

La surmortalité des jeunes adultes : réalité individuelle ou artefact des inégalités sociales ?

REMUND Adrien*

■ Résumé

La surmortalité des jeunes adultes est un phénomène démographique largement connu mais encore mal compris. Les théories explicatives qu'il a suscitées suivent souvent une interprétation endogène, inspirées des théories (neuro-)psychologiques de l'adolescence, supposant une hausse temporaire du risque de décès individuel au cours de la transition à l'âge adulte. Or, la bosse de surmortalité observée sur le taux de décès par âge agrégés pourrait alternativement être due à la présence de sous-populations non-observées engendrant une modification progressive de la composition de la cohorte sous l'effet de la force de sélection. Cette contribution propose de formuler un modèle statistique à fragilité discrète cohérent avec cette seconde hypothèse, de manière à tester sa pertinence. Ce type de modèle étant plus difficile à estimer que les autres modèles de mortalité homogènes, deux algorithmes d'optimisation sont comparés afin de ne conserver que le plus efficace. Enfin, ce modèle est appliqué au cas réel des cohortes nées en Suisse entre 1876 et 1981. Cela permet notamment de conclure que dans les cohortes de naissance les plus récentes, il suffit de supposer que la sous-population vulnérable couvre moins d'un pourcent de la cohorte totale pour générer au niveau agrégé une bosse de surmortalité telle qu'observée dans la réalité. Ce résultat suggère que la surmortalité des jeunes adultes pourrait parfaitement être le résultat de la présence de fortes inégalités de mortalité, plutôt que d'une augmentation du risque de décès commune à tous les individus.

Introduction

La conception dominante de la surmortalité des jeunes adultes, en démographie, est celle d'un phénomène individuel lié au développement psychologique propre à l'adolescence. Celle-ci serait naturellement tourmentée marqué par de nombreux conflits avec les pairs, la famille et les jeunes eux-mêmes. Il s'agirait donc d'un processus de nature individuelle qui engendrerait notamment une hausse temporaire du risque de décès par causes externes et qui, de manière agrégée, s'exprimerait par une bosse de surmortalité au moment de l'entrée à l'âge adulte.

Dès l'ouvrage fondateur de Hall (1904), certains comportements ont été identifiés comme typiques de l'adolescence, notamment un pic de dépression et de criminalité, une tendance à rechercher les sensations fortes et une sensibilité aux médias et aux pressions sociales (Arnett, 2007 : 87). Plus tard, le développement de la psychanalyse a renforcé l'idée de crise d'adolescence qui serait due à la résurgence de conflits œdipiens (Freud, 1968, Perret-Catipovic et Ladame, 1997). Pour certains psychanalystes,

* Université de Genève, Institut d'études démographiques et du parcours de vie et Pôle de recherche national LIVES.

l'absence de conflit dans cette période de la vie serait inquiétante (Blos, 1962 : 52), voire pathologique (Freud, 1968).

Plus récemment, cette perception de l'adolescence a été renouvelée dans la littérature neuropsychologique par la théorie du *Cerveau Adolescent* (Adolescent Brain). D'après ses promoteurs, il y aurait des signes neurologiques d'une phase de développement particulière dans le cerveau des adolescents, notamment une évolution en cloche de la matière grise dans le lobe frontal, pariétal et temporal (Giedd, *et al.*, 1999), qui correspondrait à un processus de remodelage synaptique. Certains auteurs ont souligné le fait que ces changements neurologiques sont localisés dans des régions du cerveau connues pour être liées au contrôle des impulsions, au jugement et à la prise de décision (Lenroot et Giedd, 2006 : 723). D'autres ont également suggéré que les adolescents seraient moins capables d'estimer les risques à cause du développement désynchronisé de leurs capacités cognitives et émotionnelles (Steinberg, 2005). La rencontre des théories psychanalytiques de l'adolescence et de ces observations neurologiques a renforcé une conception radicale du comportement adolescent qui a amené certains à affirmer de manière assez brutale que « les adolescents sont des preneurs de risque » (Spear, 2000 : 421). Selon ce dernier auteur, cette spécificité du comportement adolescent s'étend même au-delà de l'espèce humaine, comme chez certains singes ou rats.

Prises ensemble, ces théories ont posé trois postulats selon lesquels l'adolescence est (1) une phase particulière de la vie, (2) un phénomène universel et (3) systématiquement problématique (Sercombe, 2010). De plus, il s'agirait d'une étape de vie universelle car endogène au développement de chaque personne. Bien que ce point de vue ait été critiqué à de multiples reprises par d'autres psychologues (Arnett, 1999, Bessant et Watts, 2012, Males, 2009, Offer et Schonert-Reichl, 1992, Petersen, 1993) et des anthropologues (Choudhury, 2010, Dasen, 2000, Mead, 1928), il a conservé une certaine popularité, non-seulement aux yeux du public, mais également parmi les professionnels de la santé et les chercheurs d'autres disciplines (Offer, *et al.*, 1981).

En démographie, où l'intérêt pour les jeunes adultes s'est surtout exprimé à travers le concept de transition à la vie adulte, les études de mortalité s'en sont souvent tenues à la description de ce qui est souvent appelé bosse des accidents (accident hump) avec la diffusion du modèle de Heligman et Pollard (1980). Cette expression réductrice sous-entend que la bosse de surmortalité des jeunes adultes serait la conséquence d'une prédisposition aux accidents dont les causes ne sont pas explicitement exprimées mais s'apparentent à l'hypothèse endogène décrite ci-dessus. En termes démographiques, cela implique que tous les individus passent par une phase d'augmentation de la prise de risque, d'agitation émotionnelle et de conflit social due à leur processus de maturation cérébrale. Ces évolutions auraient pour conséquence d'augmenter temporairement leur risque individuel de décès (figure 1).

Cette hypothèse est par exemple dominante dans une récente étude de l'évolution séculaire de la maturité sexuelle masculine, justifiée par l'allégation que « la prise de risque et la surmortalité (la « bosse des accidents ») sont les signatures des jeunes hommes » (Goldstein, 2011 : 1). Par son caractère généralisateur, cette affirmation sous-entend que la bosse observée dans la mortalité des jeunes adultes est due au fait que tous les individus passent par une phase d'excès de risque de décès. Cela correspond à ce que Vaupel et Yashin qualifient d'explication de niveau 1, dans laquelle « le changement observé est produit par un changement correspondant au niveau individuel » (Vaupel et Yashin, 2006 : 272).

Pourtant, la nature endogène et universelle de la surmortalité des jeunes adultes ne peut pas être testée puisque le risque individuel de décès est latent. En effet, on ne peut observer directement pour un individu qu'une condition dichotomique (vivant ou mort) et non une propension latente à décéder. Seules des mesures agrégées, comme les taux de mortalité par âge, peuvent être construites, mais cela implique de faire l'hypothèse que la population observée est homogène, c'est-à-dire que le taux moyen s'applique à tous les individus de la même manière. En suivant cette approche, nous avons montré (Remund, 2015) que, contrairement à ce qu'implique l'hypothèse endogène, la surmortalité des jeunes adultes

