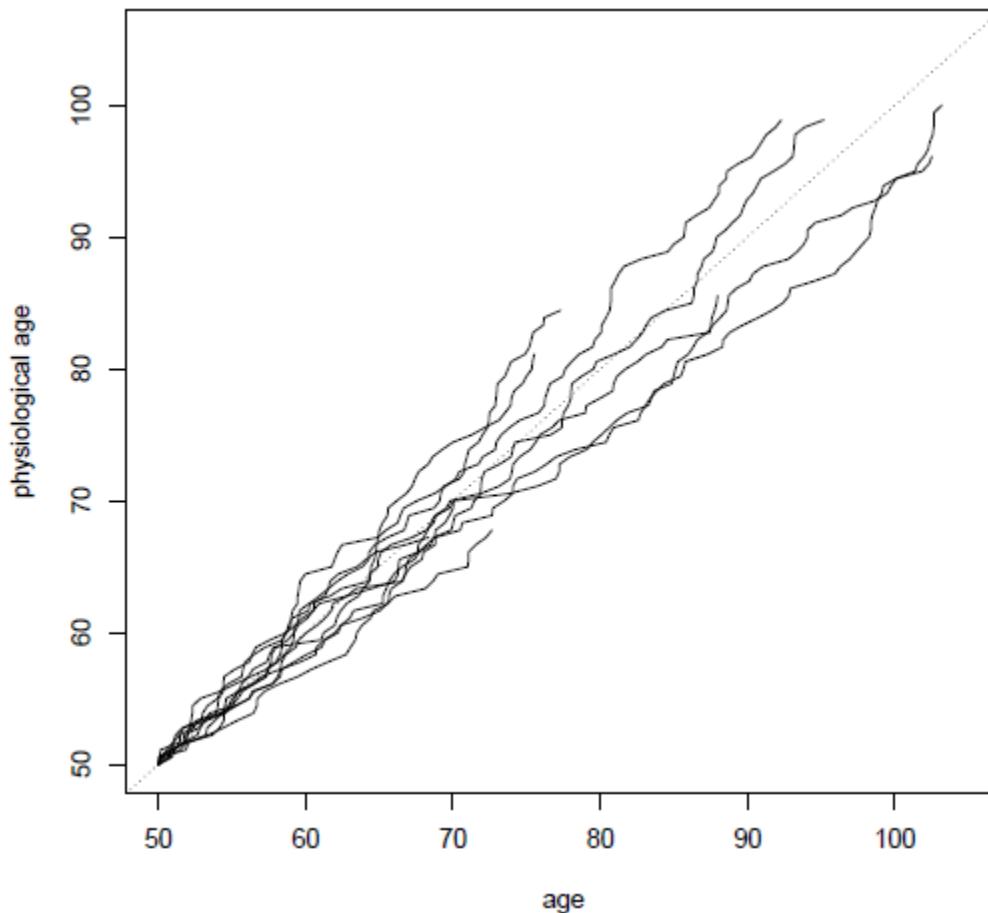


*En anglais seulement.

Nous pouvons constater une autre perspective de notre modèle de vieillissement en observant le comportement des trajectoires du processus. Nous pouvons y parvenir en simulant et en traçant plusieurs trajectoires. Pour simuler une trajectoire, nous supposons que la personne est à l'état 1 (âge physiologique de 50 ans) à 50 ans. Cette personne demeurera à l'état 1 pendant une période distribuée de façon exponentielle à un taux de $\lambda + h_1$. Nous pouvons générer cette variable aléatoire exponentielle. À la fin du séjour à l'état 1, la personne décèdera avec une probabilité de $h_1/(\lambda + h_1)$ et elle passera à l'état 2 avec une probabilité de $\lambda/(\lambda + h_1)$. Nous pouvons générer une variable aléatoire uniforme (0,1) pour déterminer si la personne décèdera à ce moment. Dans la négative, nous générons le temps passé à l'état 2 (variable exponentielle avec un taux de $\lambda + h_2$) et nous continuons jusqu'au décès de la personne.

Nous avons simulé dix trajectoires du processus et nous les avons tracées dans le graphique 7. Encore une fois, l'état du processus a été transformé en âge physiologique comme il est décrit ci-dessus. Le graphique illustre encore une fois la variabilité de l'âge physiologique à divers âges, mais elle montre également la mesure dans laquelle chaque trajectoire se détache du vieillissement uniforme au fil du temps, c'est-à-dire que nous observons la nature « ondulée » de trajectoires individuelles.

Graphique 7 : Dix trajectoires simulées à partir de 50 ans jusqu'au décès d'après le MVP étalonné à l'aide des données sur les femmes de Channing House



*En anglais seulement.