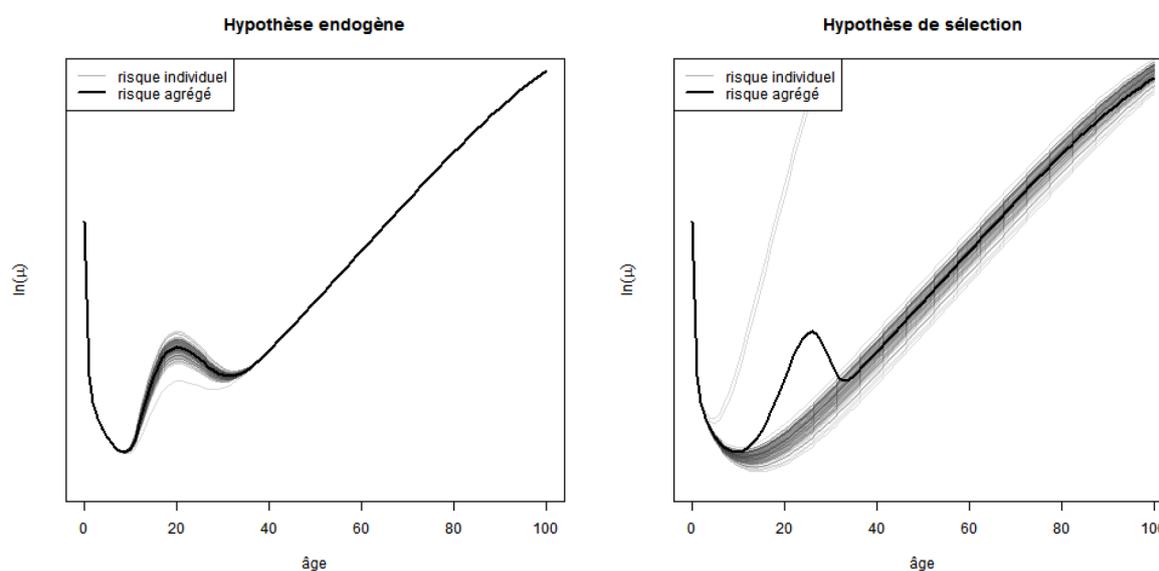
1. n'est pas un phénomène universel commun à toutes les populations humaines ;
2. s'étend bien après les âges identifiés comme neurologiquement critiques (12 à 16 ans) ;
3. n'est pas systématiquement due à des causes de décès externes.

Les résultats de cette première analyse remettent en cause l'hypothèse endogène. Premièrement, la surmortalité des jeunes adultes n'est pas un phénomène universel, près de 25 % des populations féminines et 10 % des populations masculines n'ont pas montré, dans le passé, de bosse de surmortalité significative. Ces périodes sans hausse de la mortalité ont été observées dans des contextes historiques particuliers (les décennies 1950 à 1970 pour les femmes, la cohorte de 1930 pour les hommes). Deuxièmement, le pic de surmortalité est presque tout le temps situé à 20 ans, et la bosse s'étend souvent au-delà de 40 ans, excluant le développement neurologique comme principale variable explicative de la surmortalité des jeunes adultes. Troisièmement, une analyse des causes de mortalité montre que la soi-disant bosse des jeunes a en fait été causée par la tuberculose pulmonaire avant la Seconde Guerre mondiale, et n'a été associée aux accidents (de circulation) et aux suicides que dans les dernières décennies.

Cet échec de l'hypothèse endogène amène à considérer une explication alternative à la surmortalité des jeunes adultes, ne reposant pas sur l'idée que tous les individus passent par une phase d'augmentation de leur probabilité de décès. Cela requiert de considérer la possibilité que, dans cette tranche d'âge, la population ne soit pas homogène. L'hétérogénéité non-observée peut en effet produire au niveau agrégé des évolutions qui n'existent pas au niveau individuel. En fonction de la composition des sous-populations présentes et de leurs risques de décès respectifs, divers effets sont envisageables notamment lié à des effets de structure de la population (Vaupel et Yashin, 1985). En effet, ce type d'effet de composition correspond à ce que Vaupel appelle l'explication de niveau 2, c'est-à-dire que « le changement observé est un artefact d'un changement dans la structure de la population (c'est-à-dire un changement dans la composition d'une cohorte hétérogène) » (Vaupel et Yashin, 2006 : 272). Selon cette approche, il est possible d'expliquer l'existence d'une bosse de surmortalité par un effet de sélection agissant de manière inégale entre différentes sous-populations plus ou moins vulnérables (figure 1).

L'objectif de cette étude est d'évaluer la plausibilité de cette hypothèse de sélection, en explicitant d'abord ses conditions, prémisses sur lesquelles il sera possible de concevoir un modèle de mortalité tenant compte de la présence de plusieurs sous-populations, et enfin en appliquant ce modèle à une population réelle afin d'évaluer son pouvoir explicatif. Ensemble, ces étapes montreront dans quelle mesure et dans quelles conditions l'hypothèse de sélection est capable de reproduire la bosse de surmortalité observée dans des populations réelles, permettant ainsi d'évaluer sa pertinence.

Figure 1. Représentation schématique de deux hypothèses de génération de la surmortalité des jeunes adultes



Conception du modèle

En démographie, il est habituellement d'usage de considérer les populations étudiées comme homogènes, c'est-à-dire de supposer que les individus d'une cohorte sont égaux dans leur risque de décès. Cette approche est par exemple utilisée lors du calcul de tables de mortalité ou d'autres indicateurs de mortalité. Or, depuis la fin des années 1970, ce paradigme théorique a été progressivement bousculé¹. La publication fondatrice de ce mouvement est probablement celle de Vaupel, Manton et Stallard (1979), qui distingue la force de mortalité individuelle (μ_i) de celle observée sur une cohorte ($\bar{\mu}$). Depuis, de nombreuses publications ont repris et étendu cette approche, utilisant pour la plupart la même conception : une distribution continue de la fragilité individuelle (z_i), dont l'effet est supposé proportionnel. Le risque individuel de décès évolue donc avec l'âge de la manière suivante.

$$\mu_i(x) = z_i \cdot \mu(x)$$

Où z_i est un coefficient multiplicatif, par définition latent donc non-observé, indiquant la probabilité relative de décéder de l'individu i par rapport à un individu standard, et $\mu(x)$ décrit l'évolution commune du risque de décès avec l'âge (Vaupel, *et al.*, 1979 : 440). Notons que dans ce modèle, la fragilité d'un individu est considérée comme constante au cours de la vie, une hypothèse indispensable sous peine de ne pas pouvoir distinguer l'effet de la distribution de z_i et $\mu(x)$. Lorsque la fragilité est modélisée en supposant des risques proportionnels comme ci-dessus, la distribution de z_i est habituellement continue et issue d'une distribution gamma. D'autres classes de distribution ont été étudiées (Finkelstein et Esau-

¹ Cette notion est déjà présente chez certains auteurs dès le début du XX^e siècle, par exemple chez Greenwood qui écrit : « we determine from the data the mean force of mortality experienced by a group of N persons, we find this mean value increases geometrically with age, [but] is this consistent with the individual hypothesis ? » (Greenwood, 1928). Cependant, la formalisation mathématique de cette question n'apparaît qu'à la fin des années 1970.

Iova, 2006, Missov et Finkelstein, 2011), mais la distribution gamma reste toutefois la distribution de référence.

Ce double intérêt d'une part sur les modèles à fragilité continue et d'autre part sur les modèles à risques proportionnels, a occulté d'une certaine manière la variété des distributions possibles. Ce choix s'explique probablement par un intérêt particulier pour la mortalité des personnes âgées, et en particulier les effets de plateau de mortalité observés chez les centenaires (Gampe, 2010, Horiuchi et Wilmoth, 1998).

Pourtant, la variété des modèles théoriques possibles est infinie. En particulier, une distribution discrète de la fragilité conjuguée à une divergence des risques de décès autorisent l'existence d'une bosse de surmortalité semblable à celle que l'on observe dans certaines populations humaines. Ce phénomène s'explique de la manière suivante : dans les premières années de vie, la force de mortalité moyenne est tirée vers le haut par la sous-population la plus vulnérable, mais rapidement les membres de cette dernière subissent un risque de décès tel que son effectif se réduit radicalement et la force de mortalité moyenne converge alors vers celle la sous-population la plus robuste. Au cours de cette évolution, la force de mortalité moyenne suit temporairement une trajectoire descendante, créant au niveau agrégé une bosse de surmortalité qui n'existe pas au niveau individuel.

Traiter la fragilité comme une variable discrète déterminant l'appartenance à certaines sous-populations revient à traiter la force de mortalité agrégée comme une distribution mélangée. En statistique, ce type de phénomène est analysé grâce aux modèles de mélange (mixture models) qui consistent à postuler l'existence de sous-distributions au sein de la distribution générale. Appliqués à l'analyse de survie, les modèles de mélange ont donné naissance à des variantes des modèles de Cox (Hougaard, 1995), notamment les modèles de « cure » où une partie de la population est considérée comme « immunisé » face à un certain événement (Farewell, 1982). Ces modèles sont cependant conçus pour l'analyse de données individuelles (modèles de survie) et non de taux agrégés, données plus facilement disponibles.