4. Autre analyse de notre MVP proposé

En modélisation, nous avons pour principe de base que la (durée de la) vie est le résultat de deux forces importantes : la force du vieillissement (c.-à-d. le processus biologique intrinsèque) et la force de la mort (c.-à-d. les forces externes du stress environnemental, y compris les accidents, l'accessibilité de la nutrition et des soins médicaux, etc.). En outre, nous souhaitons modéliser deux forces à l'aide d'une approche « stochastique »; plus précisément, nous adoptons un cadre de modèle markovien d'une distribution de type coxien. Nous définissons un nombre fini d'états représentant différents niveaux de capacité physiologique et d'intensité de transition qui déterminent le taux de progression à travers les états. Ces états (et leur étiquetage correctement transformé) sont interprétés comme l'âge physiologique d'une personne, car ils sont utilisés pour classifier la capacité physiologique d'une personne au moment choisi. Ainsi, nous avons naturellement intégré la notion d'hétérogénéité au processus dynamique de durée de vie en introduisant le soi-disant « *âge physiologique* ».

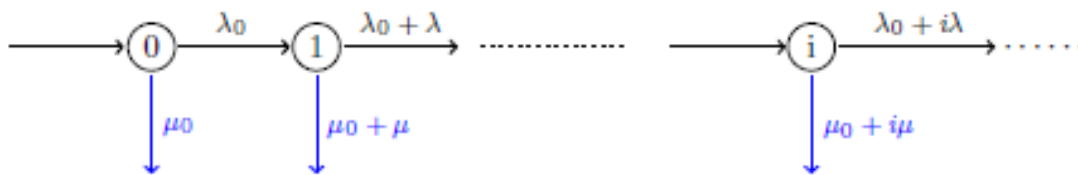
Nous avons mentionné à la section 1 que, depuis le milieu du XX^e siècle, de nombreux chercheurs ont commencé à recueillir des données longitudinales sur diverses variables physiologiques. La constatation la plus frappante réside dans le fait que le recul de ces fonctions physiologiques à la suite du processus sous-jacent du vieillissement suit une trajectoire à la fois lente, uniforme et passablement linéaire au fil de l'âge, ce qui laisse entendre que selon notre hypothèse concernant un taux uniforme d'augmentation de l'âge physiologique, l'effet du vieillissement peut être exprimé comme une fonction croissante de l'âge physiologique.

Malgré l'attrait de notre approche, d'autres stratégies peuvent servir à construire une distribution de type coxien. Nous abordons l'une de ces stratégies ci-après et nous montrons que notre modèle est presque équivalent sur le plan de la distribution de la durée de vie qui en découle.

4.1. Le modèle markovien linéaire double de Le Bras

Nous décrivons maintenant un modèle élaboré par Szilard (1959) et Le Bras (1976). Szilard propose une théorie portant sur le processus du vieillissement, qui suppose que les chromosomes subissent des mutations avec des taux constants dans les cellules. Si une cellule est soumise à un trop grand nombre de mutations, elle cesse de fonctionner. Lorsqu'un certain pourcentage de cellules cessent de fonctionner dans le corps humain, celui-ci meurt. De plus, Le Bras a ajouté une hypothèse selon laquelle la mutation chromosomique héréditaire dans le corps humain suit un processus de Markov. Pour un nouveau-né, chaque cellule fluctue avec le taux λ_0 au départ. Ensuite, de nouvelles mutations se produisent avec un taux $i\lambda$ supplémentaire, proportionnel au nombre total de mutations qui se sont produites dans cette cellule, et chaque cellule meurt avec un taux $i\mu$, également proportionnel à i . Ainsi, l'état i d'un processus de Markov représente l'état dans lequel les cellules ont accumulé des mutations i au total. Ensuite, le taux de transition de l'état i à l'état $i + 1$ est $\lambda_0 + i\lambda$ et le taux de transition de l'état i à l'état absorbant, c'est-à-dire au décès, est $i\mu$.

Graphique 8: Diagramme du modèle linéaire double de Le Bras



Le modèle de Le Bras a été abordé par Yashin et coll. (1994) dans la forme suivante : $\lambda_i = \lambda_0 + i\lambda$ et $\mu_i = \mu_0 + i\mu$, où μ_0 est le taux initial de sortie. Le diagramme du modèle est présenté dans le graphique 8. Ici, le modèle Le Bras a un nombre infini d'états. En suivant les étapes de la section 2, on peut calculer la probabilité de transition

$$P_i(t) = \frac{e^{-(\mu_0 + \lambda_0)t}}{i!} \left(\frac{\lambda(1 - e^{-(\lambda + \mu)t})}{\lambda + \mu} \right)^i \prod_{k=1}^i \left(\frac{\lambda_0}{\lambda} + k - 1 \right). \quad (14)$$

Compte tenu du fait que $P_i(t)$ suit un modèle de série binomiale, il est possible de vérifier que

$$S(t) = \sum_{i=0}^{\infty} P_i(t) = e^{-(\lambda_0 + \mu_0)t} \left(\frac{\lambda + \mu}{\mu + \lambda e^{-(\lambda + \mu)t}} \right)^{\frac{\lambda_0}{\lambda}} \quad (15)$$

Le modèle de Le Bras est l'une des premières tentatives qui ont permis d'intégrer des hypothèses physiques/biologiques à un cadre mathématique. Il est donc important de revoir son approche pour nous aider à mieux comprendre comment notre modèle proposé peut être utile.