Comme nous l'avons déjà indiqué, il est impossible d'estimer à la fois la distribution de z_i et celle de $\mu(x)$. Il nous faut donc faire un choix. Nous avons décidé de donner la priorité à l'étude de la distribution de z_i plutôt qu'à celle de $\mu(x)$ car la question des populations vulnérables est moins connue que celle de la forme de la force de mortalité, qui fait d'ailleurs l'objet d'un large consensus. Nous faisons donc l'hypothèse que chacune des sous-populations suit une loi de Gompertz, c'est-à-dire que sa force de mortalité progresse de manière exponentielle avec l'âge. Dans chaque sous-population, nous postulons donc que la force de mortalité est régie par deux paramètres, α et β , qui contrôlent respectivement le risque de décès à la naissance et le rythme de vieillissement (*rate of ageing*). Algébriquement, si l'on suppose l'existence de deux sous-populations, le risque de décès des individus dans chaque sous-population est le suivant.

$$\mu_1(x) = \alpha_1 \cdot e^{\beta_1 x} \quad \text{et} \quad \mu_2(x) = \alpha_2 \cdot e^{\beta_2 x}$$

En utilisant les propriétés de la loi de Gompertz, on peut en outre affirmer que la probabilité de survie de chaque sous-population évolue selon la relation suivante.

$$I_1(x) = e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 x} - 1)} \quad \text{et} \quad I_2(x) = e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 x} - 1)}$$

La force de mortalité agrégée s'obtient en calculant la moyenne pondérée des deux sous-populations, qui dépend non seulement de la probabilité de survie au sein de chaque sous-population, mais également de la part de chaque sous-population dans la population initiale exprimée de 0 à 1, qu'on notera π_1 et π_2 . Ainsi, la force de mortalité agrégée est régie par la fonction suivante².

$$\bar{\mu}(x) = \frac{\mu_1(x) \cdot l_1(x) \cdot \pi_1 + \mu_2(x) \cdot l_2(x) \cdot \pi_2}{l_1(x) \cdot \pi_1 + l_2(x) \cdot \pi_2} = \frac{\alpha_1 \cdot e^{\beta_1 \cdot x} \cdot e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + \alpha_2 \cdot e^{\beta_2 \cdot x} \cdot e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2}{e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2}$$

Ce modèle peut être généralisé à un plus grand nombre de sous-populations. Cela peut être pertinent lorsque la force de mortalité agrégée présente d'autres déviations par rapport à la loi de Gompertz que celle représentée par la bosse de surmortalité des jeunes adultes. C'est le cas également lorsque la bosse de surmortalité présente un aplatissement après son sommet (Kostaki, 1992). Afin de prendre en compte ces possibilités, nous avons également construit un modèle à trois sous-populations dont l'expression algébrique est la suivante³.

$$\begin{aligned} \bar{\mu}(x) &= \frac{\mu_1(x) \cdot l_1(x) \cdot \pi_1 + \mu_2(x) \cdot l_2(x) \cdot \pi_2 + \mu_3(x) \cdot l_3(x) \cdot \pi_3}{l_1(x) \cdot \pi_1 + l_2(x) \cdot \pi_2 + l_3(x) \cdot \pi_3} \\ &= \frac{\alpha_1 \cdot e^{\beta_1 \cdot x} \cdot e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + \alpha_2 \cdot e^{\beta_2 \cdot x} \cdot e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2 + \alpha_3 \cdot e^{\beta_3 \cdot x} \cdot e^{-\frac{\alpha_3}{\beta_3} \cdot (e^{\beta_3 \cdot x} - 1)} \cdot \pi_3}{e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2 + e^{-\frac{\alpha_3}{\beta_3} \cdot (e^{\beta_3 \cdot x} - 1)} \cdot \pi_3} \end{aligned}$$

Ce modèle est original dans la mesure où l'estimation d'une fragilité discrète n'a, à notre connaissance, encore jamais été réalisée en démographie. Des exemples numériques ont certes été utilisés comme illustration, mais nous proposons ici d'estimer la valeur des paramètres permettant de générer les taux de mortalité observés dans des populations réelles. En particulier, cela nous informerait sur la proportion d'individus appartenant à la sous-population vulnérable. Certaines limitations doivent toutefois être prises en compte.

Premièrement, ce modèle ne peut être appliqué qu'à des données de cohortes et non de périodes sous peine de confondre les effets d'âge et de période (Frost, 1939). Deuxièmement, ce modèle ne considère ni la migration externe, ni la migration entre les sous-populations. En d'autres termes, nous conservons l'hypothèse d'une fragilité fixe pour chaque individu, telle qu'envisagée initialement. Certains modèles ont intégré la possibilité de changer de niveau de fragilité au cours du temps (Gavrilov et Gavrilova, 1991, Le Bras, 1976, Vaupel, *et al.*, 1988) mais ceux-ci introduisent alors deux types de risque, le risque de décès et le risque de changer de fragilité, ce qui rend le modèle quasiment impossible à estimer car il existe une infinité de combinaisons de ces deux risques permettant de générer les taux observés. Bien qu'il soit préférable dans l'absolu de permettre aux individus de passer d'un groupe à l'autre, nous

² Bien que cette formule contienne six paramètres, seuls cinq doivent être estimés, puisque π_2 peut être exprimé comme la résultante de π_1 , c'est-à-dire que $\pi_2 = 1 - \pi_1$.

³ Bien que cette formule contienne neuf paramètres, seuls huit doivent être estimés, puisque π_3 peut être exprimé comme la résultante de π_1 et π_2 , c'est-à-dire que $\pi_3 = 1 - \pi_2 - \pi_1$.

allons retenir ici un modèle où l'appartenance à une sous-population est déterminée une fois pour toutes.

Un sous-produit de ce modèle est qu'il permet d'estimer de manière originale l'espérance de vie à la naissance perdue en raison de la surmortalité des jeunes adultes. Il est en effet possible d'estimer l'espérance de vie pour chacune des sous-populations ainsi que sur la population totale. L'espérance de vie calculée sur la population totale \bar{e}_0 représente le niveau de mortalité général, alors que l'espérance de vie dans chaque sous-population e_{o_1} et e_{o_2} indiquent la longévité moyenne au sein de chaque groupe. Or, l'espérance de vie de la sous-population la plus robuste peut également être interprétée comme celle qui aurait été atteinte par l'ensemble de la population en l'absence de la sous-population la plus fragile. Par conséquent, la différence entre l'espérance de vie calculée pour la population totale et celle calculée pour la sous-population la plus robuste permet d'estimer la perte d'espérance de vie induite par la présence de la sous-population vulnérable. Autrement dit, cela permet d'estimer le gain potentiel d'espérance de vie qui serait possible si la sous-population vulnérable suivait la force de mortalité de la sous-population majoritaire.

Si l'on considère l'existence de trois sous-populations différentes, un choix doit être effectué concernant la position de la sous-population intermédiaire. En effet, dans certains cas cette dernière peut être assimilée à une deuxième sous-population vulnérable. Dans d'autres cas, elle peut être considérée comme une variante de la sous-population majoritaire, qui vit par exemple aussi longtemps mais connaît un rythme de vieillissement différent. Il est particulièrement délicat de fixer une règle générale dans cette situation. Dans cette étude, nous avons choisi de comptabiliser la troisième sous-population avec la sous-population la plus vulnérable lorsque les espérances de vie de ces deux sous-populations diffèrent de moins de dix ans. Dans le cas contraire elle est considérée comme faisant partie de la population générale.