Autres remarques au sujet de l'approche de Markov et du modèle de Le Bras :

- En supposant également que $\mu \ll \lambda$, la fonction de risque du modèle de Le Bras peut faire l'objet d'une approximation par

$$\mu(t) = \left(\mu_0 - \frac{\mu\lambda_0}{\lambda} \right) + \frac{\mu\lambda_0}{\lambda} e^{(\lambda + \mu)t},$$

qui équivaut au modèle de mortalité à trois paramètres de Gompertz-Makeham, $\mu(t) = a + be^{ct}$, avec

$$a = \mu_0 - \frac{\mu\lambda_0}{\lambda}, \quad b = \frac{\mu\lambda_0}{\lambda}, \quad c = \lambda + \mu.$$

Pour plus de détails, voir Yashin et coll. (1994). L'importante contribution du modèle de Le Bras est la double structure linéaire qui décrit le processus de vieillissement (le modèle linéaire pour λ_i) et l'effet de détérioration (le modèle linéaire pour μ_i) qui pourrait donner lieu à une forme Gompertz d'augmentation exponentielle du modèle de mortalité.

- Dans Yashin et coll. (1994), les auteurs discutent du modèle Le Bras. On a constaté qu'à partir d'une hypothèse de fragilité fixe, il est possible d'en arriver au même modèle de mortalité. On a donc fait valoir que, dans l'analyse statistique des données, les résultats et les conclusions dépendent non seulement des données, mais aussi des hypothèses de base sur le mécanisme qui a généré les données. En d'autres termes, en réalité, l'utilisation des données sur la durée de vie ne suffit pas à elle seule à faire la distinction entre les différents mécanismes qui produisent les modèles de mortalité observés. Des données plus poussées doivent être utilisées pour valider l'hypothèse.
- Comme nous en avons discuté précédemment, les gens meurent pour de nombreuses raisons, et la majorité des causes sont en fait externes. Il est donc difficile d'accepter un modèle (p. ex. le modèle de Le Bras) qui explique le vieillissement comme seule cause de décès. Un autre phénomène moderne de la vie dans de nombreux pays développés réside dans le fait que les soins médicaux avancés permettent maintenant aux gens ordinaires de vivre beaucoup plus longtemps dans un état physique fragile qu'auparavant. Nous aimerions que notre modèle puisse tenir compte de la différence entre la force intrinsèque du vieillissement et la force externe de la mort.

Pour toutes ces raisons, nous préférons la structure de notre modèle à celle du modèle Le Bras.

4.2. Une étude de simulation – Passage du modèle de Le Bras au MVP

Nous souhaitons déterminer si notre modèle MVP peut saisir les caractéristiques probabilistes du modèle Le Bras.

Pour étudier cette question, nous avons simulé 5 000 observations au cours de la vie à partir du modèle Le Bras avec les paramètres indiqués au tableau 1. Nous ajustons ensuite notre modèle aux données simulées. Avant d'estimer les autres paramètres à l'aide de la méthode d'EMV, nous définissons le paramètre de durée de vie ψ comme suit :

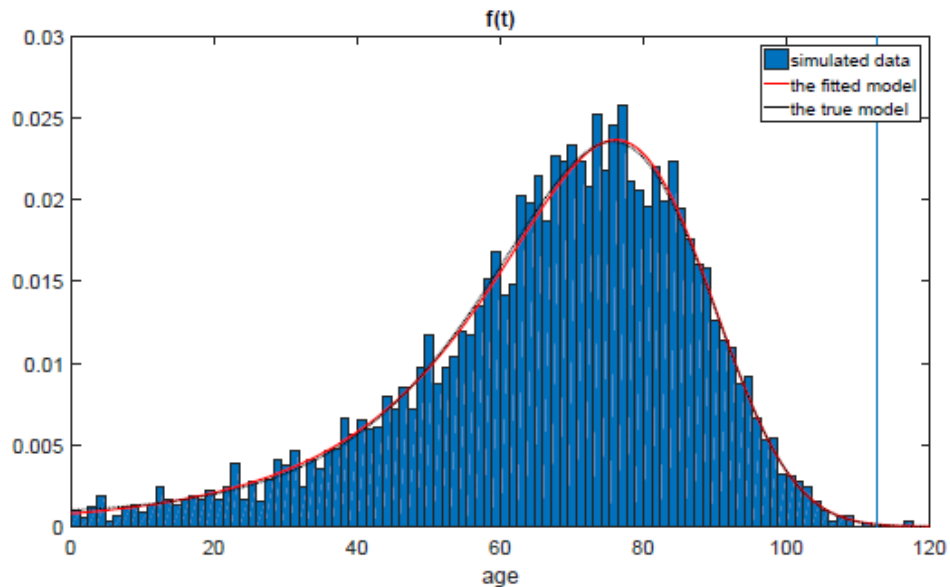
$$\psi = \widehat{TVaR}_{0,999}(T) \quad (16)$$

où $\widehat{TVaR}_{1-\alpha}(T)$ est une estimation empirique de $TVaR_{1-\alpha}(T)$, obtenue à partir des données simulées. Puisque notre échantillon simulé est de 5 000, $\widehat{TVaR}_{0,999}(T)$ est la moyenne des cinq observations les plus importantes. Nous obtenons $\psi = \widehat{TVaR}_{0,999}(T) = 112,55$. Nous estimons les autres paramètres au moyen de la méthode d'EMV. Le graphique 9 présente un histogramme des données simulées, la fonction de densité de probabilité (fdp) du modèle Le Bras et la fdp du MVP ajusté. L'EMV de m est déterminée en ajustant le modèle pour des valeurs m différentes, mais fixes à l'aide de la méthode d'EMV et en comparant leurs valeurs respectives de log-vraisemblance négative (LVN). Comme l'indique le tableau 2, lorsque $m = 225$, nous obtenons la valeur de LNV la plus faible.