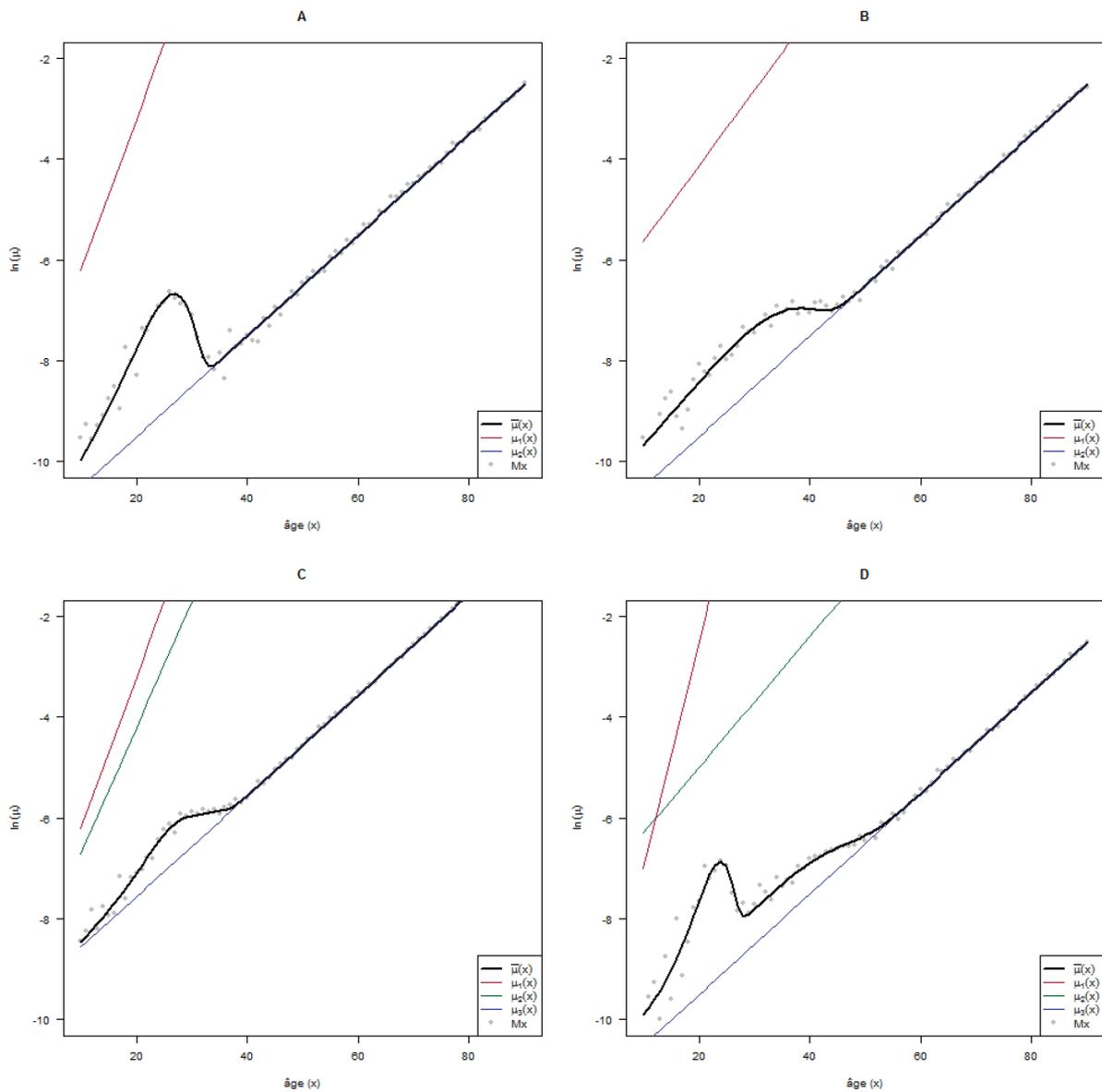
Fonction des moindres carrés et questions d'estimation

L'estimation des paramètres peut être faite par minimisation des moindres carrés. Dans le but de corriger l'hétéroscédasticité inhérente à toute estimation des taux de mortalité par âge couvrant une telle gamme de valeurs, l'usage de pondérations est indispensable. Ces poids devraient être approximativement égaux à la variance des résidus, qu'on peut approcher par l'inverse du nombre de décès dans chaque classe d'âge (Brillinger, 1986). En conséquence, la fonction des moindres carrés peut-être définie de la manière suivante.

$$RSS = \sum (M_x - \bar{\mu}(x))^2 \cdot \frac{1}{D_x}$$

Où M_x est le taux de mortalité mesuré à l'âge x , $\bar{\mu}(x)$ est le taux de mortalité agrégé estimé à l'âge x et D_x est le nombre absolu de décès enregistré à l'âge x . Etant donné la complexité de la fonction $\bar{\mu}(x)$, il est pertinent d'étudier le comportement de la fonction des moindres carrés afin de déterminer notamment dans quelle mesure elle respecte les exigences de convexité habituelles. Pour cela, l'approche la plus pertinente est de la tester par des simulations, qui ont l'avantage de permettre de comparer les « vraies » valeurs des paramètres utilisés avec les valeurs estimées.

Figure 2 . Quatre simulations de mortalité basées sur l'hypothèse de sélection



Pour cet exercice, nous avons généré quatre jeux de données reflétant chacun une hypothèse différente de distribution de l'hétérogénéité non-observée à l'intérieur de la population totale (figure 2). Le premier cas (A) représente une situation dans laquelle deux sous-populations sont présentes et divergent de manière marquée. Cette situation donne naissance à une bosse de surmortalité « classique » (symétrique et unimodale).

Le deuxième jeu de données (B) correspond à une situation dans laquelle la sous-population minoritaire connaît un rythme de vieillissement (β_1) à peine supérieur à celui de la sous-population majoritaire (β_2). Ce deuxième scénario implique une faible divergence et la formation d'un plateau de mortalité plus que d'une bosse, puisque le risque de mortalité ne diminue pas après avoir atteint son pic local.

La troisième simulation (C) postule l'existence de trois sous-populations, dont deux vulnérables subissant un rythme de vieillissement élevé. Cette situation permet de rendre compte de l'existence d'une

bosse de mortalité bimodale ou à plateau, tout en conservant une pente marquée dans la première partie de la bosse (contrairement au modèle B).

Le quatrième jeu de données (D) suppose également la présence de trois sous-populations, mais cette fois-ci seule l'une des trois connaît une progression rapide de la force de mortalité, alors que les deux autres forment ensemble une sous-population de fragilité faible. Ce scénario illustre la situation où la force de mortalité connaît une seconde déviation par rapport à la loi de Gompertz.

Une fois la force de mortalité agrégée simulée, une variabilité stochastique est introduite dans les taux de mortalité par âge. Concrètement, le nombre de décès par âge simulé (D_x) est le résultat d'un tirage aléatoire à partir d'une distribution binomiale $B(p, n)$, où p est égale à la force de mortalité calculée et n correspond à population suisse masculine du même âge en 2010 (N_x). Ce choix arbitraire correspond au cas d'un pays de faible mortalité dans laquelle le taux de mortalité subit une variabilité relativement élevée, ce qui permet de rester au plus près des conditions réelles d'utilisation du modèle. Ces décès D_x sont ensuite transformés en taux de mortalité en les divisant par la même population soumise au risque pour obtenir M_x , le taux de mortalité simulé.

Les simulations A et D représentent des situations antagonistes pour lesquelles il devrait être raisonnablement aisé de distinguer le nombre de sous-populations. Au contraire, les simulations B et C présentent une apparence similaire au niveau agrégé et ne diffèrent que sur la première partie de la bosse de surmortalité. Dans ces deux jeux de données, il devrait être plus difficile de déterminer le nombre optimal de sous-populations.

Dans toute procédure d'estimation paramétrique, il est important de connaître le comportement de la fonction objective. Cela donne une bonne indication sur les conditions dans lesquelles l'algorithme d'optimisation évolue. En effet, la plupart des algorithmes couramment utilisés fonctionnent selon l'hypothèse que la fonction objective est convexe (ou concave dans le cas d'une maximisation), c'est-à-dire que la solution se trouve au cœur d'un optimum global unique. Or, cette condition n'est pas garantie et l'est d'autant moins que le nombre de paramètres augmente et que la fonction à optimiser se complexifie. Plus la fonction objective s'écarte de cette condition, plus il est difficile pour les algorithmes classiques de localiser les valeurs optimales des paramètres.