Tableau 1 : Valeur des paramètres

Le modèle de Le Bras	λ_0	λ	μ_0	μ
	0,6	0,07	0,001	$0,4 \times 10^{-4}$
Notre MVP ajusté	h_1	h_m	λ	s
avec $m = 225$	0,0008	1,65349	1,99908	-0,11118

Graphique 9 : Histogramme de 5 000 durées de vie simulées à partir du modèle de Le Bras. Le modèle ajusté avec $m = 225$ est tracé de concert avec le modèle réel. La ligne verticale pointillée indique l'emplacement de $\psi = 112,55$



*En anglais seulement.

Le graphique 9 montre que si les taux de mortalité observés sont réellement générés à partir du modèle de Le Bras, notre modèle peut offrir une représentation presque équivalente. Nous sommes en mesure d'obtenir la même qualité d'ajustement que le modèle de Le Bras en permettant au taux de sortie h_i d'augmenter plus rapidement que de façon linéaire. Le fait que s peut être souple pour prendre n'importe quelle valeur afin de tenir compte des données (dans cet exemple $s = -0,11118$) est une caractéristique particulière de notre modèle.

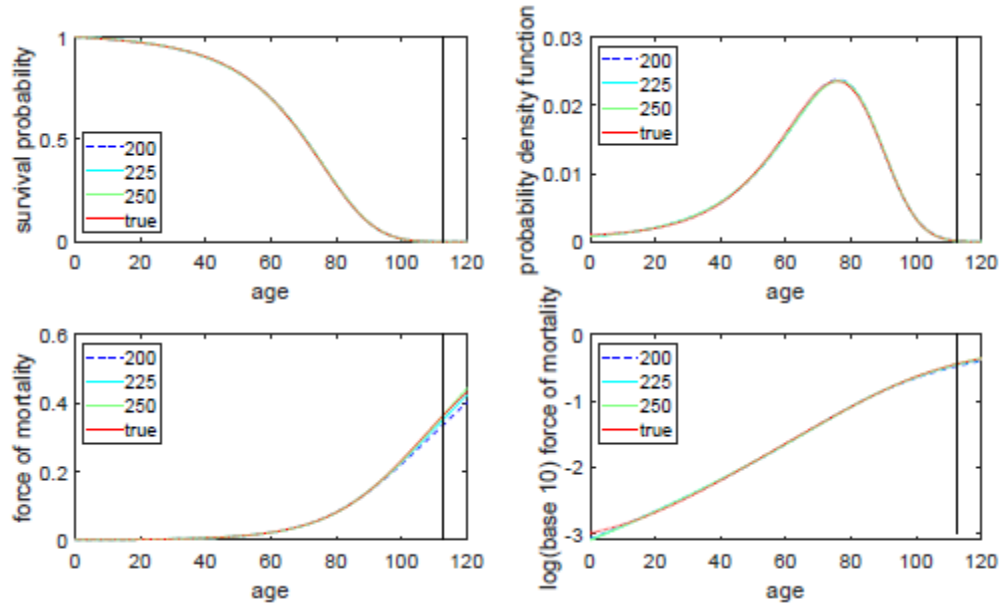
Nous avons donc trouvé une solution de rechange au modèle linéaire double de Le Bras avec un taux de transition λ constant, bien que notre modèle comporte une interprétation légèrement différente du mécanisme sous-jacent du vieillissement. Même si le modèle original de Le Bras contient un nombre infini d'états, notre MVP les convertit en états m . Le processus de vieillissement est encore modélisé et il se déplace d'un état à l'autre, et l'incidence du vieillissement est encore décrite comme une fragilité accrue au risque avec un état à indice plus élevé. La seule différence est la façon d'étiqueter les états. Dans le cadre du passage du modèle de Le Bras à notre MVP, les états antérieurs pourraient devoir être divisés en un plus grand nombre d'états, tandis que les états ultérieurs pourraient devoir être regroupés, mais pas de façon linéaire, ce qui a pour effet global que, dans notre cadre modèle, les transitions doivent se faire uniformément en raison de l'utilisation du taux de transition λ constant. Par conséquent, le modèle des variations h_i passe d'une augmentation linéaire à un modèle qui doit augmenter légèrement plus rapidement que la forme exponentielle, comme l'indique le paramètre s , qui prend une faible valeur négative.

Dans le tableau 2, nous présentons d'autres résultats, à partir de l'ajustement de notre modèle jusqu'aux données simulées. Au besoin, l'estimation du paramètre λ augmente de m . Nous avons également tracé les fonctions de survie, les fonctions de densité et les fonctions de risque de ces modèles ajustés, m allant de 200 à 250 dans le graphique 10; la ligne verticale pointillée indique l'emplacement de $\psi = 112,55$. Pour les âges jusqu'à cette valeur, les distributions sont très proches l'une de l'autre. Toutefois, pour les fonctions de risque, il existe des différences minimes, mais dignes de mention à partir de près de 100 ans.