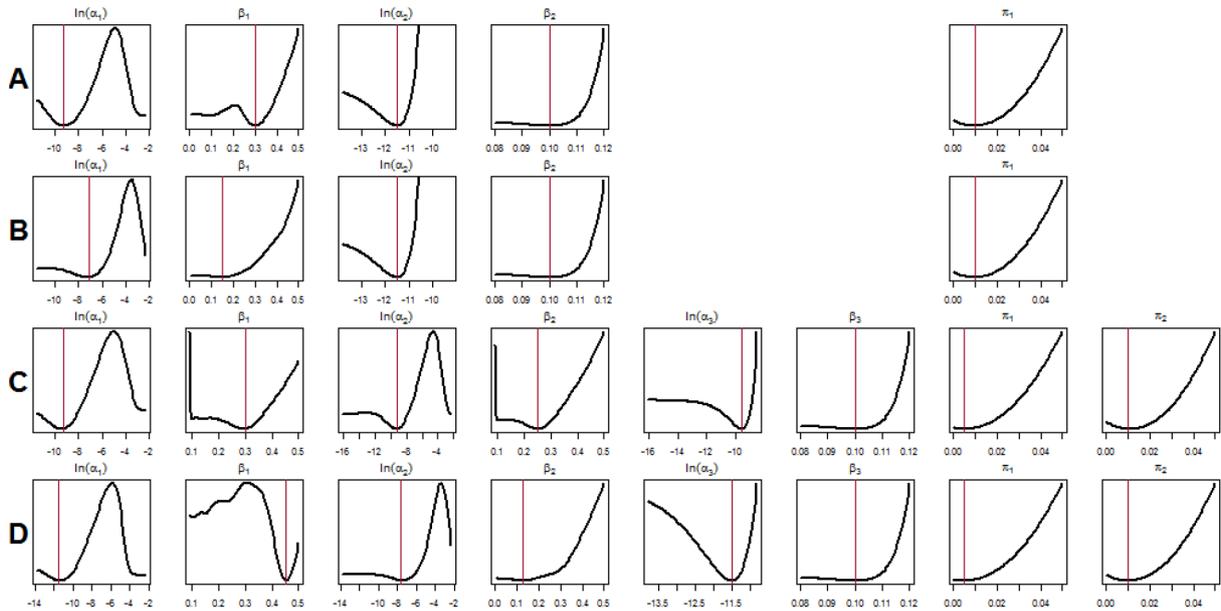
La forme d'une fonction objective est aisée à représenter lorsqu'un seul, voire deux paramètres doivent être estimés. Toutefois, notre fonction possède de cinq à huit paramètres libres, suivant que l'on suppose l'existence de deux ou trois sous-populations. Représenter graphiquement un problème à cinq ou huit dimensions étant impossible, nous avons successivement estimé l'évolution de la fonction objective selon chaque paramètre un à un, en fixant les valeurs des autres paramètres sur les vraies valeurs. Cela permet d'obtenir une bonne idée de la forme de la fonction objective dans un voisinage proche des vraies valeurs (figure 3). À chaque ligne correspond ainsi une simulation (A à D) et à chaque colonne un paramètre. Le trait vertical rouge indique la vraie valeur du paramètre en question et donc la position théorique du minimum de la fonction objective.

Cette analyse montre que la fonction des moindres carrés n'est strictement convexe que pour certains paramètres. Les cas les plus problématiques sont α_1 pour tous les jeux de données⁴, ainsi que α_2 , β_1 et

⁴ Les paramètres α sont représentés ici sous leurs formes logarithmiques pour des raisons visuelles.

β_2 pour les cas à trois sous-populations. Les autres relations sont entièrement convexes, mais n'oublions pas que n'est représentée ici qu'une portion du problème, une projection sur une seule dimension d'un problème à cinq ou huit dimensions. Ces fonctions d'apparence convexe pourraient en réalité revêtir des formes bien plus complexes en s'éloignant des solutions. La forme de la fonction de moindres carrés représente donc un défi non-négligeable pour l'estimation du modèle. Ces résultats s'expliquent par la présence de chaque paramètre à la fois au numérateur et au dénominateur de $\bar{\mu}(x)$, parfois même sous la forme d'exposant, contrairement à un modèle classique du type Heligman et Pollard (1980).

Figure 3. Evolution de la fonction des moindres carrés pour chaque jeu de données



Nous devons donc conclure de cette analyse de l'évolution de la fonction des moindres carrés pondérés que l'hypothèse de convexité n'est pas vérifiée même autour des vraies valeurs de chaque paramètre. Il existe donc plusieurs optimums locaux en plus de l'optimum général. Dans ces conditions, le choix de la méthode d'optimisation revêt une importance prépondérante. Plutôt que d'utiliser un algorithme à gradient classique, il semble donc important de choisir une méthode stochastique qui permet d'éviter au mieux les optimums locaux.

De nombreux tests sont possibles afin d'évaluer l'efficacité d'une méthode d'optimisation (pour une description détaillée, voir Remund, 2015), mais les présenter n'est pas notre objectif principal dans cette présentation et nous nous limitons ici à étudier la capacité de deux algorithmes à atteindre la solution de chacune des simulations proposées ci-dessus. Ces deux algorithmes sont d'une part celui proposé par Levenberg (1944) et Marquardt (1963), qu'on abrégera LM⁵, et d'autre part la méthode dite d'évolution différentielle, dite ED⁶, qui est une méthode *évolutionnaire* proposée entre autres par Storn (1996).

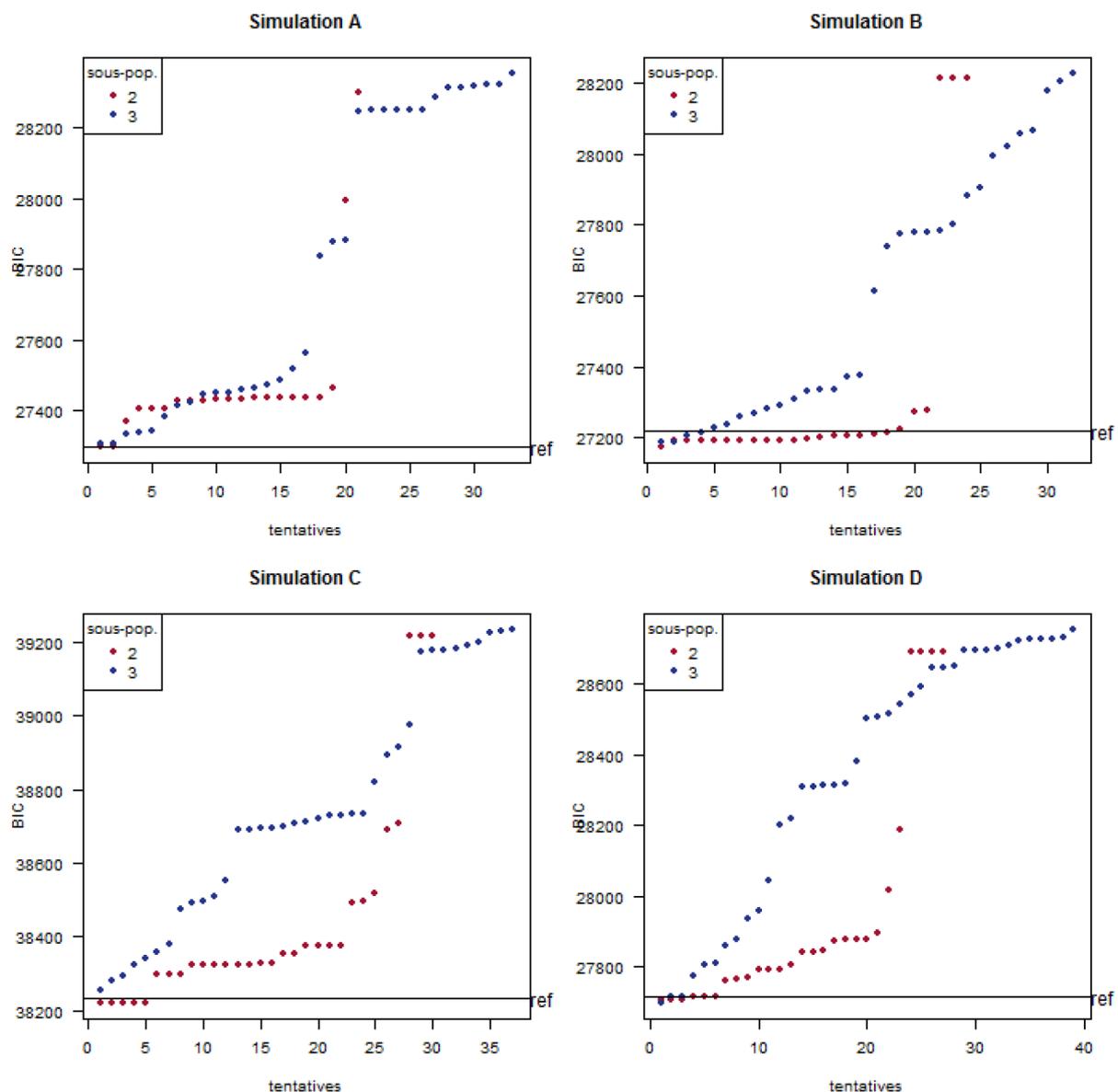
Le test imposé à chaque méthode vise à estimer leur capacité à faire abstraction des minimums locaux quelles que soient les valeurs initiales choisies dans l'espace défini par les bornes fixées. Pour cela,

⁵ La méthode LM est disponible dans la librairie R *minpack.lm* (Elzhov et al. 2013).
⁶ La méthode ED est disponible dans la librairie R *DEoptim* (Ardia et al. 2011, Price et al. 2005).

nous effectuerons une série d'estimations sur chaque jeu de données (A à D) en partant chacune des estimations de valeurs initiales issues d'un tirage au sort totalement aléatoire (issues d'une distribution uniforme). Pour chacun des quatre jeux de données, 100 valeurs initiales aléatoires seront tirées au sort et déboucheront sur 100 estimations des paramètres optimaux.

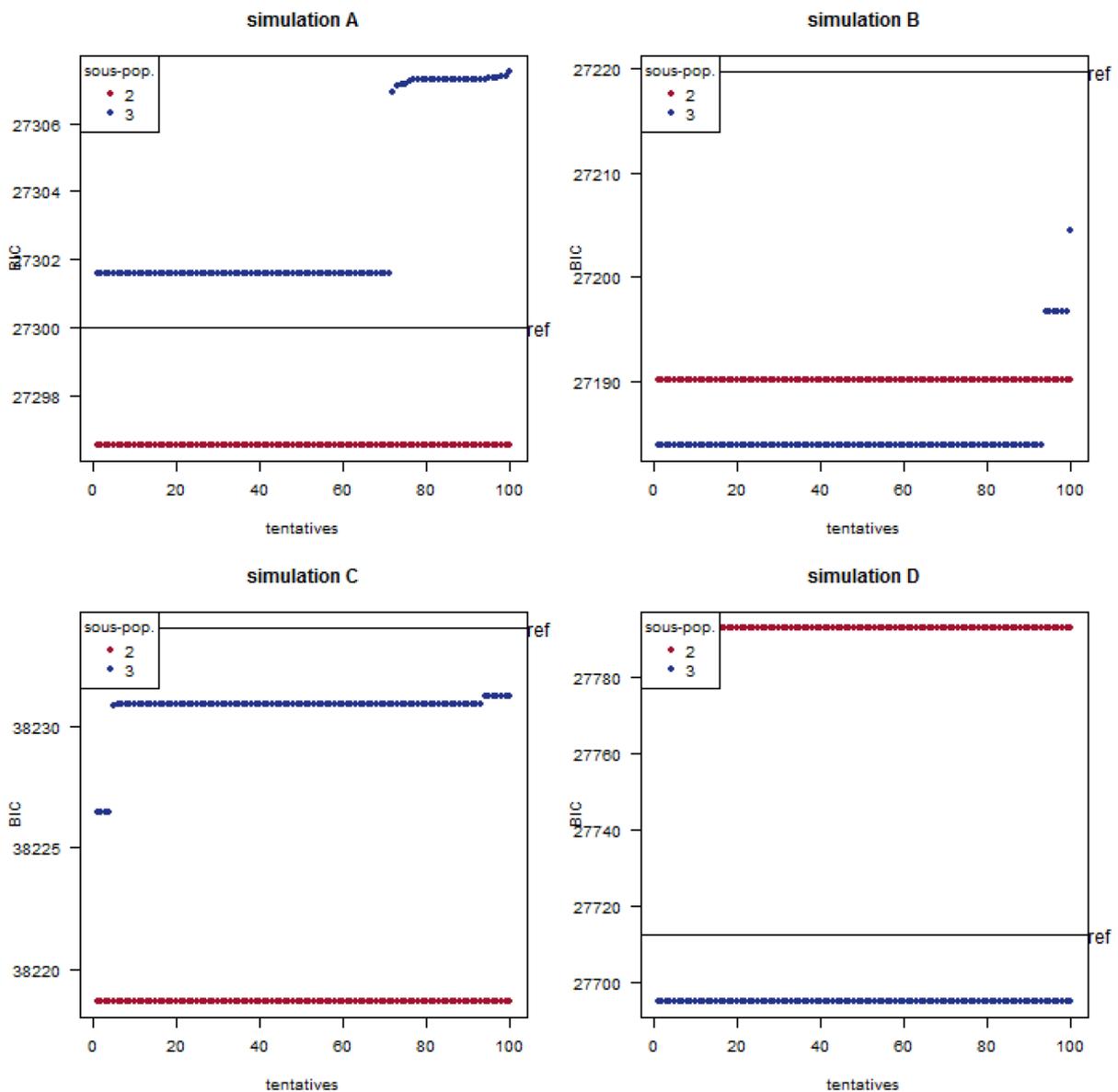
L'application de l'algorithme LM sur les quatre jeux de données débouche sur des résultats très décevants (figure 4). En effet, sur 100 estimations, seules 20 à 40 convergent suivant le jeu de données. De plus, lorsque l'algorithme converge, les résultats ne sont pas toujours bons. Certes, pour chaque jeu de données simulées, l'algorithme a au moins une fois permis de retrouver les valeurs utilisées pour les simulations, mais une probabilité de réussite de 1 sur 100 n'est de loin pas acceptable. Toutes les simulations montrent clairement l'existence de minimums locaux sur lesquels les estimations restent bloquées.

Figure 4. Capacité de l'algorithme LM à atteindre l'optimum global



L'application de l'algorithme ED⁷ permet d'obtenir de bien meilleurs résultats puisque toutes les estimations aboutissent à des résultats exploitables. De plus, quasiment toutes les estimations atteignent les valeurs fixées pour les simulations (figure 5). Cela permet, en comparant les valeurs du BIC pour chaque modèle, de déterminer combien de sous-populations doivent être utilisées. Ce choix est limpide pour les simulations A et D, pour lesquelles le modèle à deux et trois sous-populations ressortent clairement comme meilleurs. Cependant, pour les simulations B et C, le nombre de sous-populations choisi ne correspond pas à celui utilisé. Cette limitation suggère que le BIC n'est peut-être pas le meilleur critère pour identifier le bon nombre de sous-populations. Dans ce cas, un ré-échantillonnage pourrait offrir une solution pour se défaire de l'influence des valeurs aléatoires (Remund, 2015).

Figure 5 . Capacité de l'algorithme ED à atteindre l'optimum global



⁷ Nous avons utilisé pour cet exercice les paramètres suivants : NP=80, F=0.75, CR=0.9. Pour une justification de ce choix, voir Remund (2015).

La comparaison d'un algorithme à gradient et d'une méthode d'optimisation stochastique a donc permis de dégager les forces et faiblesses de chacune de ces techniques. La première est trop sensible au choix des valeurs initiales, qui, si elles sont mal choisies, peuvent résulter soit en une absence de convergence, soit en une incapacité à atteindre l'optimum global. La seconde, par contre, converge tout le temps et évite les optimums locaux. Il est toutefois nettement plus lent et la question de la détermination du nombre de sous-populations reste ouverte. La méthode ED permet dans tous les cas d'atteindre une qualité d'approximation élevée du taux de mortalité observé grâce au modèle proposé. Ces résultats sont suffisamment convaincants pour envisager une application à une population réelle.