Tableau 2 : Résultats d'estimation à l'aide de valeurs m différentes selon 5 000 durées de vie simulées à partir de la distribution des limites du modèle de Le Bras. La première colonne indique la LVN. La dernière colonne présente la limite de la fonction de risque qui en découle $h(t)$ lorsque $t \rightarrow \infty$.

<i>LVN</i>	h_1	h_m	λ	s	m	$\min(\lambda + h_1, h_m)$
21631,884	0,00081	2,00325	1,77637	-0,12373	200	1,77718
21631,826	0,0008	1,83922	1,86518	-0,11829	210	1,83922
21631,806	0,0008	1,70795	1,95400	-0,11336	220	1,70795
21631,713	0,0008	1,65349	1,99908	-0,11118	225	1,65349
21631,813	0,00079	1,60064	2,04282	-0,10886	230	1,60064
21631,843	0,00078	1,51076	2,13164	-0,10469	240	1,51076
21631,889	0,00078	1,43544	2,22046	-0,10089	250	1,43544

Graphique 10 : Fonction de survie ajustée - $S(t)$, fonction de densité de probabilité - $f(t)$, fonction de risque - $h(t)$ et fonction de risque logarithmique (base 10). Chaque graphique renferme quatre courbes qui correspondent au modèle ajusté avec $m = 200, 225$ et 250 , de même que le modèle véritable. La ligne verticale pointillée indique l'emplacement de $\psi = 112,55$.



*En anglais seulement.

En examinant de plus près les fonctions de risque dans la partie inférieure gauche du graphique 10, nous constatons que l'ajustement ne change pas symétriquement à mesure que m s'éloigne de sa valeur optimale. Le modèle avec $m = 250$ s'ajuste mieux que celui avec $m = 200$. En outre, le modèle avec $m = 250$ produit le taux de risque le plus élevé à l'extrémité droite de la durée de vie. Dans le cas des distributions par étapes, il s'agit d'une propriété bien connue :

$$\lim_{t \rightarrow \infty} h(t) = \min_{i=1, \dots, m} \{d_1, d_2, \dots, d_m\},$$

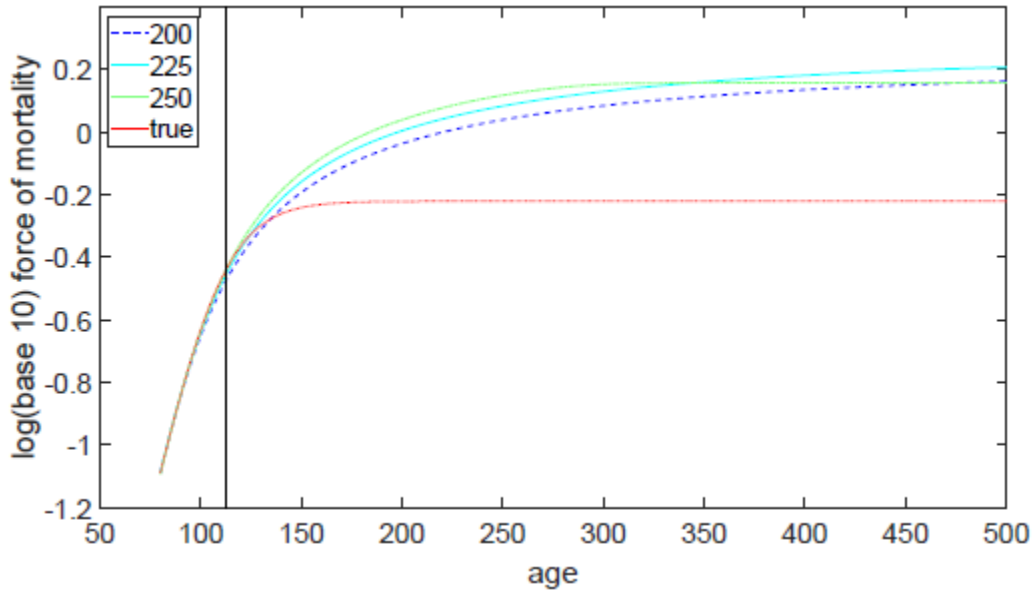
où les d_i sont les valeurs propres de la matrice d'intensité de la transition $\mathbf{\Lambda}$. Dans notre MVP, nous avons

$$\lim_{t \rightarrow \infty} h(t) = \min\{\lambda + h_1, h_m\},$$

car les valeurs propres $\lambda + h_i$ augmentent, sauf dans le cas du dernier élément de la matrice $\mathbf{\Lambda}$. La limite de $h(t)$ pour chaque modèle ajusté est fournie à la dernière colonne du tableau 2, et le modèle avec $m = 250$ présente la valeur la plus faible de cette limite.