Application à une population réelle

Pour cet exercice, nous avons choisi d'utiliser les cohortes des hommes et femmes nés en Suisse entre 1876 et 1981. Ces dates correspondent aux données disponibles dans la Human Mortality Database. Les populations exposées au risque, les nombres de décès ainsi que les taux de décès par âge ont été extraits de la base de données. Les données ont ensuite été tronquées à gauche au niveau de l'âge affichant le taux de mortalité le plus bas, ceci dans le but d'éviter la phase de baisse de la mortalité durant l'enfance, connue sous le nom d'ontogenescence (Levitis, 2011). Les données ont été également censurées à droite après 90 ans afin d'éviter le plateau de mortalité observé aux âges avancés (Thatcher, *et al.*, 1998).

Pour les cohortes non-éteintes, les taux de mortalité inconnus ont été extrapolés à l'aide d'une loi de Gompertz. Pour les cohortes nées avant 1960, les deux paramètres α et β ont été imputées en fonction de la portion connue de la force de mortalité connue après 30 ans. Pour les cohortes plus jeunes, dont la trajectoire après 30 ans est trop courte pour permettre une imputation fiable, une méthode semi-fixe a été utilisée en fixant $\beta = 0.1$ et en estimant α afin de garantir la continuité de la série. Cette méthode peut paraître cavalière, mais permet d'analyser des cohortes non-éteintes sans interférer grandement dans le calcul de la bosse de surmortalité.

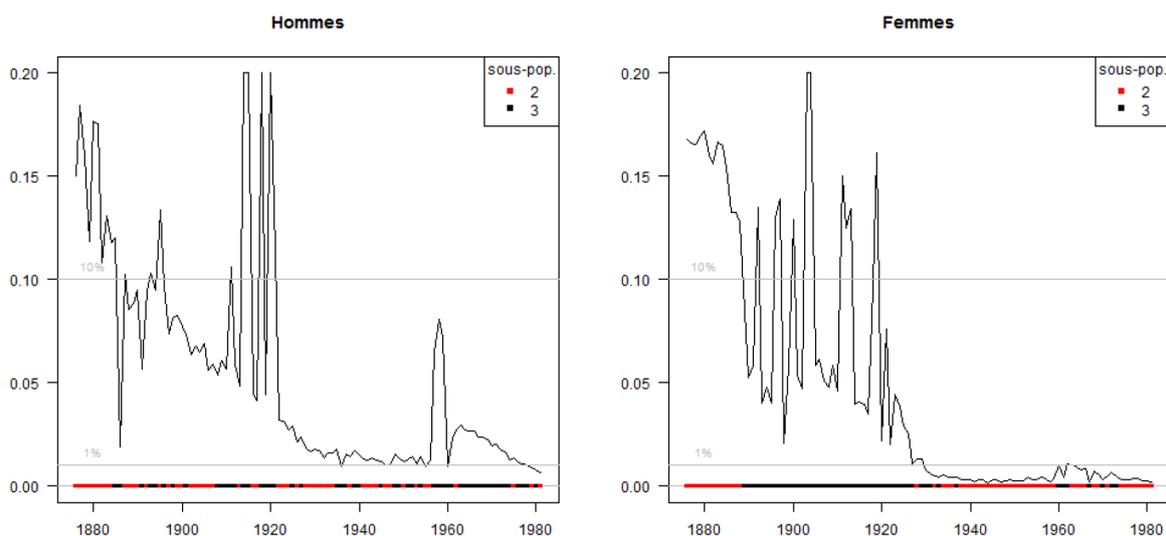
Pour chaque cohorte, nous avons estimé trois modèles avec respectivement une, deux et trois sous-populations. La pertinence de chaque modèle est comparée par la valeur du BIC. Si plus de 30 unités séparent les deux modèles, alors celui avec le BIC le plus élevé est conservé. Sinon, un ré-échantillonnage des taux de mortalité par âge est effectué en utilisant une loi binomiale de probabilité égale au taux observé afin de départager les deux meilleurs modèles. Cela permet d'exclure le rôle de la variabilité aléatoire dans le choix du nombre idéal de sous-populations.

Sur les 2×10^6 cohortes, les modèles à deux ou trois sous-populations ont toujours été préférés au modèle à une seule population. Cela montre que quelle que soit la cohorte, un modèle avec hétérogénéité est toujours capable de surpasser un modèle de Gompertz simple. Tant chez les hommes que chez les femmes, les modèles à deux ou trois sous-populations ressortent avec environ la même fréquence (figure 6). Notons que sans la méthode de ré-échantillonnage, le modèle à trois sous-populations aurait été choisi beaucoup plus souvent.

En ce qui concerne la proportion d'individus appartenant au groupe vulnérable, il est possible de l'approcher en observant l'évolution des paramètres π . Pour chaque sexe, cette proportion a diminué au cours

du temps, passant d'environ 15 % à moins de 1 %, voire 1‰ dans les dernières générations (figure 6). Cela signifie donc, si l'on suit l'hypothèse de sélection, qu'une part toujours plus importante de la population se confond dans la population générale.

Figure 6 . Proportion d'individus dans le groupe vulnérable au cours du temps



De plus, deux pics sont clairement identifiables autour des cohortes nées entre 1910 et 1920 chez les hommes, ainsi que celles nées autour de 1960 pour les deux sexes. Si le second pic correspond probablement aux cohortes frappées de plein fouet par le SIDA, le premier est difficilement explicable d'un point de vue épidémiologique. Notons également que la période pendant laquelle la sous-population vulnérable est importante (cohortes nées avant 1930) correspond à celle pour laquelle le modèle à 3 sous-populations est préféré à celui à 2 sous-populations, particulièrement chez les femmes. Il est donc possible que cette période reflète une situation similaire à la simulation C dans laquelle un plateau de mortalité est généré par une troisième sous-population, également vulnérable mais moins que la première. Dans un second temps, la bosse reprendrait une forme plus conventionnelle par la disparition de cette sous-population intermédiaire, qui laisserait place à une sous-population vulnérable plus petite mais plus fragile.

Dans tous les cas, ces analyses montrent qu'un groupe restreint de la population suffit à générer une bosse de surmortalité sur la force de mortalité agrégée. Dans les cohortes les plus récentes, où la bosse de surmortalité reste non-négligeable, un groupe d'individus vulnérables représentant à peine 1‰ de la cohorte suffirait à expliquer le phénomène. Cela signifie que la sous-population à la source de cette surmortalité peut être confinée à des profils très marginaux, au sens statistique et sans doute social du terme.

Conclusion

L'image traditionnelle renvoyée par la surmortalité des jeunes adultes, héritée de la conception psychanalytique de la crise d'adolescence et de la théorie neuropsychologique du *Cerveau Adolescent*, consiste

à présenter ce phénomène comme inévitable et universel. En démographie, cette perception est souvent acceptée sans discussion, ce qui démontre un manque d'intérêt pour la question. Cette vision est pourtant critiquable à la fois théoriquement et empiriquement. Une interprétation alternative consiste à concevoir la bosse de surmortalité des jeunes adultes comme le résultat de la présence de sous-populations particulièrement vulnérables qui augmentent temporairement le risque de décès, mais comme ces populations vulnérables meurent rapidement, le risque moyen de mortalité retrouve le niveau de la population générale.

L'intérêt de cette nouvelle hypothèse justifie que l'on clarifie les conditions dans lesquelles un tel processus peut naître. Nous avons choisi pour cela de développer un modèle à hétérogénéité discrète, c'est-à-dire qui prend en compte la possibilité que les individus appartiennent à différentes sous-populations dont le risque de décès diverge avec l'âge. Ce type de modèle étant hautement non-linéaire, nous avons montré que pour l'estimer il est nécessaire de faire appel à des techniques d'optimisation stochastiques comme, par exemple, un algorithme d'évolution différentielle. Dans ces conditions, un tel modèle est capable de refléter fidèlement l'existence d'une bosse de surmortalité sans avoir besoin de supposer qu'elle existe au sein d'aucune sous-population.