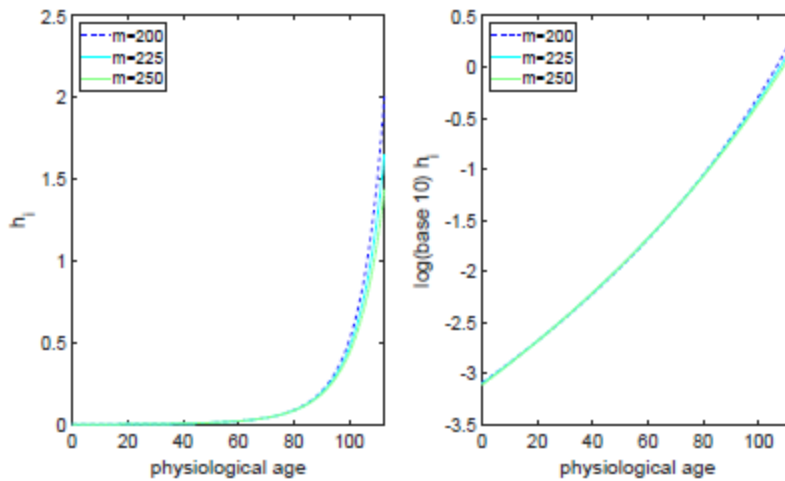
Dans le graphique 11, nous présentons le logarithme de la fonction de risque des modèles ajustés entre 80 et 500. Bien que le graphique montre des différences intéressantes dans le comportement aux extrémités, ces différences se produisent bien au-delà de la tranche d'âge qui est pertinente d'un point de vue pratique.

Graphique 11 : Fonction de risque log (base 10) ajustée, prolongée jusqu'à l'âge de 500 ans. Chaque graphique renferme quatre courbes qui correspondent au modèle ajusté avec $m = 200$, 225 et 250, de même que le modèle véritable. La ligne verticale pointillée indique l'emplacement de $f \psi = 112,55$.



*En anglais seulement.

Graphique 12 : Le graphique de gauche montre le taux de sortie h_i avec $m = 200$, 225 et 250. Le graphique de droite montre le taux log-sortie avec $m = 200$, 225 et 250. Ces deux taux sont tracés par rapport à l'âge physiologique $\frac{i-1}{m-1} \psi$.



*En anglais seulement.

Dans le graphique 12, nous traçons h_i par rapport à l'âge physiologique $\frac{i-1}{m-1} \psi$. Le modèle de h_i sur $[0, \psi]$ s'avère très stable avec différentes valeurs de m . Il s'agit d'une autre bonne caractéristique du modèle proposé, qui indique que le mécanisme sous-jacent ne varie pas trop avec m .

Jusqu'à présent dans notre analyse, nous avons démontré que certains aspects clés de notre modèle du vieillissement sont assez stables selon les variations de m . Notre valeur estimative de m a été déterminée en maximisant la probabilité. Toutefois, la probabilité est assez plate pour des valeurs de m proches du maximum. Il n'y a donc que de petites différences dans la distribution de la durée de vie qui résultent de différentes valeurs de m . Cela nous amène à nous demander dans quelle mesure le modèle est touché par la valeur de m et comment déterminer cette valeur.

Bien que la distribution de la durée de vie soit relativement insensible à m , la progression de l'âge physiologique avec l'âge civil avancé est affectée par m . Plus particulièrement, la variabilité de l'âge physiologique est sensiblement touchée par m . En effet, comme $m \rightarrow \infty$, cette variabilité disparaît et l'âge physiologique converge vers l'âge civil. Plus précisément, lorsque m augmente, l'âge physiologique attendu d'une personne vivante à l'âge civil $t \in (0, \psi)$ approche $\frac{t}{\psi} m$ et la variabilité diminue. Ce phénomène est illustré dans le graphique 13, où un échantillon de dix modèles simulés pour chacune des quatre valeurs différentes de m : 25, 100, 225, et 1 000 est tracé. En effet, nous supposons qu'avec $m \rightarrow \infty$, $\mathbb{E}(Y_t | Y_t \in E) = \frac{t}{\psi} m$ et $Var(Y_t | Y_t \in E) \rightarrow 0$.

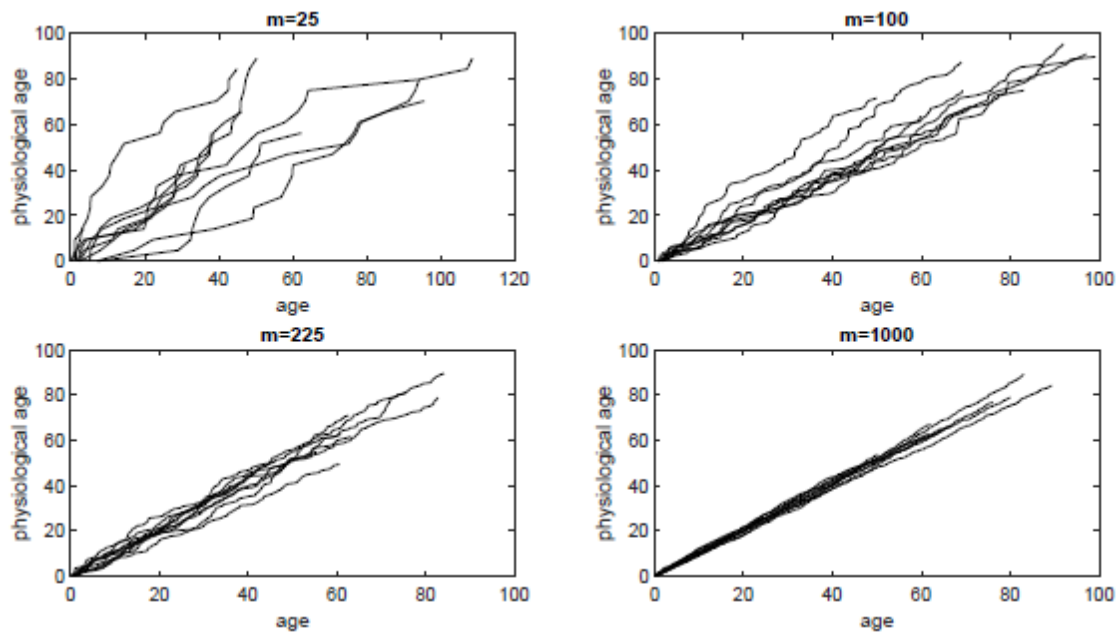
En outre, la fonction de risque qui en découle adopte le comportement de nos valeurs h_i , c'est-à-dire qu'à mesure où m augmente (tout en ne modifiant pas h_1 , h_m et s), nous observons que la fonction de risque, qui représente le taux attendu de sortie, représenté par $h^m(t) = \mathbb{E}(h_{Y_t} | Y_t \in E)$, approche

$$h(t) = \begin{cases} \left(\left(1 - \frac{t}{\psi} \right) h_1^s + \frac{t}{\psi} h_m^s \right)^{1/s} & s \neq 0, \\ h_1^{1-\frac{t}{\psi}} h_m^{\frac{t}{\psi}} & s = 0. \end{cases} \quad (17)$$

Parallèlement, la variabilité de h_{Y_t} s'amenuise de plus en plus à mesure que m s'accroît. Il s'agit d'une propriété intéressante du MVP, dans laquelle le rôle du paramètre m est étroitement lié au niveau d'incertitude relatif à la détermination de l'effet du vieillissement et(ou) de son interprétation.

Le graphique 13 montre un échantillon de dix trajectoires simulées pour chacune des quatre valeurs différentes de m : 25, 100, 225 et 1 000. Nous observons qu'à mesure où m augmente, la variabilité de l'âge physiologique diminue, et l'âge physiologique converge vers l'âge civil. Cela laisse à entendre que m ne devrait pas être trop élevée, sans quoi le modèle ne reflétera pas convenablement la variabilité de l'âge physiologique. Évidemment, m ne peut pas non plus être trop faible, sinon la variabilité de l'âge physiologique sera trop grande.

Graphique 13 : Échantillon de dix trajectoires simulées du MVP ajusté pour chacune des quatre valeurs différentes de m , en supposant que les paramètres h_1 , h_m et s demeurent fixes.



*En anglais seulement.

Cela nous ramène à la question de savoir comment choisir m . Cette valeur peut être choisie pour refléter l'opinion sur la variabilité de l'âge physiologique. Toutefois, si l'on dispose de données qui comprennent de l'information sur une ou plusieurs variables de la santé liées à l'âge physiologique en plus des durées de vie des personnes, on peut les utiliser pour estimer m avec les autres paramètres. Il s'agit de l'approche privilégiée si des données adéquates sont disponibles.

5. Résumé et conclusions

Nous avons présenté un modèle mathématique pour le vieillissement. Ce modèle a une interprétation très naturelle, car on suppose que les gens vieillissent en progressant dans un certain nombre d'états et en passant une période aléatoire dans chaque état. Le modèle est rendu encore plus intuitif en transformant l'étiquette d'état en une échelle qui peut être interprétée comme l'âge physiologique. Comme on pourrait s'y attendre, le taux de mortalité augmente à mesure que l'on passe d'un état à l'autre. Toutefois, la force du vieillissement et la force de la mortalité sont des composantes distinctes du modèle.

Nous supposons que le taux de vieillissement est constant dans le temps, de sorte que les gens progressent uniformément à travers les états. Le modèle des taux de mortalité par état est déterminé à l'aide d'une structure à trois paramètres qui offre une grande souplesse et qui comprend des tendances linéaires et exponentielles comme cas spéciaux. Une fois le nombre d'états du modèle précisé, le taux de vieillissement est déterminé de manière à ce que les gens n'atteignent l'état final que très près de la fin de la vie, qui doit être précisé.

Bien que la forme fonctionnelle choisie pour les taux de mortalité par état soit attrayante, sa souplesse est limitée. Plus précisément, la forme ne peut produire qu'un seul modèle monotone de taux avec une pente constante, croissante ou décroissante. Cette forme simple ne peut donc pas prendre la forme d'une baignoire ou d'une bosse liée à un accident pour ce qui est de la force de la mortalité. Toutefois, nous croyons qu'il pourrait être possible de combiner les fonctions de la forme que nous proposons afin d'accroître la souplesse. Il s'agit d'une piste de recherche future.