En outre, en appliquant ce modèle à une population réelle, celle des cohortes nées en Suisse entre 1876 et 1981, nous avons montré qu'il permet de tirer des conclusions intéressantes concernant l'évolution de la proportion de personnes concernées par cette vulnérabilité. En effet, selon cette interprétation, alors que dans les cohortes nées avant 1930 environ la proportion de ces individus vulnérables était relativement élevée (10 à 15 %) et composée souvent d'individus moins vulnérables, les cohortes plus récentes ont connu une concentration sur un plus petit nombre d'individus particulièrement fragiles. Cette évolution est cohérente avec certaines théories qui expliquent la persistance des différentiels socioéconomiques de santé par la prépondérance de caractéristiques personnelles, notamment psychologiques, dont les causes peuvent être liées également à la pauvreté (Mackenbach, 2012 : 765). Des conditions psychologiques comme la schizophrénie, les troubles bipolaires, l'addiction à certaines drogues ou la participation à des activités criminelles sont autant de marqueurs potentiels de cette vulnérabilité extrême qui méritent d'être explorés.

Bibliographie

- Ardia D., Boudt K., Carl P., Mullen K. M., Peterson B. G. 2011. Differential Evolution with DEoptim, *R Journal*, 3 (1).
- Arnett J. J. 1999. Adolescent storm and stress, reconsidered, *American psychologist*, 54 (5), 317-326.
- Arnett J. J. 2007. Suffering, selfish, slackers? Myths and reality about emerging adults, *Journal of Youth and Adolescence*, 36 (1), 23-29.
- Bessant J., Watts R. 2012. The mismeasurement of youth: why adolescent brain science is bad science, *Contemporary Social Science*, 7 (2), 181-196.
- Blos P. 1962. *On adolescence, a psychoanalytic interpretation*, New York, Free Press of Glencoe.
- Brillinger D. R. 1986. A Biometrics Invited Paper with Discussion: The Natural Variability of Vital Rates and Associated Statistics, *Biometrics*, 42 (4), 693-734.
- Choudhury S. 2010. Culturing the adolescent brain: what can neuroscience learn from anthropology?, *Social Cognitive and Affective Neuroscience*, 5 (2-3), 159-167.

- Dasen P. 2000. Rapid Social Change and the Turmoil of Adolescence: A Cross-Cultural Perspective, *International Journal of Group Tensions*, 29 (1-2), 17-49.
- Elzhov T. V., Mullen K. M., Spiess A.-N., Bolker B. 2013. Package 'minpack.lm'.
- Farewell V. T. 1982. The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors, *Biometrics*, 38 (4), 1041-1046.
- Finkelstein M., Esaulova V. 2006. Asymptotic behavior of a general class of mixture failure rates, *Advances in Applied Probability*, 244-262.
- Freud A. 1968. Adolescence, in *Adolescence: contemporary studies*, éd. par A. E. Winder et D. Angus, New York, American Book.
- Frost W. H. 1939. The age selection of mortality from tuberculosis in successive decades, *American Journal of Epidemiology*, 141 (1).
- Gampe J. 2010. Human mortality beyond age 110, 219-230, in *Supercentenarians*, éd. par H. Maier, J. Gampe, B. Jeune, J.-M. Robine et J. W. Vaupel, Springer.
- Gavrilov L. A., Gavrilova N. S. 1991. The biology of life span: a quantitative approach.
- Giedd J. N., Blumenthal J., Jeffries N. O., Castellanos F. X., Liu H., Zijdenbos A., Paus T., Evans A. C., Rapoport J. L. 1999. Brain development during childhood and adolescence: a longitudinal MRI study, *Nature neuroscience*, 2 (10), 861-863.
- Goldstein J. 2011. A Secular Trend toward Earlier Male Sexual Maturity: Evidence from Shifting Ages of Male Young Adult Mortality, *PLoS ONE*, 6 (8).
- Greenwood M. 1928. *Laws of mortality from the biological point of view*, *The Journal of Hygiene*, 28 (3).
- Hall S. 1904. *Adolescence: Its psychology and its relations to physiology, anthropology, sociology, sex, crime, religion and education*, New York, D. Appleton and company.
- Heligman L. Pollard J. H. 1980. The age pattern of mortality, *Journal of the Institute of Actuaries*, 107.
- Horiuchi S., Wilmoth J. R. 1998. Deceleration in the age pattern of mortality at older ages, *Demography*, 35 (4), 391-412.
- Hougaard P. 1995. Frailty models for survival data, *Lifetime Data Analysis*, 1 255-273.
- Kostaki A. 1992. A nine-parameter version of the Heligman-Pollard formula, *Mathematical Population Studies*, 3 (4), 277-288.
- Le Bras H. 1976. Lois de mortalité et age limite, *Population (French Edition)*, 655-692.
- Lenroot Rhoshel K., Giedd J. N. 2006. Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging, *Neuroscience & Biobehavioral Reviews*, 30 (6), 718-729.
- Levenberg K. 1944. A method for the solution of certain problems in least squares, *Quarterly of Applied Mathematics*, 2 164-168.
- Levitis D. A. 2011. Before senescence: the evolutionary demography of ontogenesis, *Proceedings of the Royal Society B: Biological Sciences*, 278 (1707), 801-809.
- Mackenbach J. P. 2012. The persistence of health inequalities in modern welfare states: The explanation of a paradox, *Social Science & Medicine*, 75 (4), 761-769.
- Males M. I. 2009. Does the Adolescent Brain Make Risk Taking Inevitable?: A Skeptical Appraisal, *Journal of Adolescent Research*, 24 (1), 3-20.

- Marquardt D. W. 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11 (2), 431-441.
- Mead M. 1928. *Coming of age in Samoa; a psychological study of primitive youth for western civilisation*, New York, W. Morrow & Company.
- Missov T. I., Finkelstein M. 2011. Admissible mixing distributions for a general class of mixture survival models with known asymptotics, *Theoretical Population Biology*, 80 (1), 64-70.
- Offer D., Ostrov E., Howard K. I. 1981. The mental health professional's concept of the normal adolescent, *Archives of General Psychiatry*, 38 (2), 149.
- Offer D., Schonert-Reichl K. A. 1992. Debunking the Myths of Adolescence: Findings from Recent Research, *Journal of the American Academy of Child & Adolescent Psychiatry*, 31 (6), 1003-1014.
- Perret-Catipovic M., Ladame F. 1997. *Adolescence et psychanalyse une histoire* sous la dir. de Maja Perret-Catipovic et François Ladame, Lausanne Paris, Delachaux et Niestlé.
- Petersen A. C. 1993. Presidential Address: Creating Adolescents: The Role of Context and Process in Developmental Trajectories, *Journal of Research on Adolescence*, 3 (1), 1-18.
- Price Kenneth V., Storn Rainer M., Lampinen J. A. 2005. *Differential evolution a practical approach to global optimization*.
- Remund A. 2015. *Jeunesses vulnérables? Mesures, composantes et causes de la surmortalité des jeunes adultes.*, Faculté des sciences de la société, Université de Genève, Genève.
- Sercombe H. 2010. The gift and the trap: Working the "teen brain" into our concept of youth, *Journal of Adolescent Research*, 25 (1), 31-47.
- Spear L. P. 2000. The adolescent brain and age-related behavioral manifestations, *Neuroscience & Biobehavioral Reviews*, 24 (4), 417-463.
- Spiess A. N. 2012. *A better 'nls'?*, édité par A. N. Spiess.
- Steinberg Laurence. 2005. Cognitive and affective development in adolescence, *Trends in Cognitive Sciences*, 9 (2), 69-74.
- Storn R. 1996. On the usage of differential evolution for function optimization Pp. 519-523 in *Fuzzy Information Processing Society, 1996. NAFIPS. 1996 Biennial Conference of the North American, IEEE*.
- Thatcher A. R., Kannisto V., Vaupel J. W. 1998. *The force of mortality at ages 80 to 120*, 22, Odense University Press Odense.
- Vaupel J. W., Yashin A. I. 2006. Unobserved population heterogeneity, in *Demography: analysis and synthesis*, 1, éd. par G. Caselli, J. Vallin et G. Wunsch, Amsterdam, Elsevier.
- Vaupel J. W., Manton K. G., Stallard E. 1979. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality, *Demography*, 16 (3), 439-454.
- Vaupel J. W., Yashin A. I. 1985. Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics, *The American Statistician*, 39 (3), 176-185.
- Vaupel J. W., Yashin A. I., Manton K. G. 1988. Debilitation's aftermath: Stochastic process models of mortality, *Mathematical Population Studies*, 1 (1), 21-48.