Notre modèle peut être étalonné au moyen de données sur la durée de vie à l'aide d'une estimation du maximum de vraisemblance. Toutefois, dans ce cas, nous avons beaucoup d'incertitude quant à l'estimation du nombre d'états du modèle. Cela est logique puisque le nombre d'états détermine la variabilité de l'âge physiologique; par conséquent, nous avons besoin de données contenant des informations pertinentes et explicites sur le processus de vieillissement et ses variations au sein de la population cible. Plus spécifiquement, nous aurions besoin de données longitudinales sur les variables liées à la santé observées, qui sont de bons indicateurs du vieillissement. Au stade actuel, puisque nous n'avons utilisé que les taux de mortalité pour calibrer le modèle, le modèle ajusté ne peut fournir qu'une mesure relative de l'âge physiologique. Sous réserve de la disponibilité des données relatives au vieillissement, le modèle peut être encore amélioré pour décrire les caractéristiques de vieillissement spécifiques à un état. Donc, l'un des domaines de recherche futurs est l'exploration des méthodes d'estimation des paramètres lorsque les données comportent des observations sur la santé et de l'information sur la mortalité. L'approche de Govorun et coll. (2018) pourrait être un bon point de départ.

Un autre domaine de recherche future consiste à étudier la façon de modifier le modèle pour tenir compte de différentes populations ou de différentes cohortes de naissance. Une option consiste à utiliser des taux de vieillissement différents avec des taux de mortalité identiques. Par ailleurs, on pourrait également permettre que les paramètres des taux de mortalité soient différents.

Remerciements

Les auteurs tiennent à remercier sincèrement l'Institut canadien des actuaires pour son appui financier sous forme de subvention dans le cadre de cette recherche.

Références

- Aalen, O. O., « Phase type distributions in survival analysis », *Scandinavian Journal of Statistics*, vol. 22, n° 4, 1995, pp. 447-463.
- Asmussen, S., « Exponential families generated by phase-type distributions and other Markov lifetimes », *Scandinavian Journal of Statistics*, vol. 16, n° 4, 1989, pp. 319-334.
- Asmussen, S., Nerman, O. et M. Olsson. « Fitting phase-type distributions via the em algorithm », *Scandinavian Journal of Statistics*, vol. 23, n° 4, 1996, pp. 419–441.
- Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., Harrington, H., Israel, S., Levine, M. E., Schaefer, J. D., et coll., « Quantification of biological aging in young adults », *Proceedings of the National Academy of Sciences*, vol. 12, n° 30, 2015, pp. E4104-E4110.
- Box, G. E. et D. R. Cox., « An analysis of transformations », *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, n° 2, 1964, pp. 211-252.
- Crimmins, E. M., Johnston, M., Hayward, M. et T. Seeman., « Age differences in allostatic load: an index of physiological dysregulation », *Experimental Gerontology*, vol. 38, n° 7, 2003, pp. 731-734.
- Dublin, L. I., Lotka, A. J. et M. Spiegelman., *Length of Life*, Ronald New York, 1949.
- Govorun, M., Jones, B. L., Liu, X. et D.A. Stanford., « Physiological age, health costs, and their interrelation », *North American Actuarial Journal*, vol. 22, n° 3, 2018, pp. 323-340.
- Herskind, A. M., McGue, M., Holm, N. V., Sørensen, T. I., Harvald, B. et J. W. Vaupel., « The heritability of human longevity: a population-based study of 2872 danish twin pairs born 1870–1900 », *Human Genetics*, vol. 97, n° 3, 1995, pp. 319-323.
- Jones, H. B., « A special consideration of the aging process, disease, and life expectancy », *Advances in Biological and Medical Physics*, n° 4, 1956, pp. 281-337.
- Le Bras, H., « Lois de mortalité et âge limite », *Population*, vol. 31, n° 3, 1976, pp. 655-692.
- Lin, X. S. et X. Liu., « Markov aging process and phase-type law of mortality », *North American Actuarial Journal*, vol. 11, n° 4, 2007, pp. 92-109.
- Neuts, M. F., « Explicit steady-state solutions to some elementary queueing models », *Operations Research*, vol. 30, n° 3, 1982, pp. 480-489.
- Slud, E. V. et J. Suntorncost., « Parametric survival densities from phase-type models », *Lifetime Data Analysis*, vol. 20, n° 3, 2014, pp. 459-480.
- Su, S. et M. Sherris., « Heterogeneity of Australian population mortality and implications for a viable life annuity market », *Insurance: Mathematics and Economics*, vol. 51, n° 2, 2012, pp. 322-332.
- Szilard, L., « On the nature of the aging process », *Proceedings of the National Academy of Sciences*, vol. 45, n° 1, 1959, pp. 30-45.
- Yashin, A. I., Arbeev, K. G., Ukraintseva, S. V., Akushevich, I. et A. Kulminski., « Patterns of aging-

related changes on the way to 100 », *North American Actuarial Journal*, vol. 16, n° 4, 2012, pp. 403-433.

Yashin, A. I., Vaupel, J. W. et I. A. Iachine., « A duality in aging: the equivalence of mortality models based on radically different concepts », *Mechanisms of Ageing and Development*, vol. 74 (1-2), 1994, pp 1-14.